Mapping & Modeling Migration and Displacement with “Big” Data

Alex de Sherbinin, PhD

CIESIN – The Earth Institute Columbia University

Population-Environment Research Network (PERN)

NASA Socioeconomic Data and Applications Center (SEDAC)

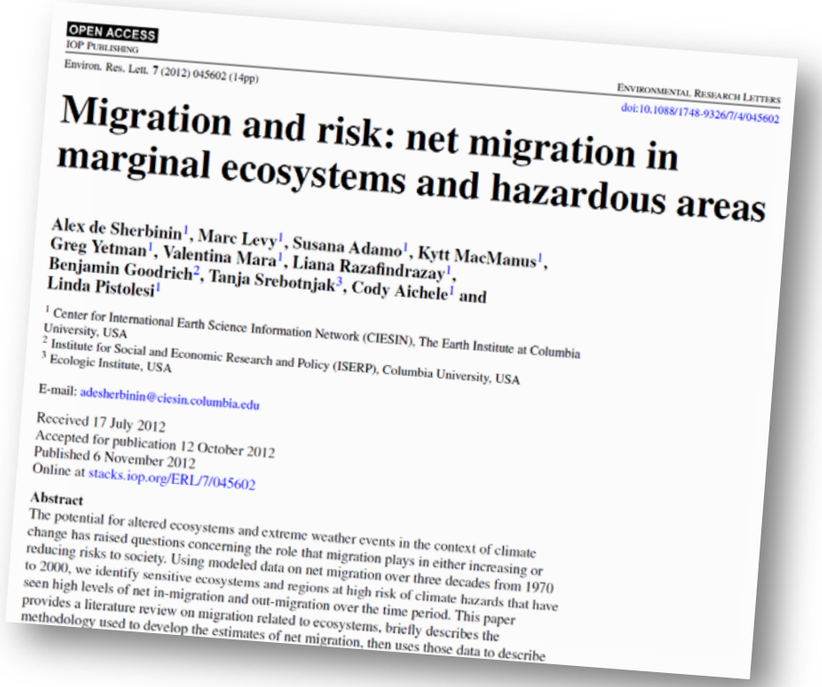
22 April 2021

Columbia Population Research Center and Data Science Institute

Lessons from the “big data” approach to modeling migration and displacement

- Migration data are scant and often not comparable (e.g. stocks, flows, differing time periods)
- Using past and future population distributions offers potential solutions
- Modeling future migration is fraught with uncertainty but plausible futures can also be better understood
- Big data for displacement monitoring – a holy grail but we’re not there yet

Illustrating these points through three projects – (1) estimating past net migration, (2) projecting future migration, and (3) using big data for displacement monitoring



OPEN ACCESS
IOP PUBLISHING

Environ. Res. Lett. 7 (2012) 045602 (14pp)

ENVIRONMENTAL RESEARCH LETTERS

doi:10.1088/1748-9326/7/4/045602

Migration and risk: net migration in marginal ecosystems and hazardous areas

Alex de Sherbinin[1], Marc Levy[1], Susana Adamo[1], Kytt MacManus[1], Greg Yetman[1], Valentina Mara[1], Liana Razafindrazay[1], Benjamin Goodrich[2], Tanja Srebotnjak[3], Cody Aichele[1] and Linda Pistolesi[1]

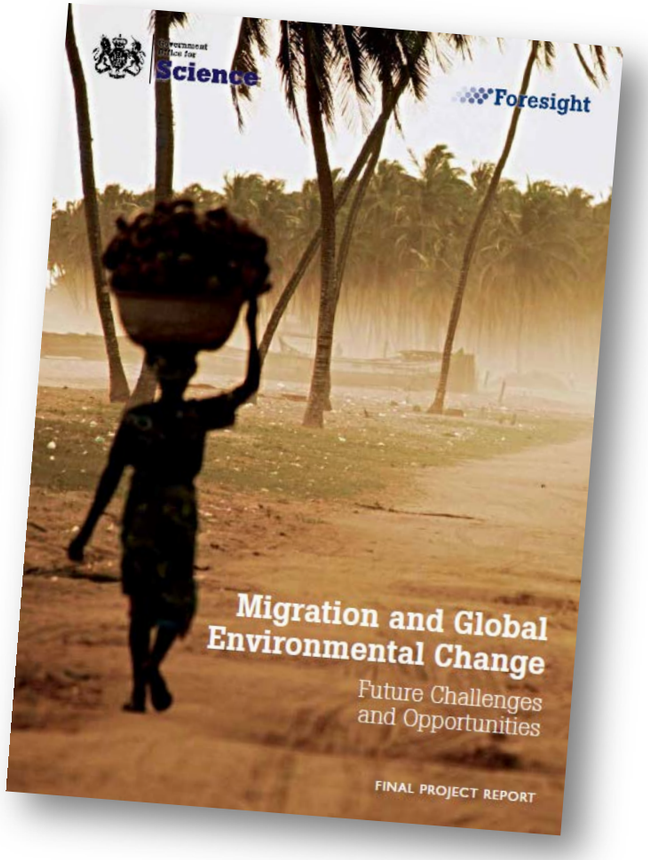
[1] Center for International Earth Science Information Network (CIESIN), The Earth Institute at Columbia University, USA
[2] Institute for Social and Economic Research and Policy (ISERP), Columbia University, USA
[3] Ecologic Institute, USA

E-mail: adesherbinin@ciesin.columbia.edu

Received 17 July 2012
Accepted for publication 12 October 2012
Published 6 November 2012
Online at stacks.iop.org/ERL/7/045602

Abstract

The potential for altered ecosystems and extreme weather events in the context of climate change has raised questions concerning the role that migration plays in either increasing or reducing risks to society. Using modeled data on net migration over three decades from 1970 to 2000, we identify sensitive ecosystems and regions at high risk of climate hazards that have seen high levels of net in-migration and out-migration over the time period. This paper provides a literature review on migration related to ecosystems, briefly describes the methodology used to develop the estimates of net migration, then uses those data to describe

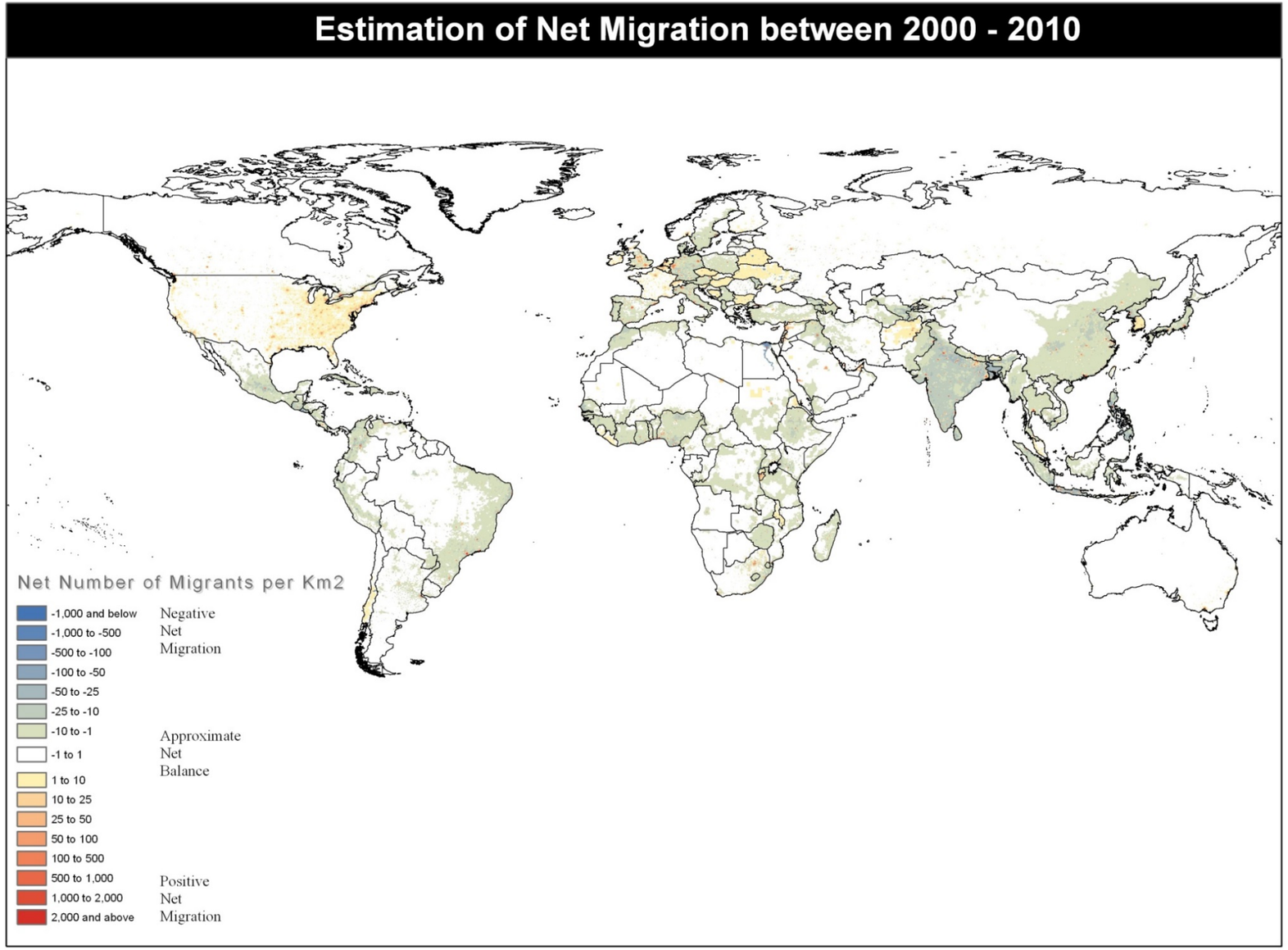


Modeling Net Migration

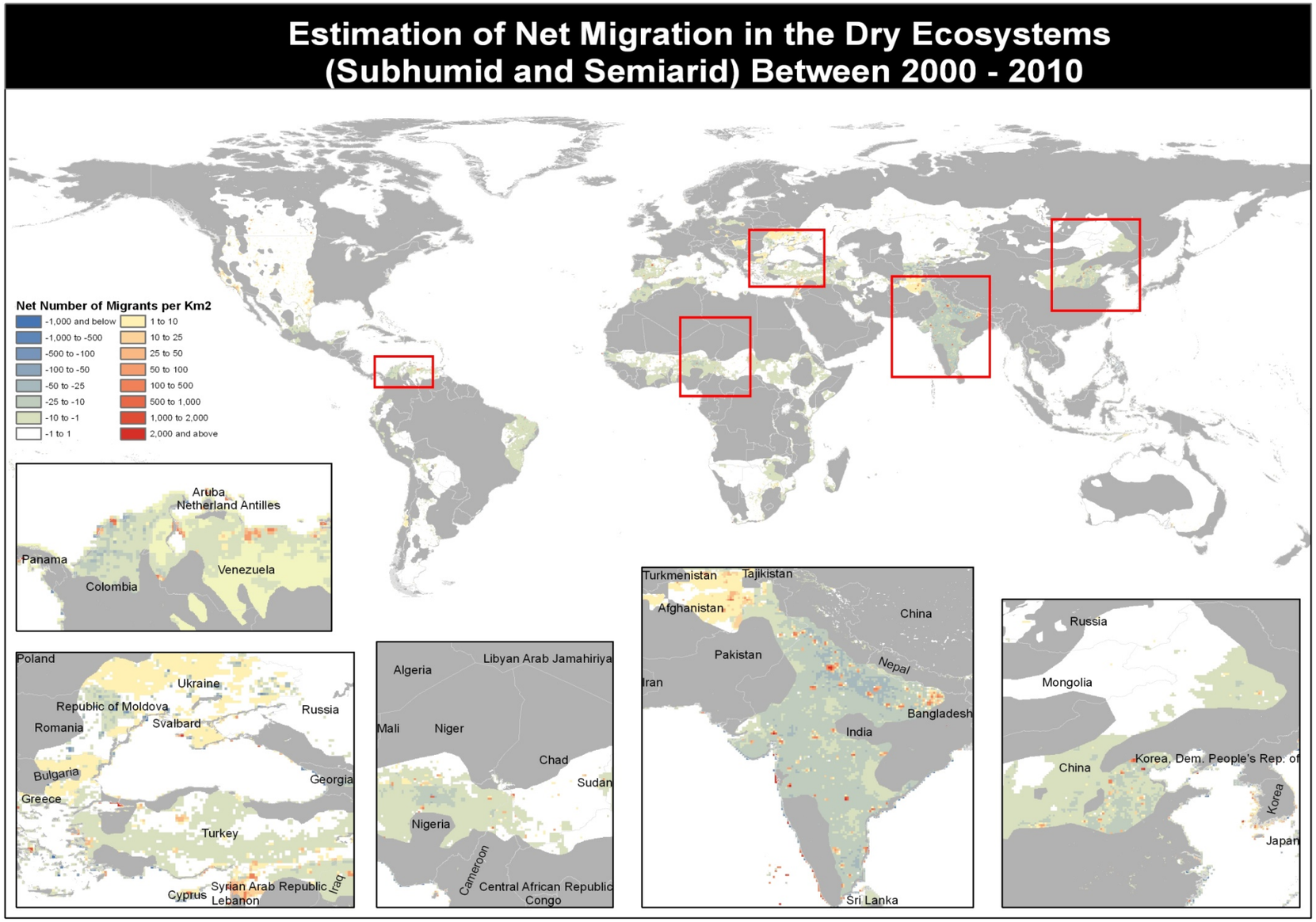
Project funded by the UK Foresight Project on Migration and Global Environmental Change

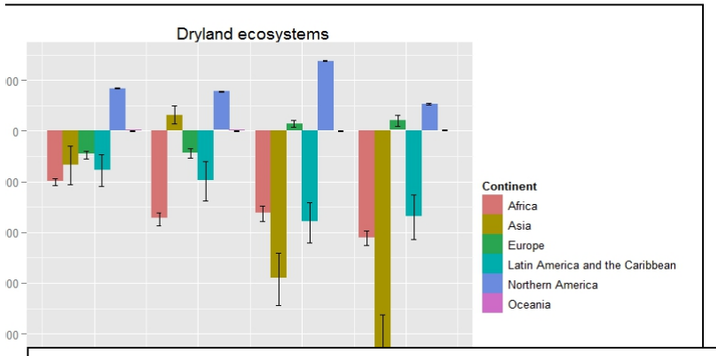
Alex de Sherbinin (PI), Marc Levy, Susana Adamo, Kytt MacManus, Greg Yetman, Vali Mara, Liana Razafindrazay, Ben Goodrich, Tanja Srebotnjak, Cory Aichele, and Linda Pistolesi.

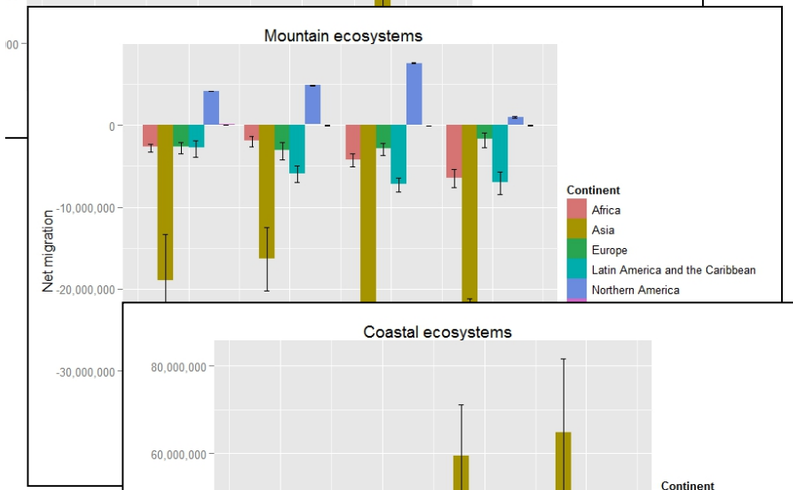
Migration data are scarce – so indirect estimation using “big” population data is sometimes the best solution for global-scale analyses

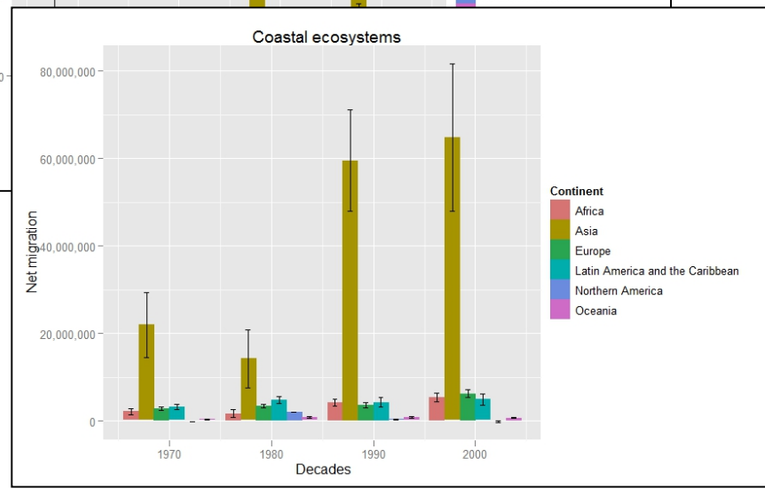


Estimation of Net Migration in the Dry Ecosystems (Subhumid and Semiarid) Between 2000 - 2010









The Approach

Using the population balancing equation:

Population Growth = (Births – Deaths) + (In-migration – Out-migration)

Or:

Population Growth = Natural Increase + Net Migration

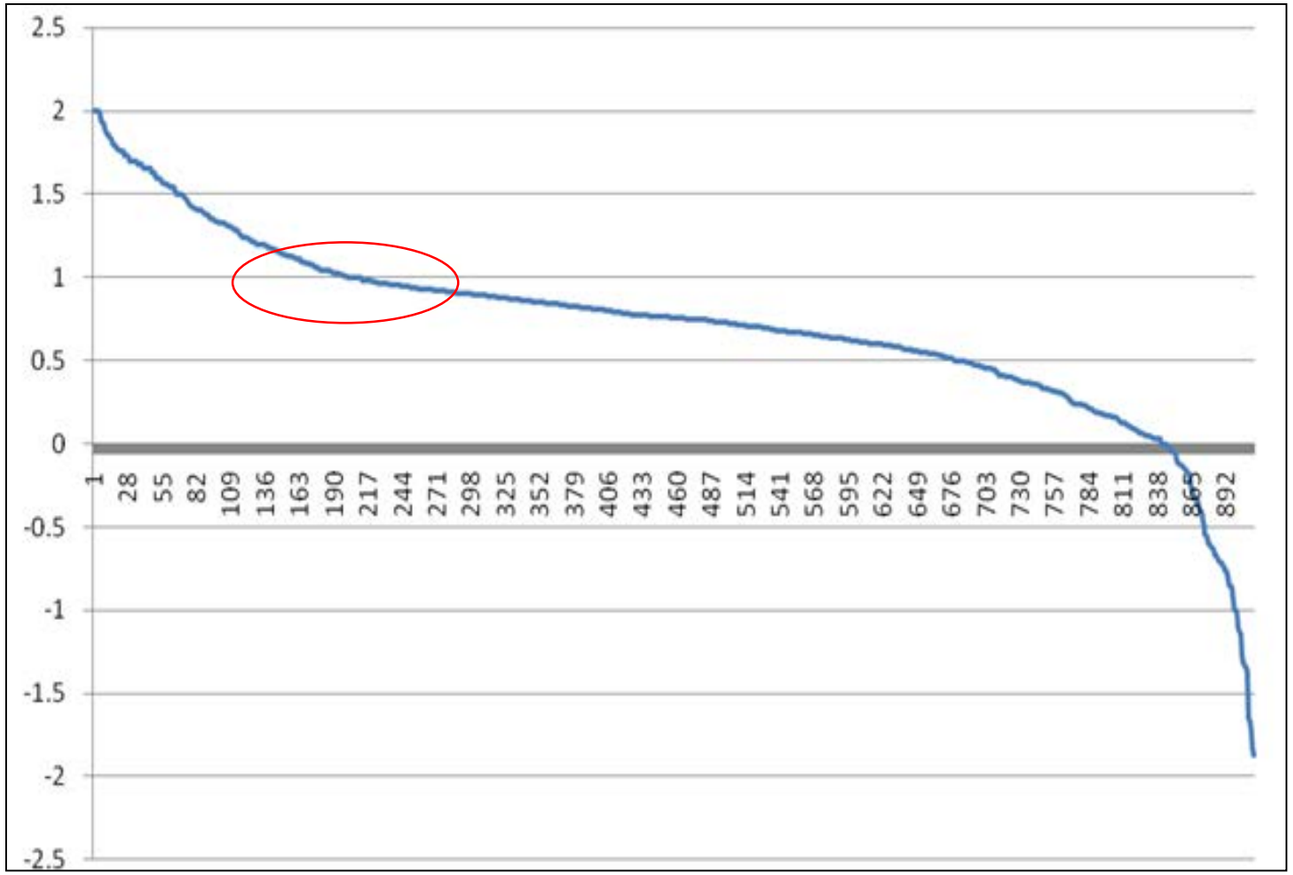
Therefore:

Net Migration = Population Growth – Natural Increase

Methodological Challenge: Natural Increase Grids

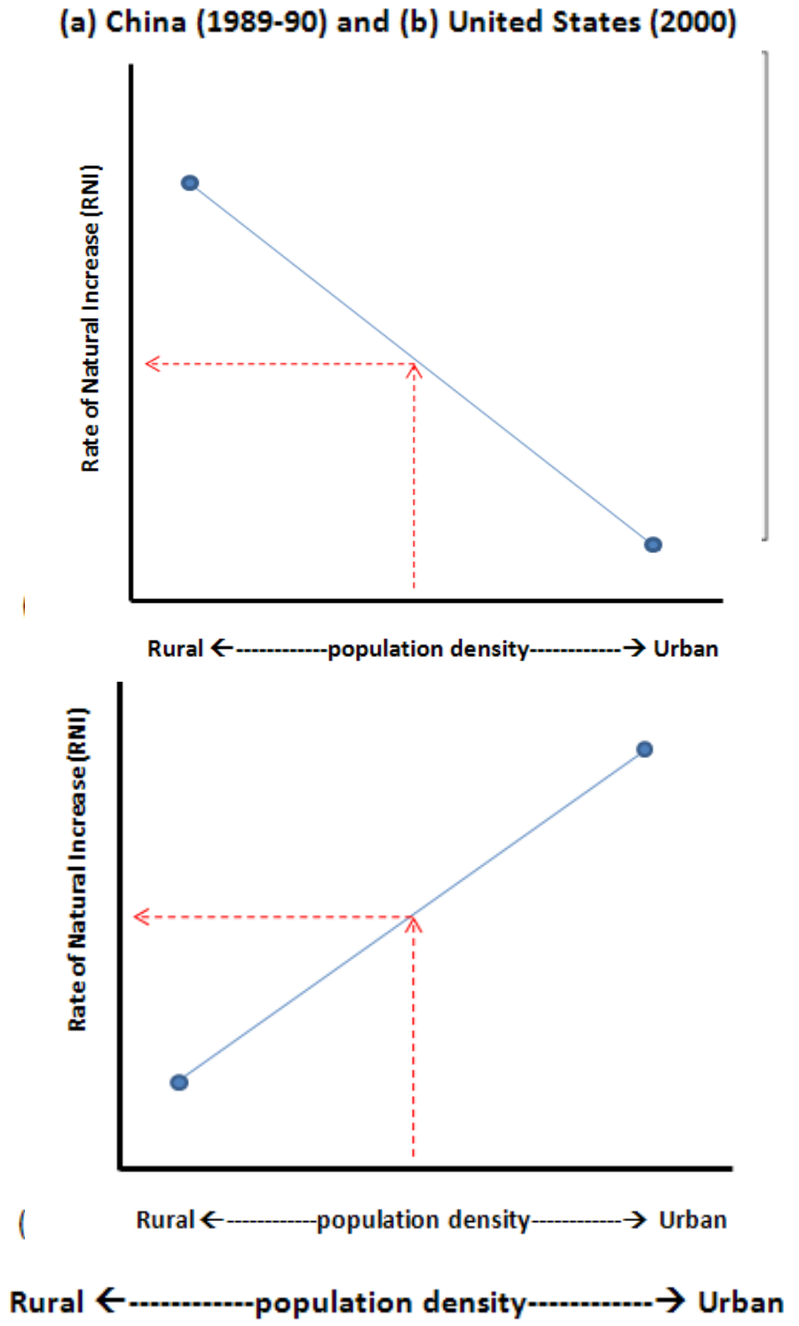
- Could have used country level rates of natural increase (RNIs)
- But we know that RNIs are not uniform within countries
- So we compiled crude birth rates and crude death rates by urban and rural area for as many countries and years as possible from the *UN Demographic Yearbooks*

Ratio of Urban to Rural Rates of Natural Increase (RNIs)

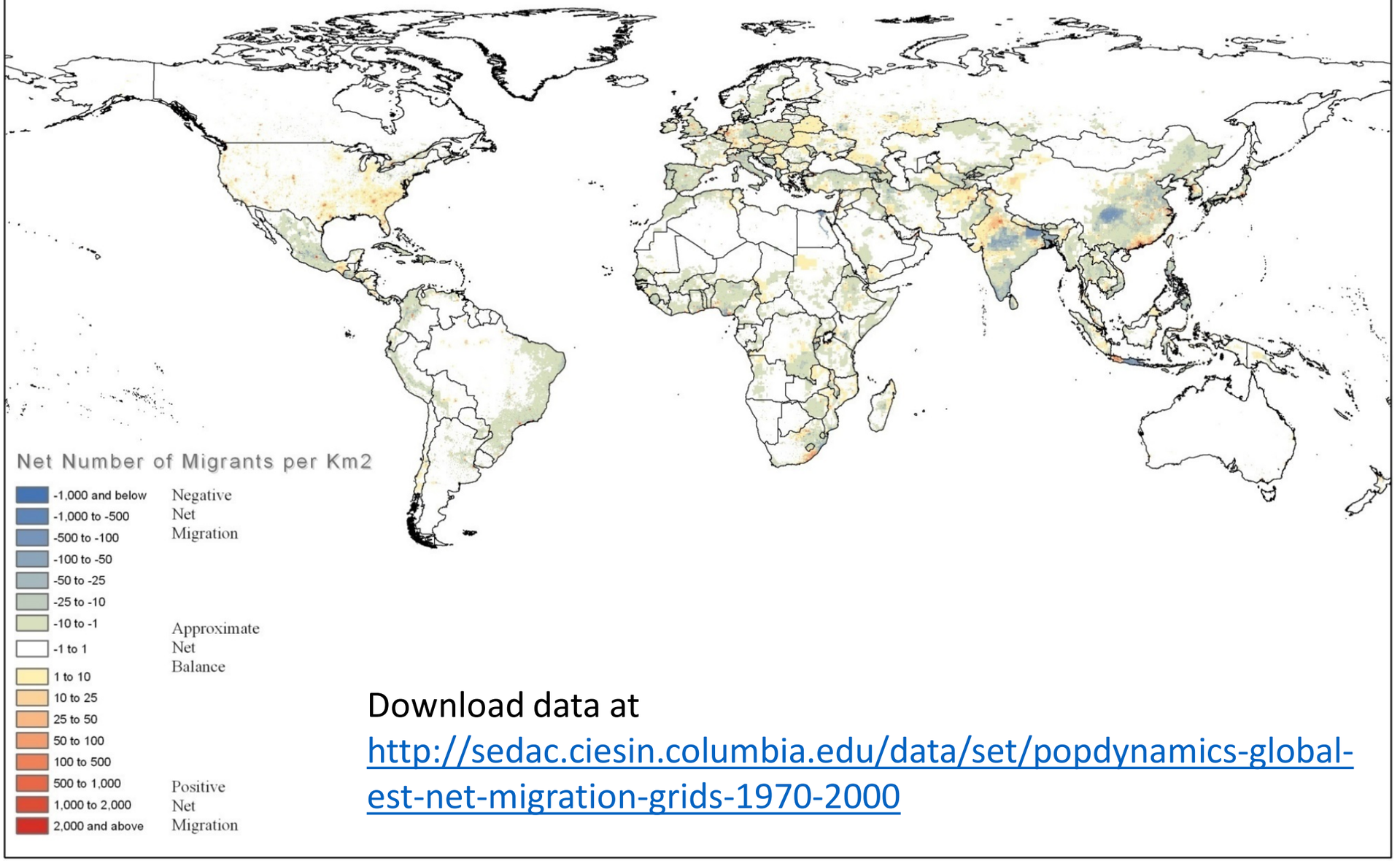


Methodological Challenge: Natural Increase Grids

- So we sought to characterize natural increase at a subnational level
  - differentiating urban and rural rates of natural increase
- Compiled urban and rural crude birth and death rates for as many country-years as possible
  - 5,016 country-year observations
- Imputed the remainder using mi and Amelia package
- Used population density grids to spatially allocate differential rates of natural increase
- Last steps:
  - RNI grid x Pop Grid (t1) = Nat. Inc. Grid
  - Pop Grid (t2) – Pop Grid (t1) – Nat Inc. Grid = Net Migration Grid for decade

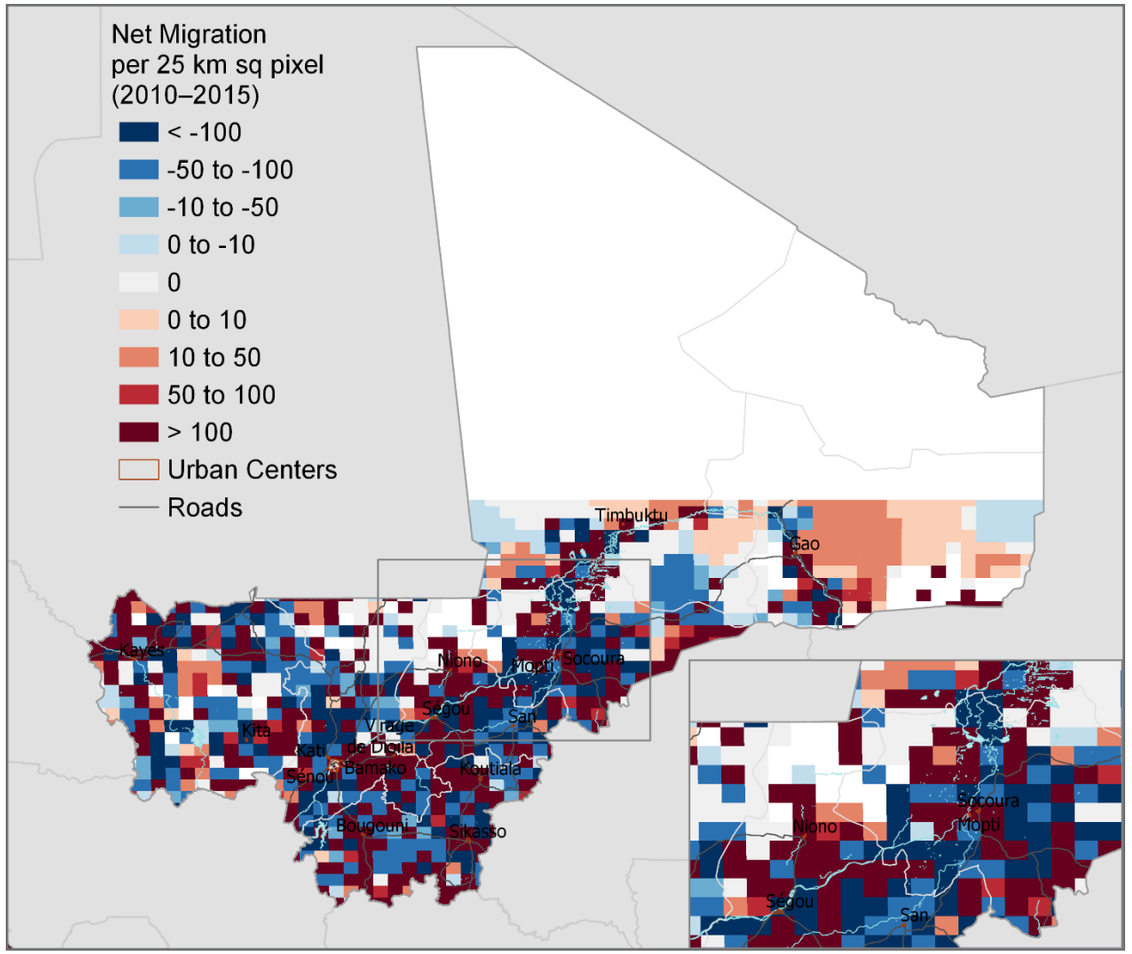


Estimation of Net Migration between 1990 - 2000



Download data at
http://sedac.ciesin.columbia.edu/data/set/popdynamics-global-est-net-migration-grids-1970-2000

Emulation is the highest form of flattery…



Download at: https://knowledge4policy.ec.europa.eu/dataset/ds00167_en

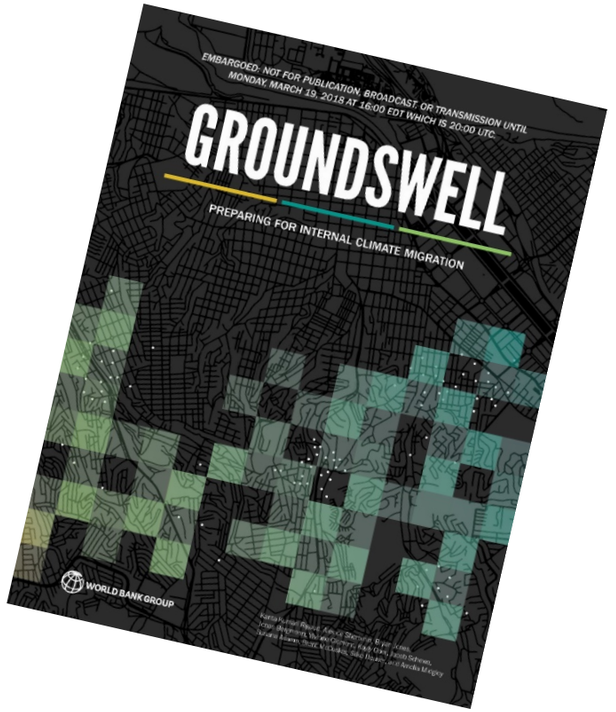
Population Gravity Modeling to Infer Migration

Project funded by the World Bank, with new work under the African Union Commission funded by the Robert Bosch Foundation

Alex de Sherbinin (PI), Susana Adamo, Jane Mills, Tricia Chai-Onn, Alyssa Fico, Malanding Jaiteh, Valentina Mara, Kytt MacManus, Kira Topik, and Haibin Xia

Groundswell: Preparing for Internal Climate Migration

- Groundswell was released in 2018, and covered three regions – Latin America, Sub-Saharan Africa, and South Asia
- Groundswell II is expected to be released in June 2021 – covering the remainder of World Bank regions



Modeling conducted by Bryan Jones, PhD
CUNY Institute for Demographic Research
CUNY Baruch College

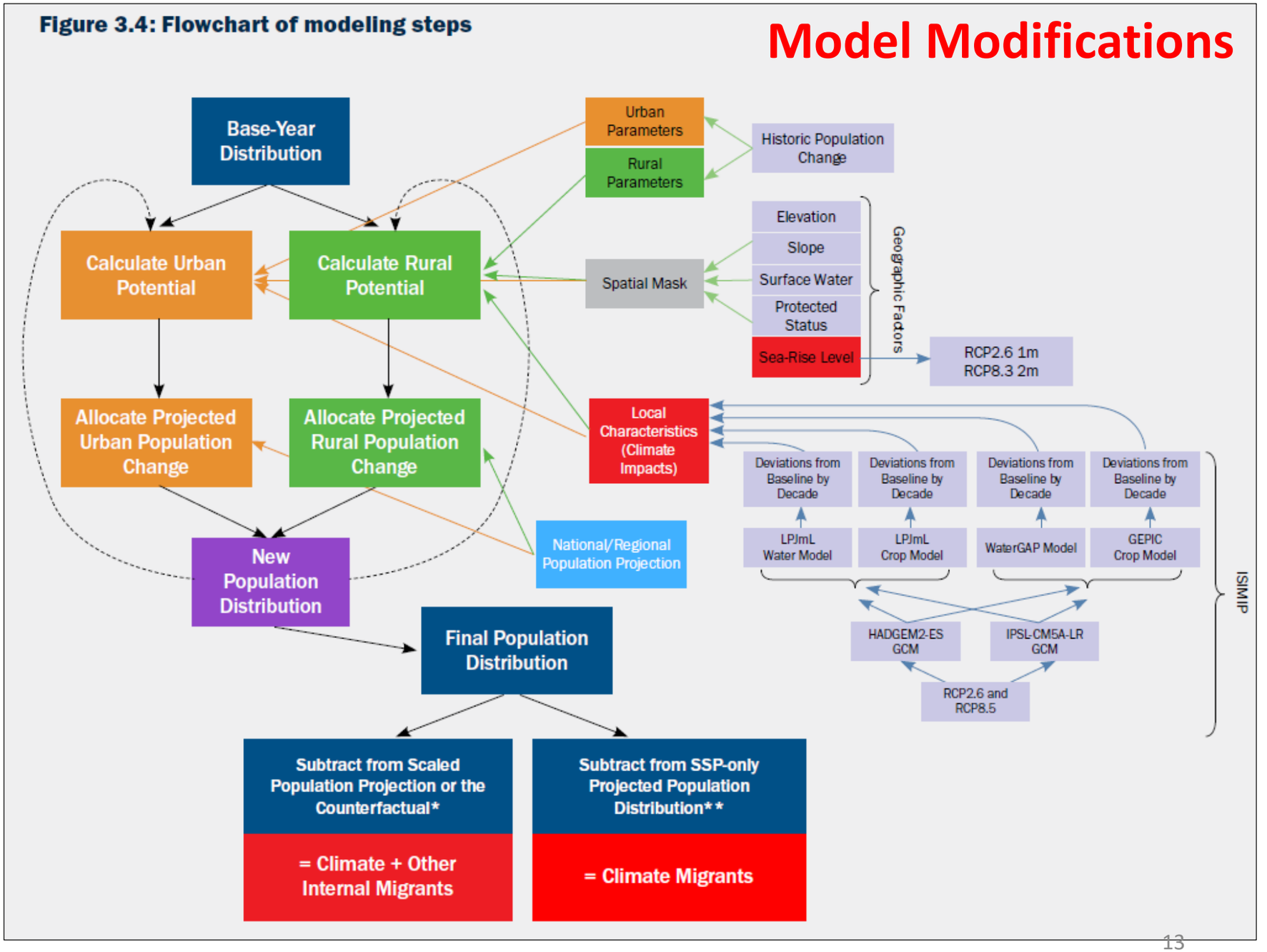
CUNY INSTITUTE FOR DEMOGRAPHIC RESEARCH

With inputs form Jacob Schewe, PhD
Potsdam Institute for Climate Impact Research (PIK)

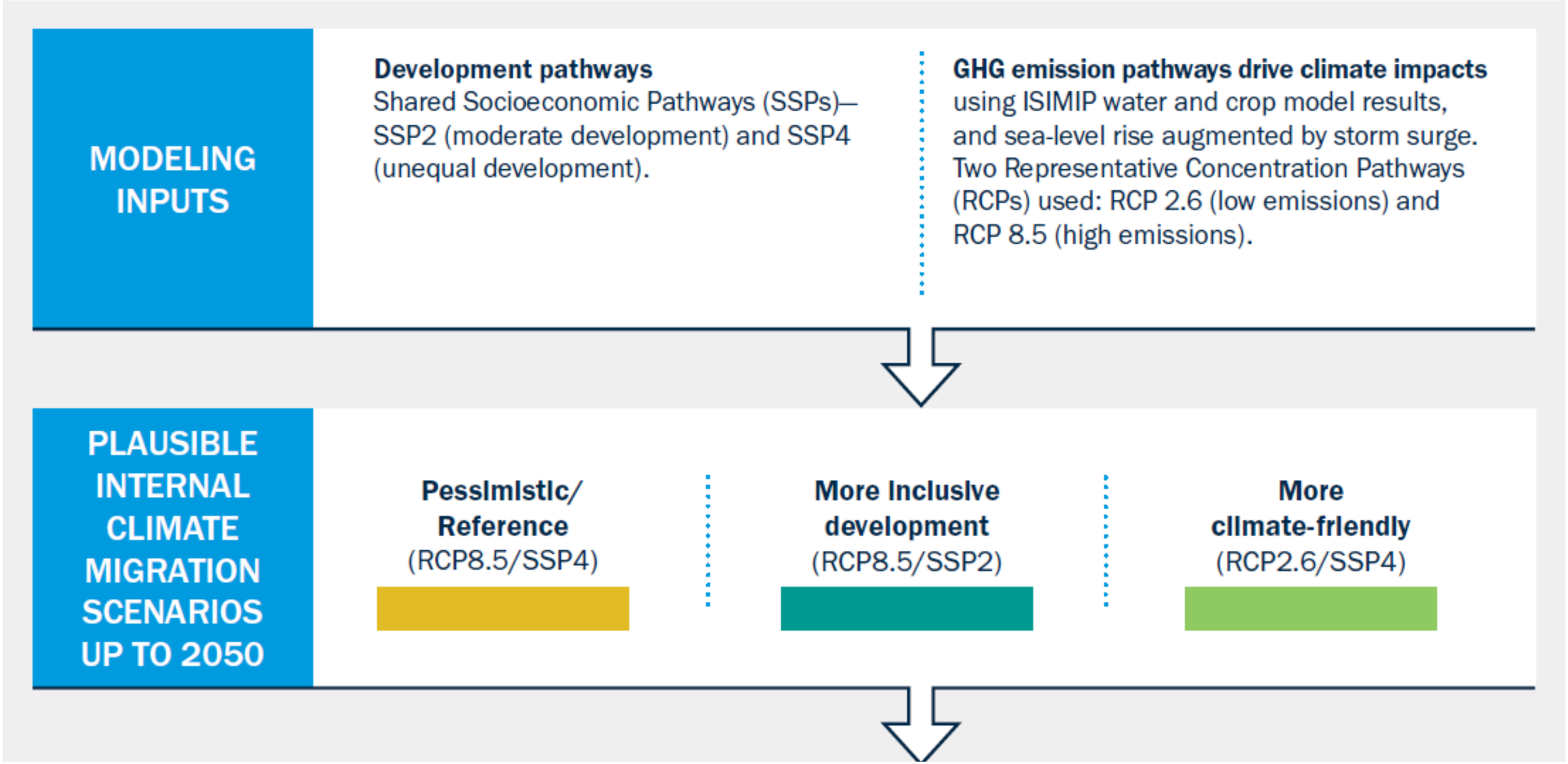
PIK

Model Modifications

- Groundswell uses a population gravity modeling approach
- Gravity models start with a baseline population
- Future populations are projected based on the empirical observation that people tend to move to urban areas owing to economic opportunities and services (attractiveness)
- **Climate impacts could be a push or pull factor that will affect the relative attractiveness of locations**

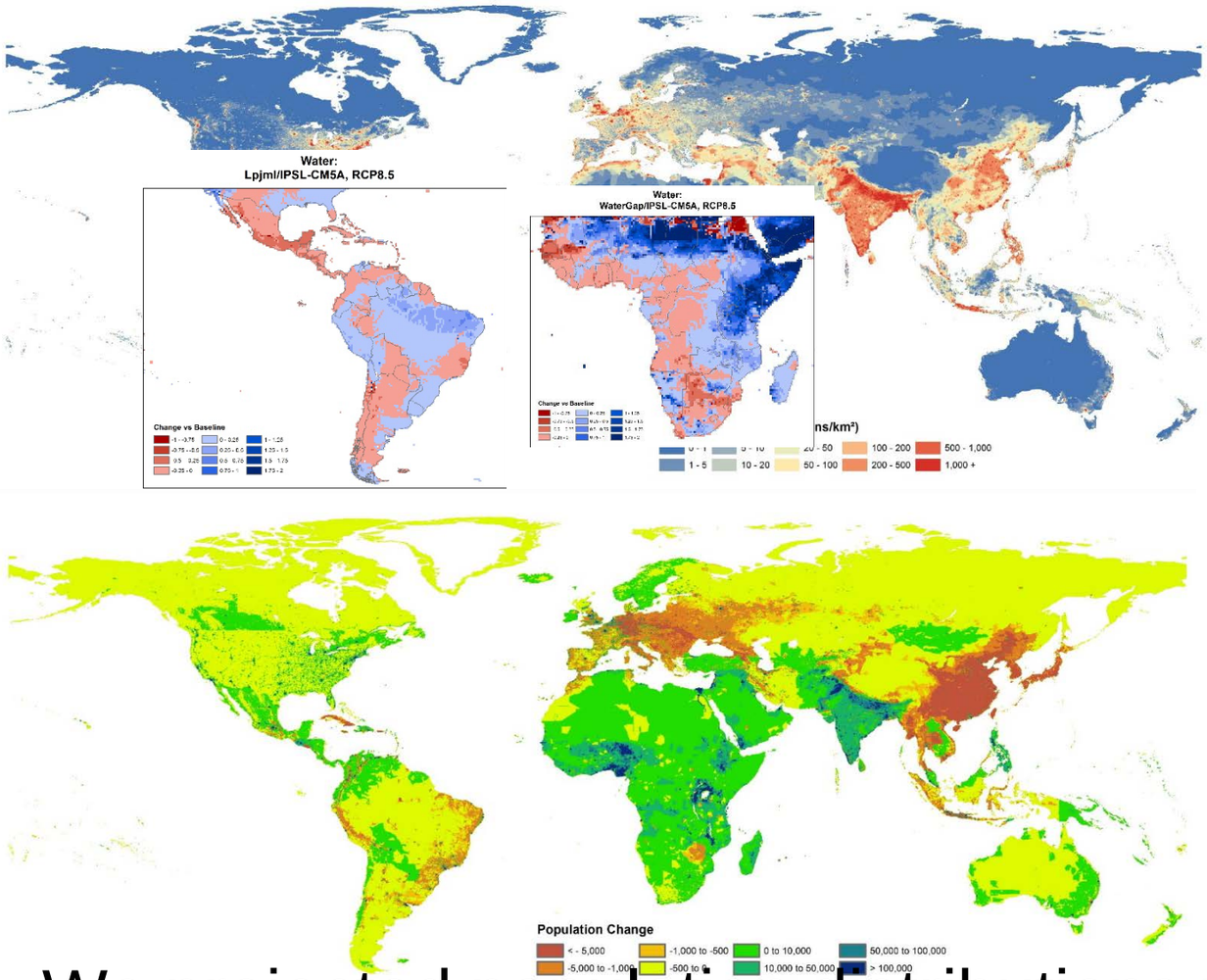


Groundswell took a scenario based approach based on combinations of development trajectories (Shared Socioeconomic Pathways) and climate impacts on crop production and water availability (from the ISIMIP project)

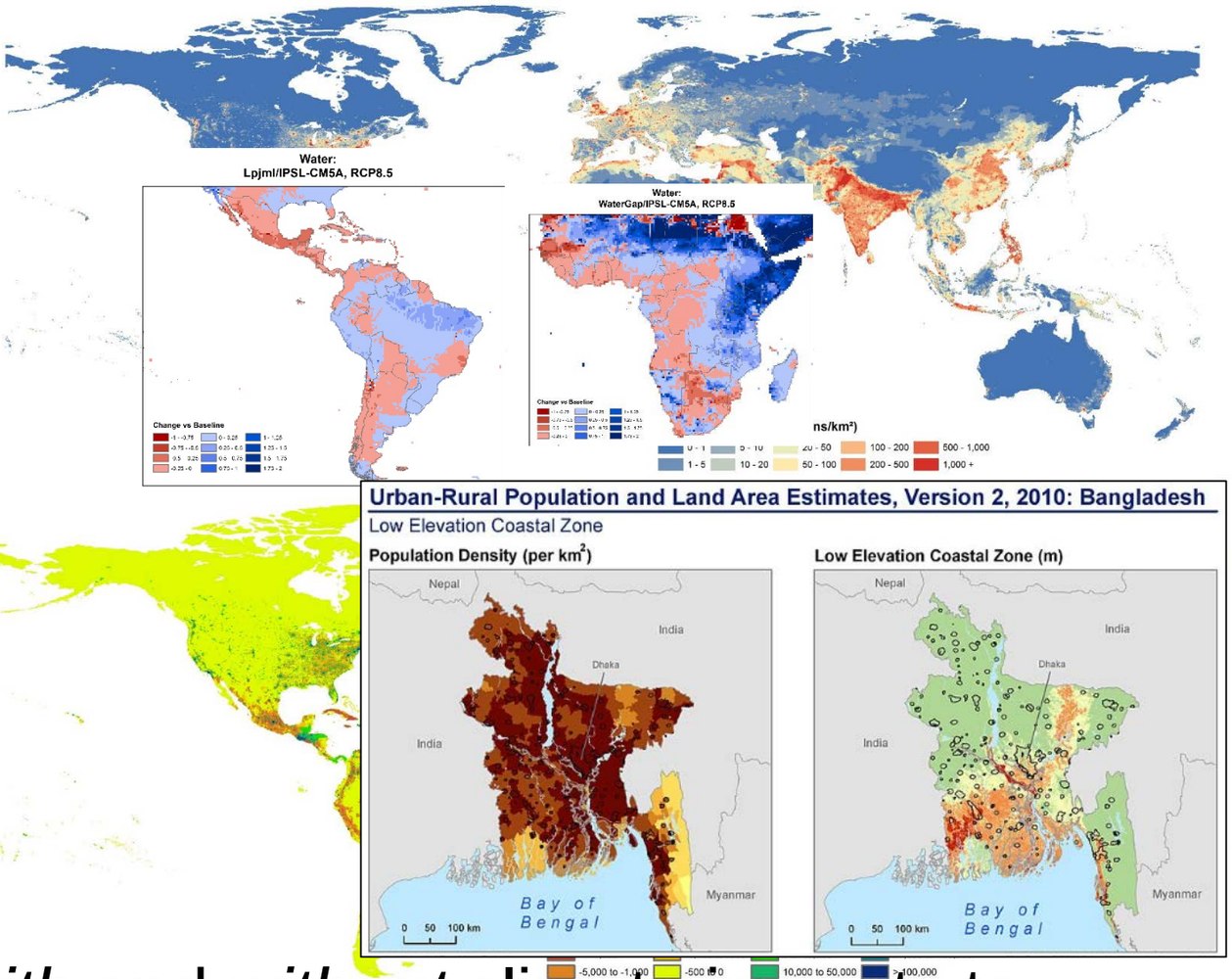


Two SSPs + Climate Impacts

SSP 2: Middle of the Road



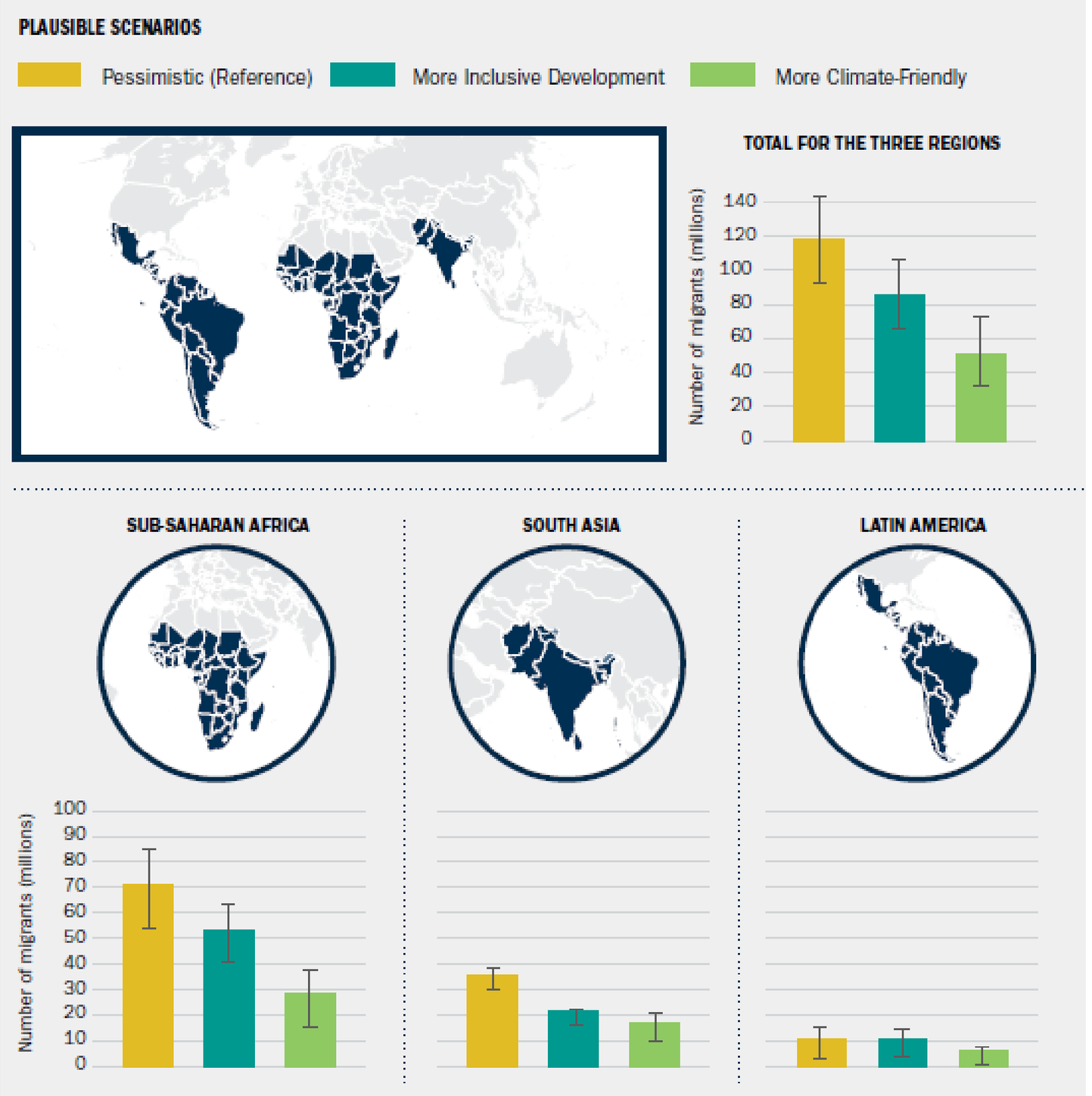
SSP 4: Inequality



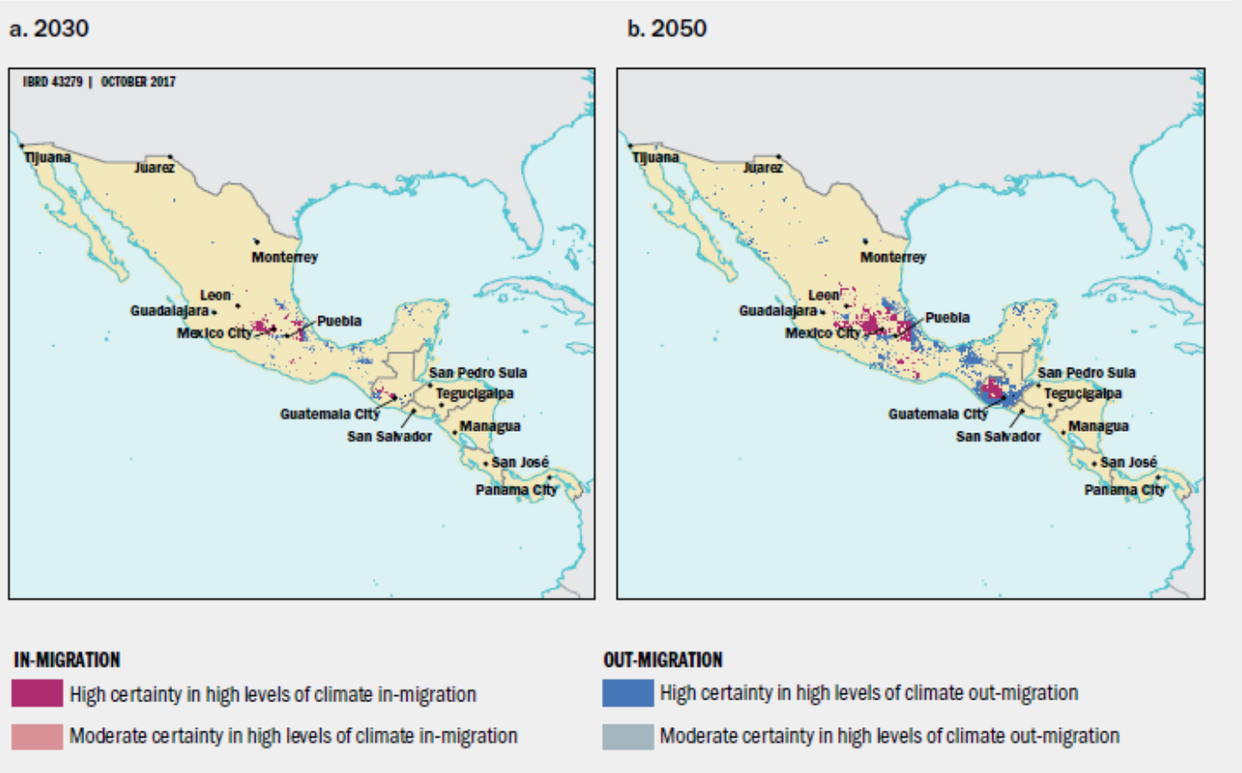
We projected population distributions *with* and *without* climate impacts to generate two future population distributions – the difference between the two was assumed to result from the fast demographic variable…. migration

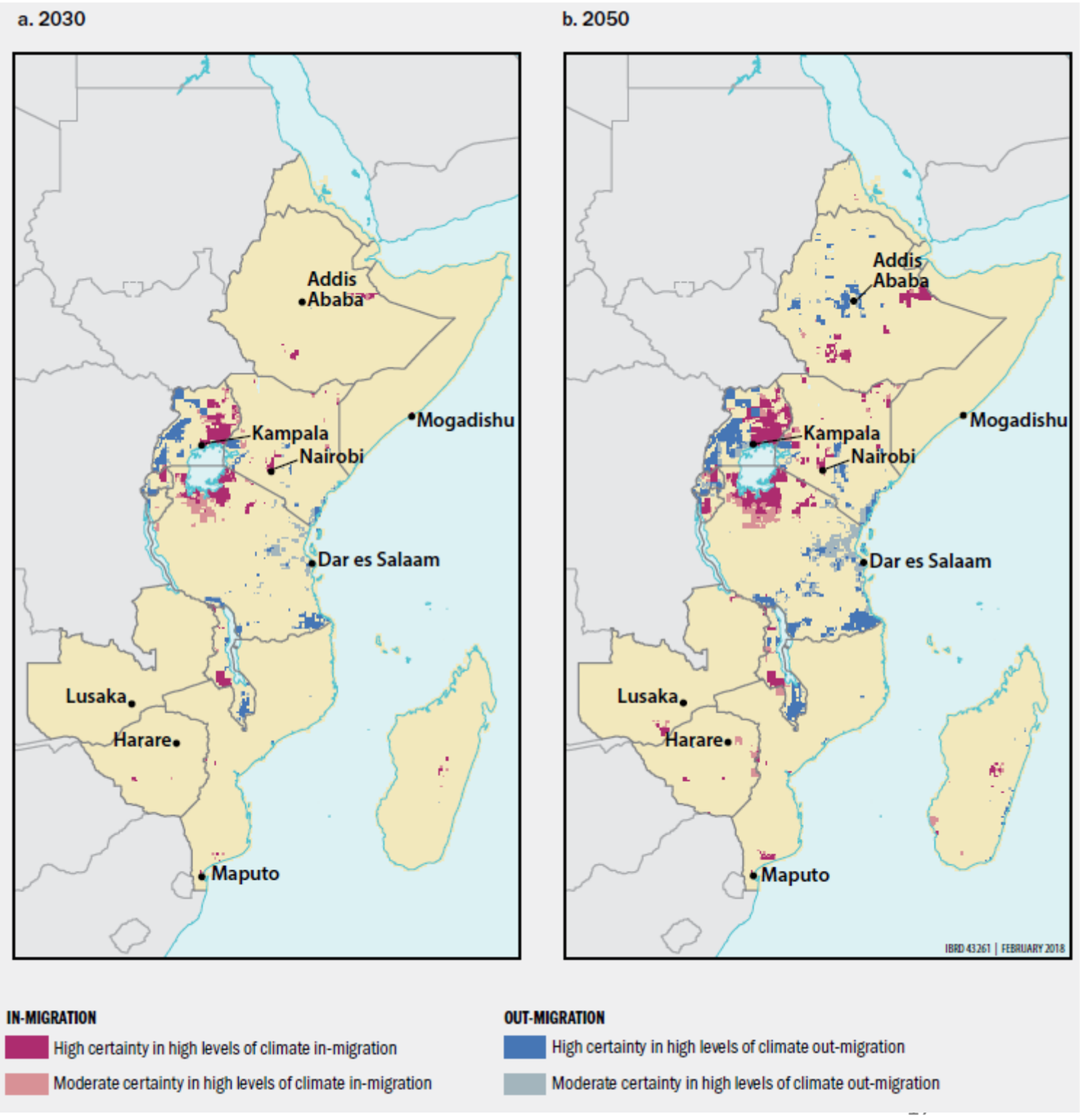
Headline Numbers

- Up to 140 million internal migrants by 2050 in the three regions
- The largest numbers are in Africa, suggesting high climate sensitivity in that region
- Numbers are lower for SSP2: more inclusive development, and lowest for the climate-friendly RCP2.6 scenario



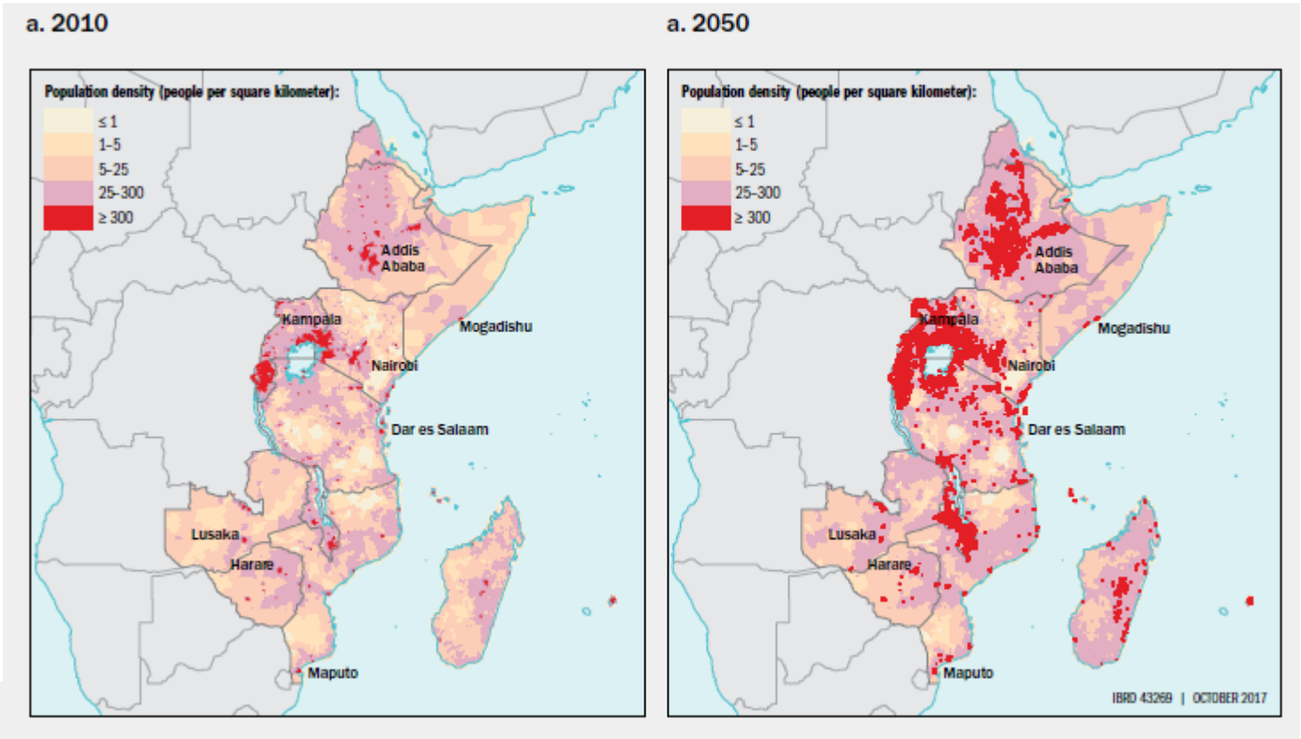
Migration Hotspots for two out of three regions
(South Asia not shown)

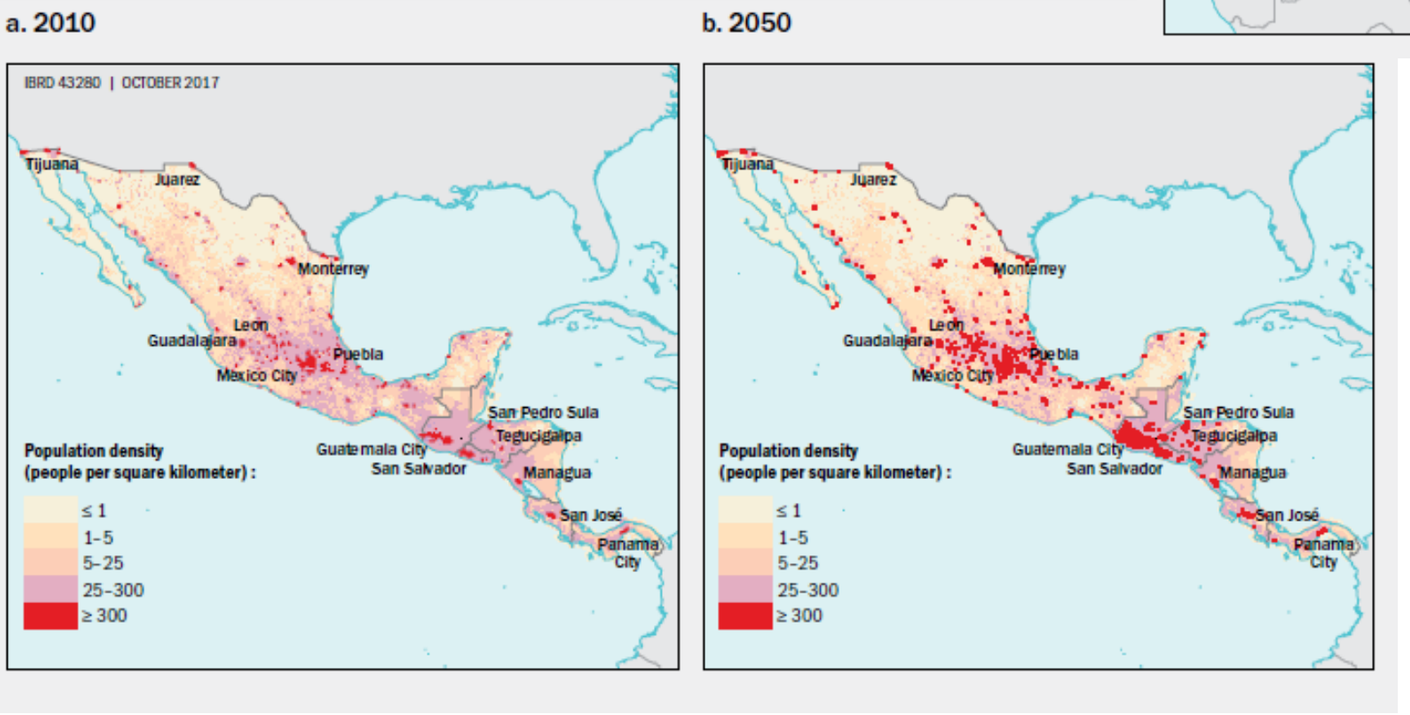




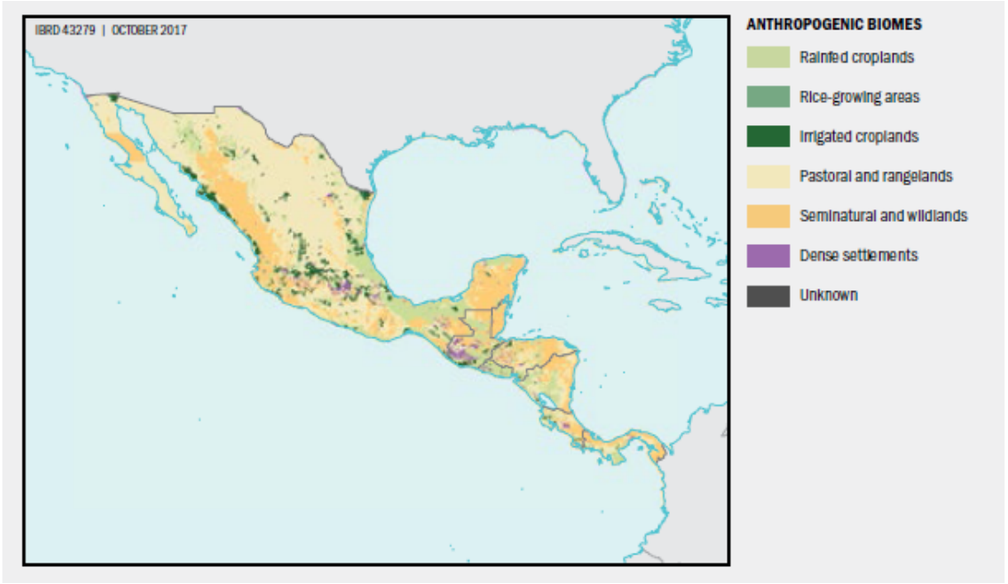
Urbanization

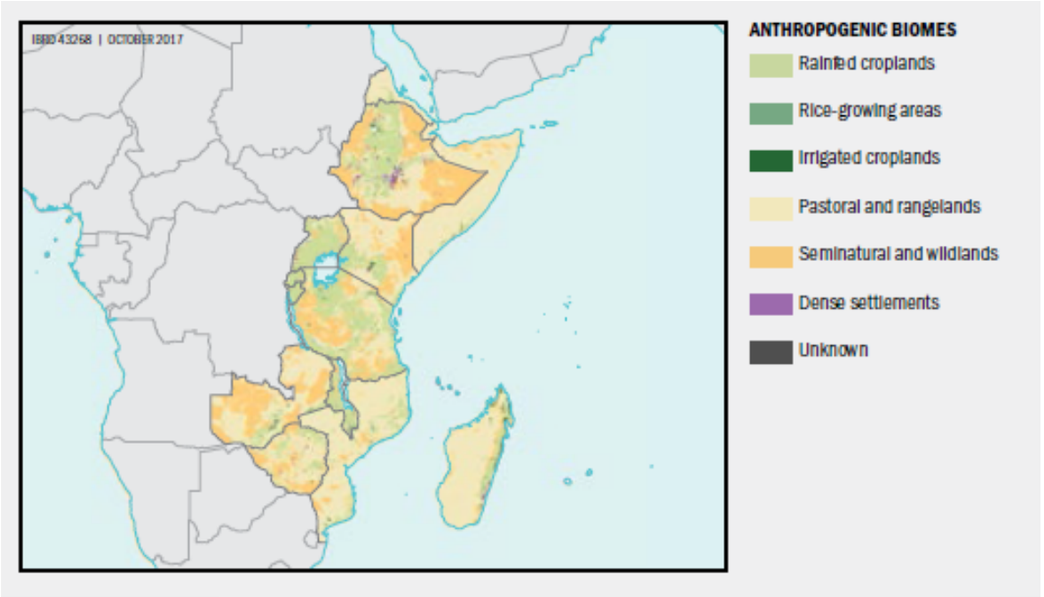
Urban Threshold: >=300 persons per sq. km.

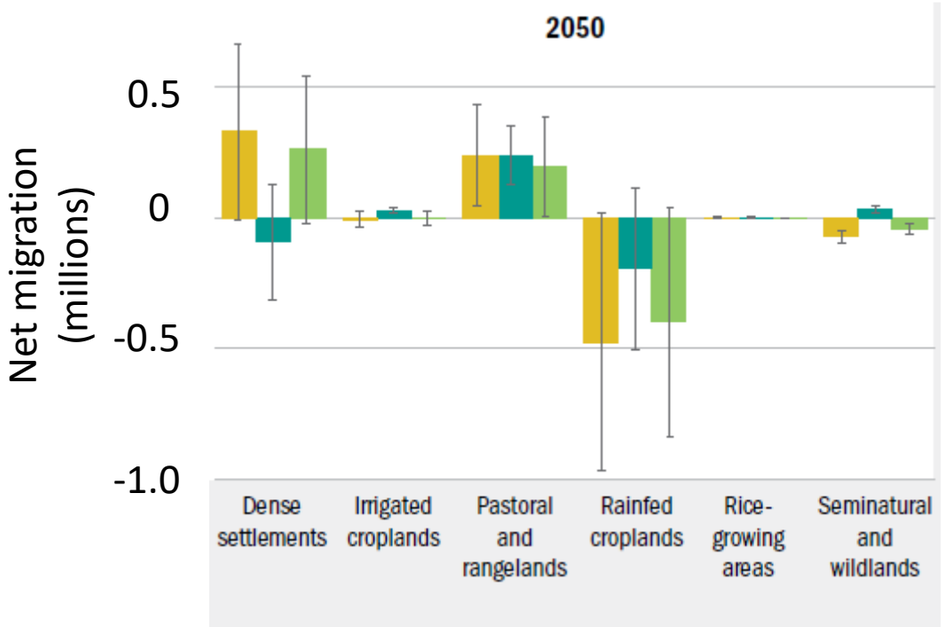


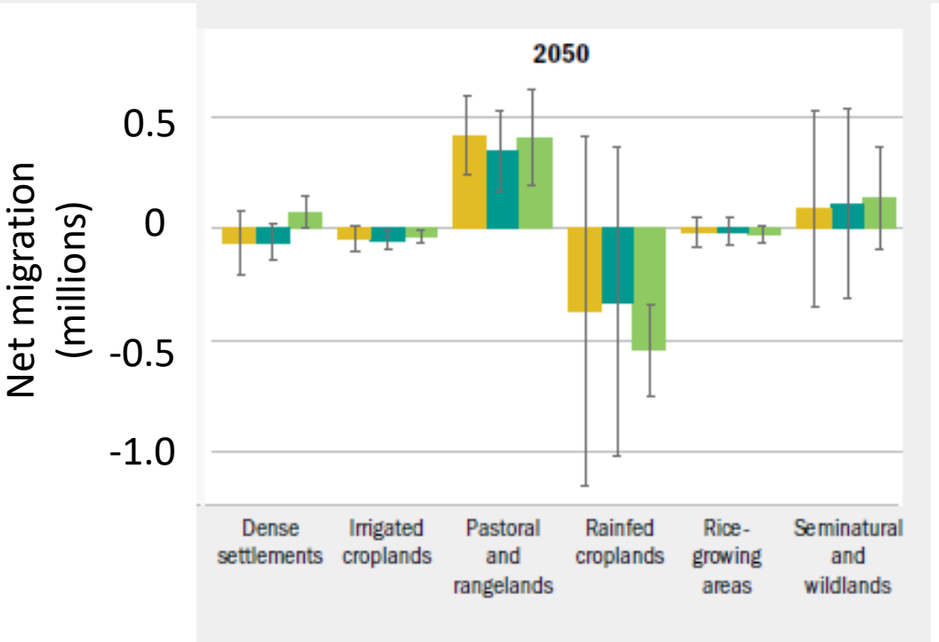


Livelihood Systems









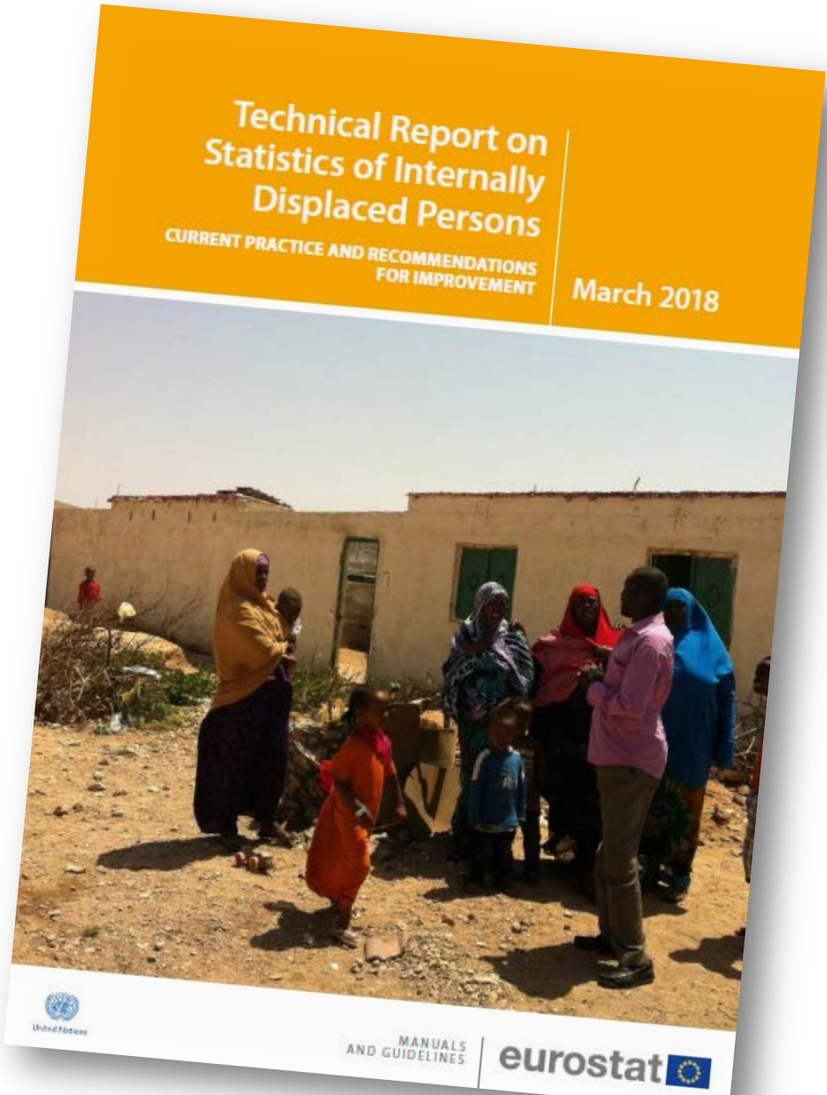
Big Data to Monitor and Quantify Disaster Displacement

Project funded by Schmidt Futures through the Data Science Institute

Bob Chen (PI), Kytt MacManus, Greg Yetman, Nanshan Li, Andrea Navarrete and Vikas Vicraman

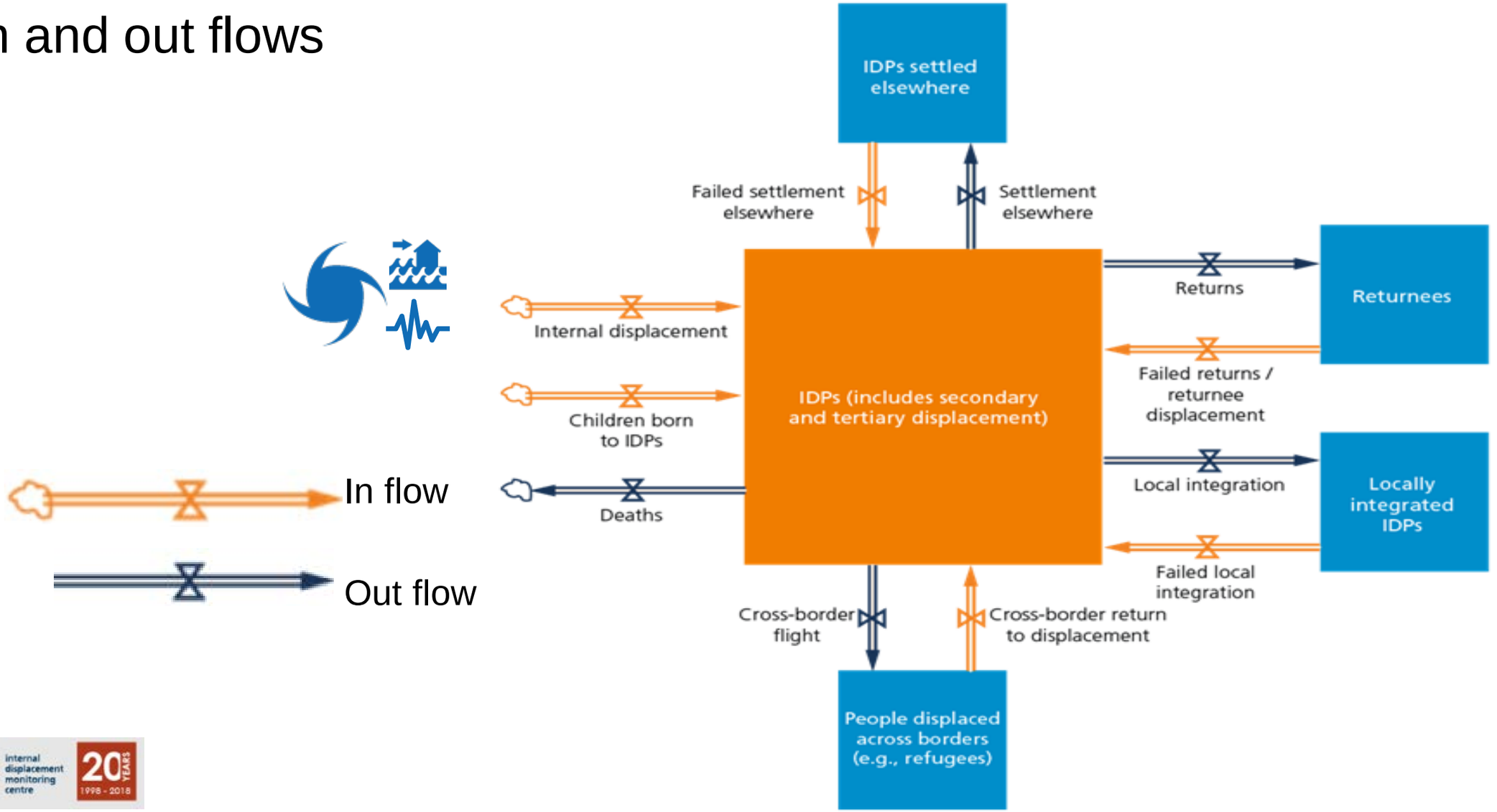
Problem Statement

- The number of internally displaced persons (IDPs) is on the rise
- The development of official definitions / statistics on IDPs has just begun
- Definitional issues are complicated
- Humanitarian groups need real-time data on displacement
- Such data are difficult to collect in challenging contexts (natural disasters, conflict)
- Novel data streams are increasingly being sought to fill the gaps



Monitoring is challenging

In and out flows

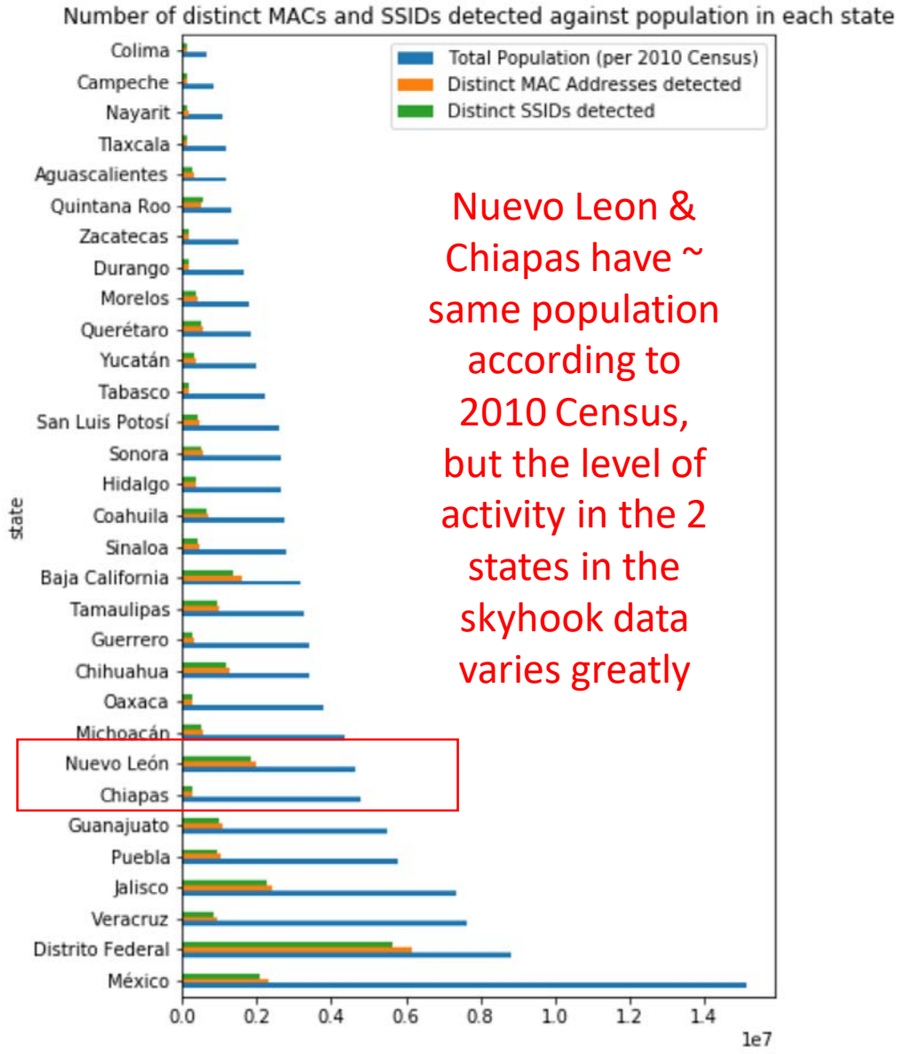


What are Skyhook data?

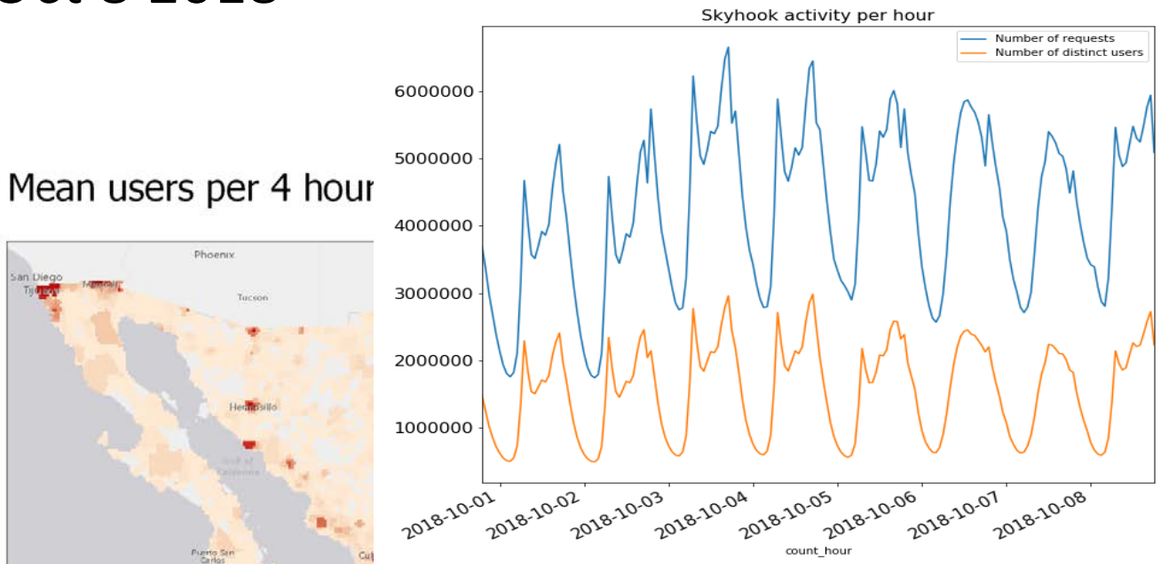
- Cross platform device location data based on wifi, GPS, and cell tower signals
  - MAC ID request and distinct user counts aggregated for 30m pixels
  - Hourly temporal resolution
- Advantages:
  - Includes smart phones, computers, laptops, and gaming devices (not just phones)
  - Does not require separate MOUs with each cell phone provider
- Disadvantages:
  - Can’t track individual devices (as one would cell phones)
  - Devices need to be data enabled (not simple flip phones) – which introduces a potentially biased sample, esp. in least developed countries

Total population by state in Mexico against device numbers

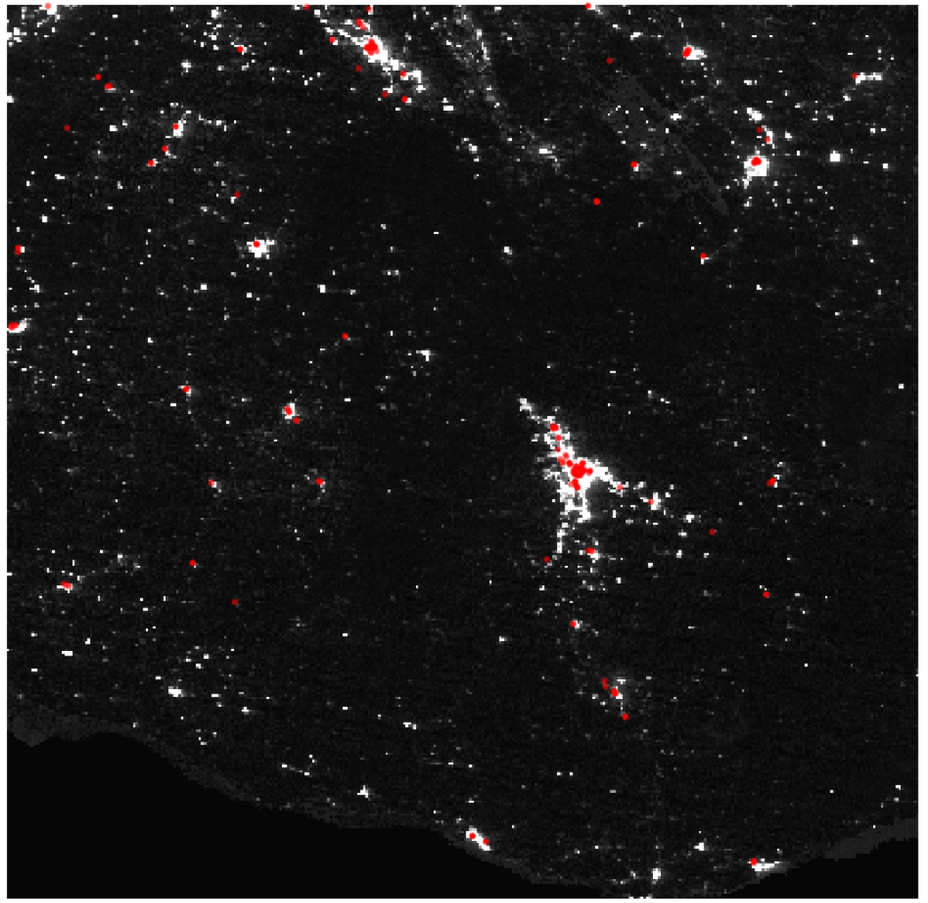
- Sample data from Mexico visualized by state exhibits wide variation between mobile activity and 2010 census population estimates



Hourly MAC ID request and distinct user counts, aggregated by 100m x 100m tiles in Mexico, from Sep 30 - Oct 8 2018



Hourly Skyhook Mobility Data & VIIRS DNB - Oaxaca, Feb 11 to 24 2020



Jakarta, Indonesia

- Major flood occurred January 1, 2020
- 400mm of rain on New Year's eve

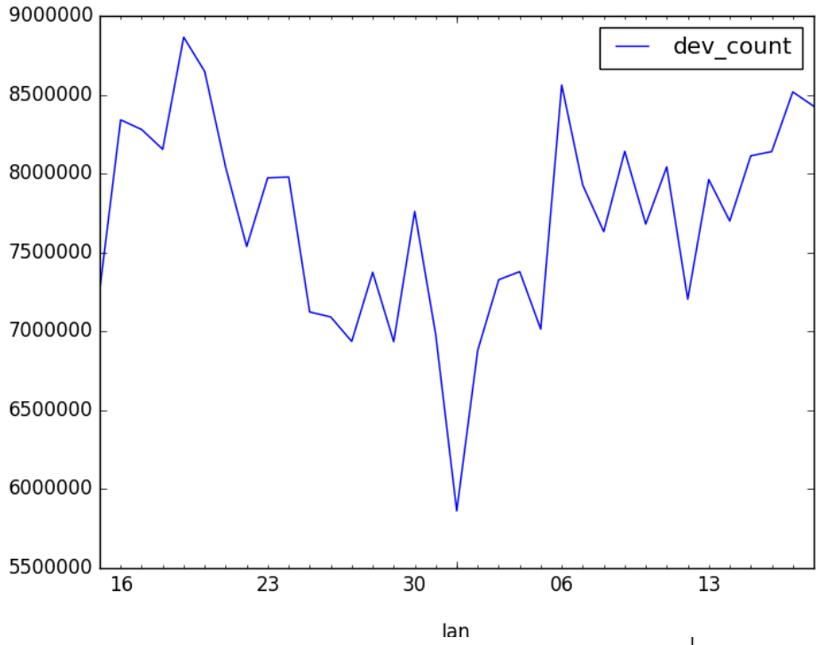


Figure 1: The total number of devices that connected to internet endpoints across Jakarta, West Java, and Banten provinces of Indonesia on each day from Dec 16, 2019 to Jan 17, 2020 from Skyhook data

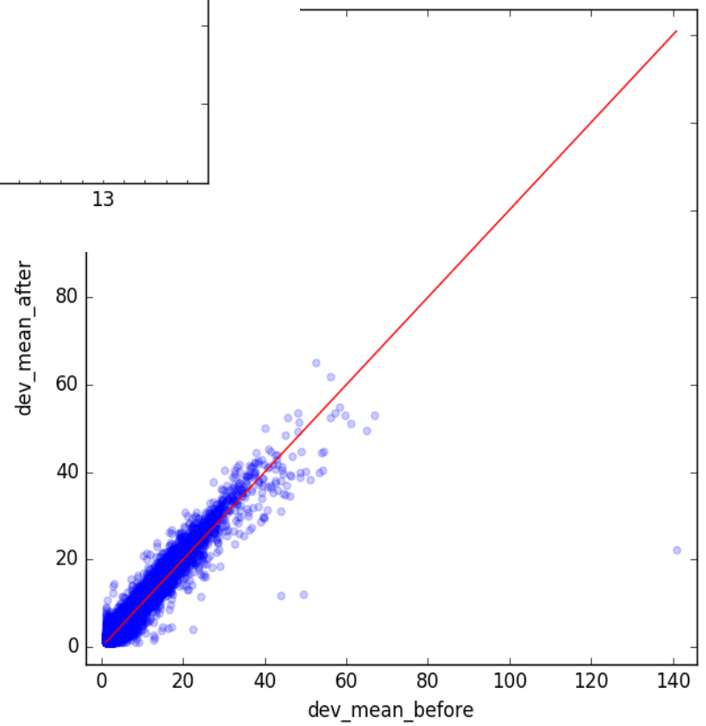
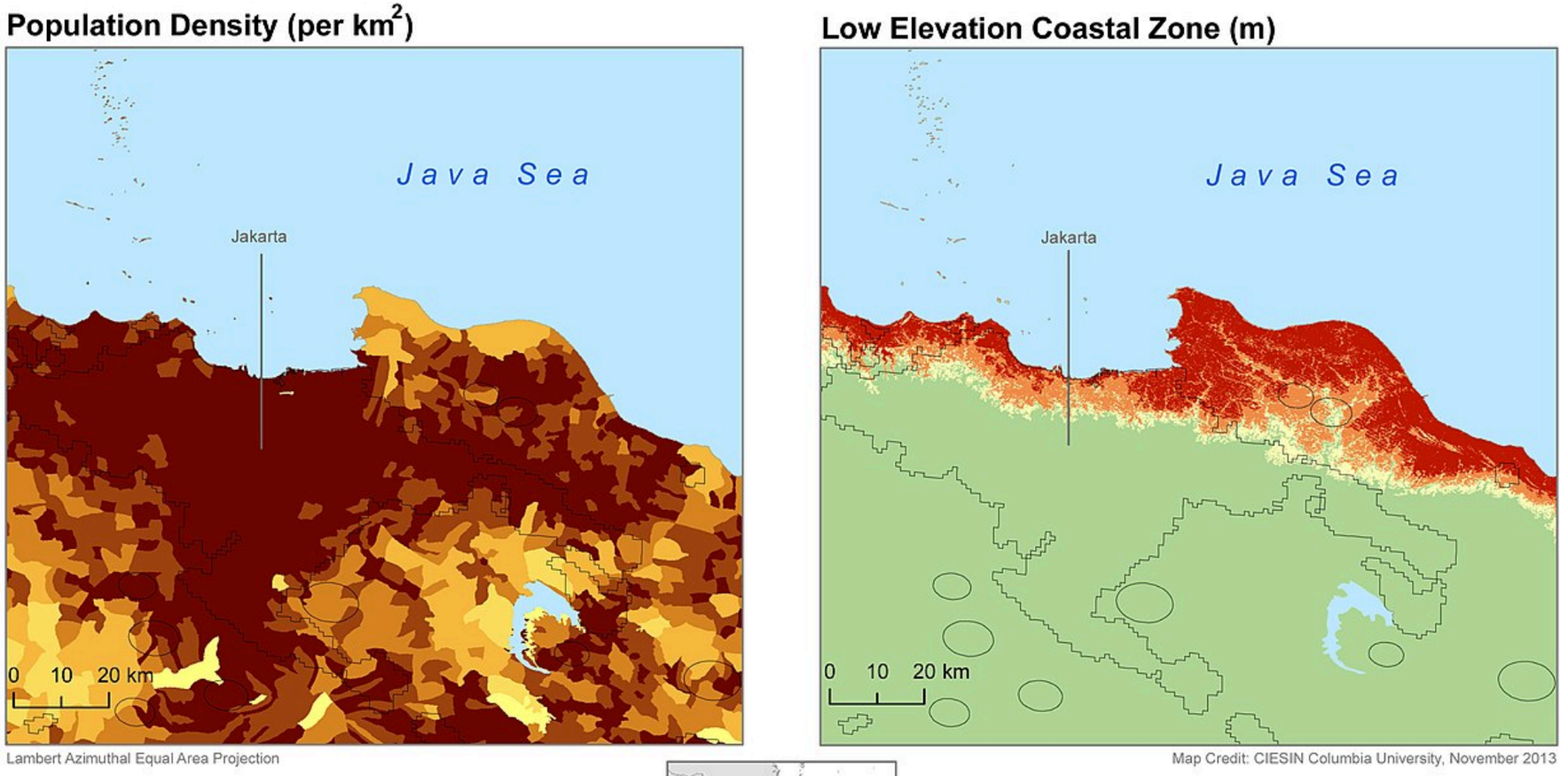


Figure 2: The average of device counts for each 100m x 100m tile in the Skyhook data before (X-axis) and after (Y-axis) the floods

Urban-Rural Population and Land Area Estimates, Version 2, 2010: Jakarta, Indonesia

Low Elevation Coastal Zone



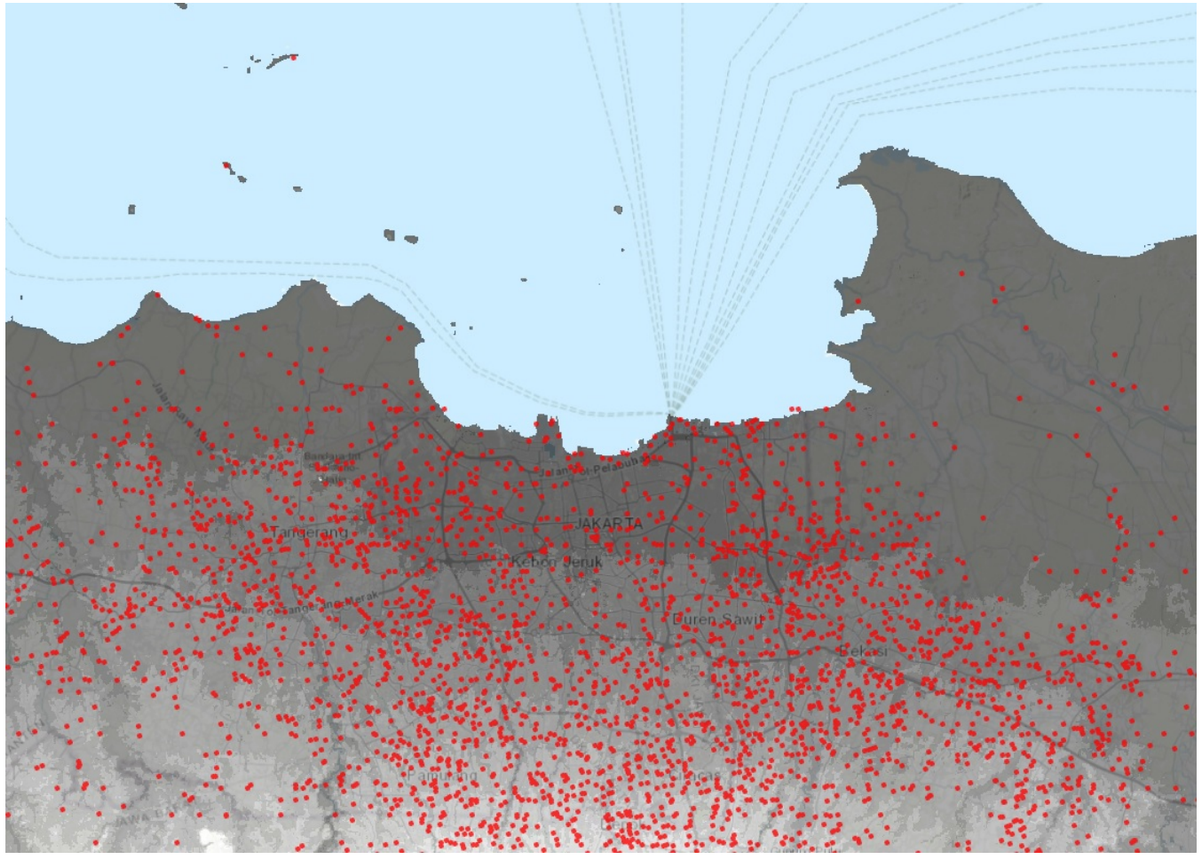
Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Data Source: Center for International Earth Science Information Network (CIESIN) /Columbia University. 2013. Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 2. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/lecz-urban-rural-population-land-area-estimates-v2.

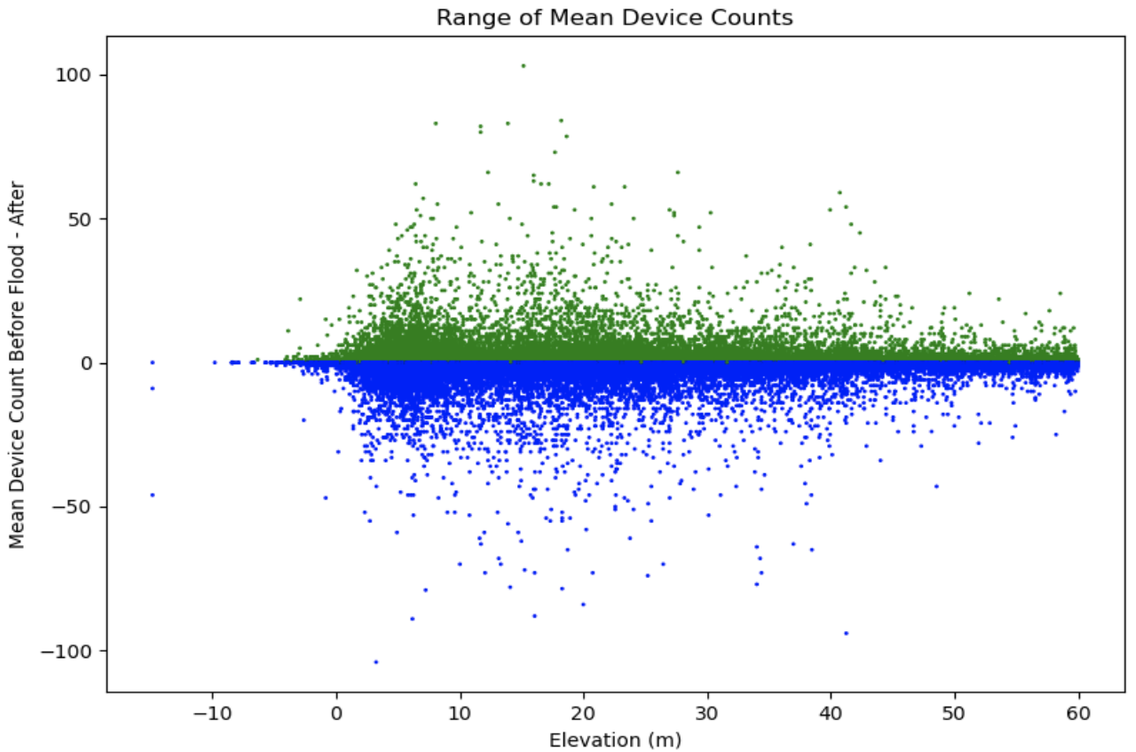
The Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates Version 2 data set provides continent-level and country-level estimates of land area and urban, rural, and total population for 202 statistical areas (countries and other UN recognized territories). Population inputs were derived from Gridded Rural-Urban Mapping Project, version 1 (GRUMPv1). Elevation data were derived from the Shuttle Radar Topographic Mission (SRTM) 90 meter data set.

Jakarta lies in a low, flat basin, averaging 7 meters above sea level and 40% of Jakarta, particularly the northern areas, is below sea level, while the southern parts are comparatively hilly. (*source:* https://en.wikipedia.org/wiki/Floods_in_Jakarta)

Change in device counts before and after flood

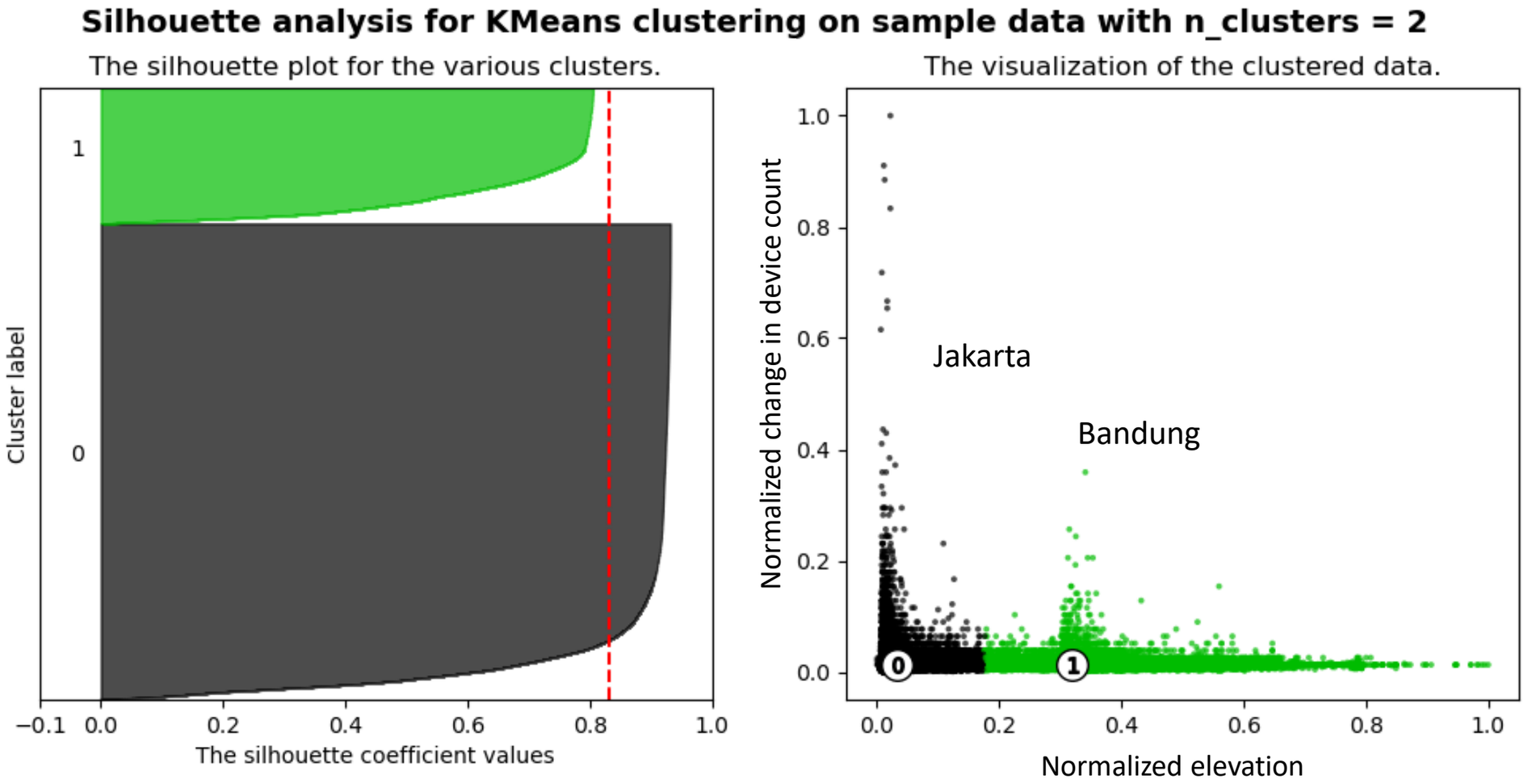


100m x 100m tiles from Skyhook where the average number of devices used increased after the floods. The red dots indicate the areas where the number of devices connecting to internet endpoints were higher after the floods when compared to the same number before the floods.



Blue dots represent a decrease in the mean device count and green represent an increase

Cluster Analysis



Building on IDMC's IDETECT: Natural Language Processing

reliefweb.int/report/mozambique/unhcr-mozambique-factsheet-cyclone-idai-june-2019

UNHCR Mozambique - Factsheet: Cyclone Idai, June 2019

News and Press Release • **Source:** UNHCR • **Posted:** 14 Jul 2019 • **Originally published:** 10 Jul 2019 • **Origin:**

View original

On 14 March 2019, Tropical Cyclone Idai made landfall as a category four Cyclone near Beira City, Sofala region, causing human losses, massive destruction of infrastructure, households, and significant amount of internal displacement. The disaster exacerbated preexisting vulnerabilities. An impressive total of 1.85 million people have been affected.

As of 1 April 2019, over 146,000 internally displaced persons (IDPs) sought refuge in 155 temporary sites across four provinces (Sofala, Manica, Zambezia, Tete). Many IDPs have since then been in the process of returning home, while others were relocated to permanent sites allocated by the Government of Mozambique (GoM). The last temporary accommodation were closed in June 2019.

Download document

(PDF | 818.41 KB)

Primary country:
Mozambique

Source:
UN High Commissioner for Refugees

Disaster:
Tropical Cyclone Idai - Mar 2019

Format:
News and Press Release

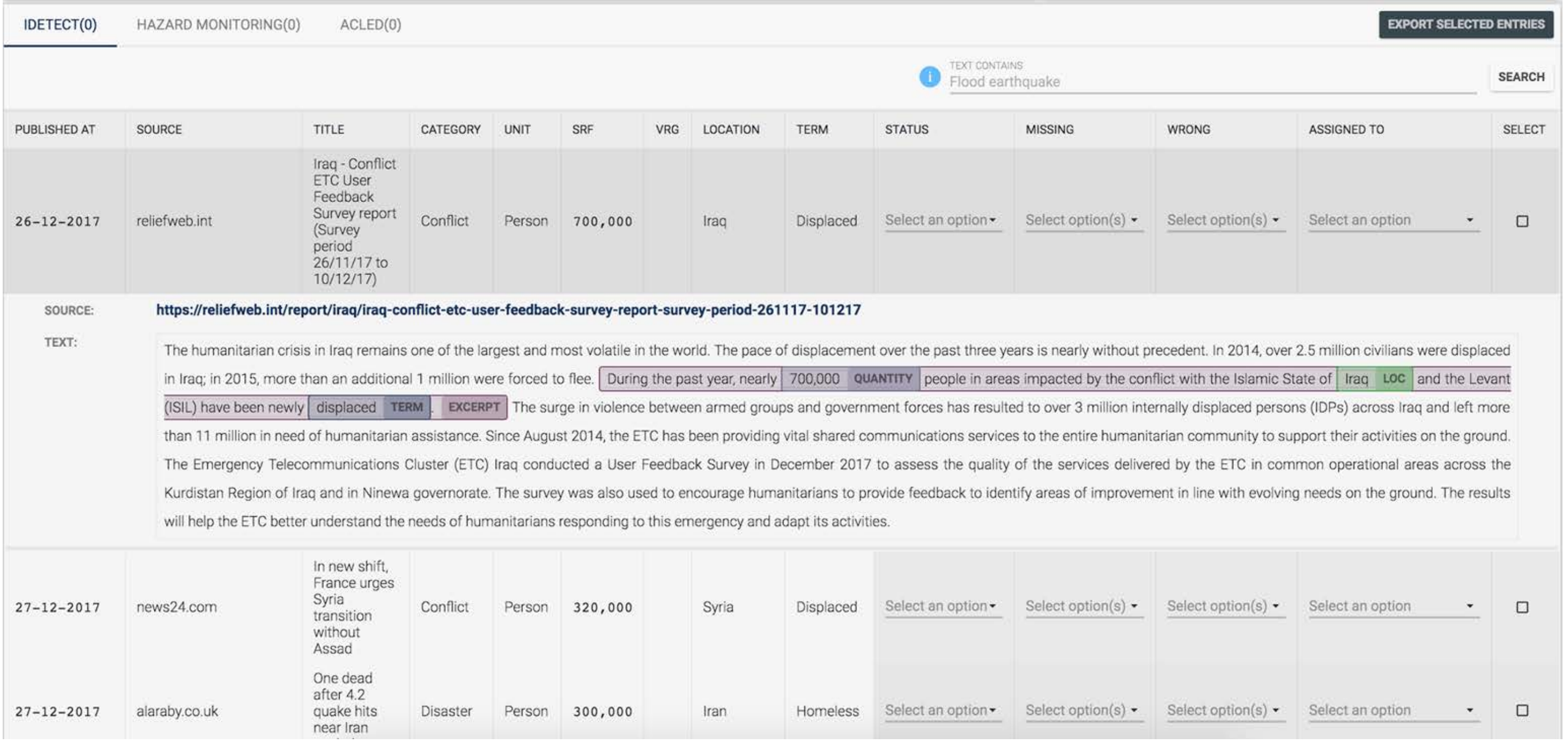
Themes:
Education / Protection and Human Rights / Shelter and Non-Food Items

When?

What?

How many?

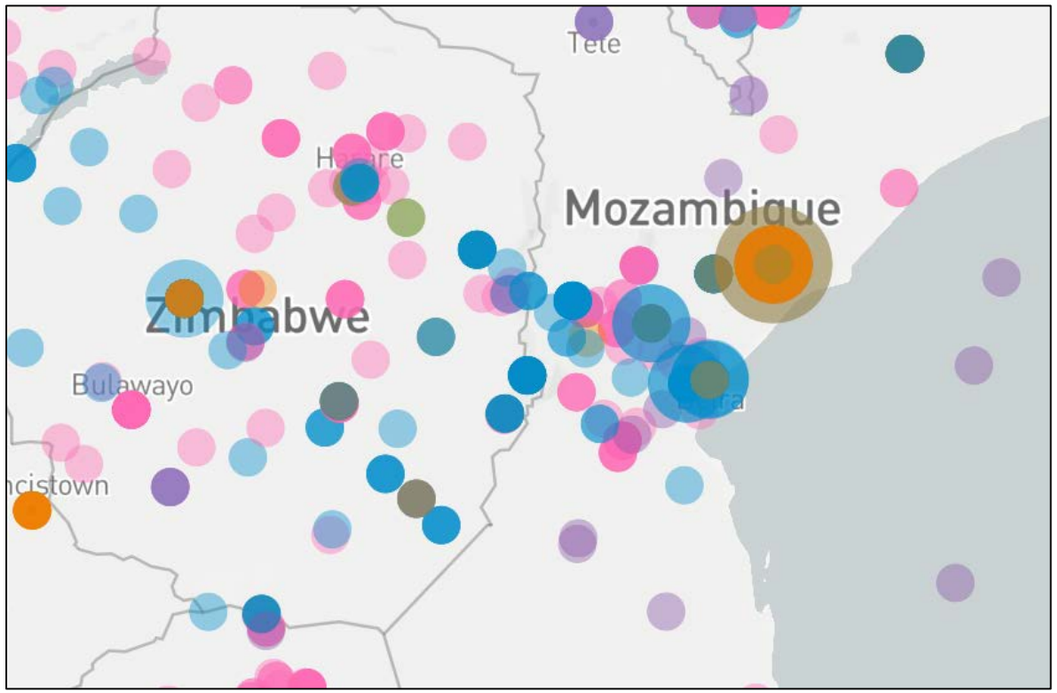
Where?



https://idmc-mm.surge.sh/

Slide courtesy of IDMC

CIESIN – DSI Collaboration (with ISI Foundation and IDMC)



Developing Annotated Resources for Internal Displacement Monitoring

Fabio Poletto
fabio.poletto@isi.it
ISI Foundation
Turin, Italy

Yunbai Zhang
yz3386@columbia.edu
Earth Institute, Columbia University
Palisades, NY, USA

André Panisson
andre.panisson@isi.it
ISI Foundation
Turin, Italy

Yelena Mejova
yelena.mejova@isi.it
ISI Foundation
Turin, Italy

Daniela Paolotti
daniela.paolotti@isi.it
ISI Foundation
Turin, Italy

Sylvain Ponserre
sylvain.ponserre@idmc.ch
Internal Displacement Monitoring Centre
Geneva, Switzerland

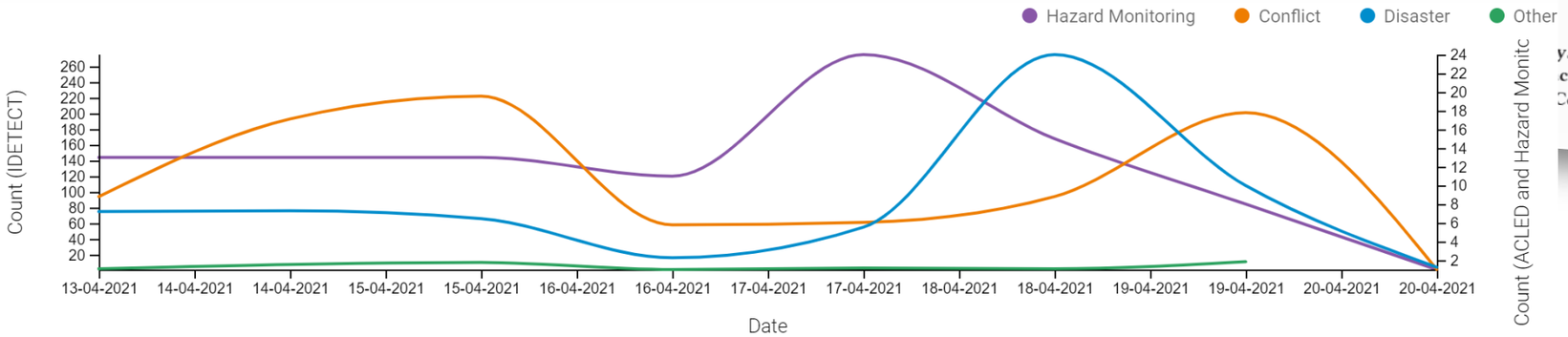
ABSTRACT

This paper describes in details the design and development of a novel annotation framework and of annotated resources for Internal Displacement, as the outcome of a collaboration with the Internal Displacement Monitoring Centre, aimed at improving the accuracy of their monitoring platform IDETECT. The schema includes multi-faceted description of the events, including cause, quantity of people displaced, location and date. Higher-order facets aimed at improving the information extraction, such as document relevance and type, are proposed. We also report a case study of machine learning application to the document classification tasks. Finally, we discuss the importance of standardized schema in dataset benchmark development and its impact on the development of reliable disaster monitoring infrastructure.

unstructured texts. We pursue this by developing a novel framework for annotating ID in online news articles and by applying it to five datasets covering three languages: English, Spanish, and French.

Internal displacement (ID) is defined as the forced movement of people, who have to leave their home or place of habitual residence, within the country they live in. The major causes of ID are armed conflicts and disasters resulting from natural hazards, but other factors such as development plans, geophysical hazards or effects of climate change may as well displace people. Unlike migrants or refugees, Internally Displaced People (IDPs) do not cross any border and remain within their country. Depending on several factors, their condition may last indefinitely, and can get worse when institutions are not able, or willing, to provide assistance and protection, or when the place where they took shelter presents other risks. Although often neglected by media, ID is a massive phenomenon: in 2019 alone, 33.4 million new displacements have been recorded across 145 countries, of which about three fourths are caused by disasters resulting from natural hazards.

This work's rationale lies in a collaboration with Internal Displacement Monitoring Center (IDMC)[1], an NGO dedicated to monitoring ID worldwide and to providing data, analyses and support



Conclusions

Lessons from the “big data” approach to modeling migration and displacement

- Migration data are scant and often not comparable (e.g. stocks, flows, differing time periods)
- Using past and future population distributions offers potential solutions
- Results at local levels must be treated with caution
- However, results on aggregate are plausible
- Modeling future migration is fraught with uncertainty
  - Future patterns are likely to be affected by economic interdependencies, conflicts, and national policies in ways that can never be fully foreseen
  - Yet, if the models cause policy makers to consider the potential migration impacts of climate change for the first time, that is a useful result
- Big data for displacement monitoring – a holy grail but we’re not there yet