ONLINE
AFRICAN SCHOOL
on Migration Statistics

2–4 December 2020

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

IOM UN MIGRATION

GMDAC IOM Global Migration Data Analysis Centre

SCB

African Union

MODULE 4

Mapping migration using alternative data sources

Marzia Rango, *Data Innovation and Capacity-Building Coordinator, Global Migration Data*
Andrew Tatem, *Professor, World Pop, University of Southampton*

PART 1

Harnessing data innovation for migration analysis

Marzia Rango
Data Innovation and Capacity-Building Coordinator, Global Migration Data

Module 4: Mapping migration using alternative data sources

Part 1: Harnessing data innovation for migration analysis_Marzia Rango

Outline

1. **What**
   - Big data, data innovation
2. **Why**
   - The global context
   - The potential
3. **How**
   - Case-study: social media data to estimate ‘migrant’ stocks
   - The challenges
   - The Big Data for Migration Alliance (BD4M)

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 1: Harnessing data innovation for migration analysis_Marzia Rango

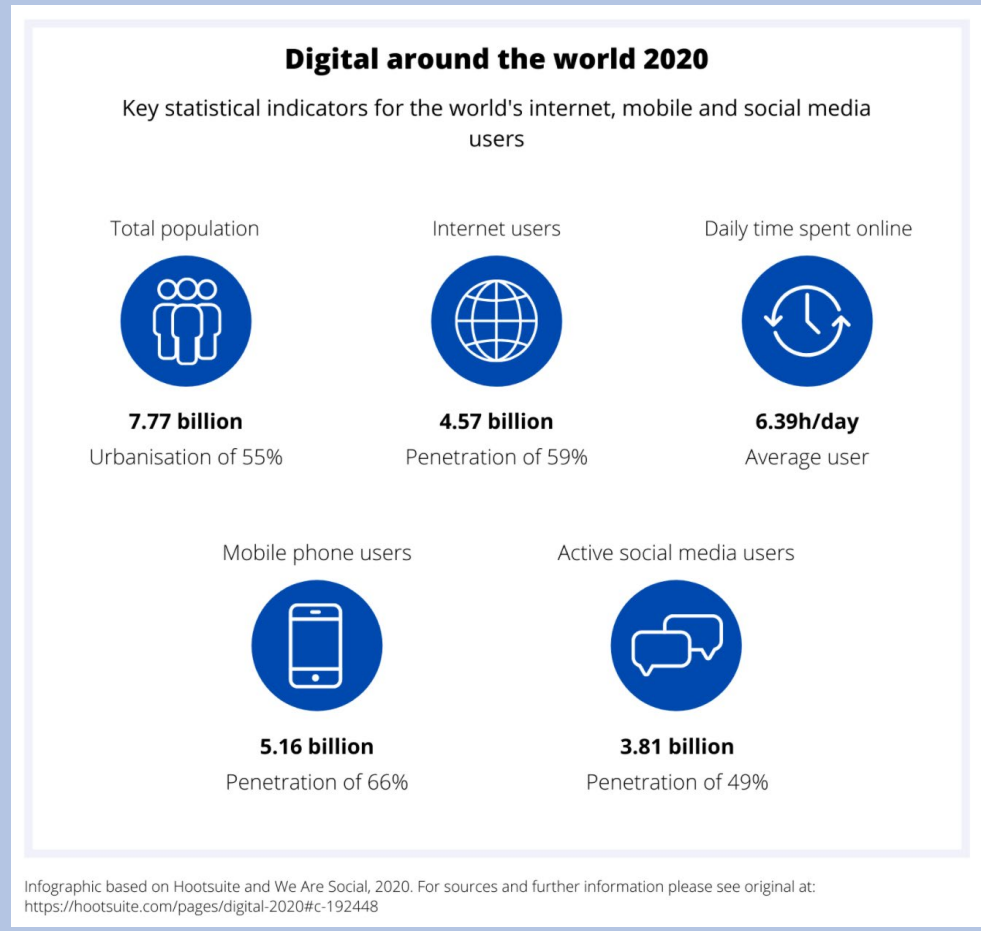
‘Big data,’ data innovation

Data innovation

- The universe of alternative data sources and methods
- Includes combination of traditional and new sources
- Sometimes also administrative/operational data

Big data

- Data generated automatically by users of mobile phones, social media, internet-based platforms, digital sensors
- Volume, velocity, variety (the “3 Vs”; Hilbert, 2013)
- Specific technology and analytical methods to extract insights and create value (de Mauro et al., 2014)



Infographic based on Hootsuite and We Are Social, 2020. For sources and further information please see original at: https://hootsuite.com/pages/digital-2020#c-192448

migrationdataportal.org/themen/big-data-migration-and-human-mobility

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 1: Harnessing data innovation for migration analysis_Marzia Rango

The global context

Huge progress made in discussions about data innovation for migration

- Renewed calls (needs) for evidence on migration: the SDGs and the GCM
- Limitations of traditional data sources
- Impact of COVID-19

Need for ethically responsible and privacy-safeguarding collaboration frameworks

- Re-using existing data (big data are already there!)
- Integrating ‘Big’ and ‘small’ data

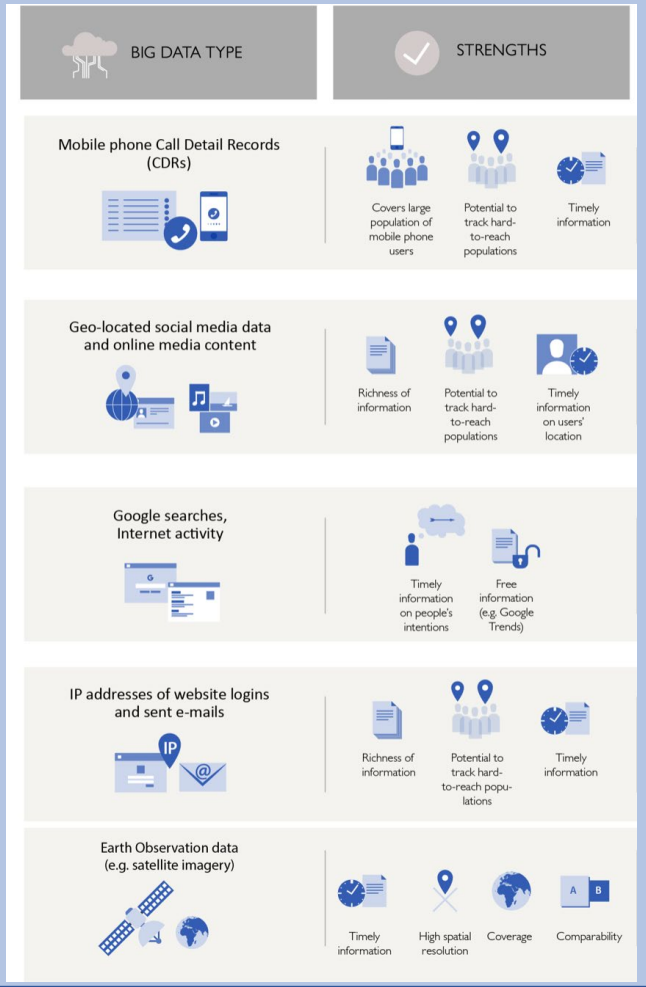
Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 1: Harnessing data innovation for migration analysis_Marzia Rango

The potential

- High spatial resolution
- High frequency of update
- Timeliness (virtually real-time)
- Wide coverage (hard-to-reach populations)
- Larger sample size compared to surveys
- Richness of information
- Relatively low cost (access!)



Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

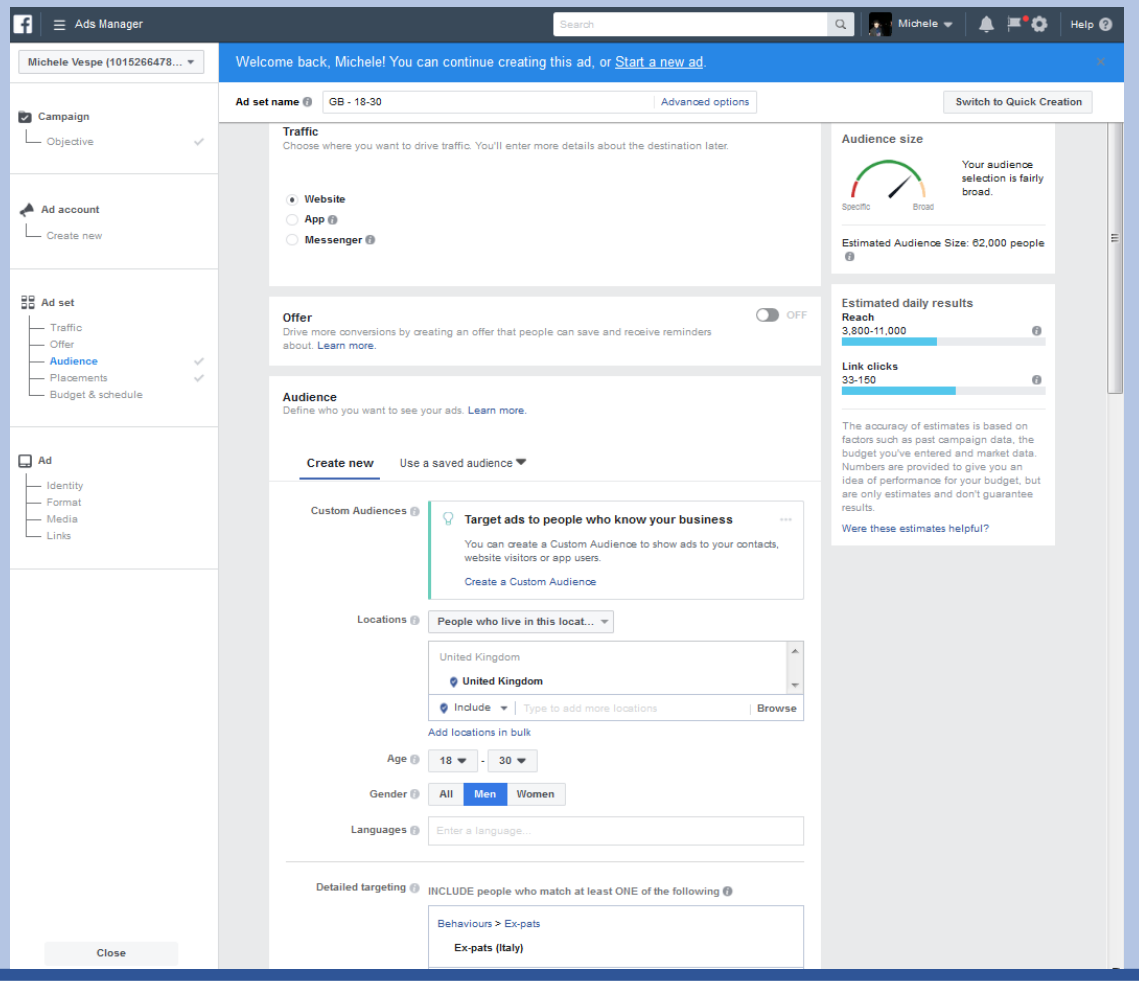
Part 1: Harnessing data innovation for migration analysis_Marzia Rango

Social media advertising platforms

- A “real time census”?
- ‘Inexpensive’
- Disaggregation → ‘used to live in country X’

BUT

- Reliability (often self-reported)
- Definitions (proprietary)
- Aggregations/rounding (proprietary)
- Penetration rate/selection bias
- Fake or double accounts



Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Understanding the bias

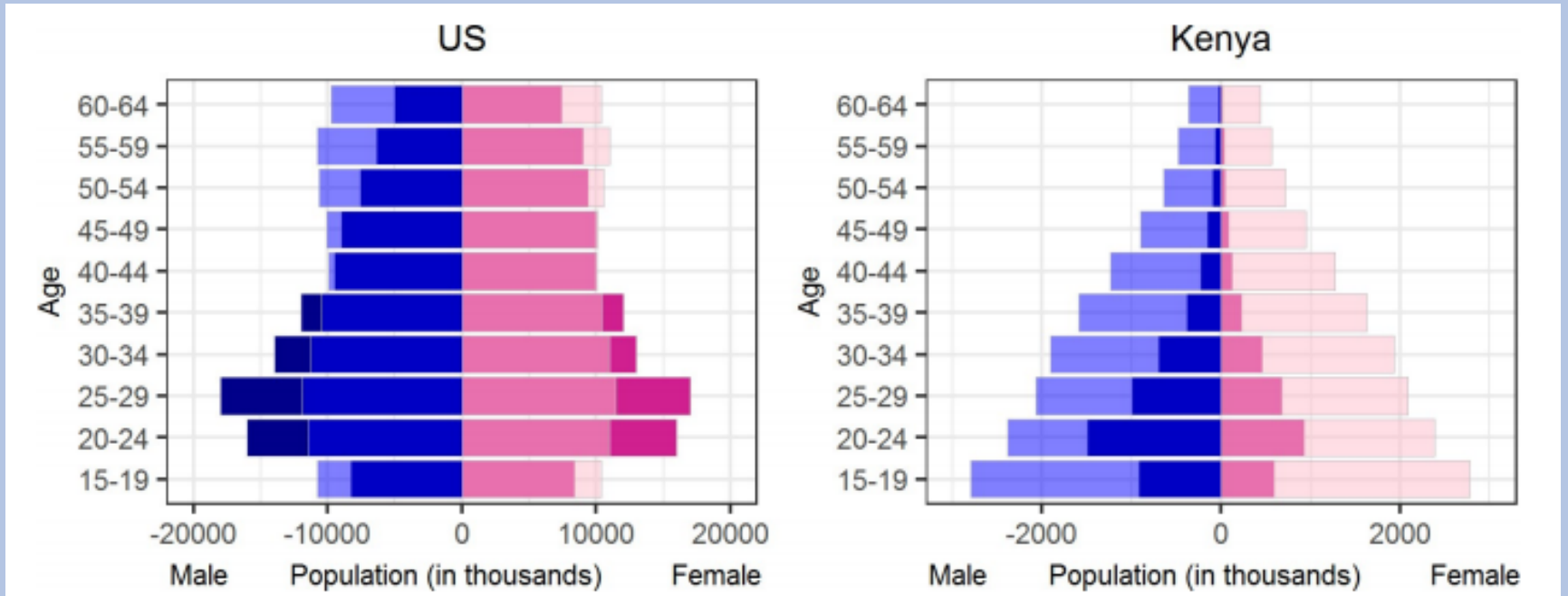


Fig 2. UNDESA population and FN users in the US and Kenya by age and sex/gender. Light shading = UNDESA population, mid shading = UNDESA & raw FN users, and dark shading = raw FN users.

Source: Spyratos S, Vespe M, Natale F, Weber I, Zagheni E, Rango M (2019) Quantifying international human mobility patterns using Facebook Network data. PLoS ONE 14(10): e0224134. https://doi.org/10.1371/journal.pone.0224134

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 1: Harnessing data innovation for migration analysis_Marzia Rango

‘Nowcasting’

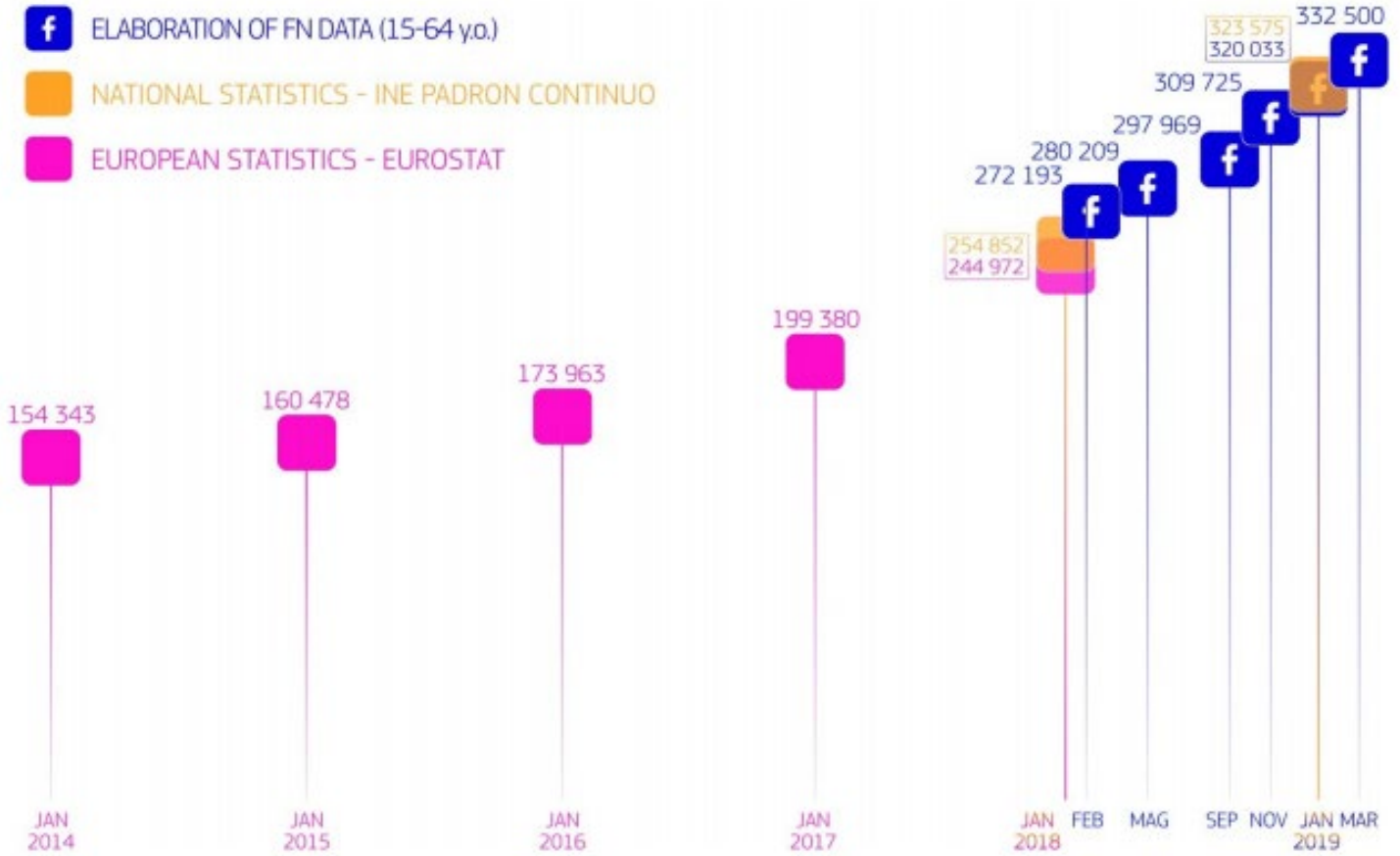


Fig 8. Stocks of international migrants and FN migrants from Venezuela in Spain. The national and Eurostat statistics refer to all ages, while the FN-derived estimates refer to ages 15–64.

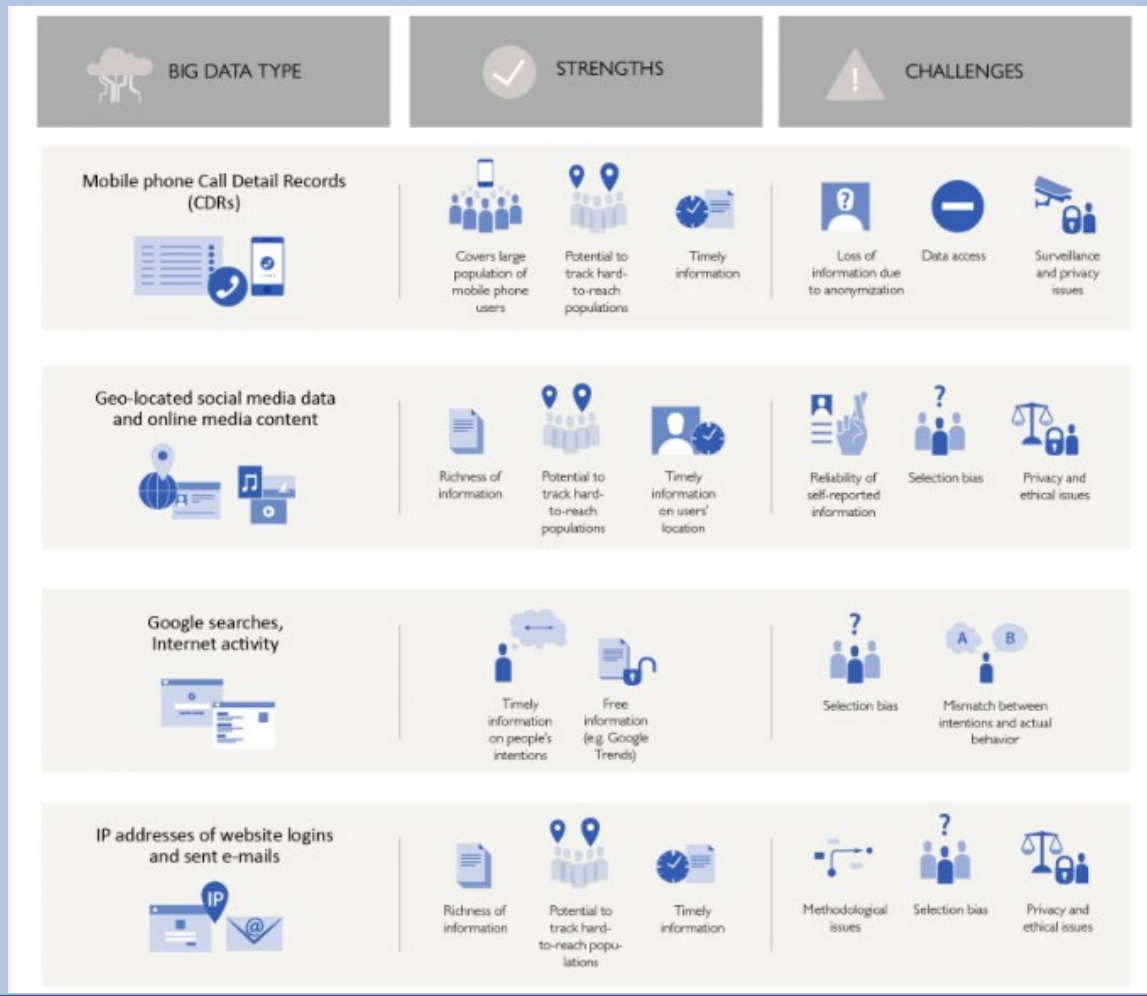
Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 1: Harnessing data innovation for migration analysis_Marzia Rango

The challenges

- Data access, continuity, independence
- Confidentiality
- **Security/Ethical issues**
- Methodology (selection bias)
- Reliability
- Definitions
- Fragmentation of applications



Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 1: Harnessing data innovation for migration analysis_Marzia Rango

The Big Data for Migration Alliance (BD4M)

Launched in 2018, convened by:

- IOM's GMDAC
- The European Commission's KCM
- The GovLab at NYU

data4migration.org/

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

The Big Data for Migration Alliance (BD4M)

A network of individuals and organizations across sectors that aims to:

- **Explore the potential of new data sources** and the combination of traditional and innovative methods
- **Ensure the ethical use** of data and the **protection of individuals' privacy**
- **Promote and facilitate new forms of partnerships** across the business, policy and scientific communities
- **Support peer-to-peer learning**, including by facilitating the sharing of good practices and building capacities

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 1: Harnessing data innovation for migration analysis_Marzia Rango

How we work

- Building the **knowledge base** for data innovation and collaboration

  → New **Data Innovation Directory**

- Identifying and **sourcing demand** and questions for data innovation

  → **100 Questions Initiative**

- **Building** data innovation **capacities**

- Promoting the use of data innovation to **address the demand** for policy-relevant analysis

- **Connecting** key stakeholders – data users, providers and policymakers

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

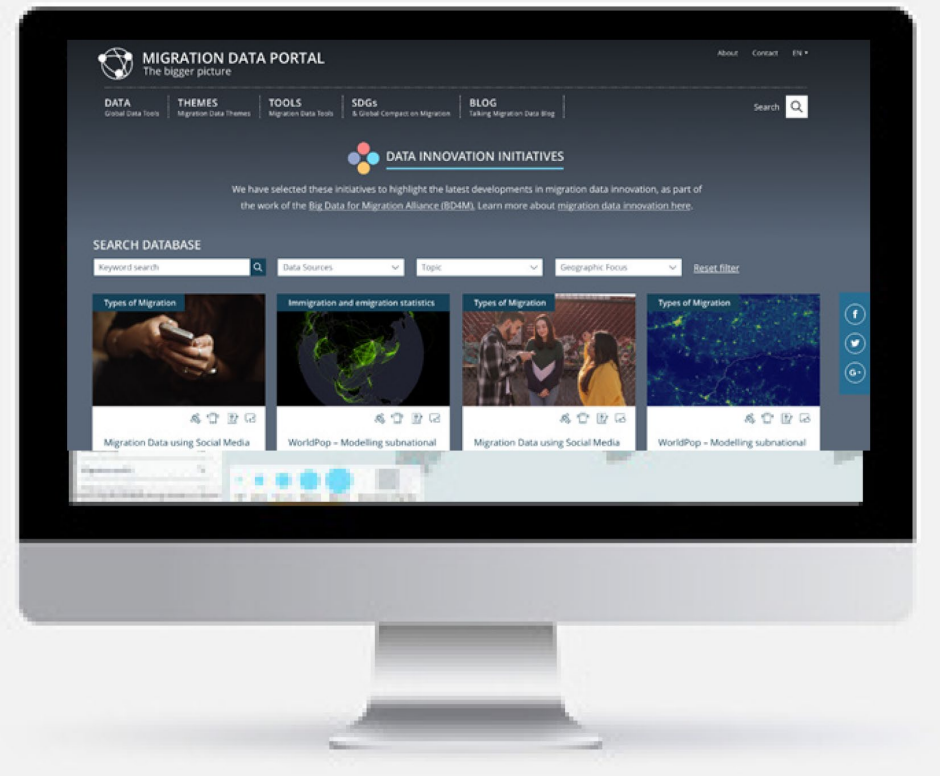
Module 4: Mapping migration using alternative data sources

Part 1: Harnessing data innovation for migration analysis_Marzia Rango

The Data Innovation Directory (DID)

- Overview of existing projects and applications of new data sources in relation to migration/mobility
- 50+ projects by more than 70 different organizations (9 related to COVID-19)
- Sources: publicly available information/interviews, validation process

→ To feature relevant projects: did@iom.int



Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

BD4M: ongoing activities

- **'Studios' around key migration questions**
  Dedicated workshops gathering representatives from businesses and policy
- **BD4M Guidebook *Harnessing Data Innovation for Migration Policy***
  With contributions from BD4M members and partners
- **BD4M webinar series (since Nov. 2020)**
- **Research project with the University of Liverpool**
  on episodes of xenophobia pre/post-COVID-19 based on Twitter data

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Contact

Marzia Rango

IOM's Global Migration Data Analysis Centre

mrango@iom.int

www.gmdac.iom.int

Twitter: @IOM_GMDAC

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

PART 2

Andrew Tatem
Professor, World Pop, University of Southampton

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Contents

- ‘Traditional’ vs ‘Novel/Big/Alternative’ data sources
- Case study 1: Geospatial data integration for subnational sex-disaggregated migration mapping
- Case study 2: Mobile phone data for migration statistics

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

TRADITIONAL VS ALTERNATIVE DATASETS FOR MIGRATION MAPPING

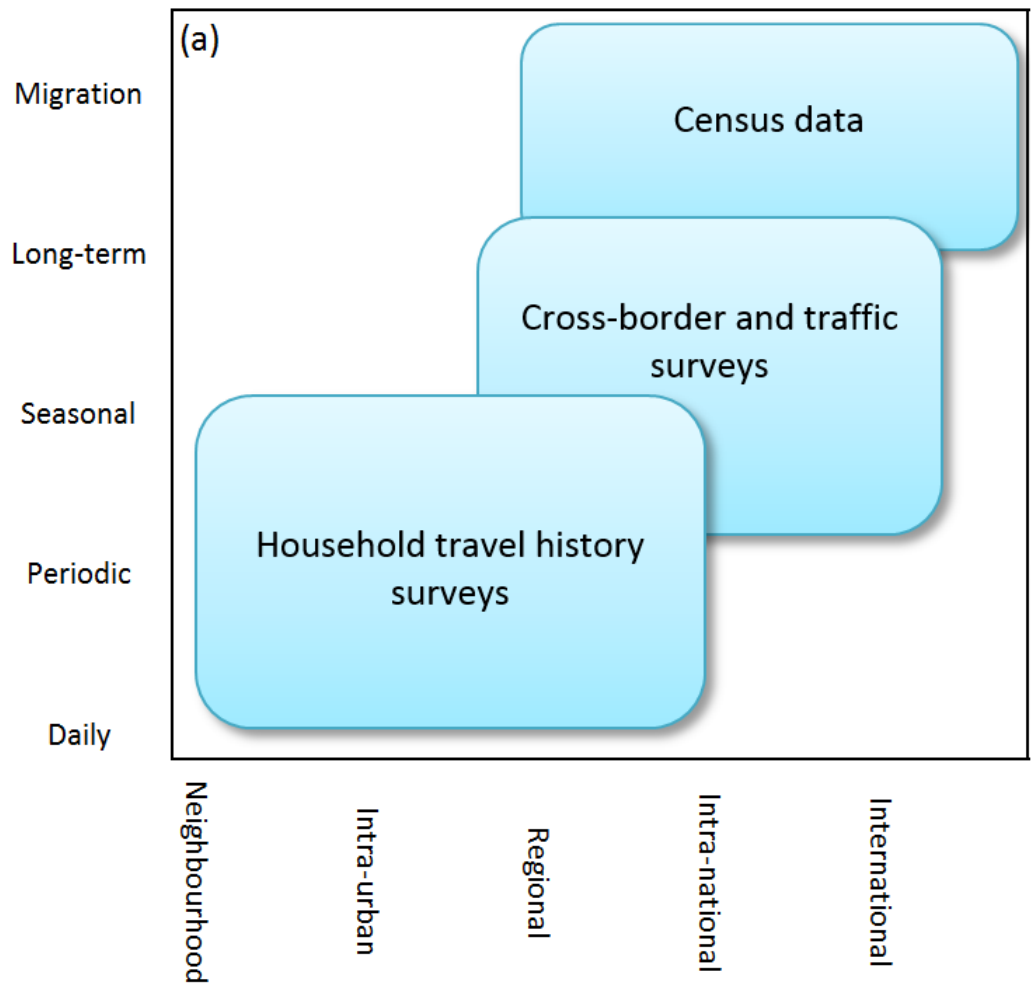
Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

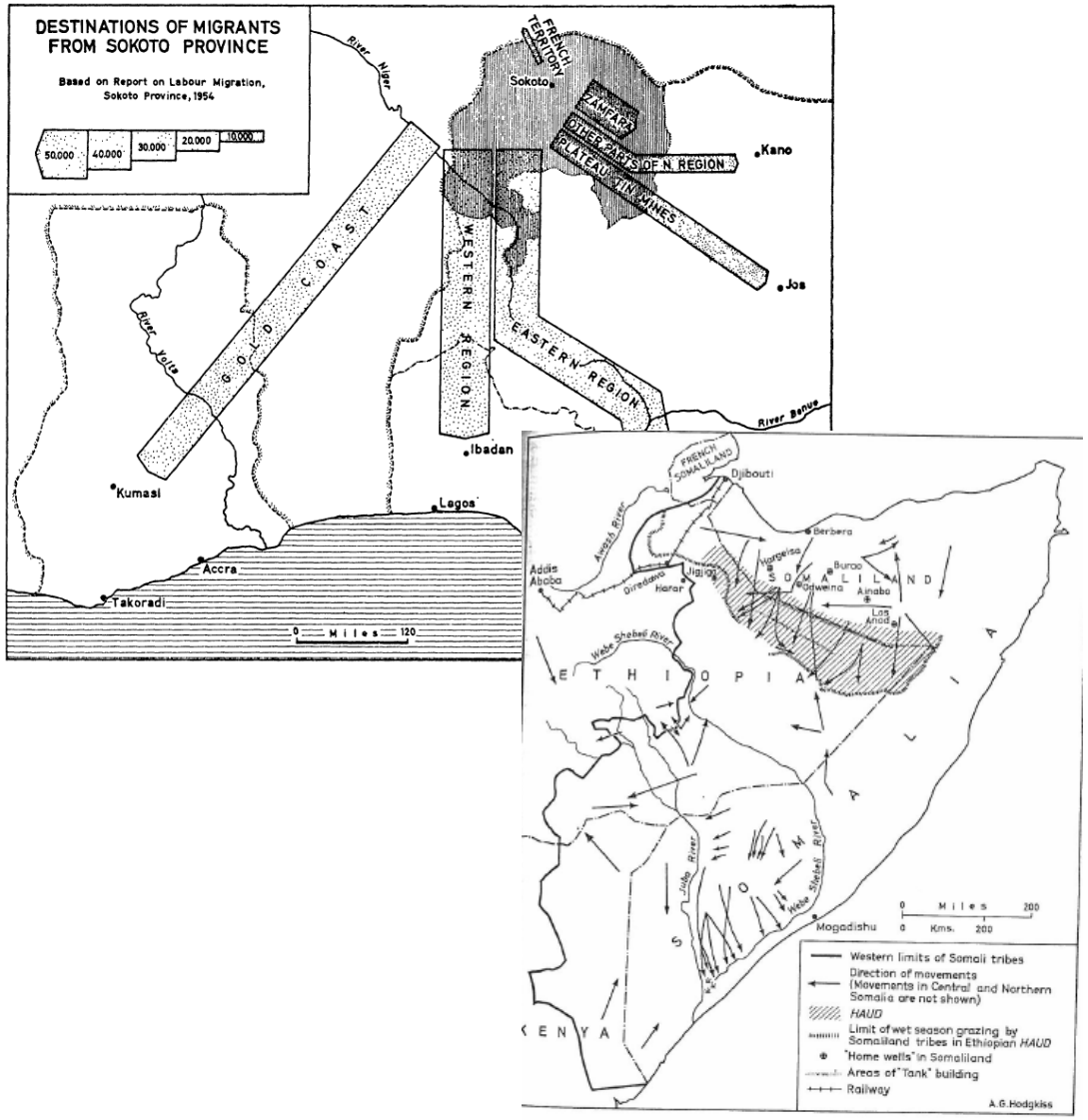
Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Data for measuring population movements/migrations

Traditional

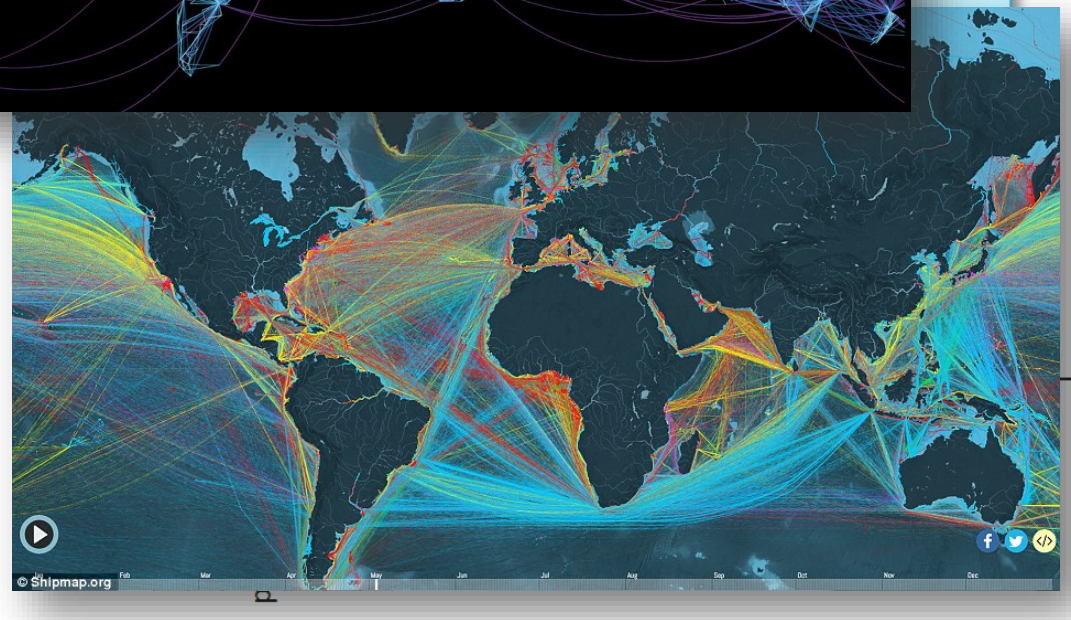


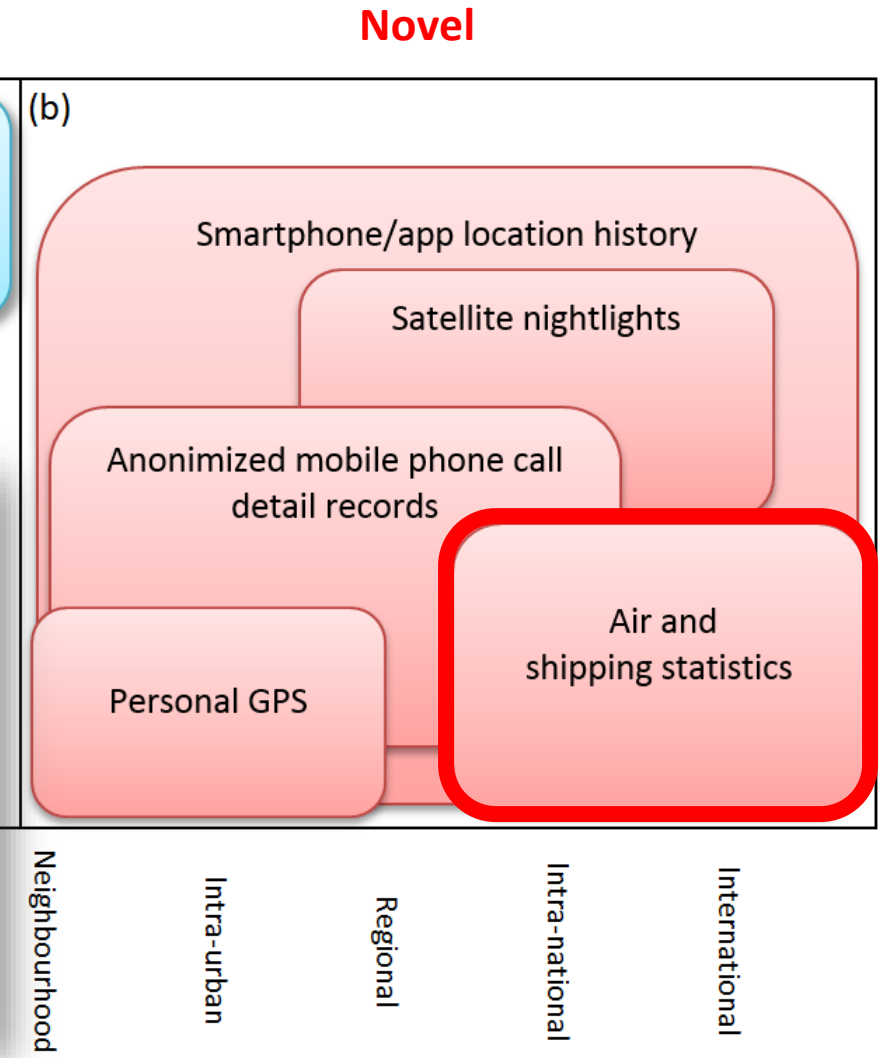


Tatem, *Int Health*, 2014; Ruktanonchai et al, *IJHG*, 2018

Prothero, JIAI 1957; Bull WHO 1961

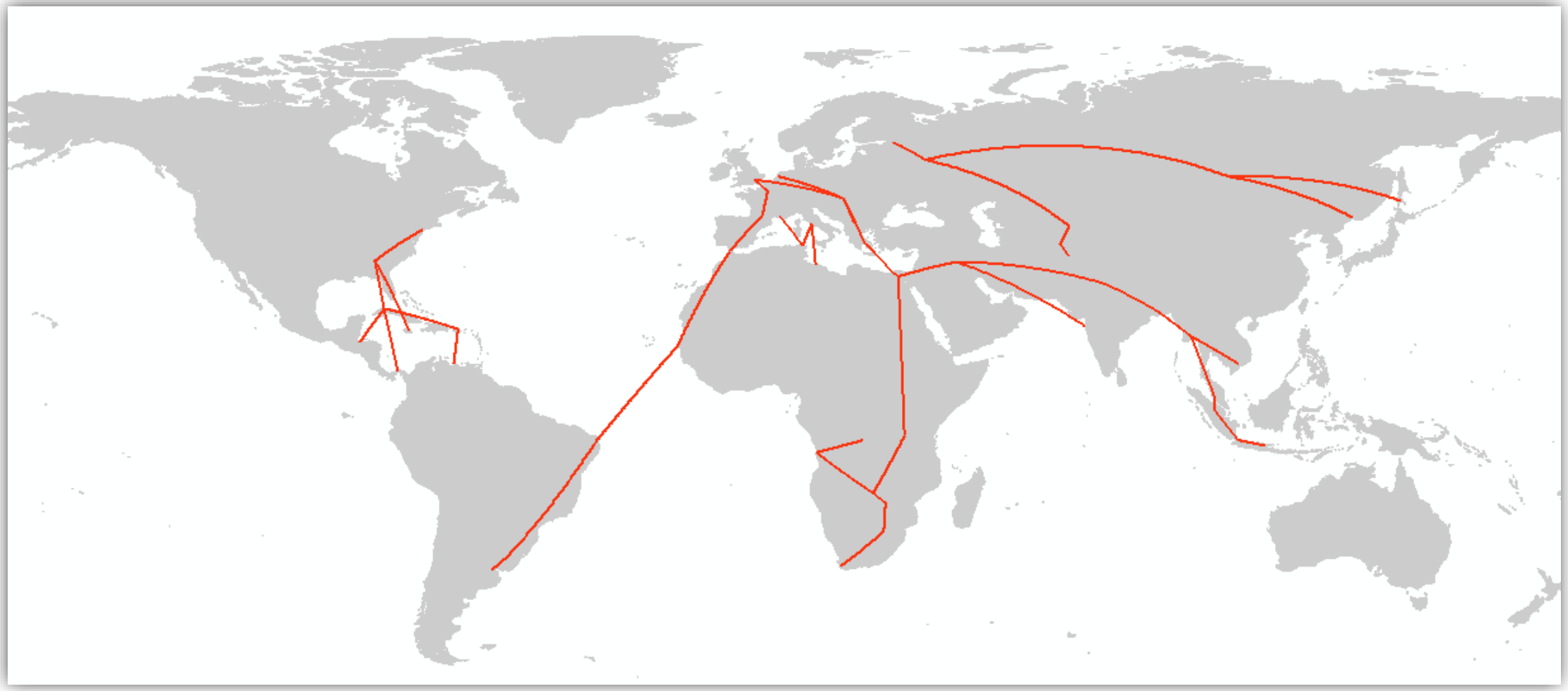
Data for measuring population movements/migrations





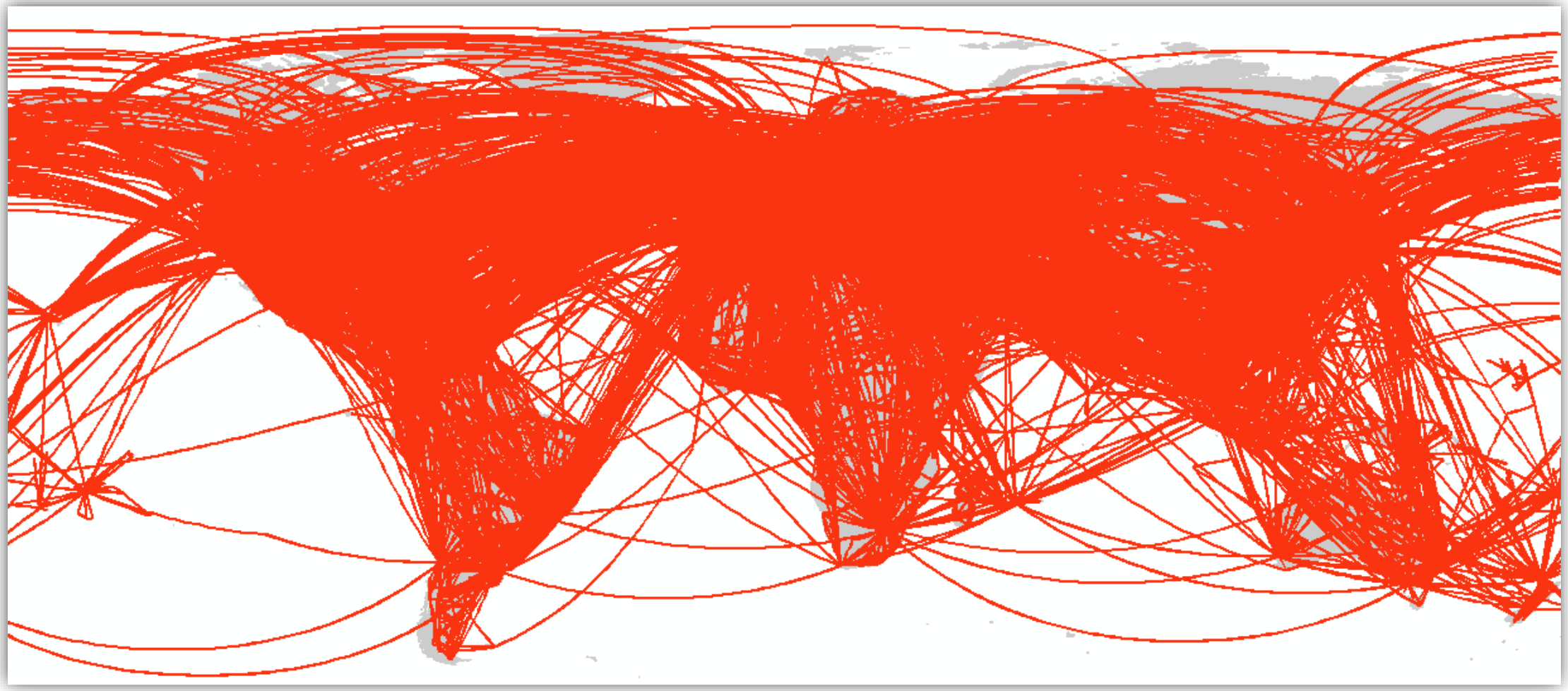
Tatem, *Int Health*, 2014; Ruktanonchai et al, *IJHG*, 2018

Global air network: 1933



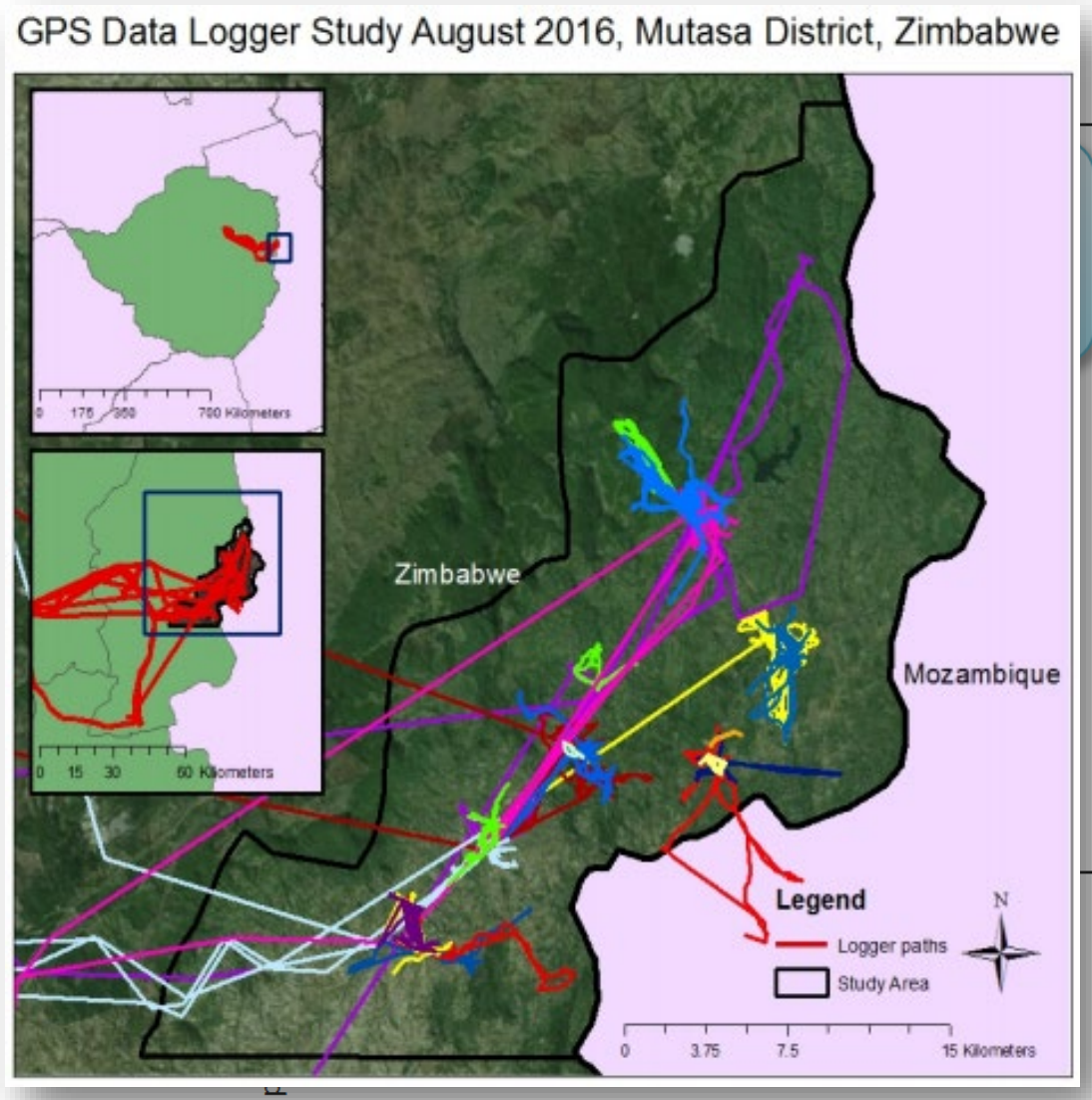
Tatem (2014) Int Health

Global air network: Today

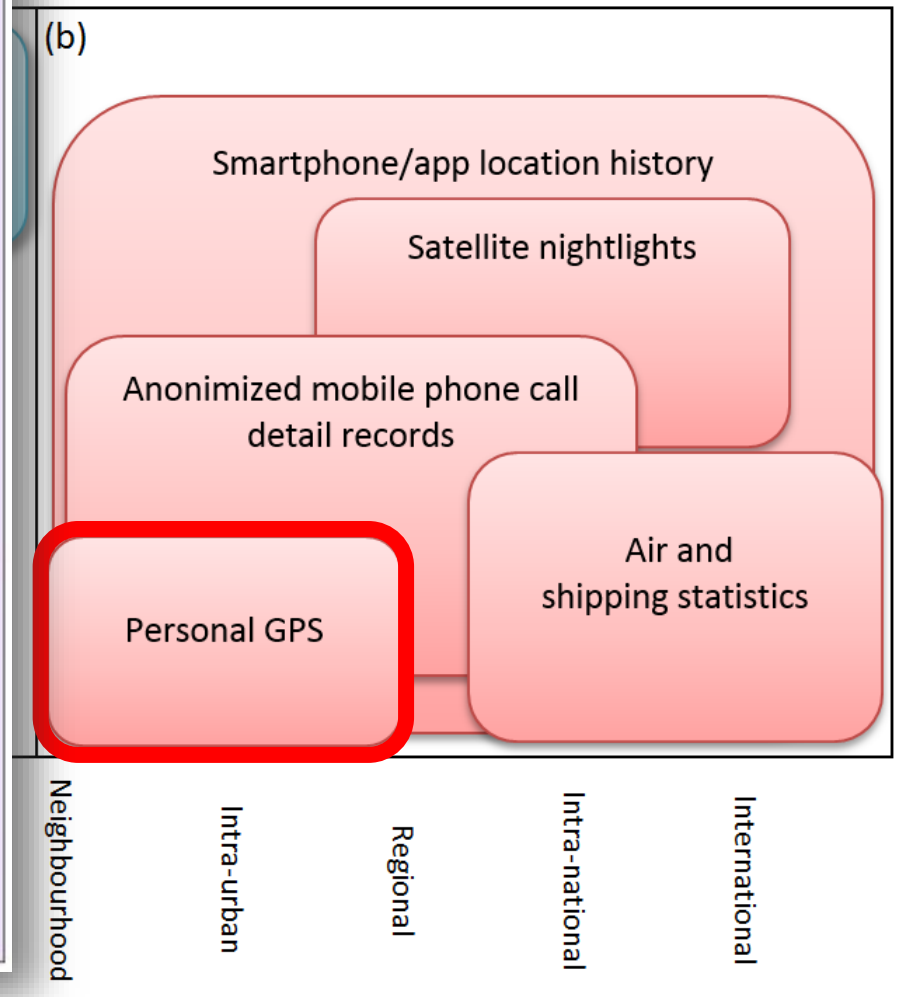


Tatem (2014) Int Health

Data for measuring population movements/migrations

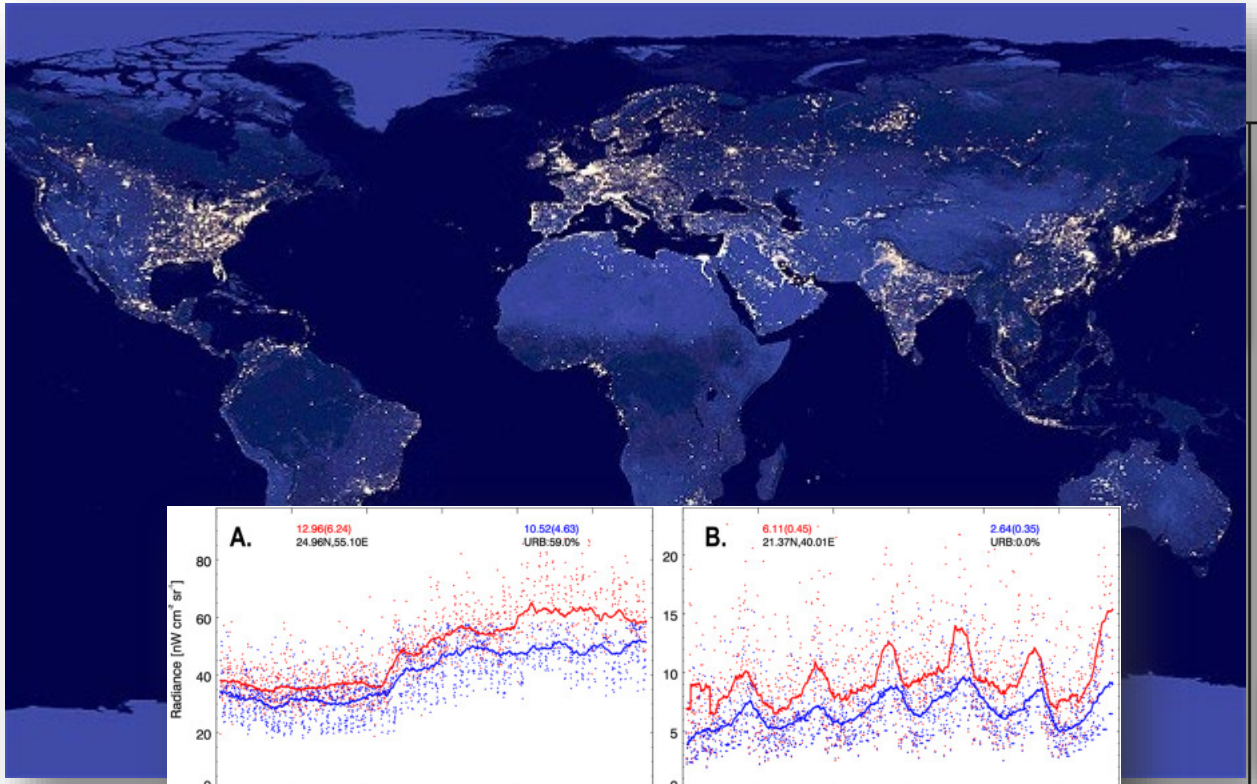


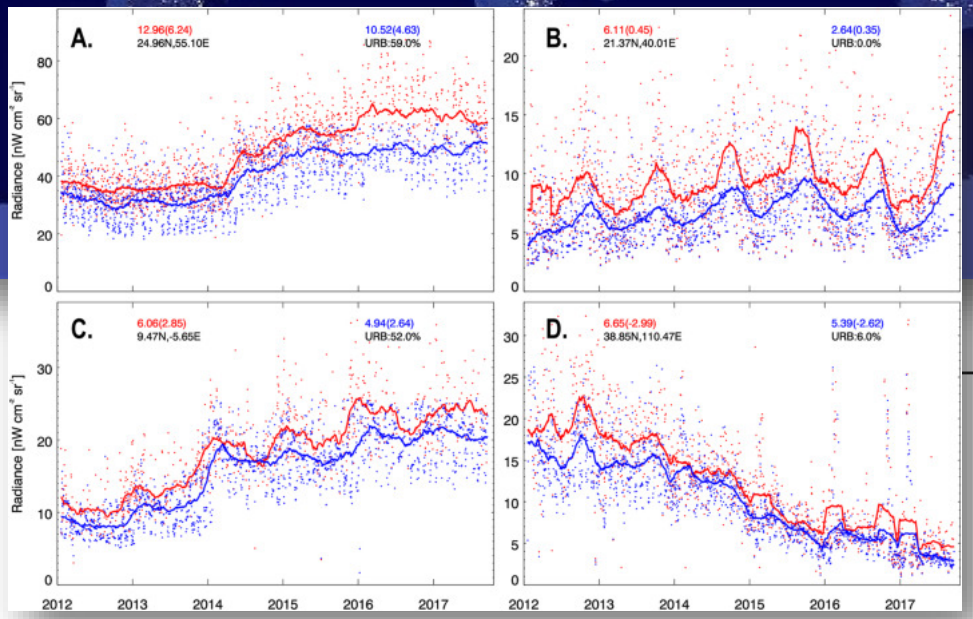
Novel

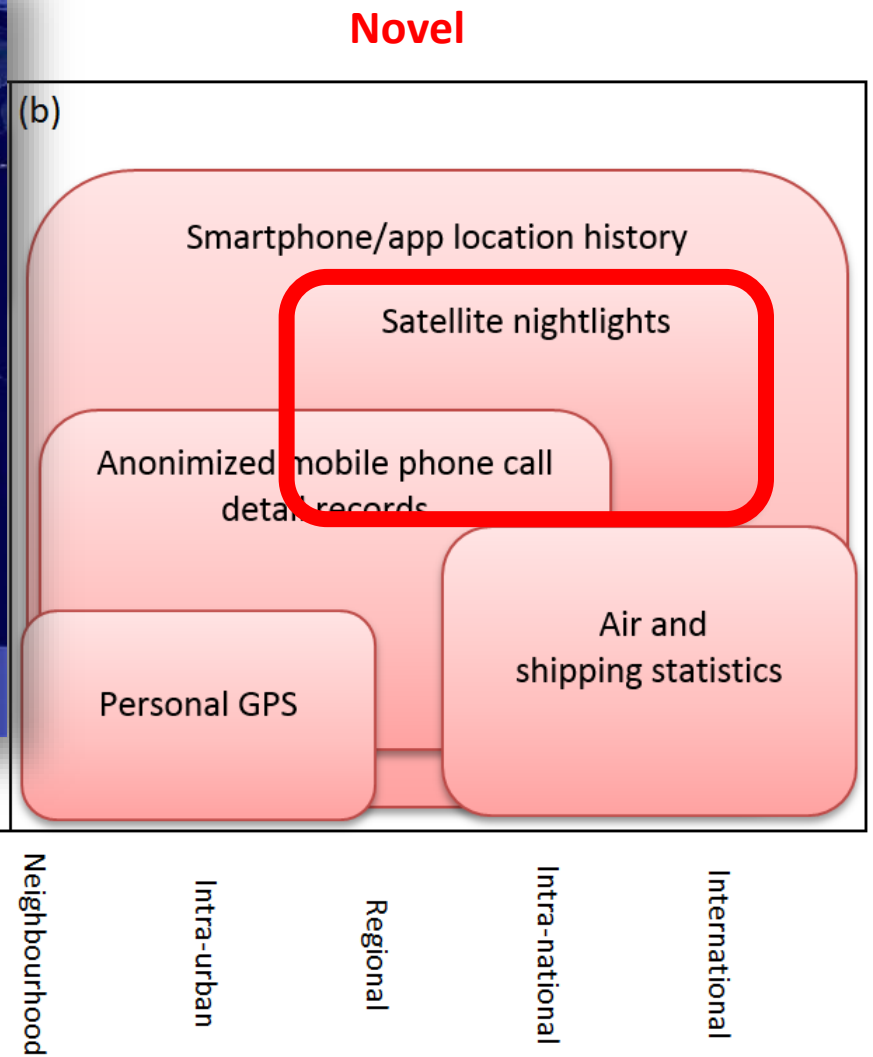


Tatem, *Int Health*, 2014; Ruktanonchai et al, *IJHG*, 2018

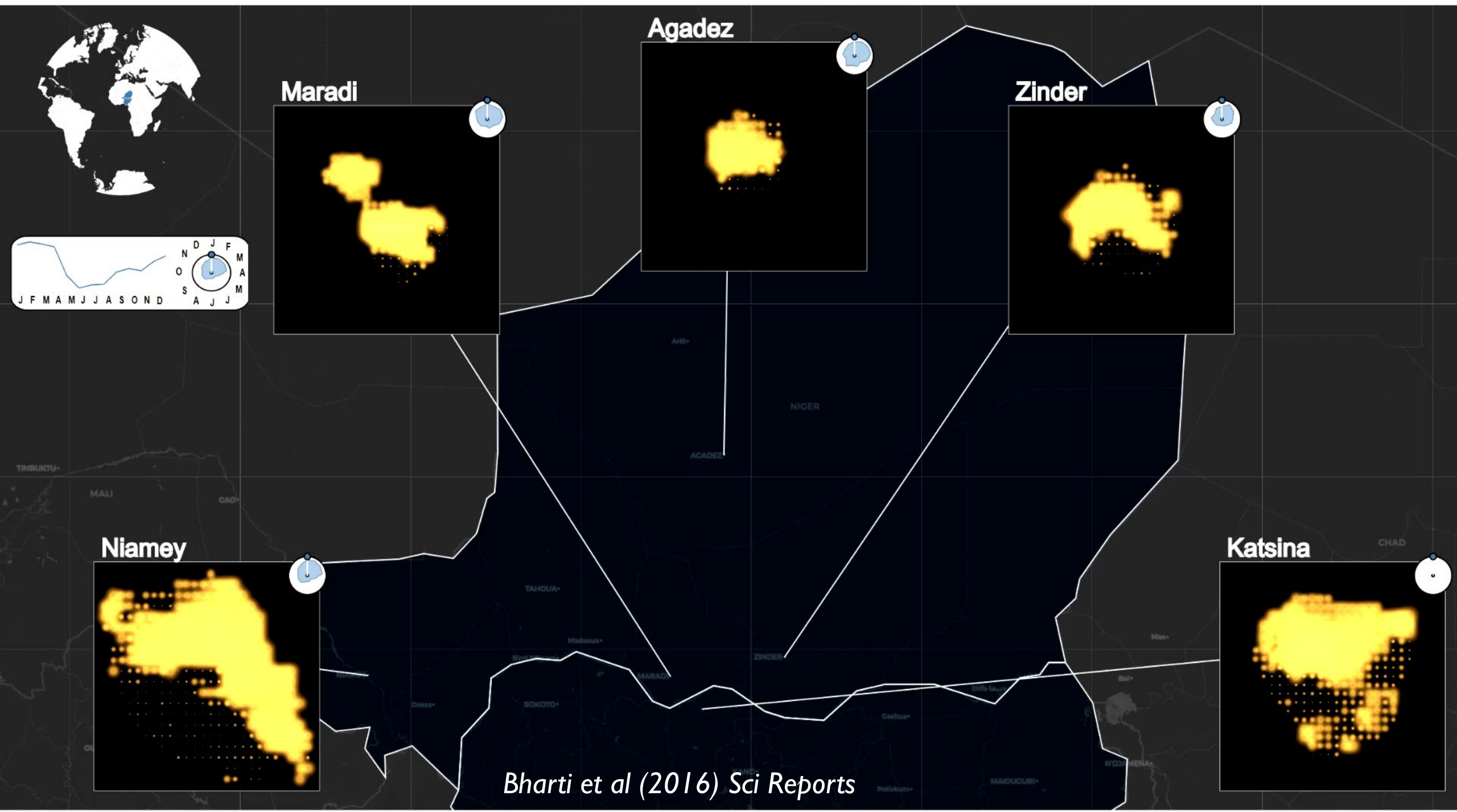
Data for measuring population movements/migrations





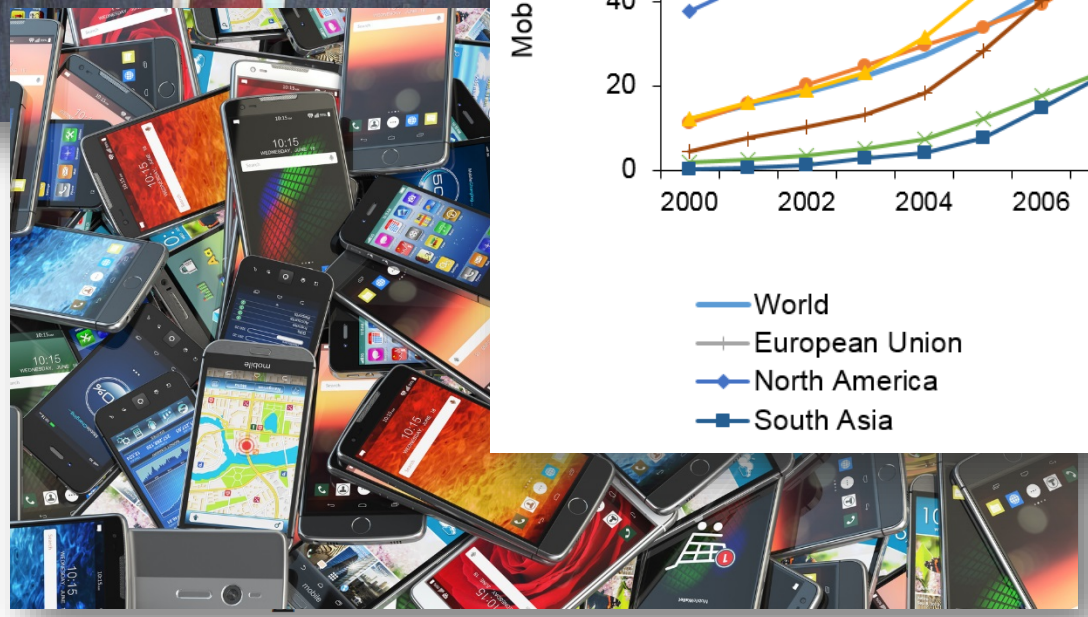


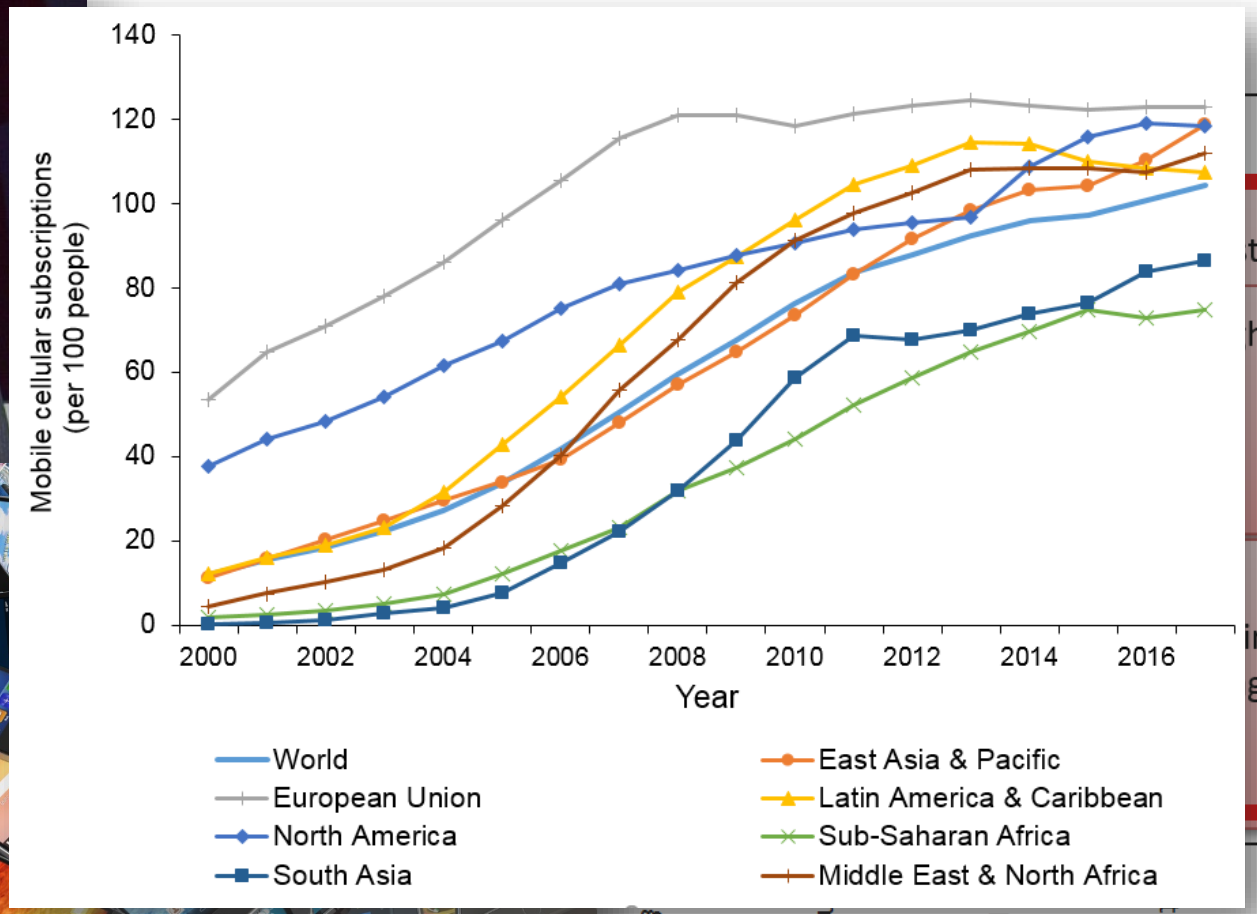
Tatem, *Int Health*, 2014; Ruktanonchai et al, *IJHG*, 2018

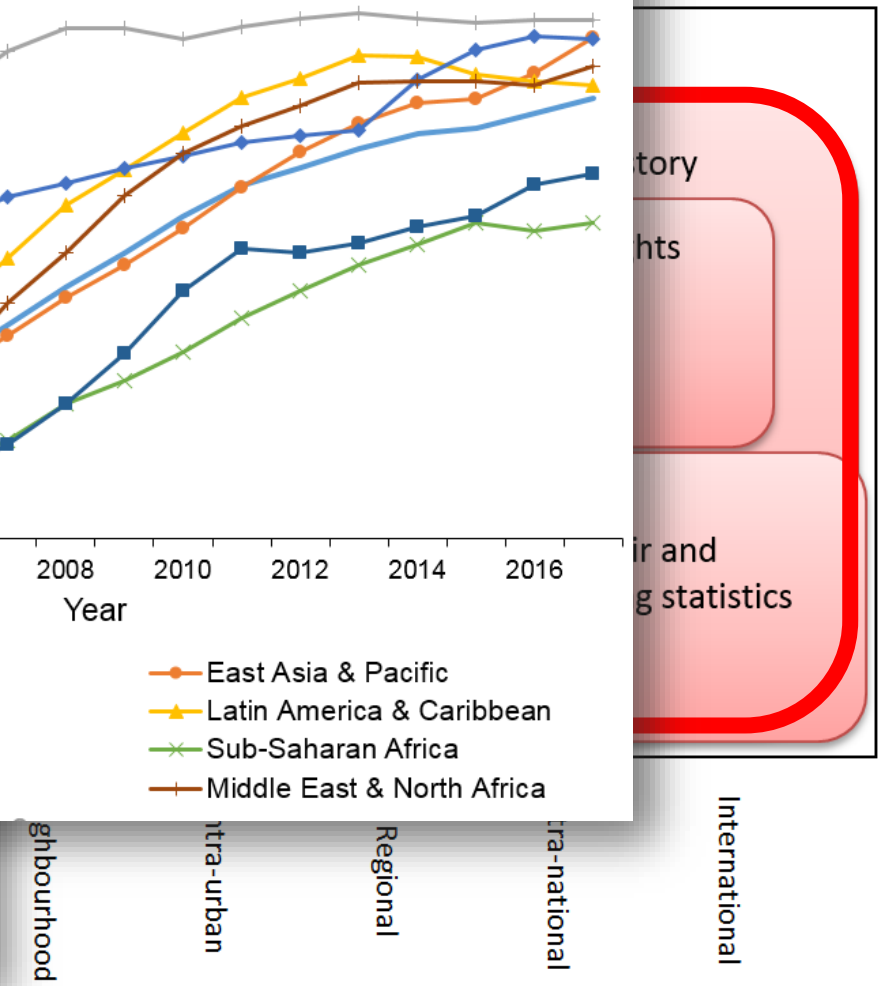


Bharti et al (2016) Sci Reports

Data for measuring population movements/migrations



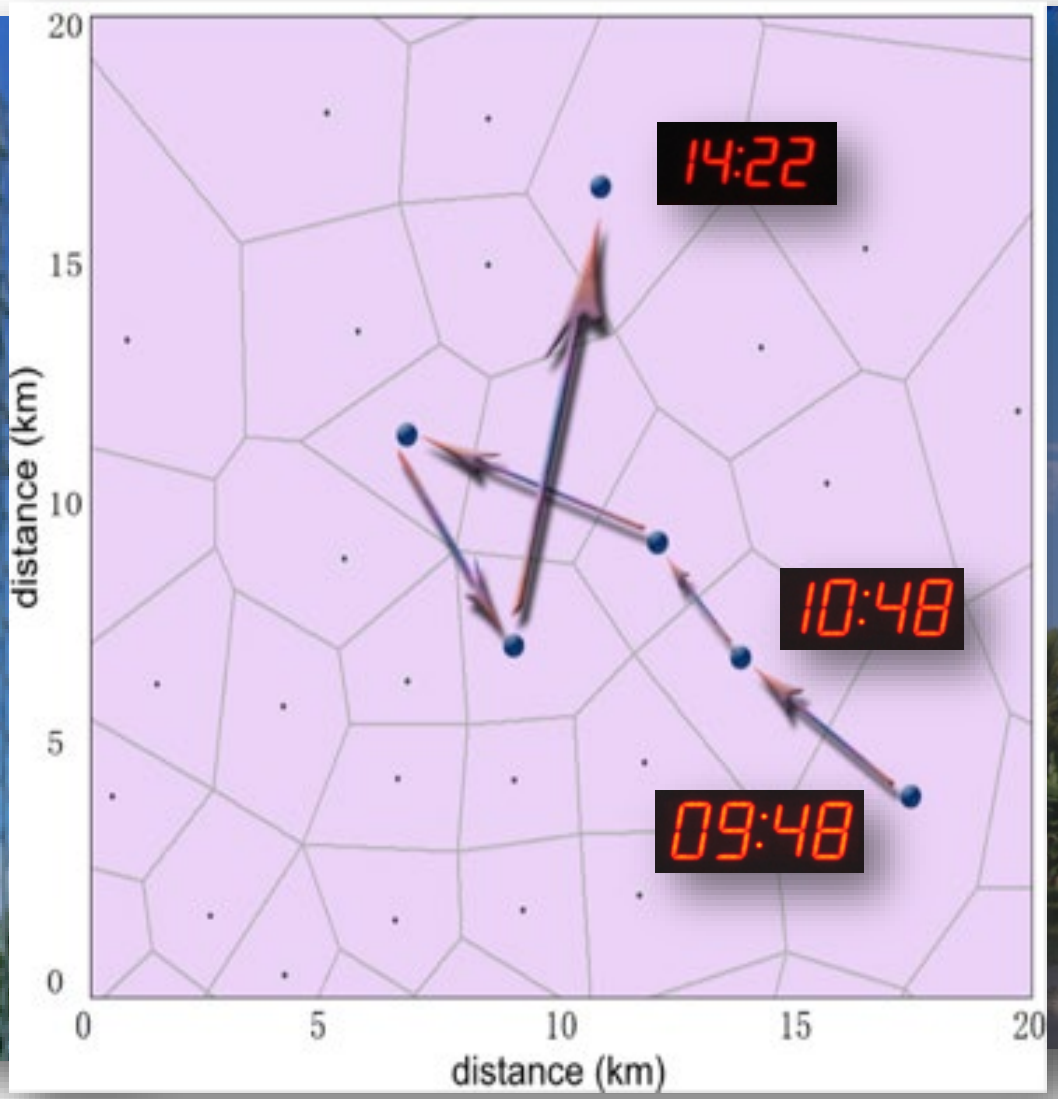




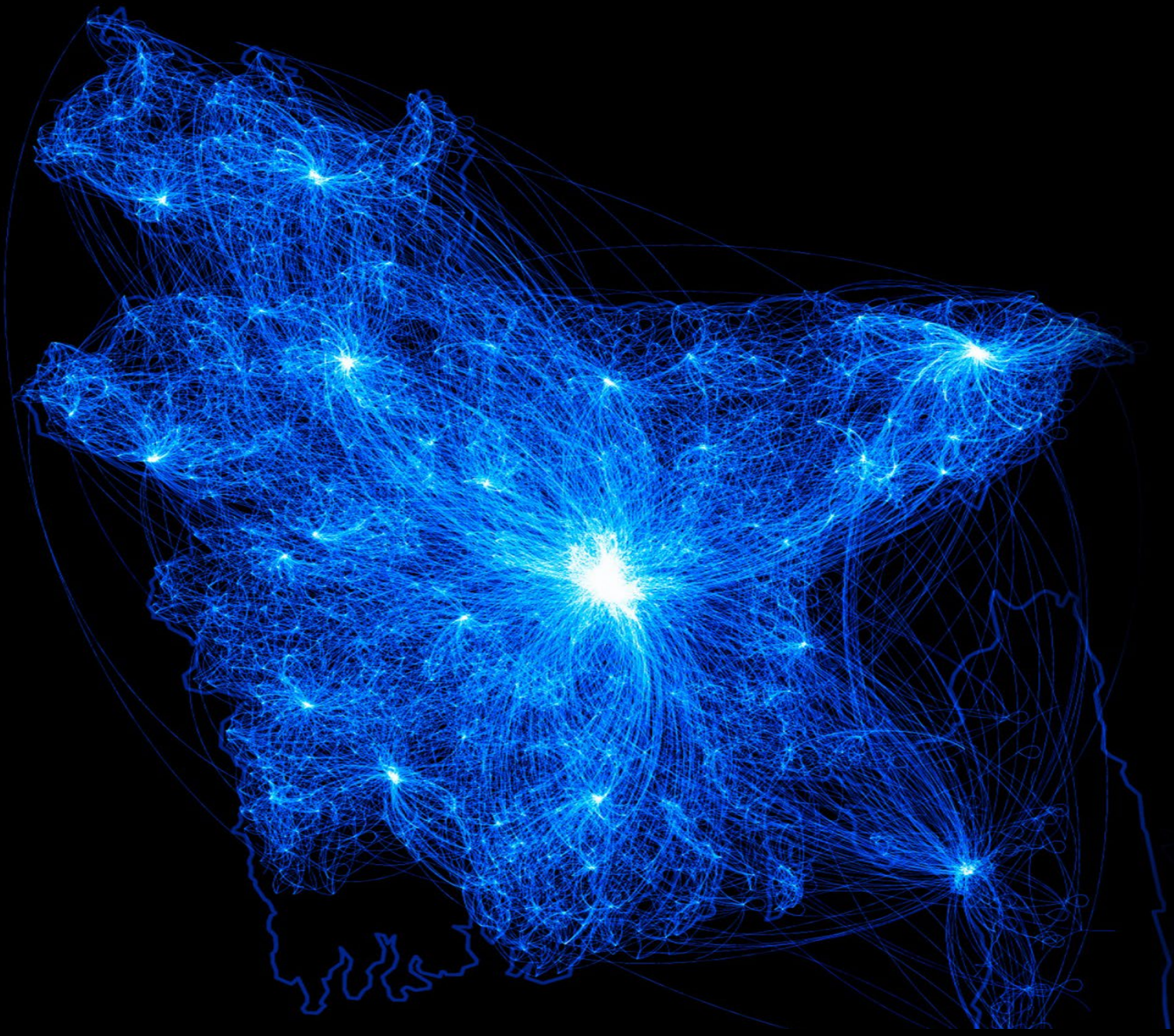
Tatem, *Int Health*, 2014; Ruktanonchai et al, *IJHG*, 2018

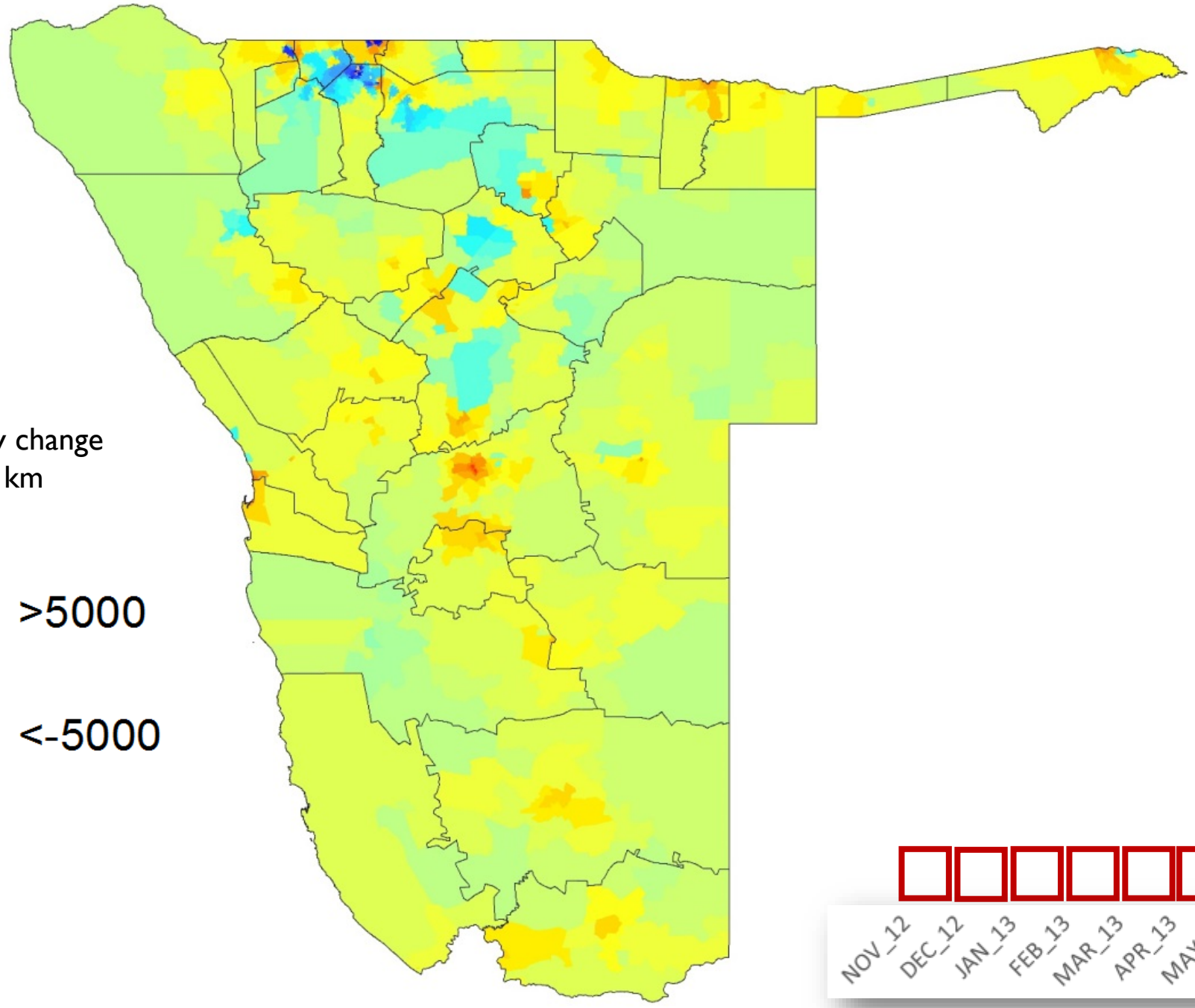
Mobile phone call detail records

09:48



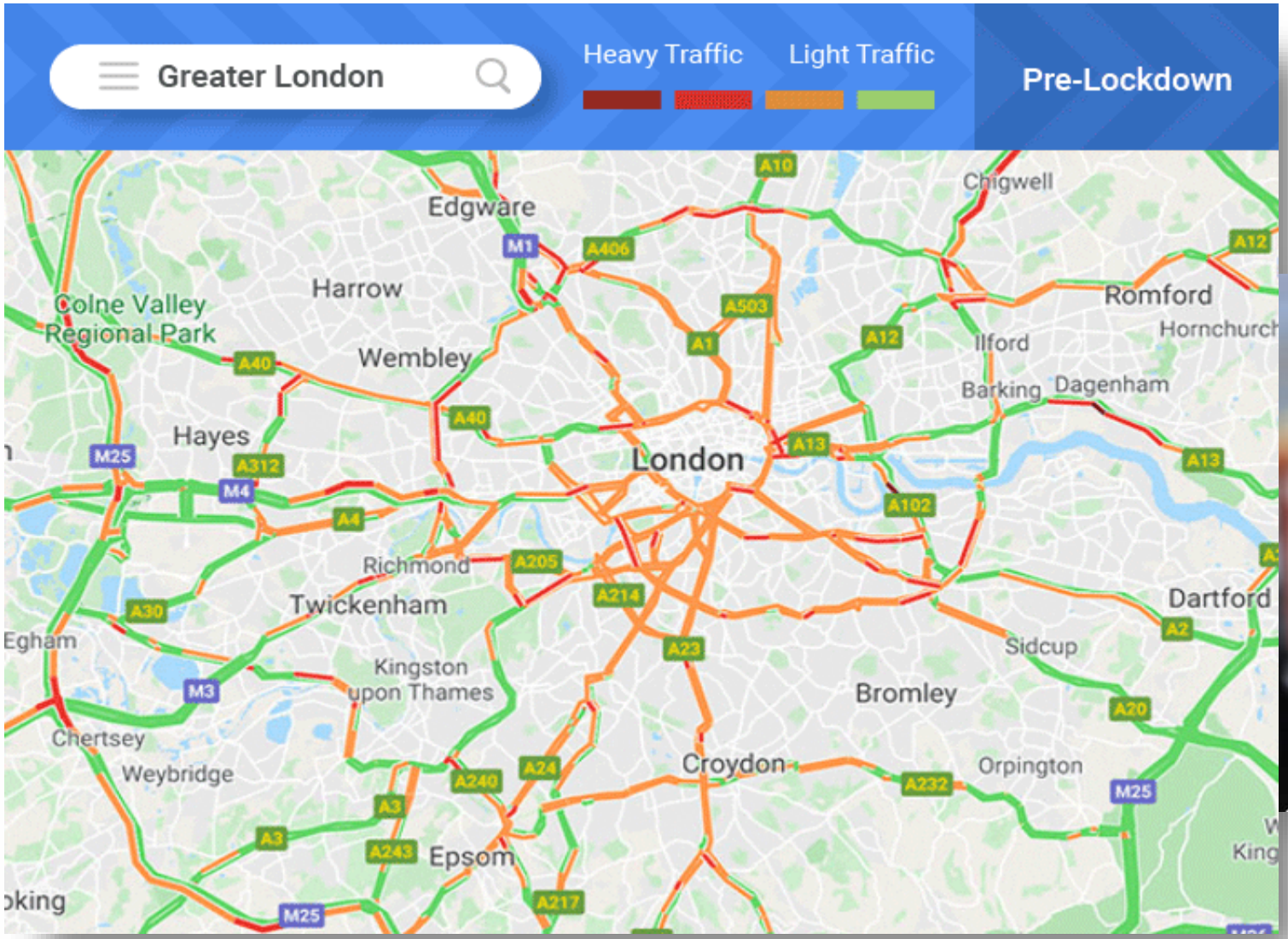
10:48

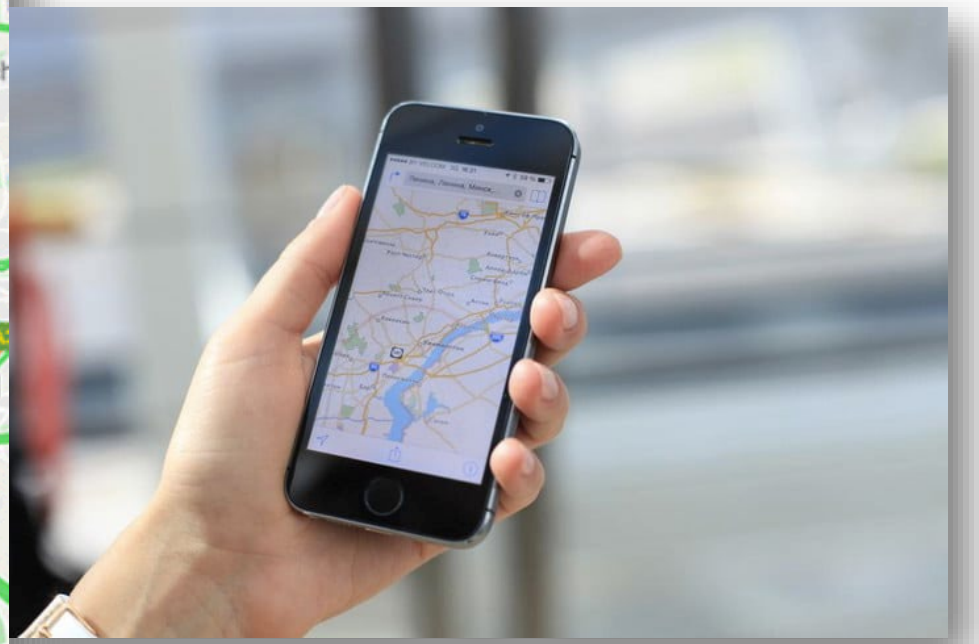



Pop density change
per square km
>5000
<-5000
NOV_12
DEC_12
JAN_13
FEB_13
MAR_13
APR_13
MAY_13
JUN_13
JUL_13
AUG_13
SEP_13

Smartphone/app location histories

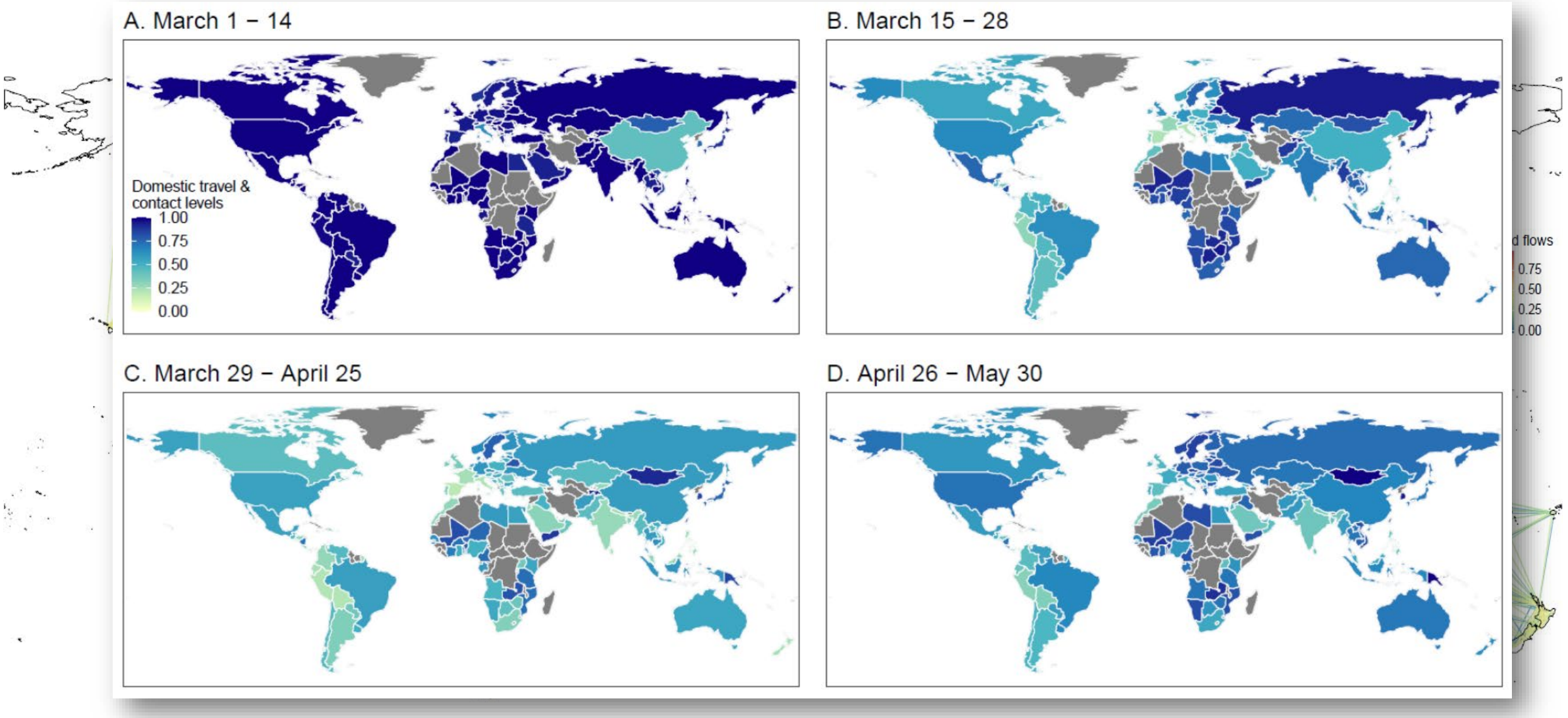
Part 2: Andy Tatem





Ruktanonchai et al (2018) IJHG

Smartphone/app location histories



Lai et al (2020) MedArXiv

Module 4: Mapping migration using alternative data sources
Part 2: Andy Tatem

https://covid19.apple.com/mobility

Maps

Mobility Trends Reports

Learn about COVID-19 mobility trends. Reports are published daily and reflect requests for directions in Apple Maps. Privacy is one of our core values, so Maps doesn't associate your data with your Apple ID, and Apple doesn't keep a history of where you've been.

https://www.google.com/covid19/mobility/

Google COVID-19 Community Mobility Reports

See how your community is moving around differently due to COVID-19

As global communities respond to COVID-19, we've heard from public health officials that the same type of aggregated, anonymized insights we use in products such as Google Maps could be helpful as they make critical decisions to combat COVID-19.

These Community Mobility Reports aim to provide insights into what has changed in response to policies aimed at combating COVID-19. The reports chart movement trends over time by geography, across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential.

FACEBOOK Data for Good

Public Datasets | Tools for Nonprofits | Impact | Approach

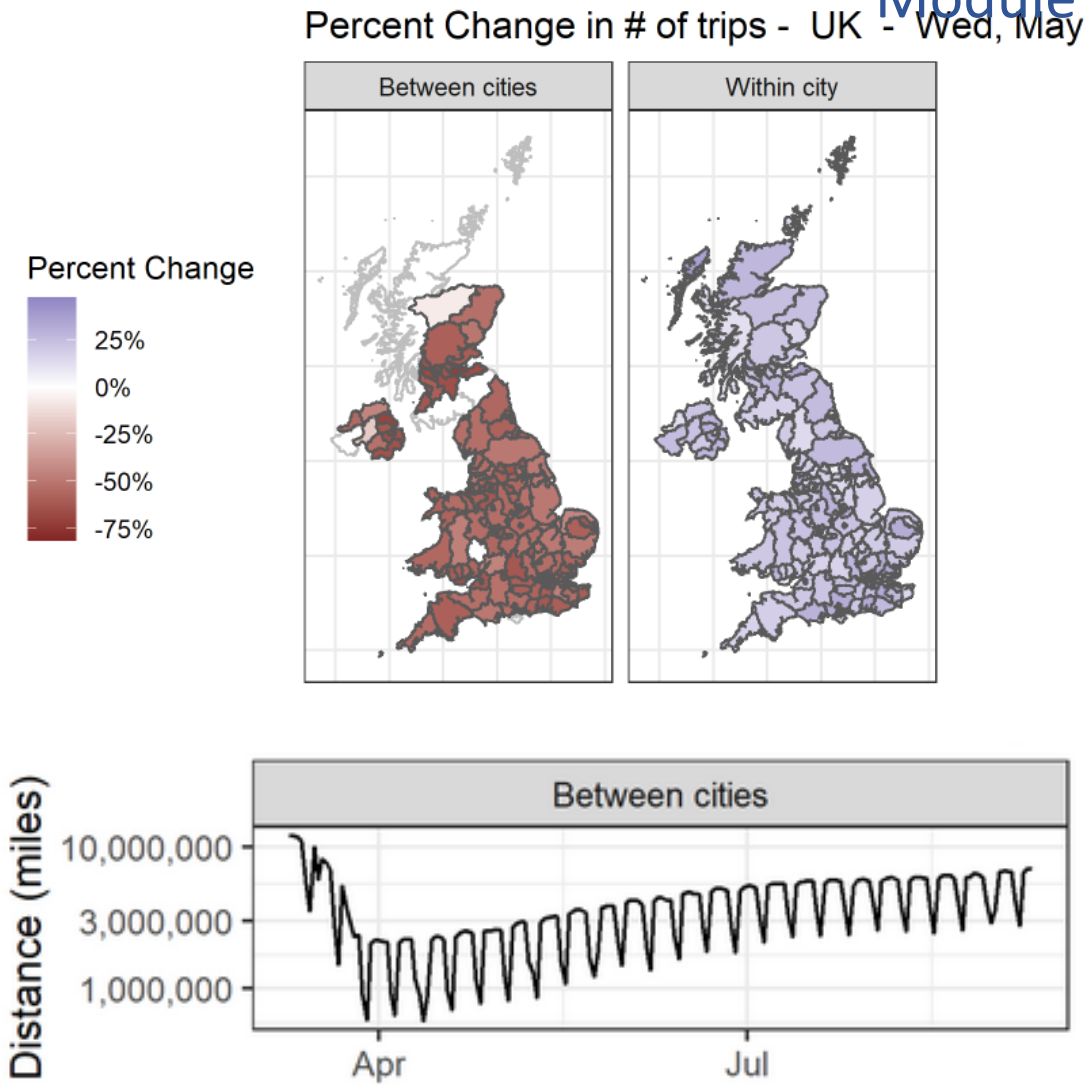
We use data to address some of the world's greatest humanitarian issues.

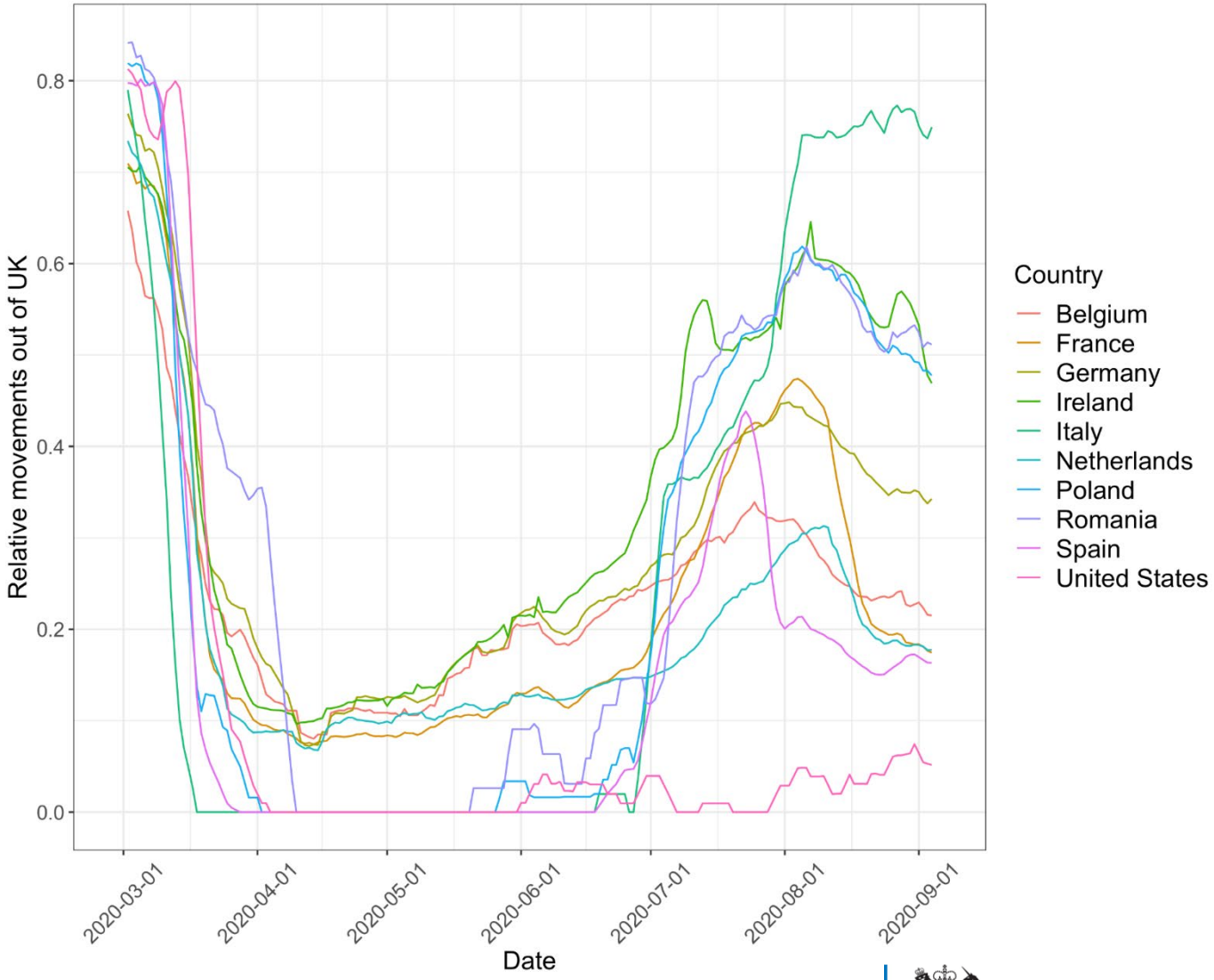
Flattening the COVID-19 curve is a challenge that takes all of us. People are distancing to protect their communities, healthcare workers are saving lives on the front lines, and public health systems are looking to put the right guidelines in place. To do that, they need better information on whether preventive measures are working and how the virus may spread. We offer maps on population movement that researchers and nonprofits are already using to understand the coronavirus crisis, using aggregated data to protect people's privacy.

https://dataforgood.fb.com/

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem





Office for National Statistics

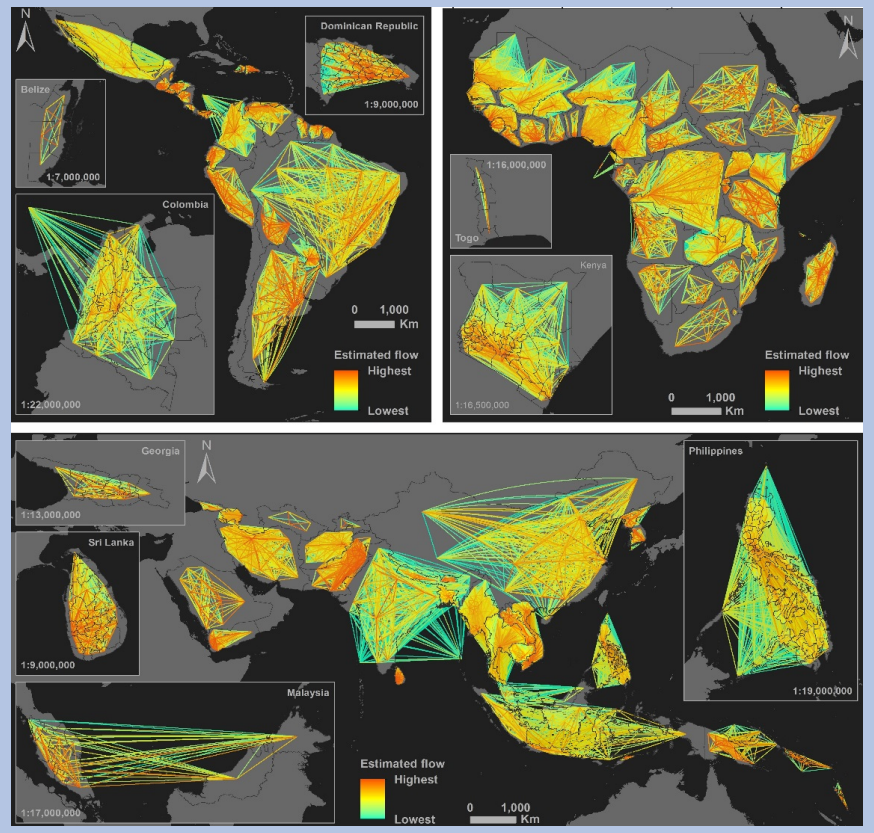
FACEBOOK Data for Good

Cabinet Office

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

GEOSPATIAL DATA INTEGRATION FOR SUBNATIONAL SEX-DISAGGREGATED MIGRATION MAPPING



www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Malaria
» Geography
» Public health
» Environmental impact

Data Descriptor: Mapping internal connectivity through human migration in malaria endemic countries

Alessandro Sorichetta[1,2,3], Tom J. Bird[1,2], Nick W. Ruktanonchai[1,2], Elisabeth zu Erbach-Schoenberg[1,2], Carla Pezzulo[1,2], Natalia Tejedor[1,2,4], Ian C. Waldock[4], Jason D. Sadler[4], Andres J. Garcia[5], Luigi Sedda[6] & Andrew J. Tatem[1,2]

MIGRATION STUDIES 2014 • 1–22 1 of 22

Modeling internal migration flows in sub-Saharan Africa using census microdata

Andres J. Garcia[†,‡,*], Deepa K. Pindolia[†,‡,§], Kenneth K. Lopiano[**,††] and Andrew J. Tatem[‡‡,§§,***]

Quantifying Global International Migration Flows

Guy J. Abel* and Nikola Sander*†

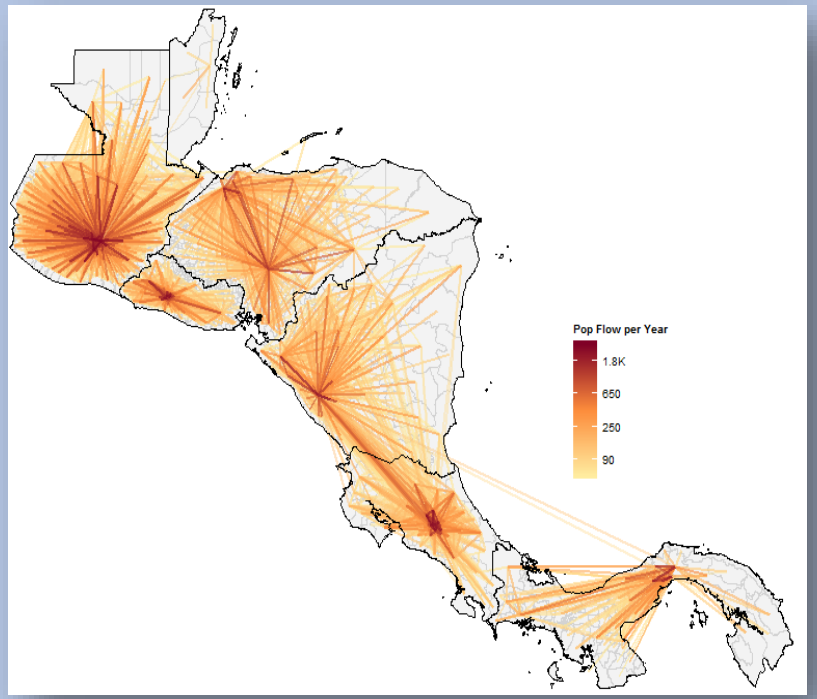
Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

BACKGROUND AND AIMS

- No consistent subnational data on migration flows across low/middle income settings exists
- Utilize census microdata and geospatial covariates to build regional models and fill data gaps
- Understand relative importance of some key migration drivers



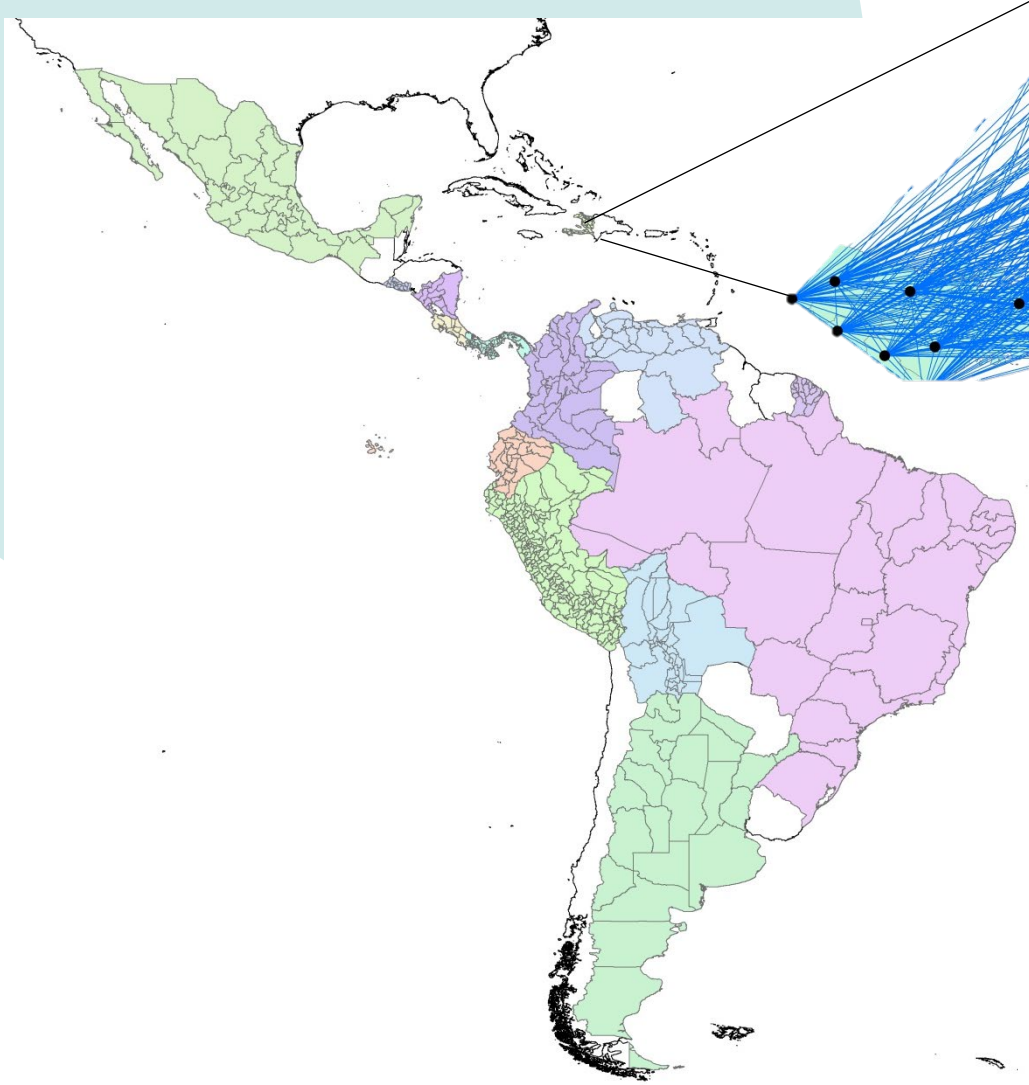
Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

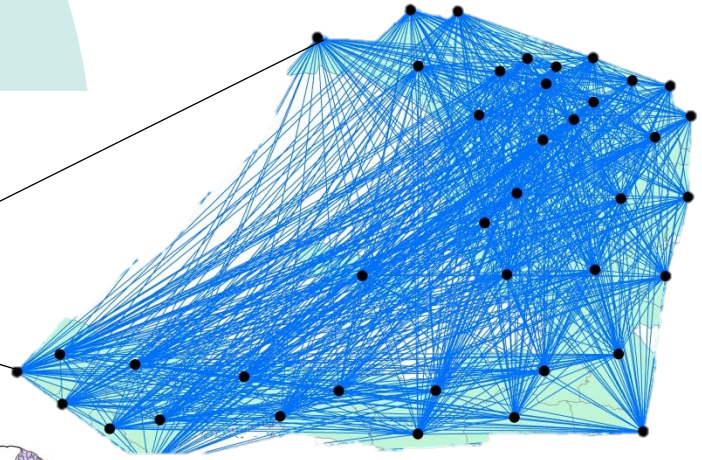
IOM UN MIGRATION | ONLINE AFRICAN SCHOOL *on Migration Statistics*

Module 4: Mapping migration using alternative data sources

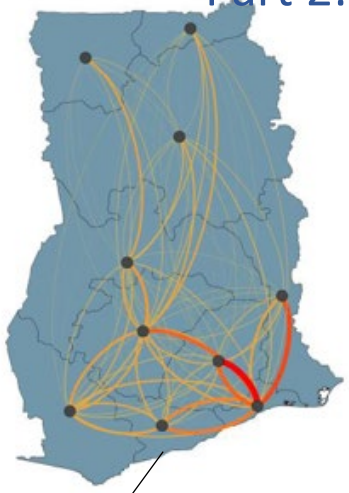
Part 2: Andy Tatem

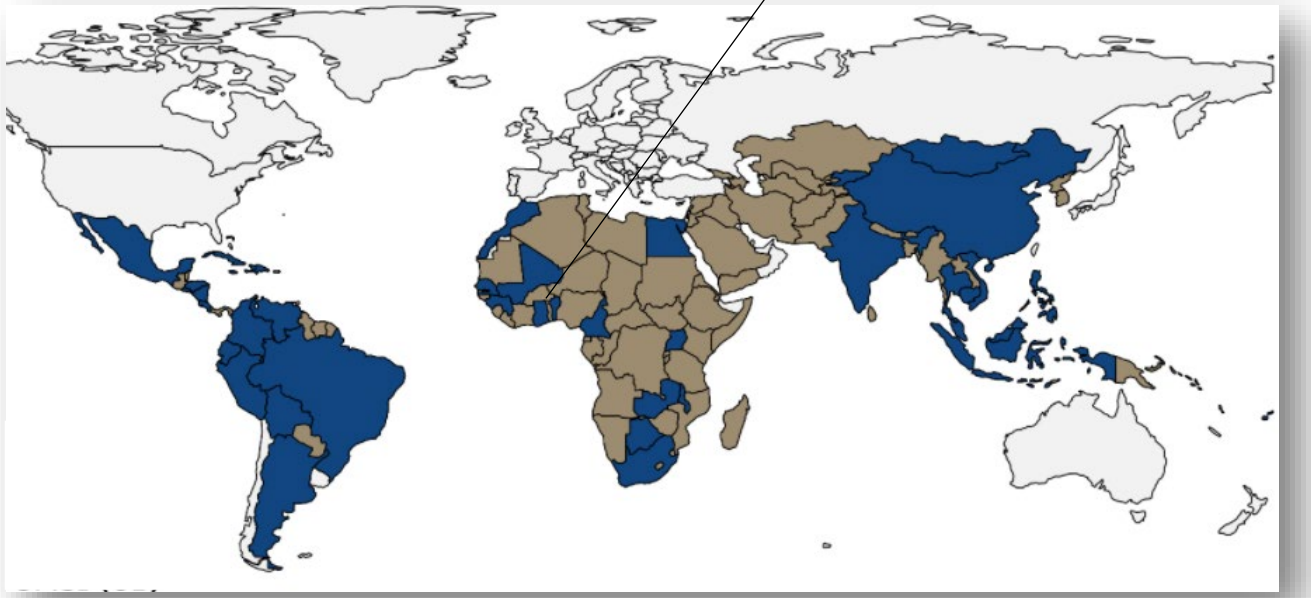
Assembling subnational migration data





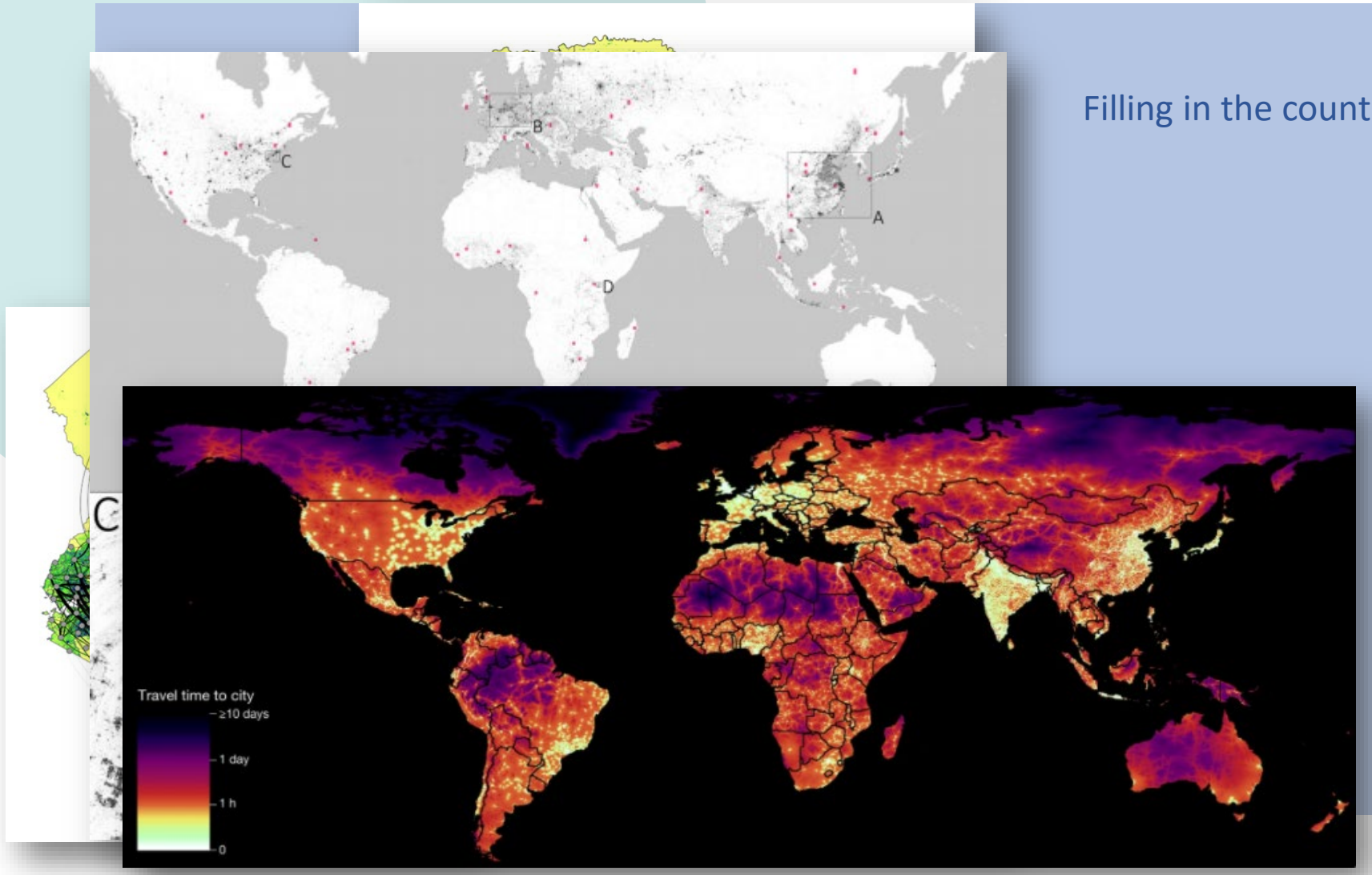
Haiti 2013





Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem



Filling in the country gaps

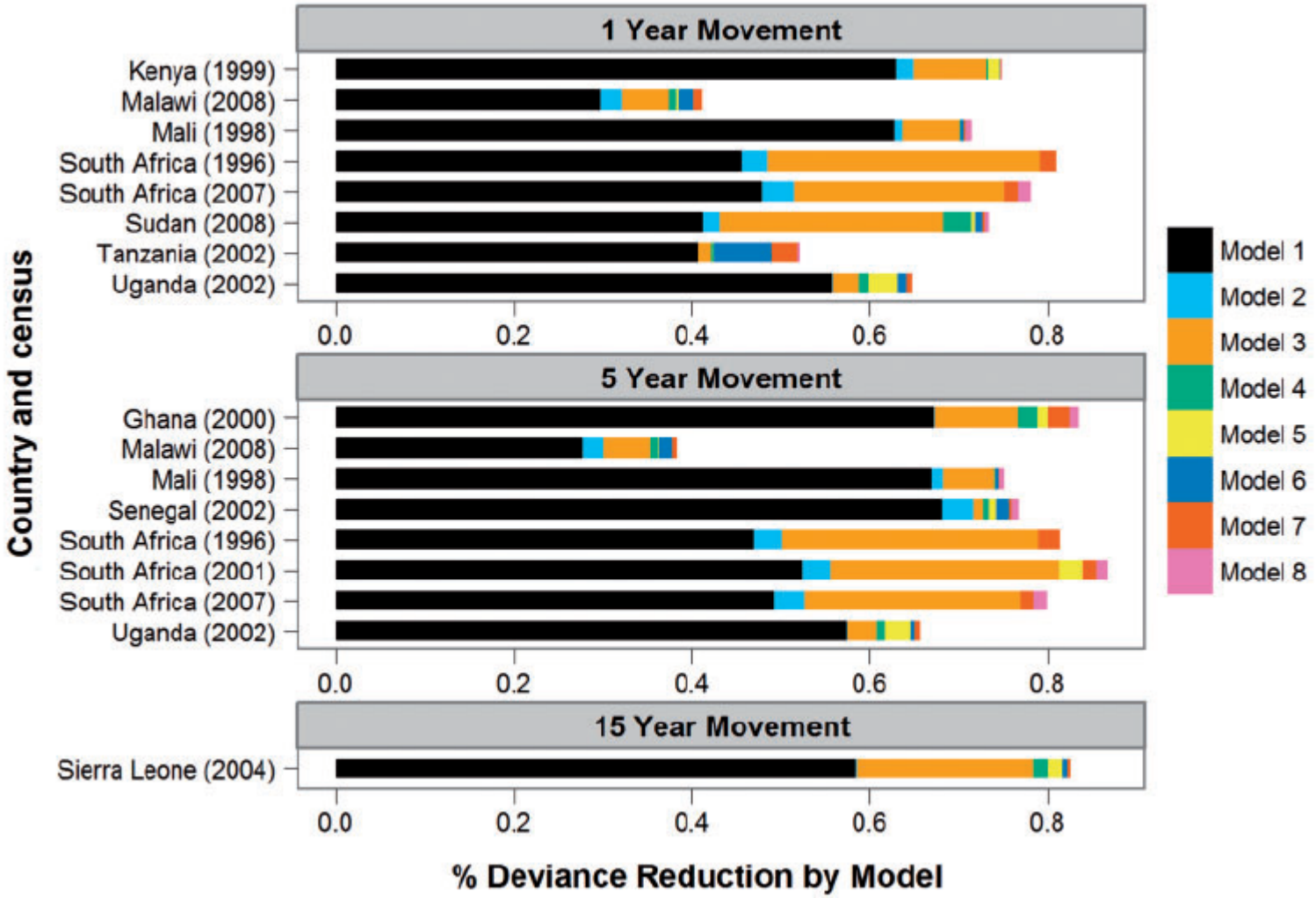
-Proportion urban
-Proportion economically active
-Proportion male
-Drought/conflict indicators
-Economic output
-Accessibility
-Ethnicity

Garcia et al (2014) Mig Studies; Sorichetta et al (2016) Nature Sci Data

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Model	Model description
Model 1	ln(POPI) + ln(POPJ) + ln(DISTIJ)
Model 2	Model 1 + CONTIG
Model 3	Model 2 + URBANPROPI + URBANPROPJ
Model 4	Model 3 + ACTIVEI + ACTIVE J
Model 5	Model 4 + MALEPROPI + MALEPROPJ
Model 6	Model 5 + MEDAGEI + MEDAGEJ
Model 7	Model 6 + DROUGHTI + DROUGHTJ
Model 8	Model 7 + RAINVARI + RAINVARJ

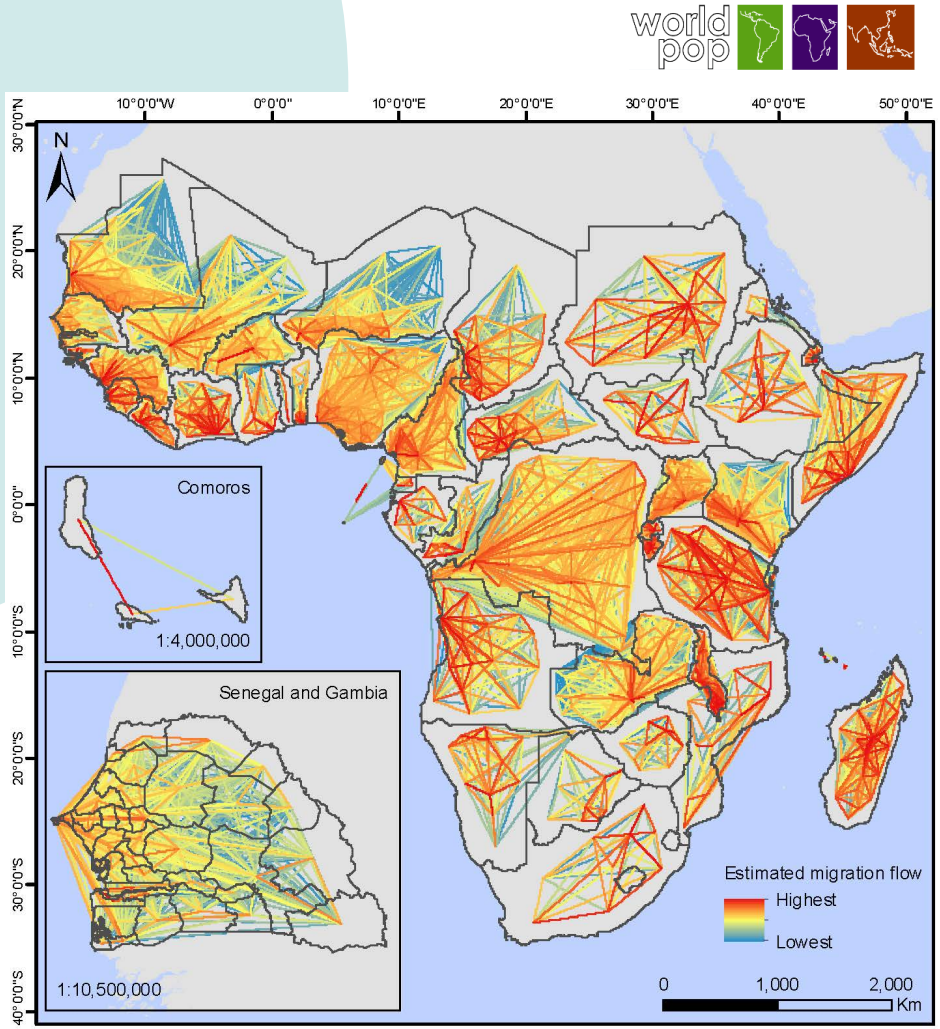


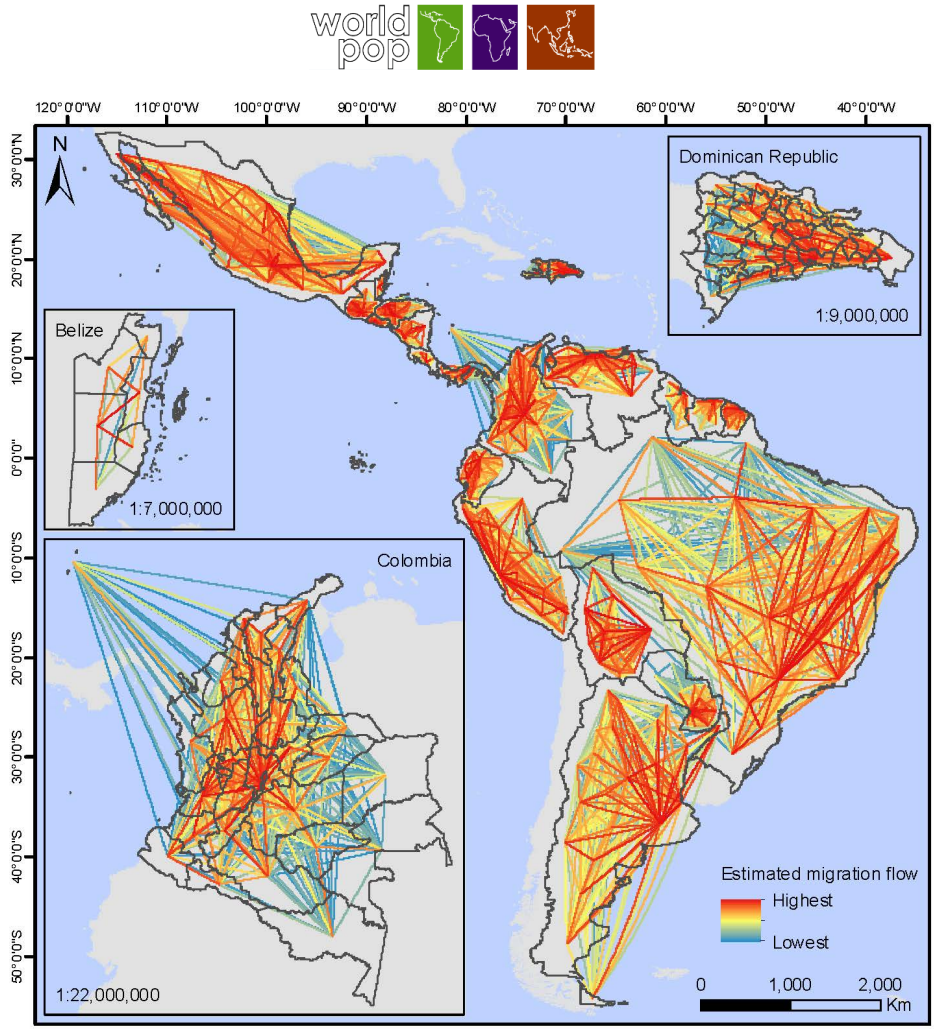
Garcia et al (2014) Mig Studies; Sorichetta et al (2016) Nature Sci Data

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Internal Migration Estimates



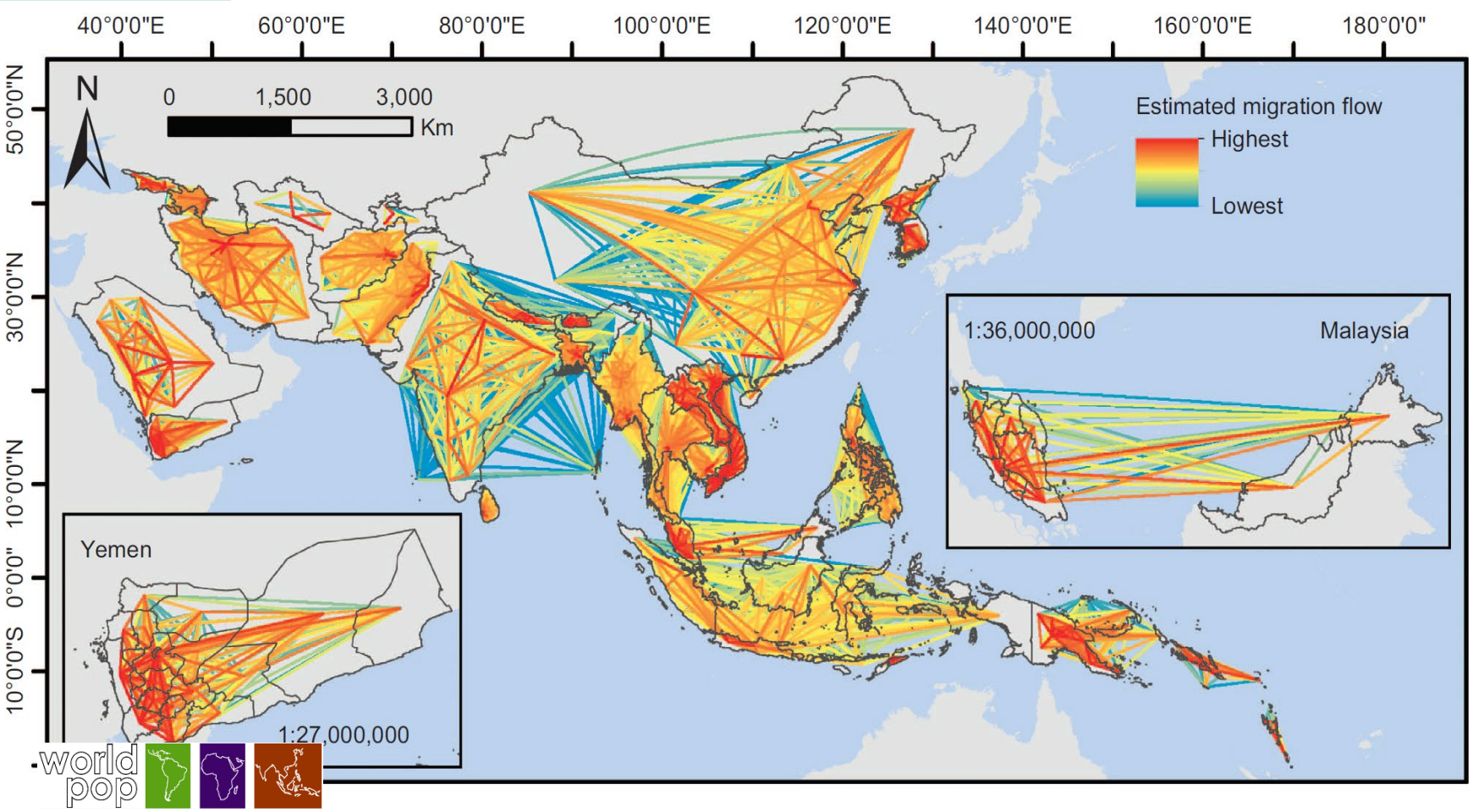


Sorichetta et. al. (2016)

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Internal Migration Estimates

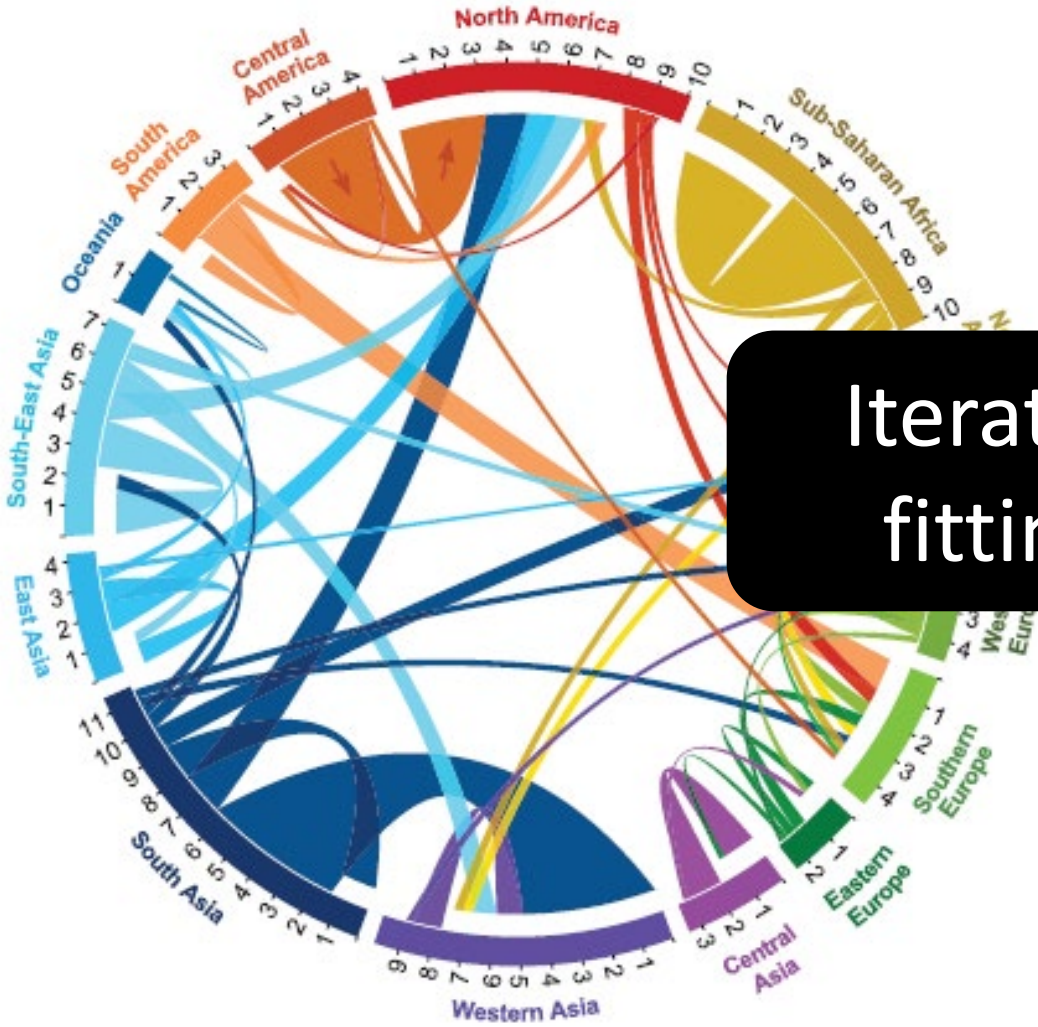


Sorichetta et. al. (2016)

Module 4: Mapping migration using alternative data sources

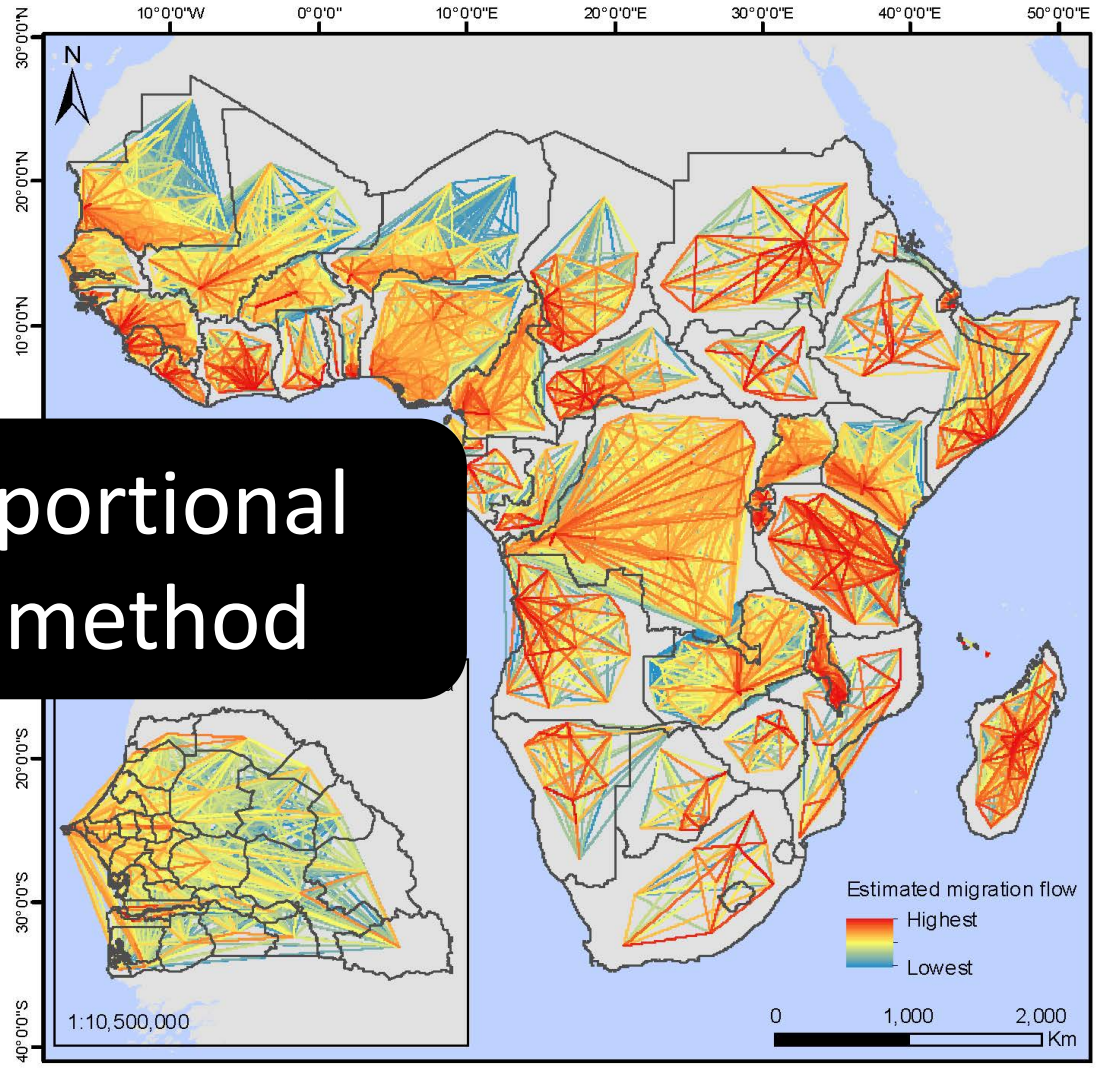
Part 2: Andy Tatem

Integrating internal and international migration data

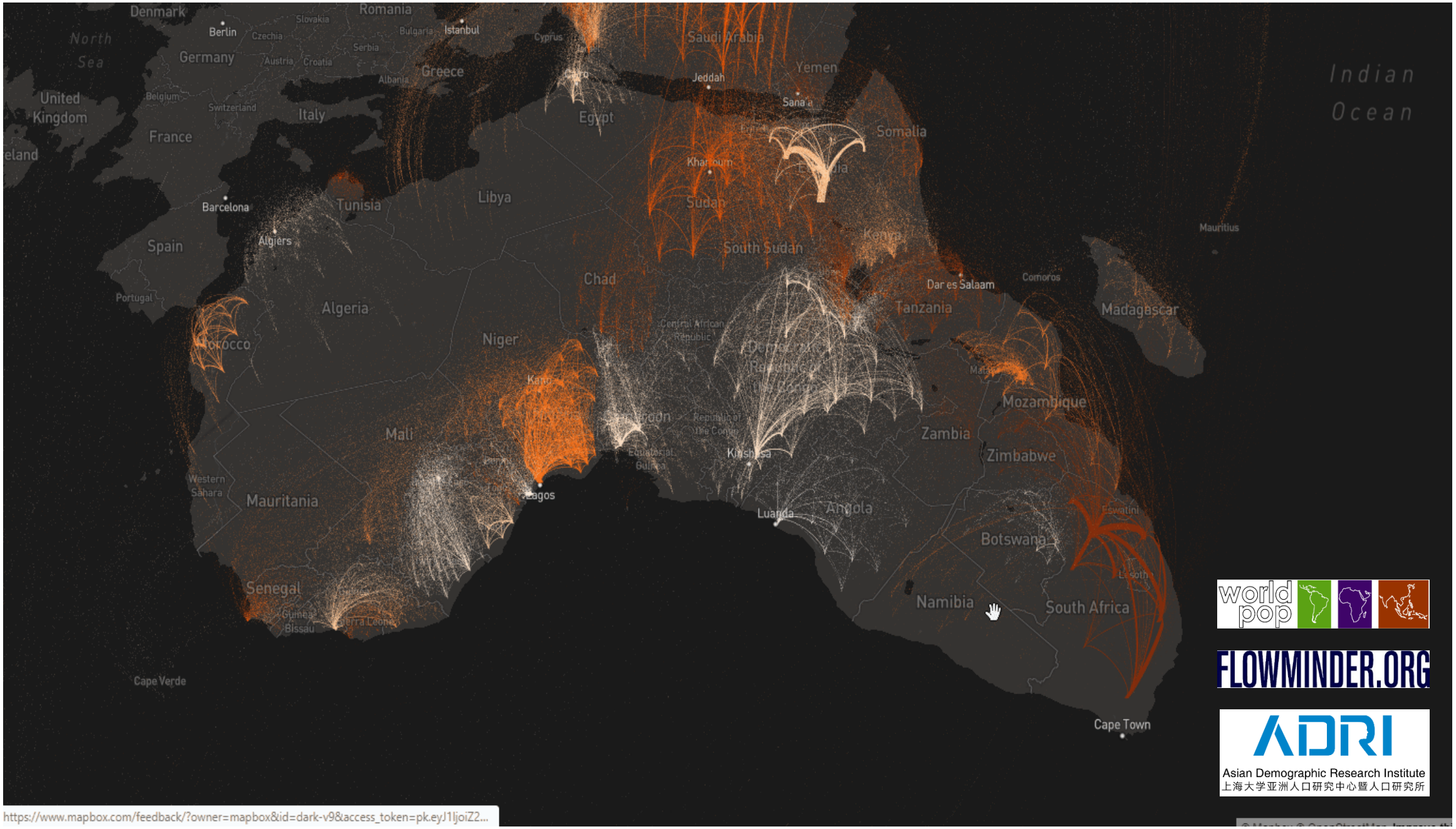


Abel & Sander 2014 (Science)

Iterative proportional fitting (IPF) method



Sorichetta et al 2016 (Nature Sci Data)



worldpop

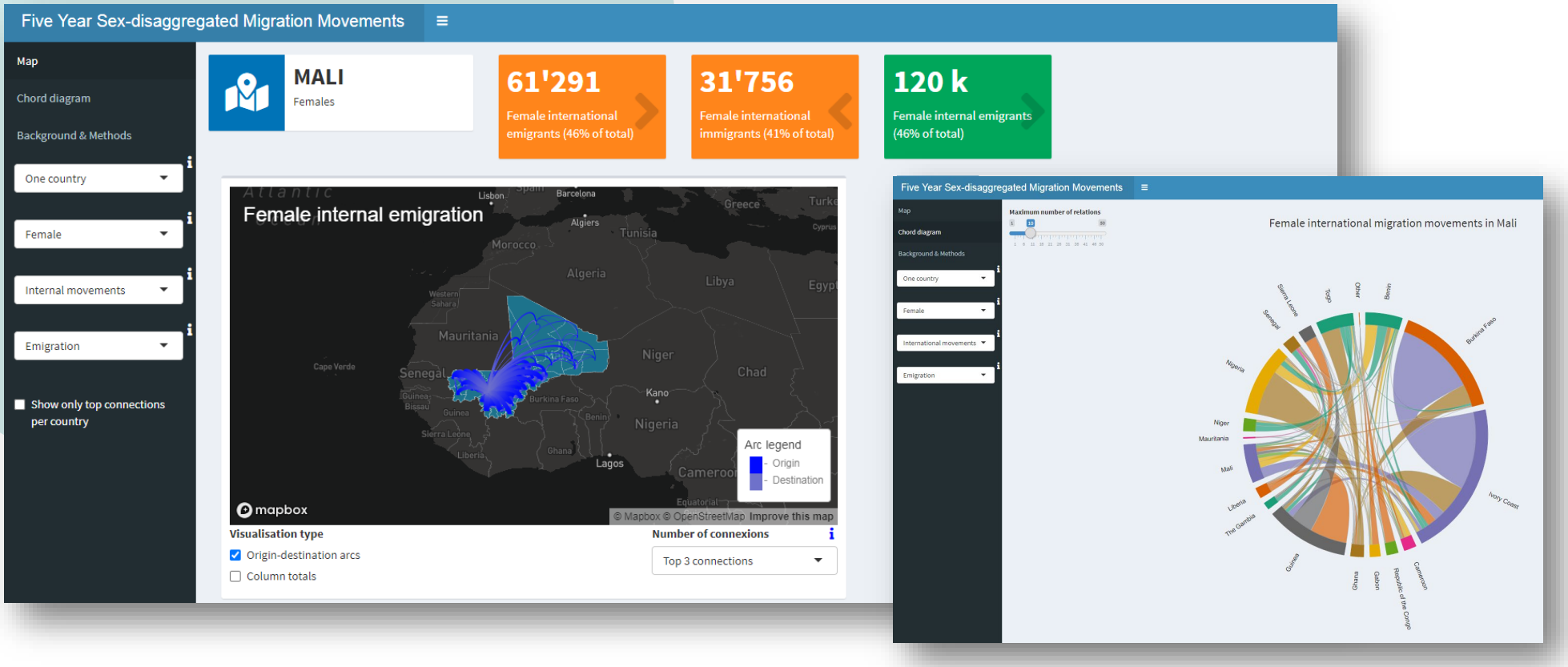
FLOWMINDER.ORG

ADRI
Asian Demographic Research Institute
上海大学亚洲人口研究中心暨人口研究所

IOM UN MIGRATION | ONLINE AFRICAN SCHOOL on Migration Statistics

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem



https://gravier.shinyapps.io/fdfa_01_dev_v4/

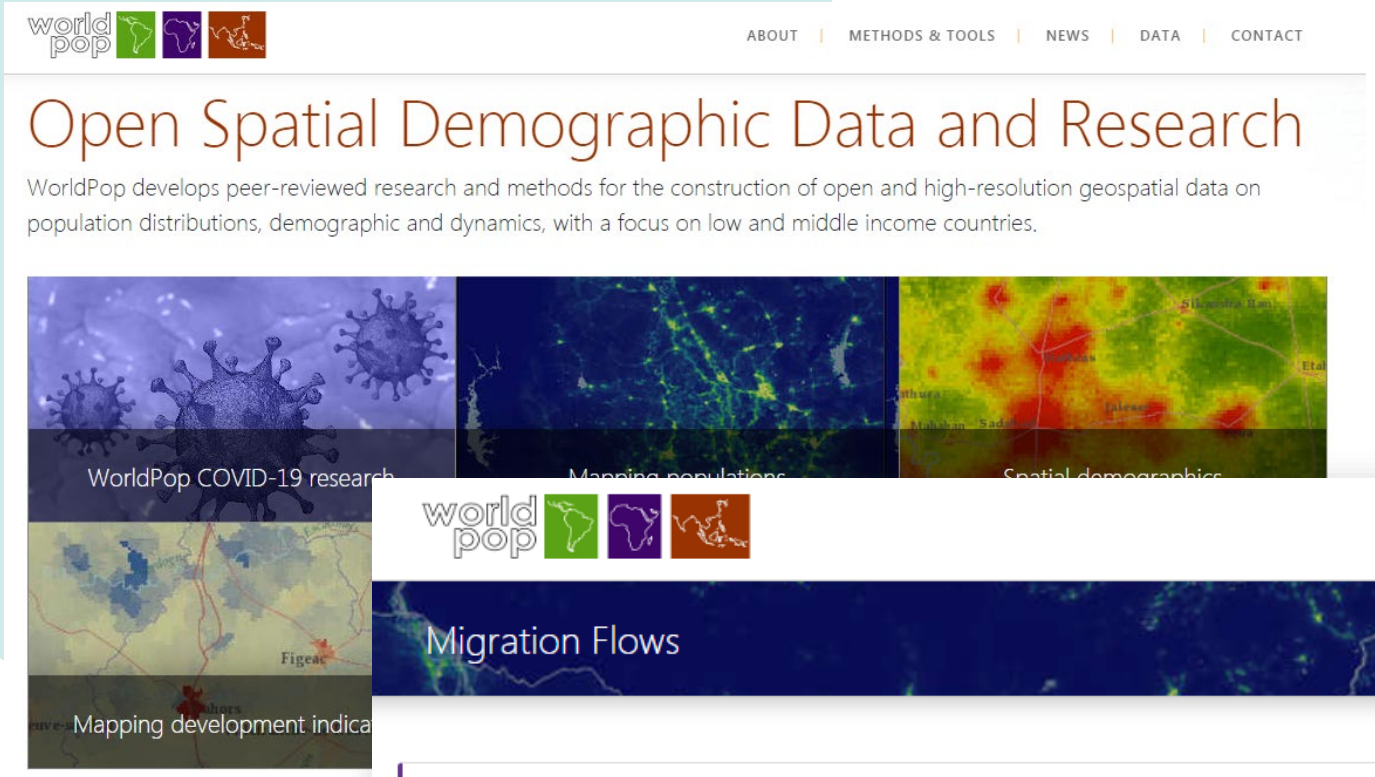
FLOWMINDER.ORG

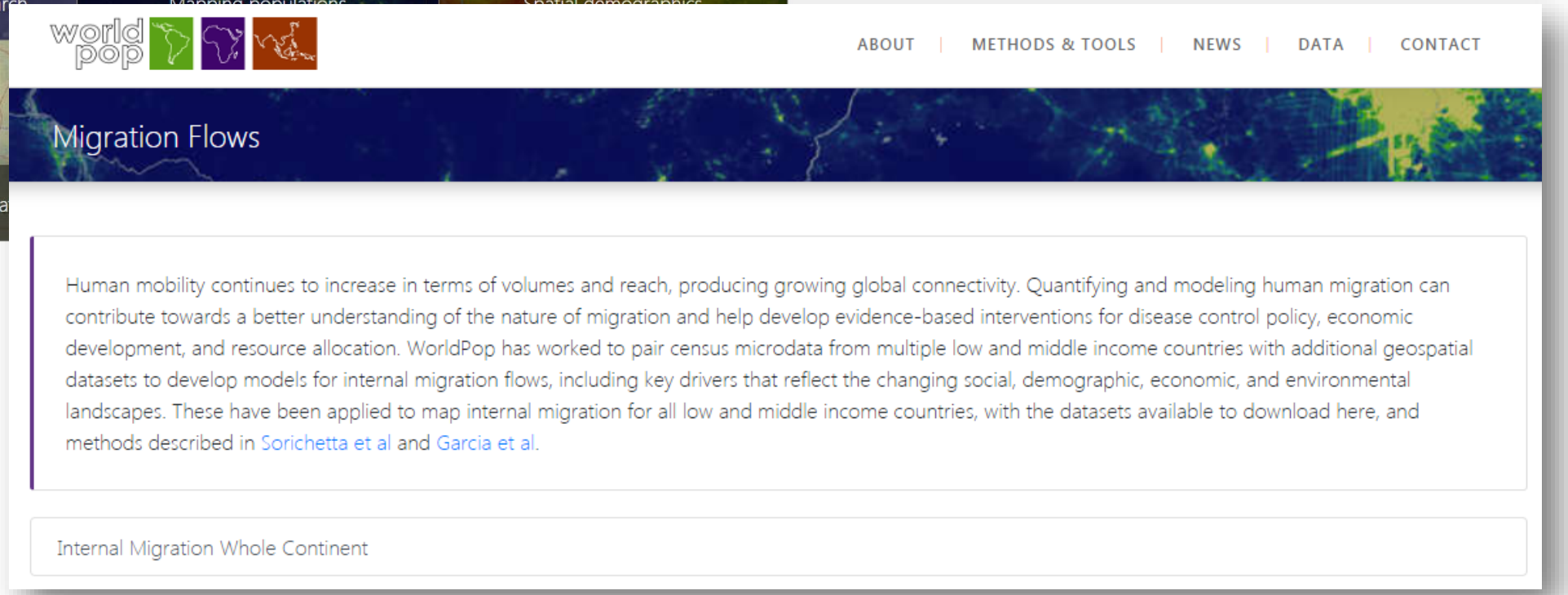
ADRI

Asian Demographic Research Institute
上海大学亚洲人口研究中心暨人口研究所

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem





https://www.worldpop.org/project/categories?id=11

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

MOBILE PHONE DATA FOR MIGRATION STATISTICS

palgrave communications
HUMANITIES | SOCIAL SCIENCES | BUSINESS

ARTICLE

https://doi.org/10.1057/s41599-019-0242-9 OPEN

Exploring the use of mobile phone data for national migration statistics

Shengjie Lai[1,2,3], Elisabeth zu Erbach-Schoenberg[1,2], Carla Pezzulo[1], Nick W. Ruktanonchai[1,2], Alessandro Sorichetta[1,2], Jessica Steele[1], Tracey Li[2], Claire A. Dooley[1,2] & Andrew J. Tatem[1,2]

PNAS

Dynamic population mapping using mobile phone data

Pierre Deville[a,b,c,1], Catherine Linard[c,d,1,2], Samuel Martin[e], Marius Gilbert[c,d], Forrest R. Stevens[f], Andrea E. Gaughan[f], Vincent D. Blondel[a], and Andrew J. Tatem[g,h,i]

[a]Department of Applied Mathematics, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium; [b]Center for Complex Network Research and Physics Department, Northeastern University, Boston, MA 02115; [c]Fonds National de la Recherche Scientifique, B-1000 Brussels, Belgium; [d]Biological Control and Spatial Ecology, Université Libre de Bruxelles, B-1050 Brussels, Belgium; [e]Université de Lorraine CNRS, Centre de Recherche en Automatique de Nancy, UMR 7039, 54518 Vandoeuvre-lès-Nancy, France [f]Department of Geography and Geosciences, University of Louisville, Louisville, KY 40292; [g]Department of Geography and Environment, University of Southampton, Southampton SO17 1BJ, United Kingdom; [h]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892; and [i]Flowminder Foundation, 17177 Stockholm, Sweden

Edited by Michael F. Goodchild, University of California, Santa Barbara, CA, and approved September 15, 2014 (received for review May 8, 2014)

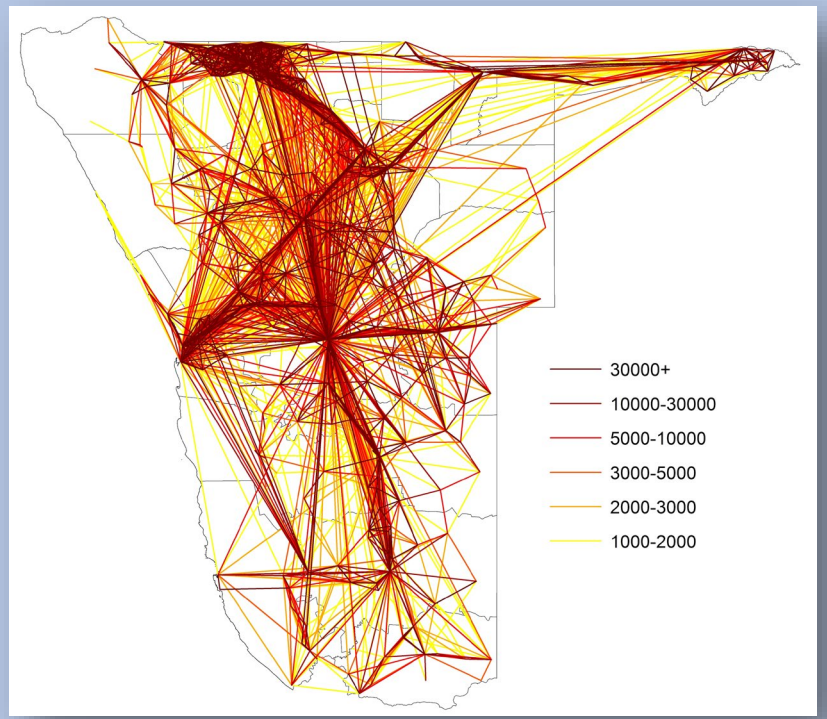
Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

BACKGROUND AND AIMS

- Population and housing censuses with migration questions are conducted every 10 years and sometimes longer
- Substantial changes to migration patterns can occur during inter-censal periods
- Mobile phone call detail records capture elements of national movements regularly and may be a valuable source for migration statistics, but are biased
- **How well do models based on CDRs line up with census-based migration statistics?**



Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

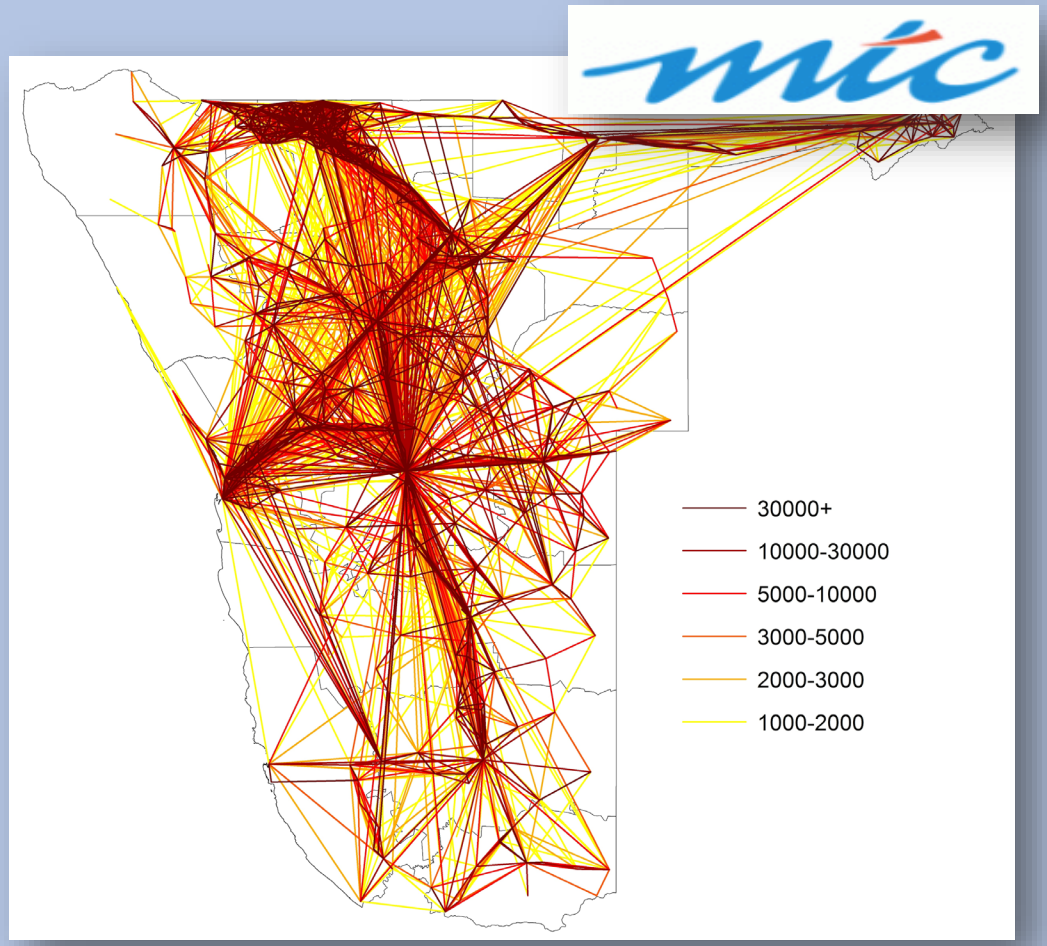
Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Call detail records

Mobile phone data:

- Dataset of 72 billion anonimized CDRs between October 2010 and April 2014 from MTC, the leading network operator in Namibia with a 76% market share.
- Processed to match as closely as possible time period and categories/geography of census questions



Lai et al. (2019) *Nature Palgrave Communications*

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

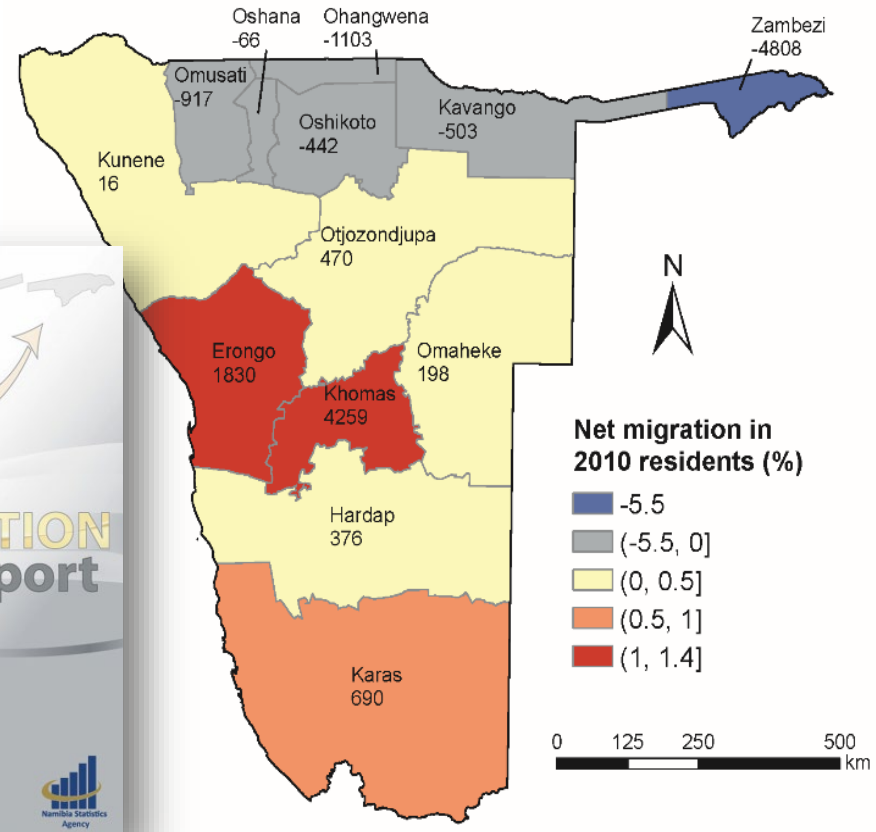
Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

2011 Namibia Census

- The 2011 Census asked about each individual's places of usual and previous residence, with a reference night of 28 August 2011:
  - where does the person usually live?
  - where did the person usually live within 12 months prior to the census?
- An individual will be considered as a short-term migrant if the area of current residence does not match the area of usual residence 12 months prior to the census.

* The place of usual residence refers to the place where a person usually lives for the most part of any year (at least six month).



Population: 2,013,671

Internal short-term migrants: 40,867

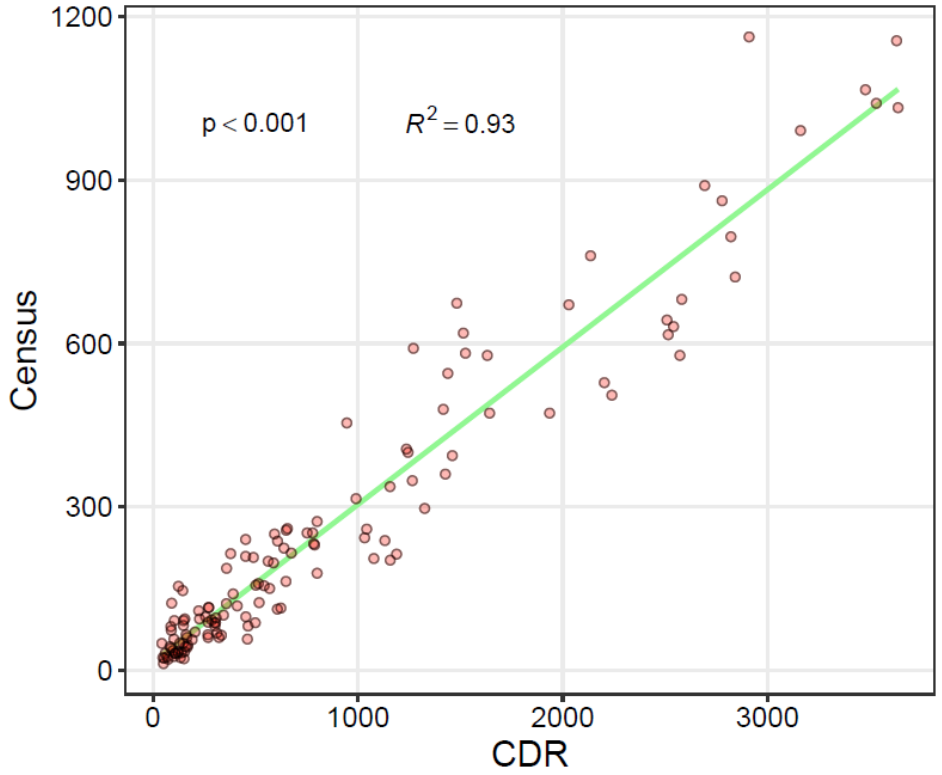
Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Lai et al. (2019) *Nature Palgrave Communications*

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Highly correlation between CDR and census-derived migrations



Lai et al. (2019) *Nature Palgrave Communications*

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Migration Prediction Models

- **CDR-based linear models (CDRLMs)**: simply using CDR-derived migrating user data alone or combined with covariates

$$MIG_{i,j} = \beta_0 + \beta_1 CDR_{i,j} + \vec{\beta}[X]$$

- **Validation and goodness-of-fit indicators**: leave-one-out-cross-validation (LOOCV) approach to calculate the root-mean-square error (RMSE). The model with the lowest RMSE was determined as the best model.

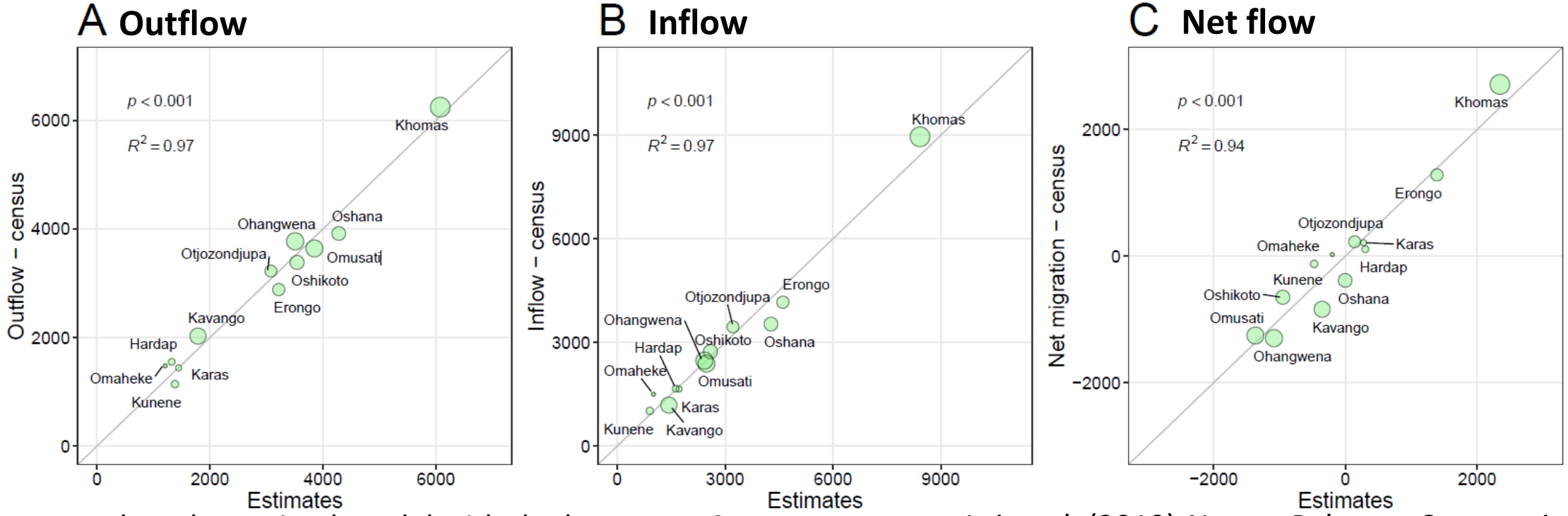
Type	Model	Independent variables	
		CDRs	Other variables
CDR-based linear model (CDRLM)	1	Yes	None
	2		$+ POP_i + POP_j + DIST_{i,j}$
	3[a]		$+ URBAN_i + URBAN_j + DIST_{i,j}$
	4		$+ RAIN_i + RAIN_j + DIST_{i,j}$

POP_i and POP_j : Population of origin i and destination j.
$URBAN_i$ and $URBAN_j$: Proportion of population living in urban areas.
$RAIN_i$ and $RAIN_j$: Annual average precipitation (mm).
$DIST_{i,j}$: Euclidean distance between centroids of origin i and destination j.
[a] Optimal model of each model family for regions except Zambezi, based on the lowest root-mean-square error (RMSE).

Lai et al. (2019) *Nature Palgrave Communications*

Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

CDRLM vs Census-derived migration



Based on the optimal model with the lowest RMSE

Lai et al. (2019) *Nature Palgrave Communications*

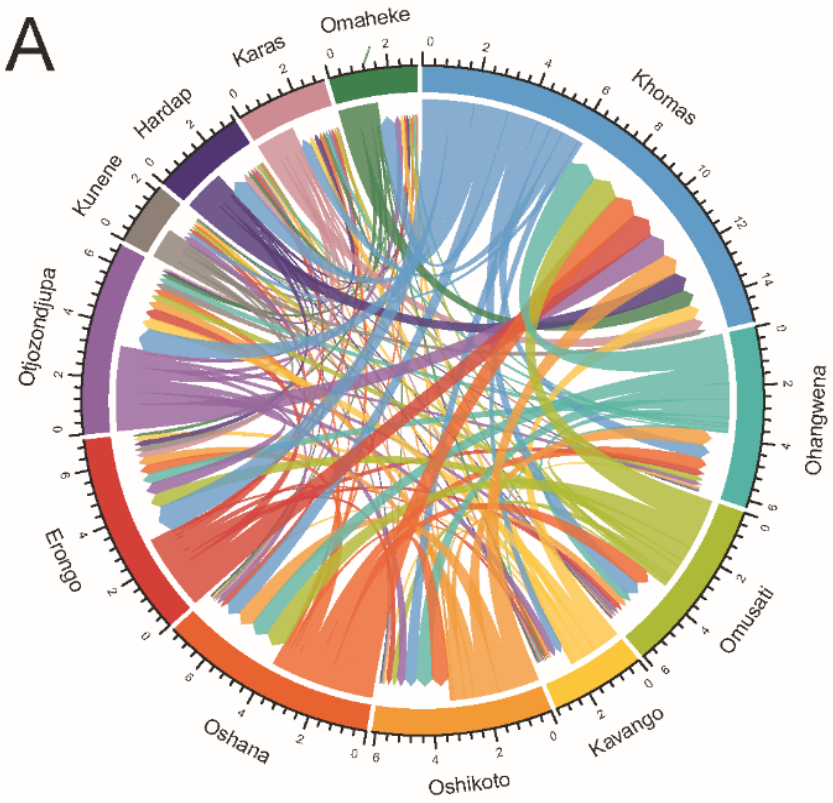
Made possible with funding from the Federal Ministry of the Interior, Building and Community, Germany, represented by the Federal Office for Migration and Refugees (BAMF).

Module 4: Mapping migration using alternative data sources

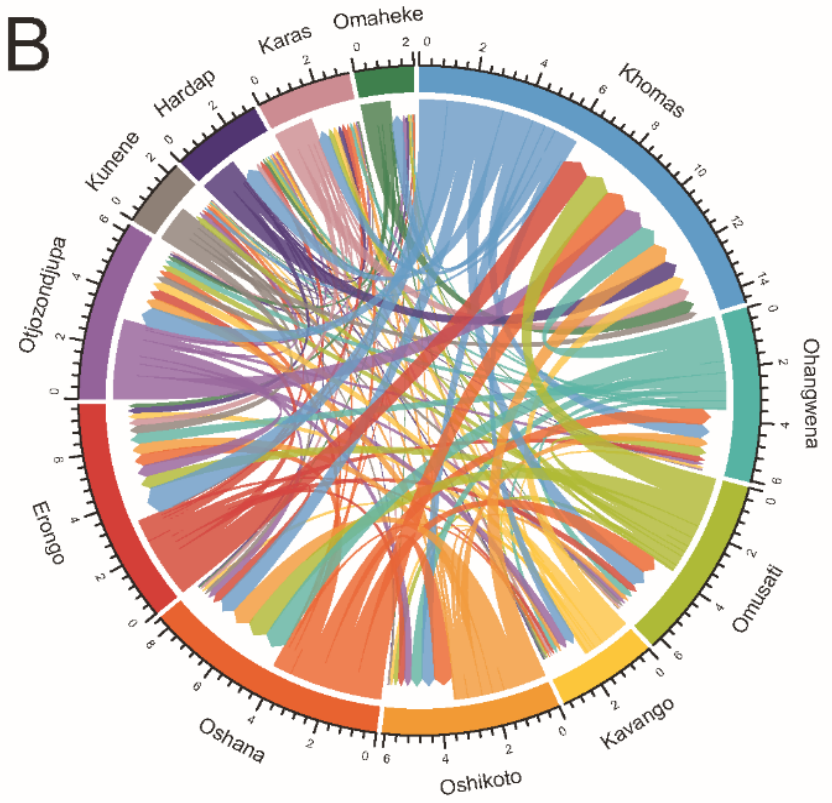
Part 2: Andy Tatem

CDRs and census data show very similar migration patterns

Census



CDRLM



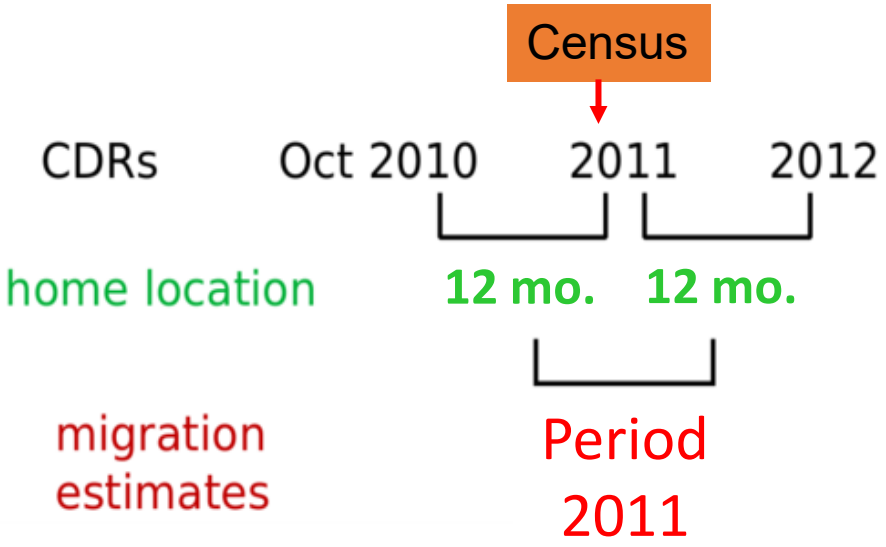
Lai et al. (2019) *Nature Palgrave Communications*

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Using mobile phone data to update migration statistics across years?

- The multiannual time series of CDRs in Namibia allows us to assess their potential to be used to update intercensal national statistics and understand the changing patterns of internal migrations by comparing the places of residence across years.
- The migrations in Period 2012 were estimated by the linear model fitted for Period 2011 and using 2012 CDRs adjusted for increasing users across years.



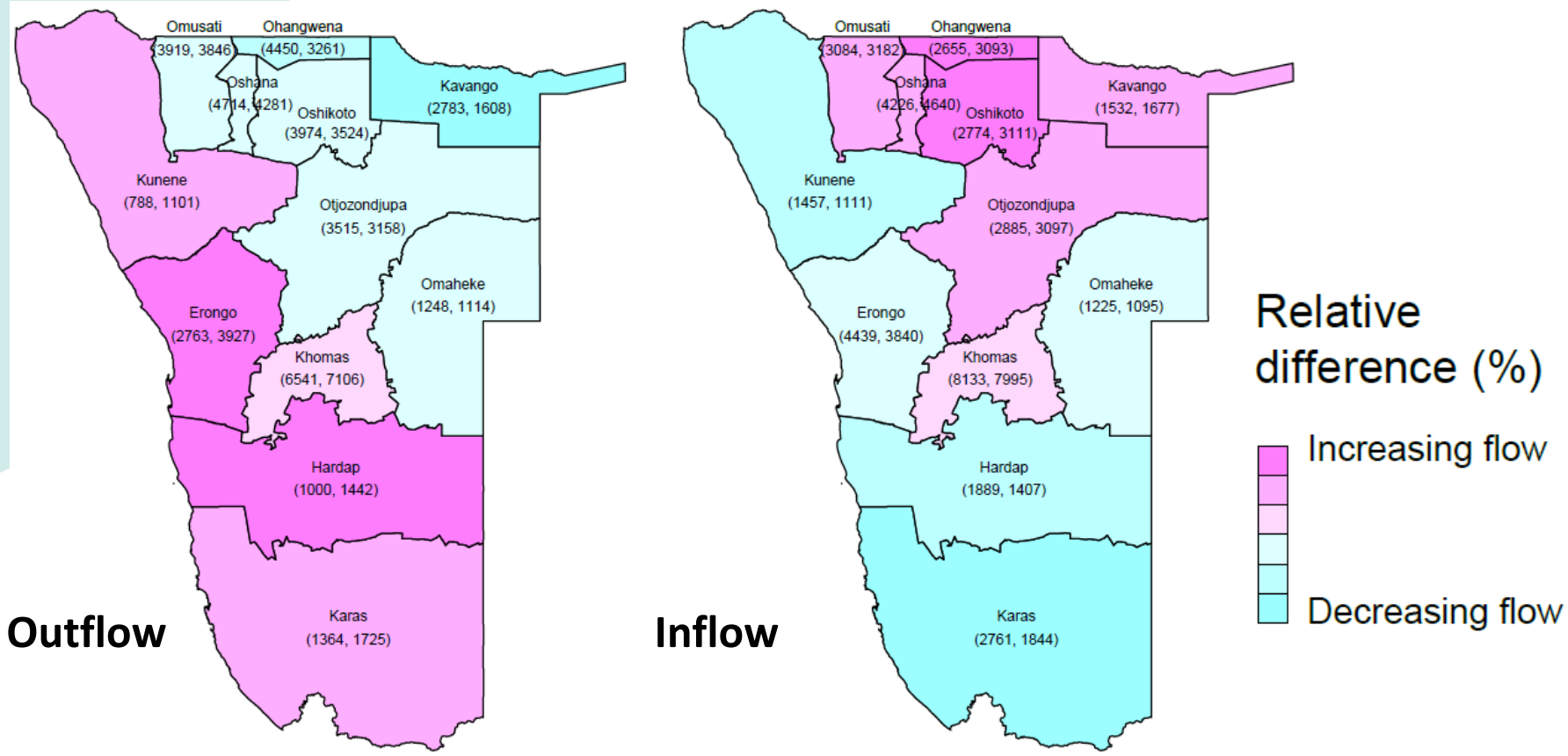
- Period 2011: 1,049,379 users (SIMs); 117,173 migrant users (11.2%)
- Period 2012: 1,238,124 users (SIMs); 144,064 migrant users (11.6%)

Lai et al. (2019) *Nature Palgrave Communications*

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Updating migration statistics across years



Migrants departing and arriving in 2012 compared to 2011

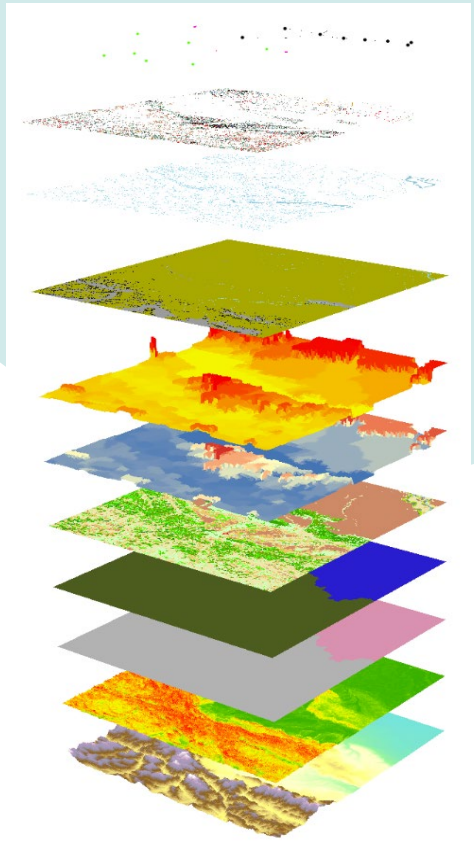
*The Zambezi region as an outlier with limited data was excluded

Lai et al. (2019) *Nature Palgrave Communications*

Module 4: Mapping migration using alternative data sources

Part 2: Andy Tatem

Key messages



Migration patterns are changing continuously and data on these are needed to build effective and equitable evidence-based policies

‘Traditional’ datasources (e.g. census, surveys) remain vital…..

….but they can be outdated, incomplete and unreliable in many low income settings.

‘New’ datasources (e.g. satellites, GPS, phones) can help compliment traditional sources – but they are not perfect: important to compare against traditional data and build models to draw on strengths of different datasources.

Local ownership and engagement with users is key for sustainable uptake and use.

Acknowledgements and further information

www.worldpop.org
@WorldPopProject

COVID-19 work:
www.worldpop.org/covid19

E-mail: A.J.Tatem@soton.ac.uk

@AndyTatem