Geographic and Remote Sensed Data

STEG Virtual Course "Data in Macro Development"

David N. Weil

Brown University and NBER

May 17, 2024

"Data! Data! Data!.. I can't make bricks without clay."

— Sherlock Holmes

"The government are very keen on amassing statistics. They collect them, add them, raise them to the n^{th} power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damn pleases."

— Josiah Stamp

Remote-Sensed Data

- Data collected from above earth.
- Historically, this included balloons, kites, airplanes, and even pigeons!
- Today it is mostly satellites, but also airplanes and drones.

…pigeons!

- The Bavarian pigeon corps, 1903

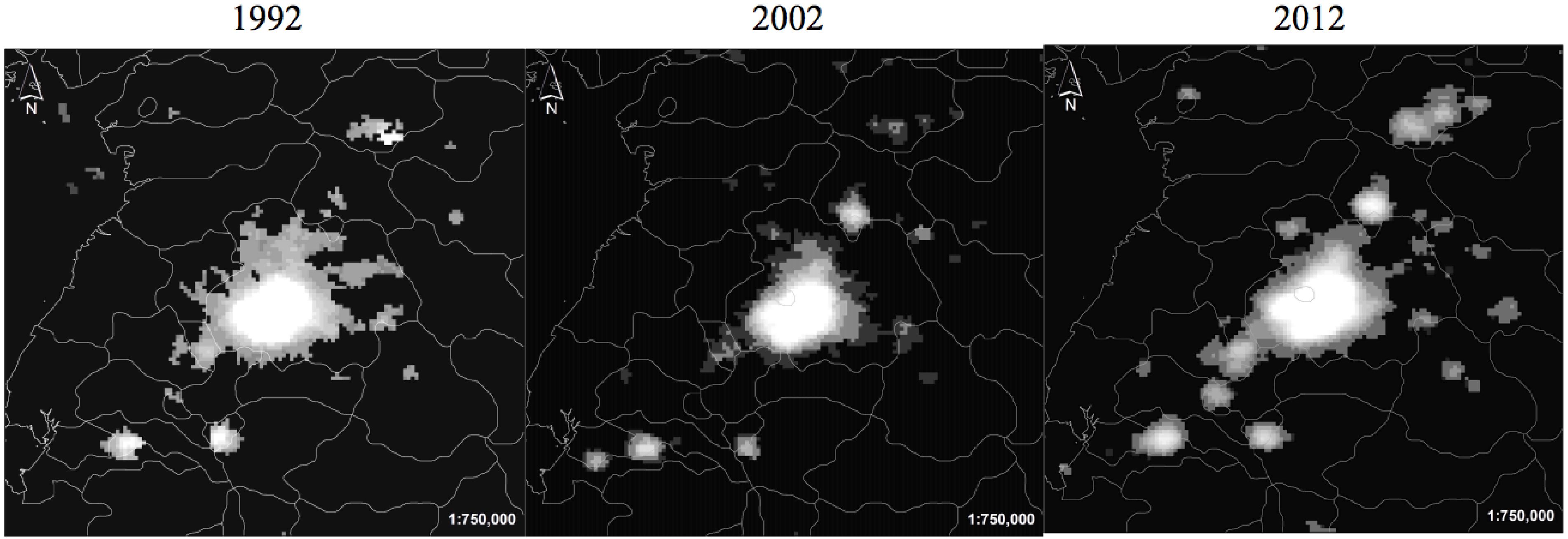
Image source: http://www.cameramuseum.ch/fr/N31124/des-pigeons-photographes.html

Advantages of Remote-Sensed Data

1. Satellite data exist where other data do not.

North Korea responds to sanctions by shifting activity toward Pyongyang, Chinese trade hubs (Lee 2018)

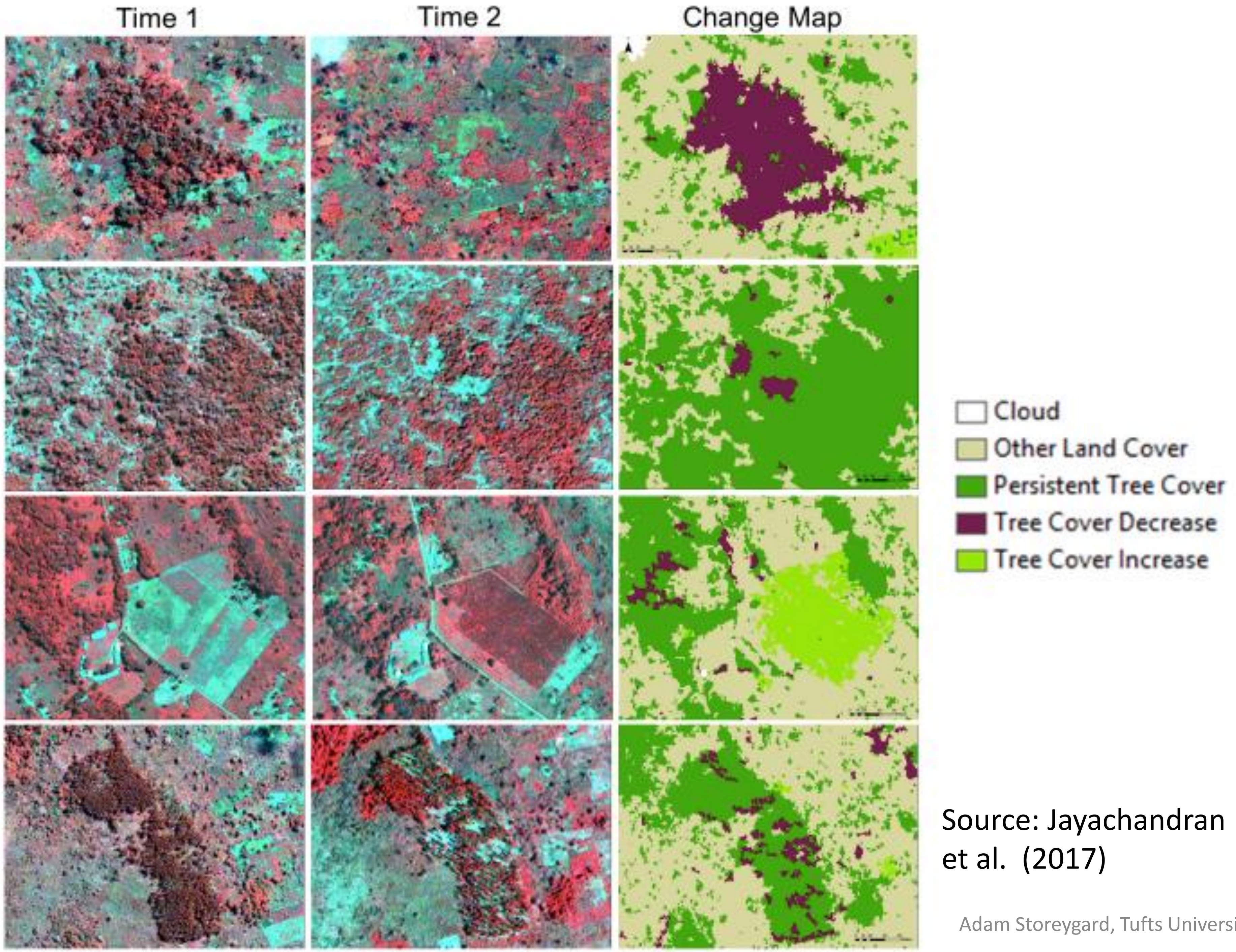
Figure 6. Lights near Pyongyang in 1992, 2002, and 2012.



Advantages of Remote-Sensed Data

1. Satellite data exist where other data do not.
2. High spatial resolution.

RCT pays households not to cut down trees. Satellite data allows for measurement down to the level of household plots.



Source: Jayachandran et al. (2017)

Agricultural output

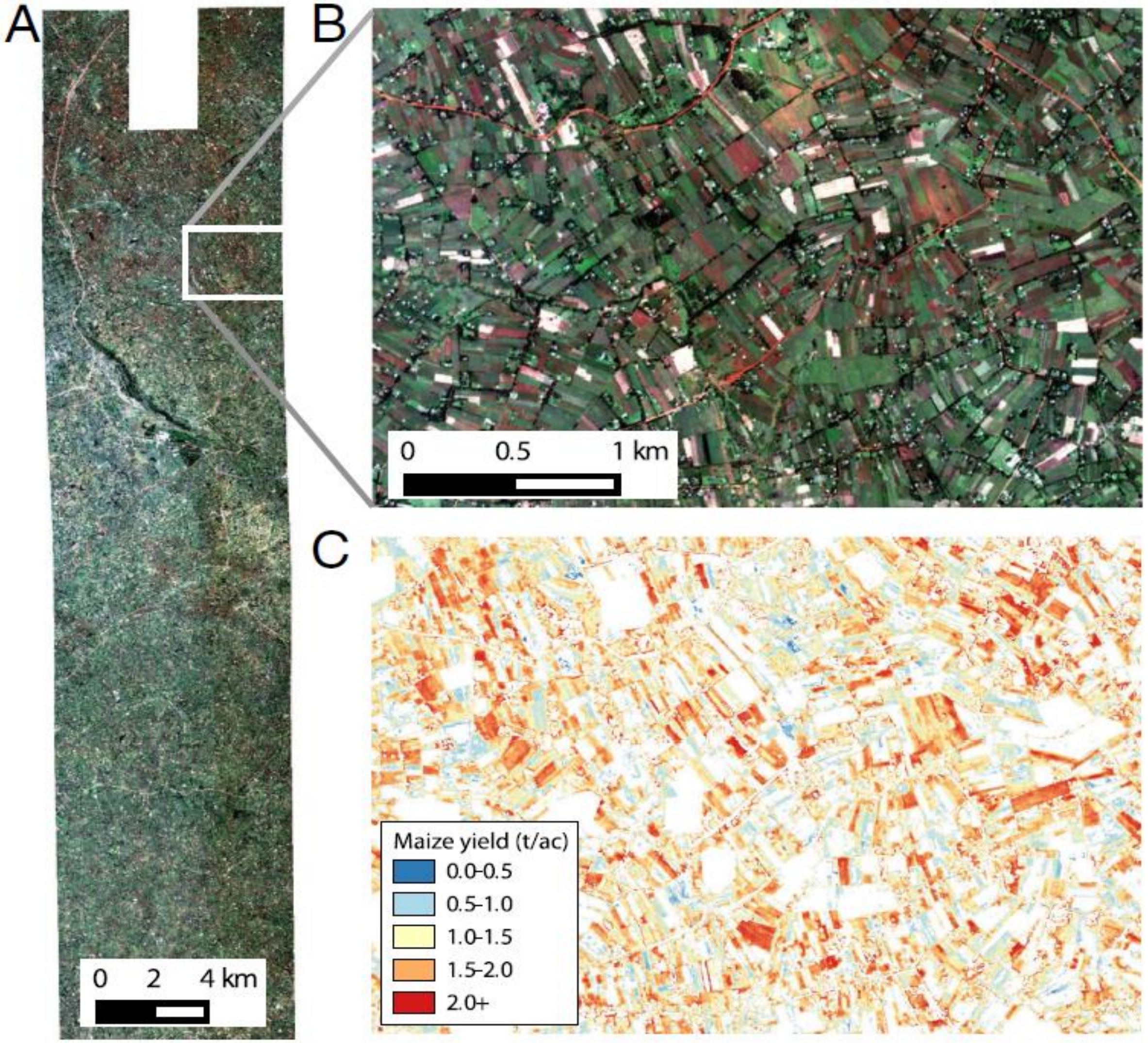


Fig. 5. Maize yield map for the study region, 2015. (*A* and *B*) One-meter image from Terra Bella of the study region (*A*) and zoom-in of that image (*B*) (see Fig. S3 for a higher-resolution version). (*C*) Yield map of the zoomed-in region for pixels classified as maize.

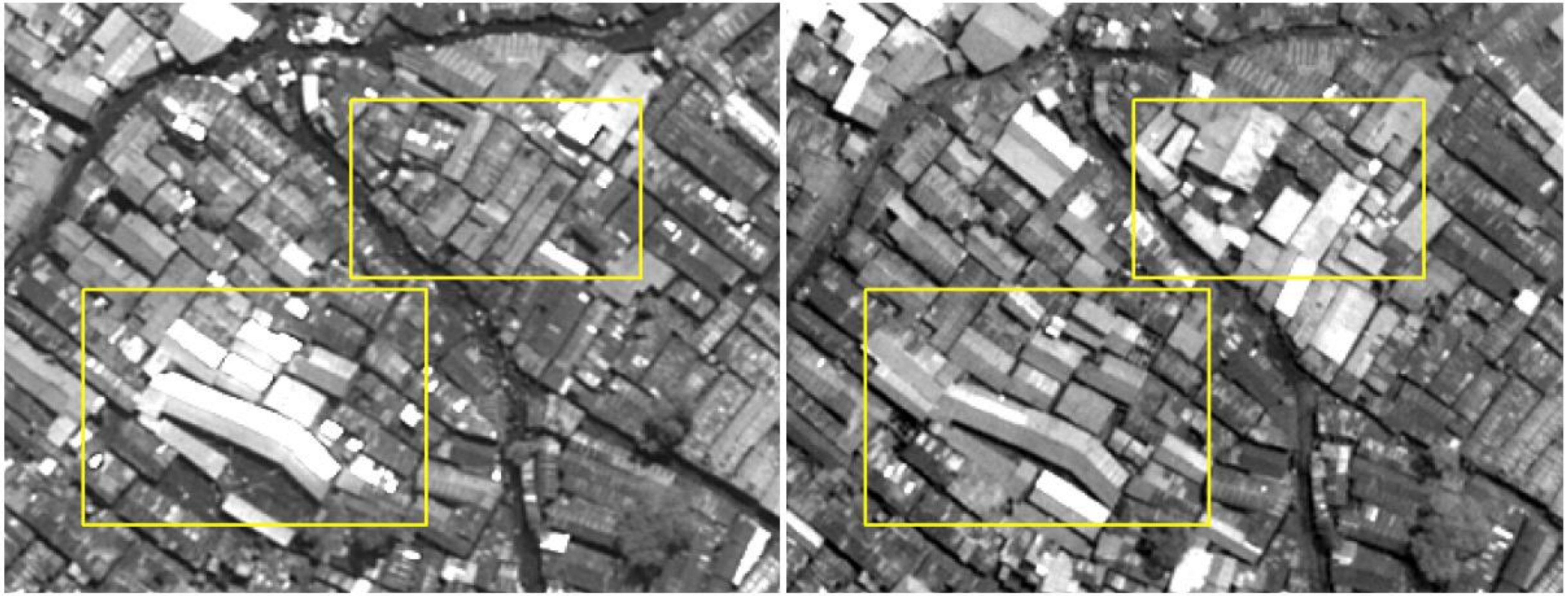
Source: Burke and Lobell (2017)

Advantages of Remote-Sensed Data

1. Satellite data exist where other data do not.
2. High spatial resolution.
3. Cheap/easy repeat measurements.

Urban land cover: roofs

Appendix Figure A2: Old and New Roofs in Kibera

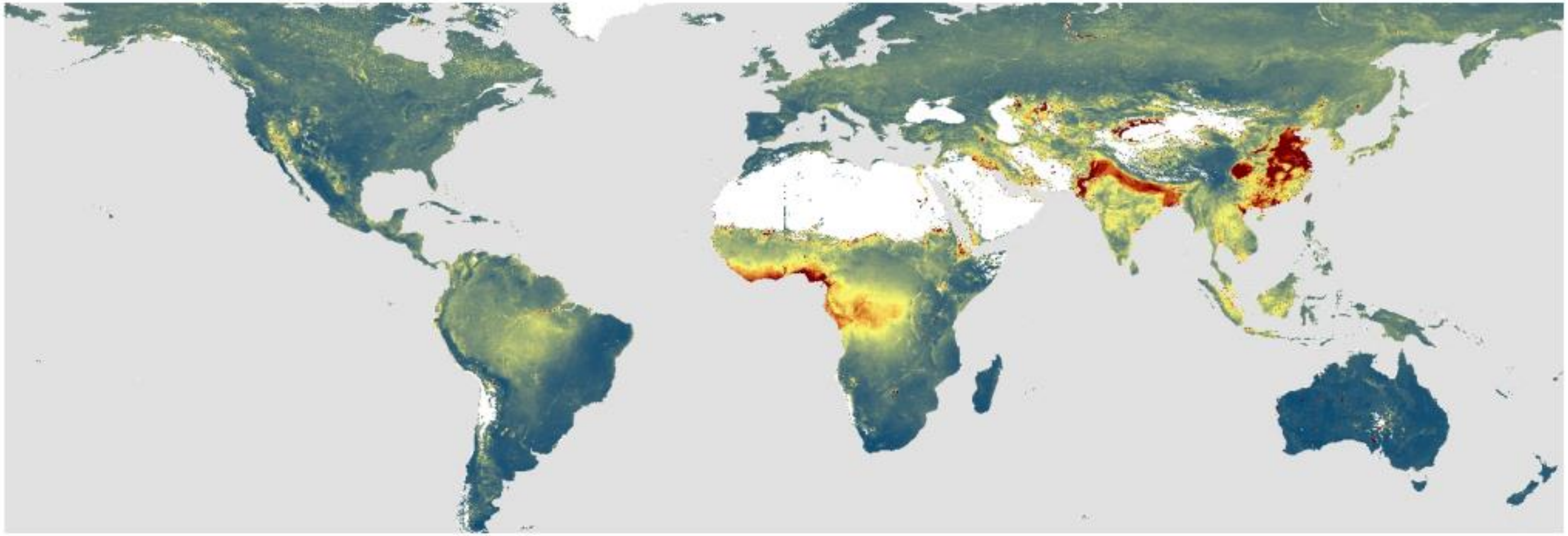


Note: Both pictures are taken over the same area of the slum with the same resolution (0.5 meters panchromatic).
The picture in the left panel was taken in July 2009 and that in the right panel in August 2012.
The yellow rectangles highlight clusters of roofs that markedly evolved over the period.
Roofs highlighted in the bottom rectangle degraded while roofs within the top rectangle were upgraded in the same timeframe.
The picture area is approximately 175 meters long and 140 meters wide.

Source: Marx et al. (2019)

Advantages of Remote-Sensed Data

1. Satellite data exist where other data do not.
2. High spatial resolution.
3. Cheap/easy repeat measurements.
4. Near-global coverage.



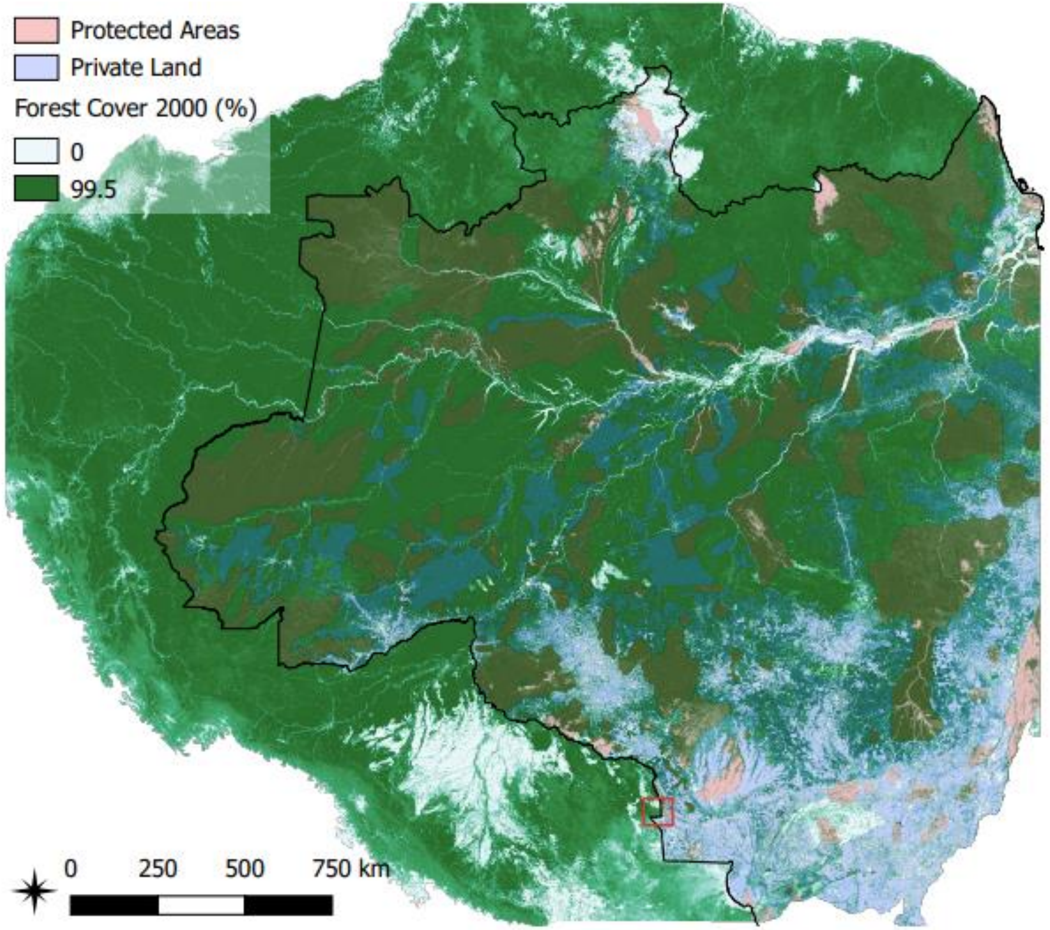
Aerosol Optical Depth, Average 2000-2014, Terra

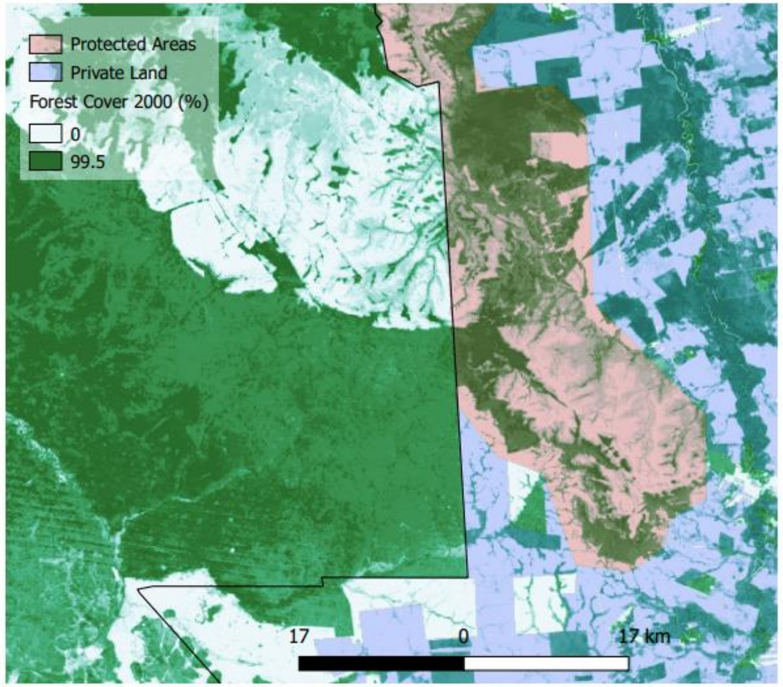
Gendron-Carrier, Nicolas, et al. “Subways and urban air pollution.” *American economic journal: Applied economics* 14.1 (2022): 164-196.

Advantages of Remote-Sensed Data

1. Satellite data exist where other data do not.
2. High spatial resolution.
3. Cheap/easy repeat measurements.
4. Near-global coverage.
5. Consistent coverage across borders.

Deforestation in Brazil vs. Neighbors

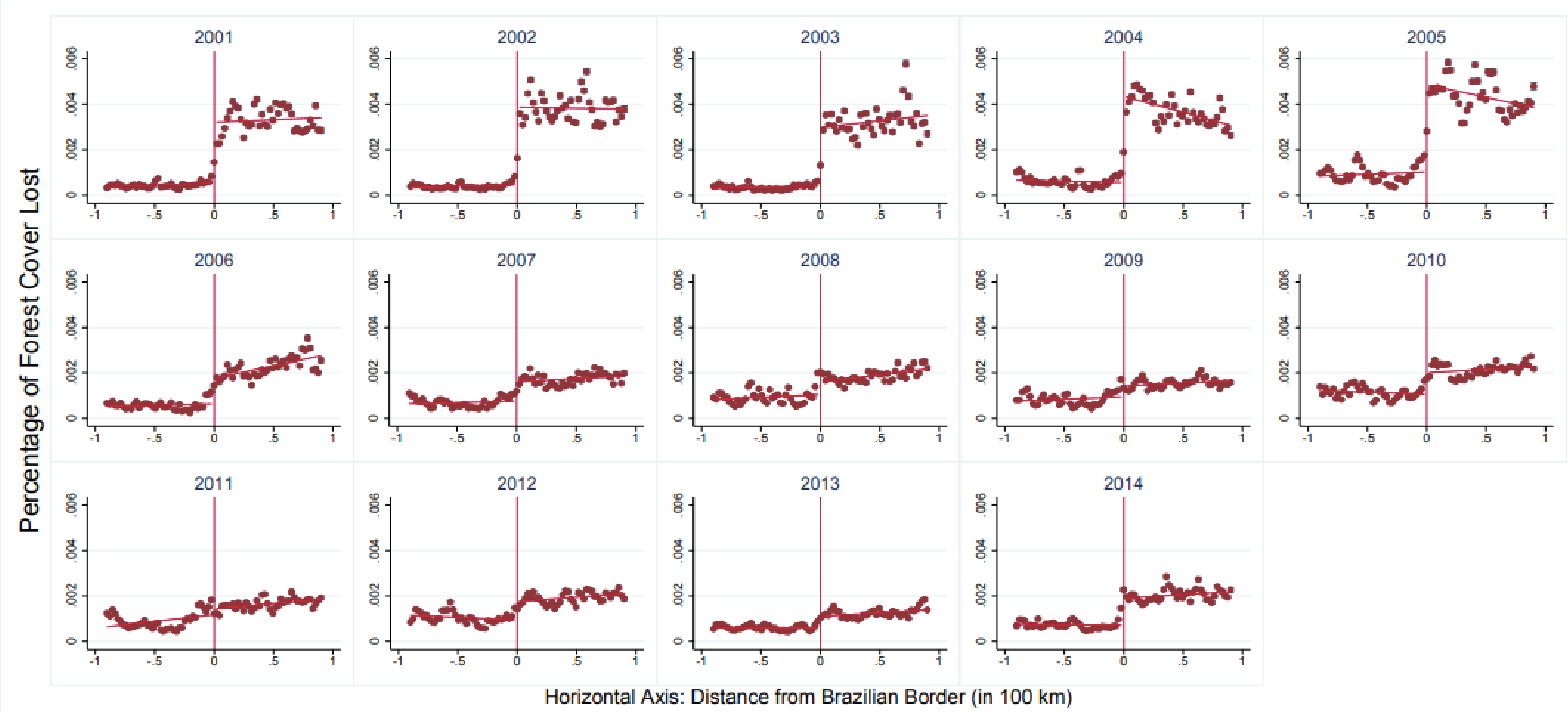




(b) Example of area of border with Bolivia

Source: Burgess, Costa and Olken (2018)

Deforestation in Brazil vs. Neighbors



Conclusion: Brazil's policy had a large impact reducing deforestation starting in 2006

Advantages of Remote-Sensed Data

1. Satellite data exist where other data do not.
2. High spatial resolution.
3. Cheap/easy repeat measurements.
4. Near-global coverage.
5. Consistent coverage across borders.
6. Independent of official reporting mechanisms.
    - Both national governments and international agencies.

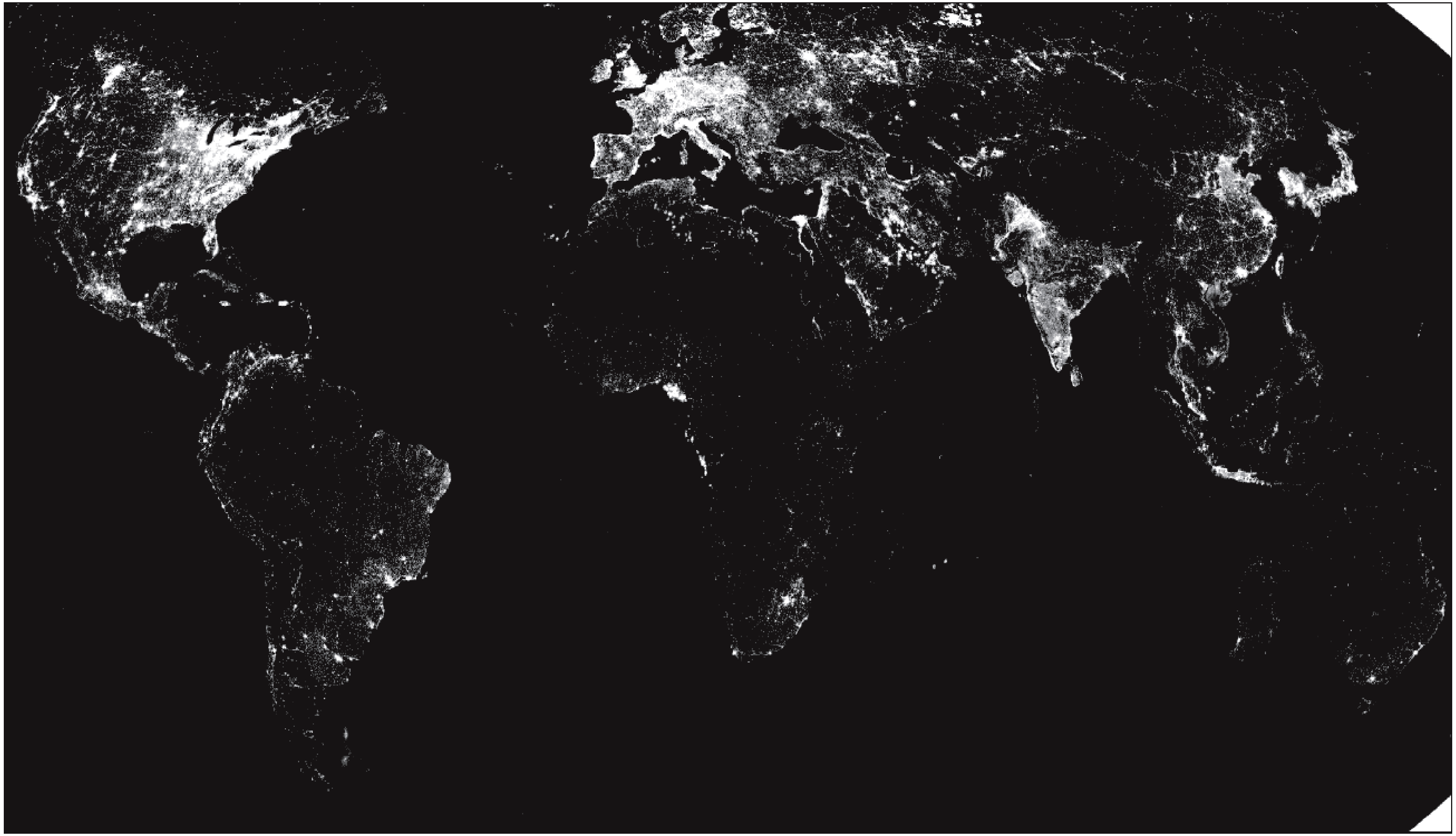
Inquiry finds World Bank officials, including now-I.M.F. chief, pushed staff to inflate China data.

An investigation focused on accusations that top World Bank officials pressured the team conducting an annual survey to improve China's standing.

Where Things Stand Today

- The quantity of remote sensed data is exploding!
  - Cheaper satellites
  - Cheaper launches
  - Competition in launch market: SpaceX, etc.
  - New launch technologies such as reusable rockets
- Accessibility is also improving very rapidly (e.g. NASA Black Marble).
- New machine learning tools have the potential to unlock even more value in this firehose of data.

Biggest Application to Date: Night Lights and GDP



Night Lights Datasets

- Defense Meteorological Satellite Program (DMSP) satellites: 1992-2013
- Visible Infrared Imaging Radiometer Suite (VIIRS): 2011-

DMSP Data (1992-2013)

- Each satellite observes whole Earth between 8:30 and 10pm local time every night
- Processing
  - Avoid sunlight, moonlight, aurorae, clouds
  - Remove forest fires, gas flares, squid fishing
- Annual composite nominal resolution is 30 arc seconds 0.86 x 0.86km at equator, with E-W distance ~ cos(latitude)
  - In practice, readings within 3-5 km are not fully independent ("blooming")
- Complicated issues with sensor calibration.
- Output is "digital number" from 0-63. Top-coded in bright areas like city centers.
  - Except for special "radiance calibrated" releases
- Error correction in processing leads to many values of zero.

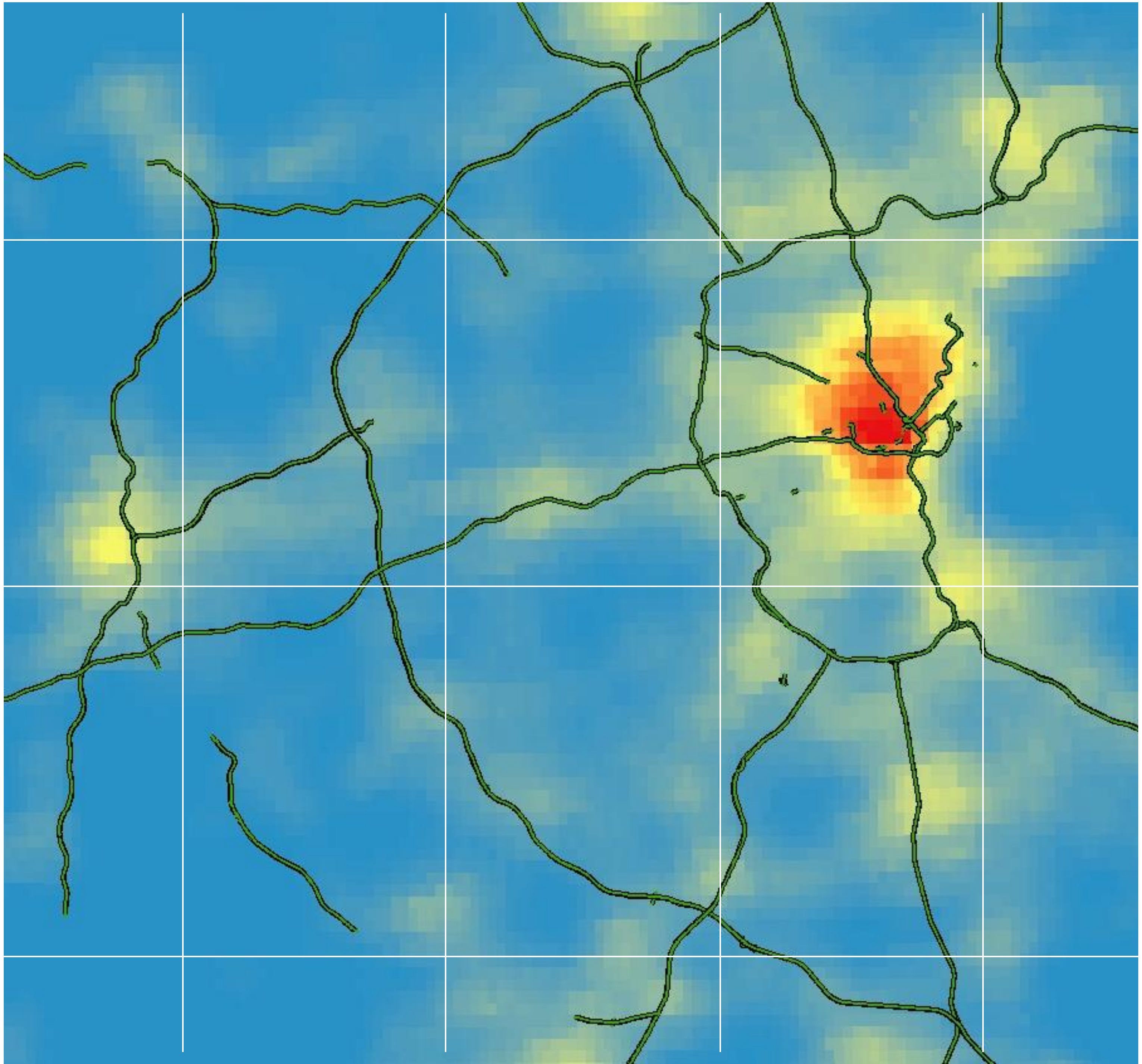
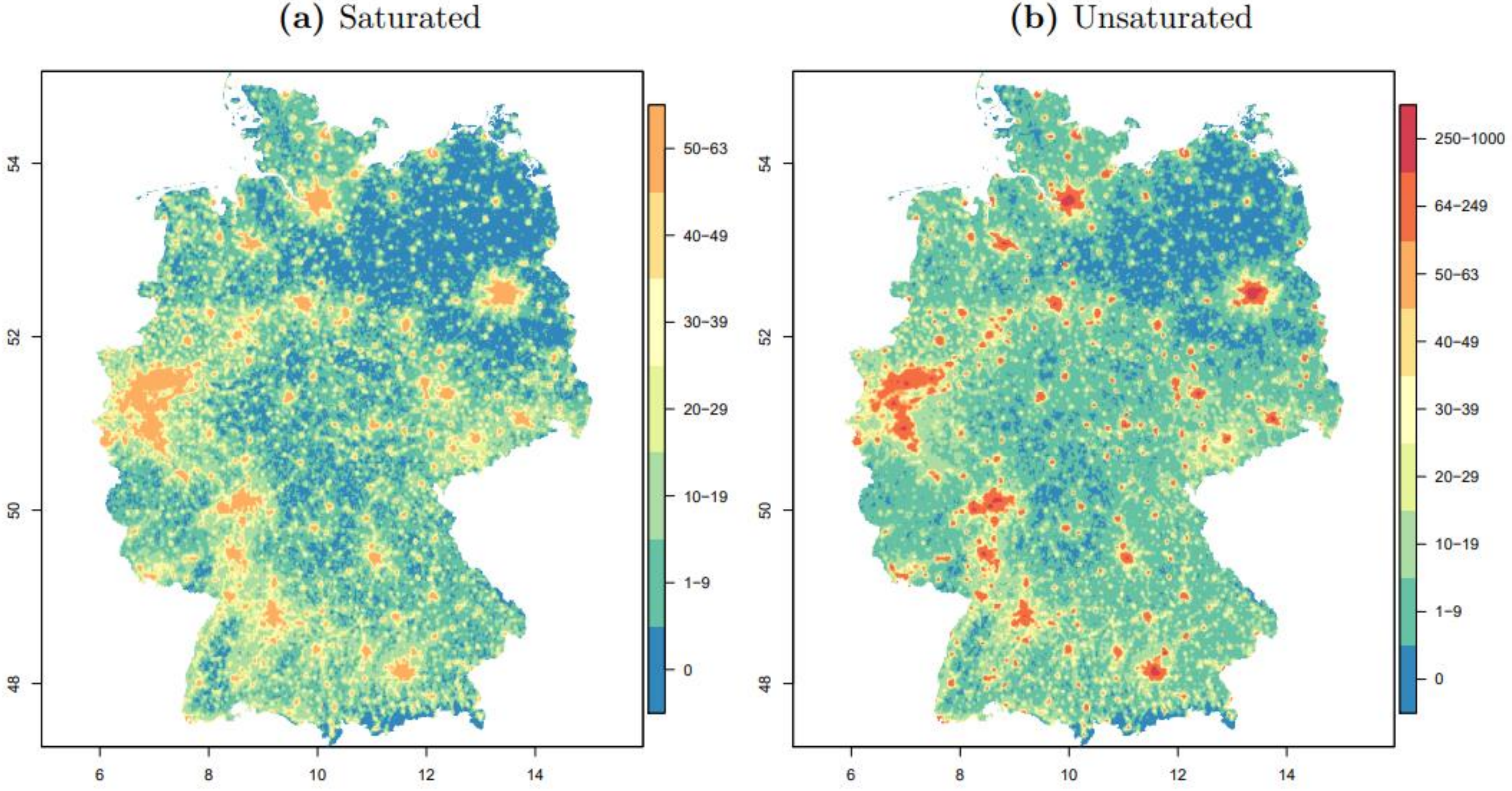
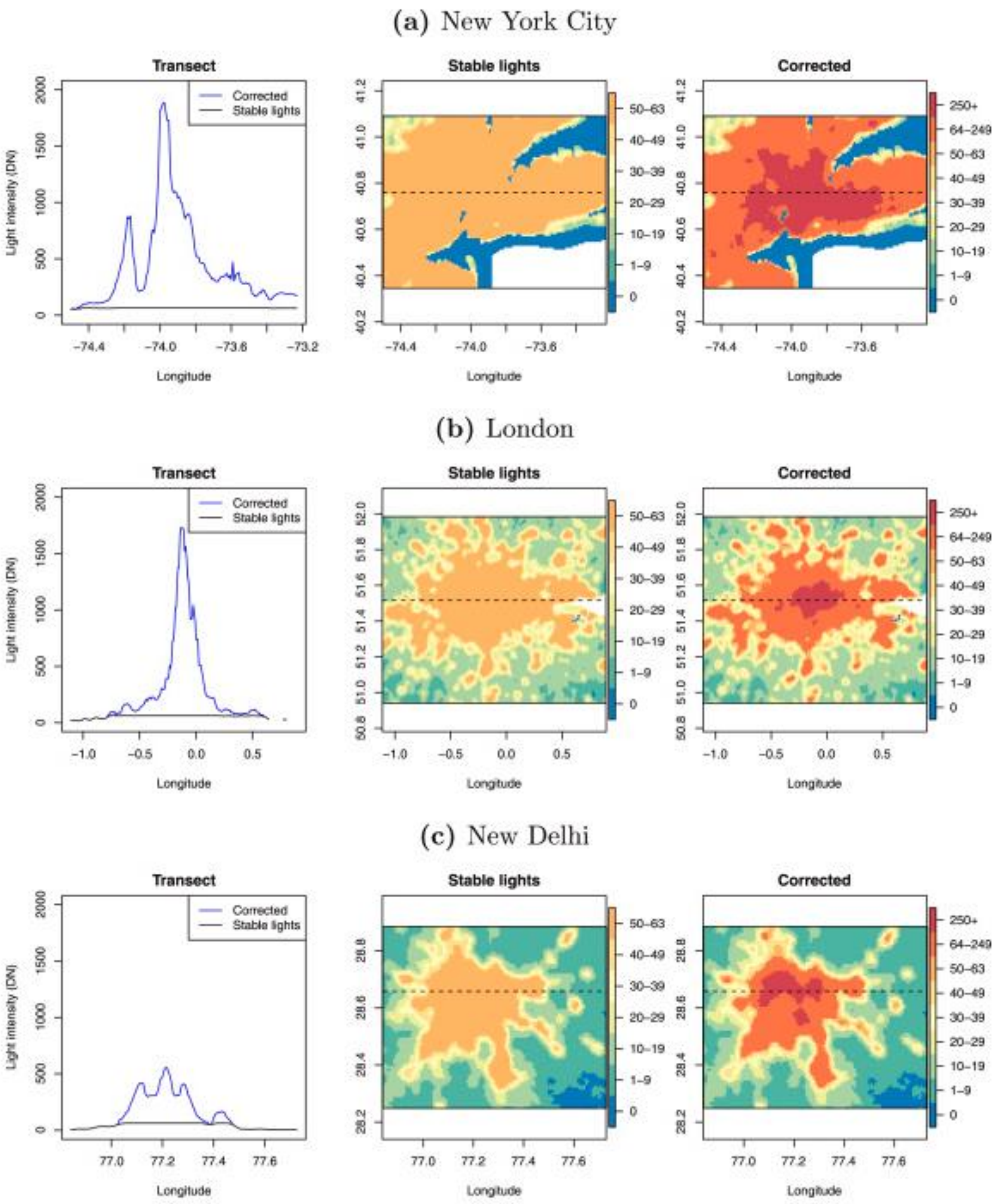


Figure 2: Germany in 1999, saturated and unsaturated lights



Bluhm and Krause, "Top Lights," *JDE* 2022

Notes: Illustration of the saturated and unsaturated data for Germany. Panel a) shows a map of Germany based on the stable lights data from satellite F121999 that is commonly used in the literature. Panel b) shows the same map using the radiance-calibrated data in 1999 from Hsu et al. (2015). Both data have been binned as shown in the legend and the color scales were adjusted so as to be comparable.



Bluhm and Krause, “Top Lights,” *JDE* 2022

DMSP Data for Selected Countries

Uncalibrated

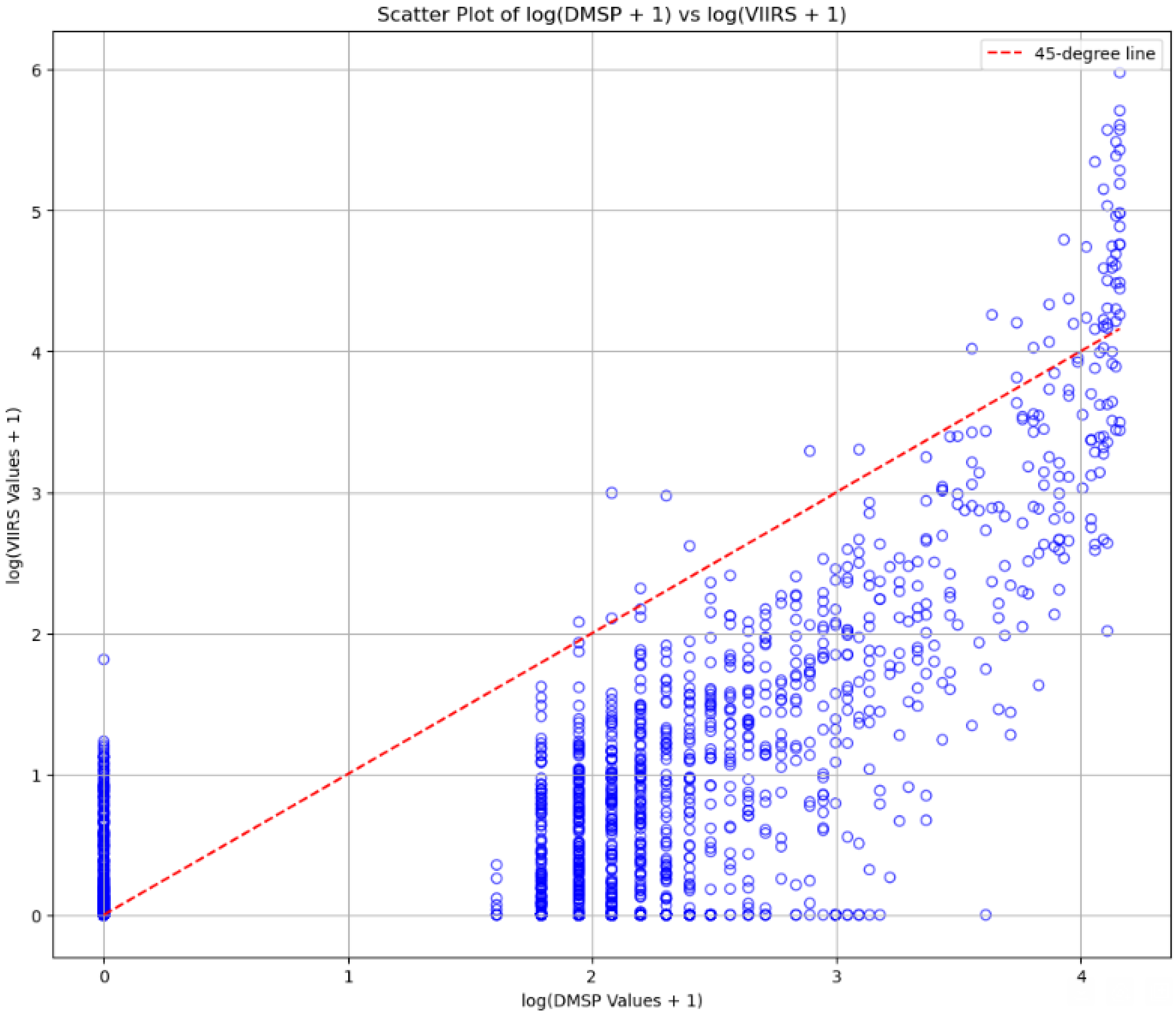
% in each cell	Bangla-desh	USA	Canada	Nether-lands	Brazil	Costa Rica	Guate-mala	Mada-gascar	Mozam-bique	Malawi
0	66.72%	71.79%	**95.24%**	**1.00%**	94.02%	59.26%	79.23%	99.73%	99.47%	97.67%
1-2	0.64%	0.10%	0.00%	0.00%	0.00%	1.06%	0.24%	0.00%	0.03%	0.00%
3-5	24.48%	9.96%	1.29%	3.45%	2.60%	24.79%	13.84%	0.15%	0.28%	0.93%
6-10	5.27%	8.83%	1.94%	24.04%	1.83%	9.26%	4.17%	0.06%	0.11%	0.85%
11-20	1.69%	4.17%	0.85%	28.83%	0.77%	3.00%	1.46%	0.03%	0.05%	0.27%
21-62	1.13%	4.62%	0.64%	41.10%	0.73%	2.33%	0.95%	0.03%	0.05%	0.27%
63	**0.06%**	**0.53%**	**0.04%**	**1.58%**	**0.06%**	**0.31%**	**0.10%**	**0.00%**	**0.00%**	**0.00%**
% area unlit	66.94%	67.68%	93.72%	1.05%	94.31%	60.70%	80.42%	99.74%	99.51%	97.15%
avg. DN	**2.0108**	**4.4622**	**0.7869**	**23.5244**	**0.6342**	**3.1401**	**1.4059**	**0.0233**	**0.0435**	**0.3010**
pop. den. (sq. km)	1080	31	3	469	21	76	105	26	23	125
percent urban	24	79	79	76	81	59	45	27	30	15
GDP p.c. PPP 05	917	37953	31232	32226	8046	8167	3905	892	546	672
GDP p. c. (2000 $)	344	33582	22657	23208	3760	4084	1693	249	252	143

VIIRS Data

- Onboard calibration: comparable physical quantities of light
- True resolution of ~750 meters
- Better atmospheric correction (clouds and more)
- No top-coding
- Less bottom-coding
- Essentially no solar contamination (~2am overpass)
- NASA will be distributing for shorter periods, even nightly
  - "Black marble"

Combining DMSP (1992-2013) and VIIRS (2011-)

- There are some papers that produce “harmonized” datasets.
  - Nechaev et al, “Cross-Sensor Nighttime Lights Image Calibration for DMSP/OLS and SNPP/VIIRS with Residual U-Net” *Remote Sensing* 2021.
  - Xi Li (Wuhan University) – several papers
- I have worries about comparability (8:30-10:00 PM vs. 2:00 AM)
- You certainly *can’t* take the lights-income elasticity measured in DMSP and apply it to VIIRS data.



This is for a random sample of 10,000 (out of 212 million) pixels.

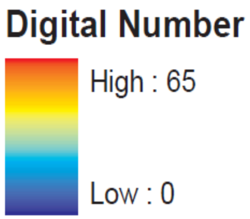
2013 data

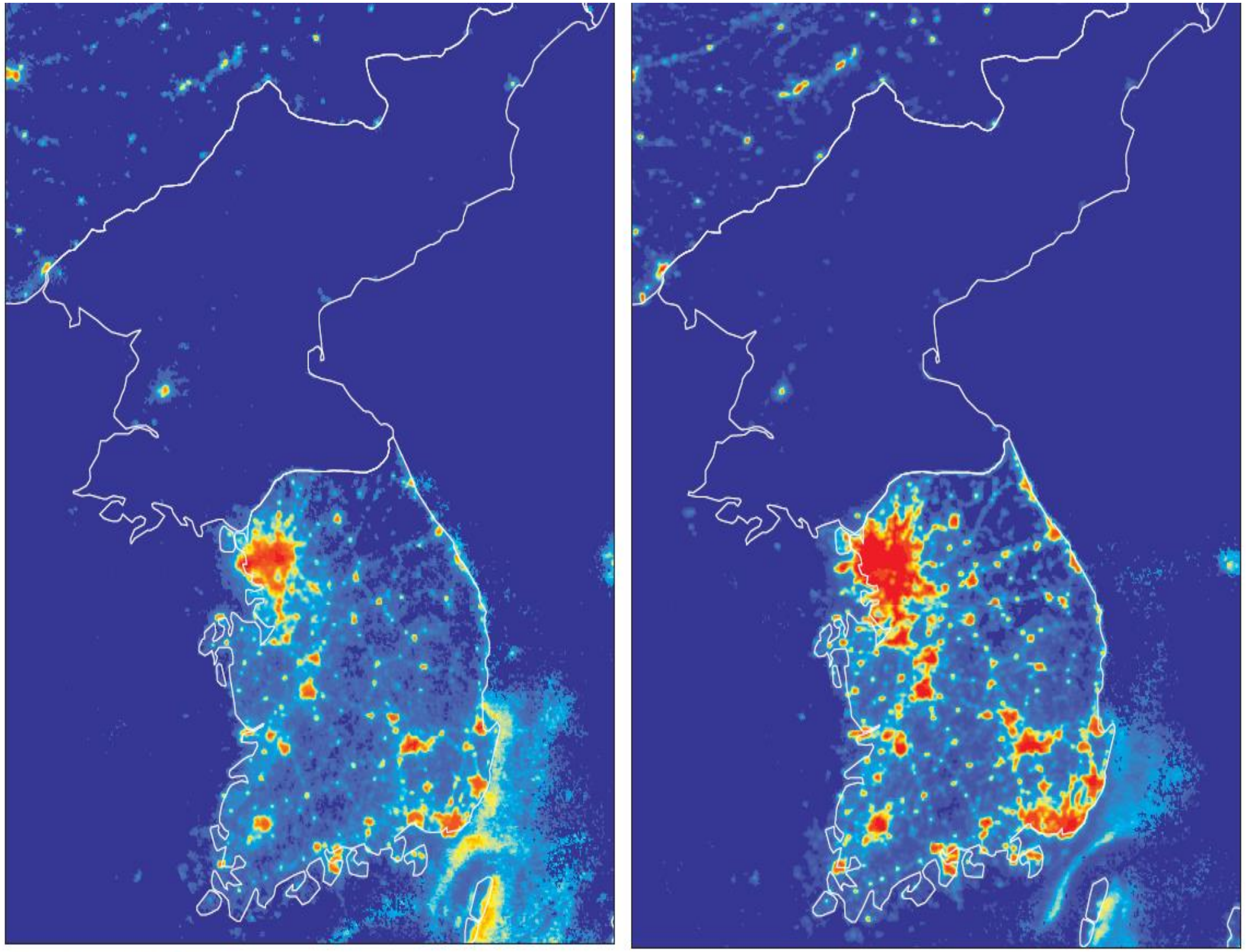
What Do Night Lights Measure?

Answer: Both population density and GDP per capita.

More specifically:

- Within countries, lights mostly measure population density.
- Between countries, lights measure both population density and income per capita.





F-10, 1992

F-16, 2008

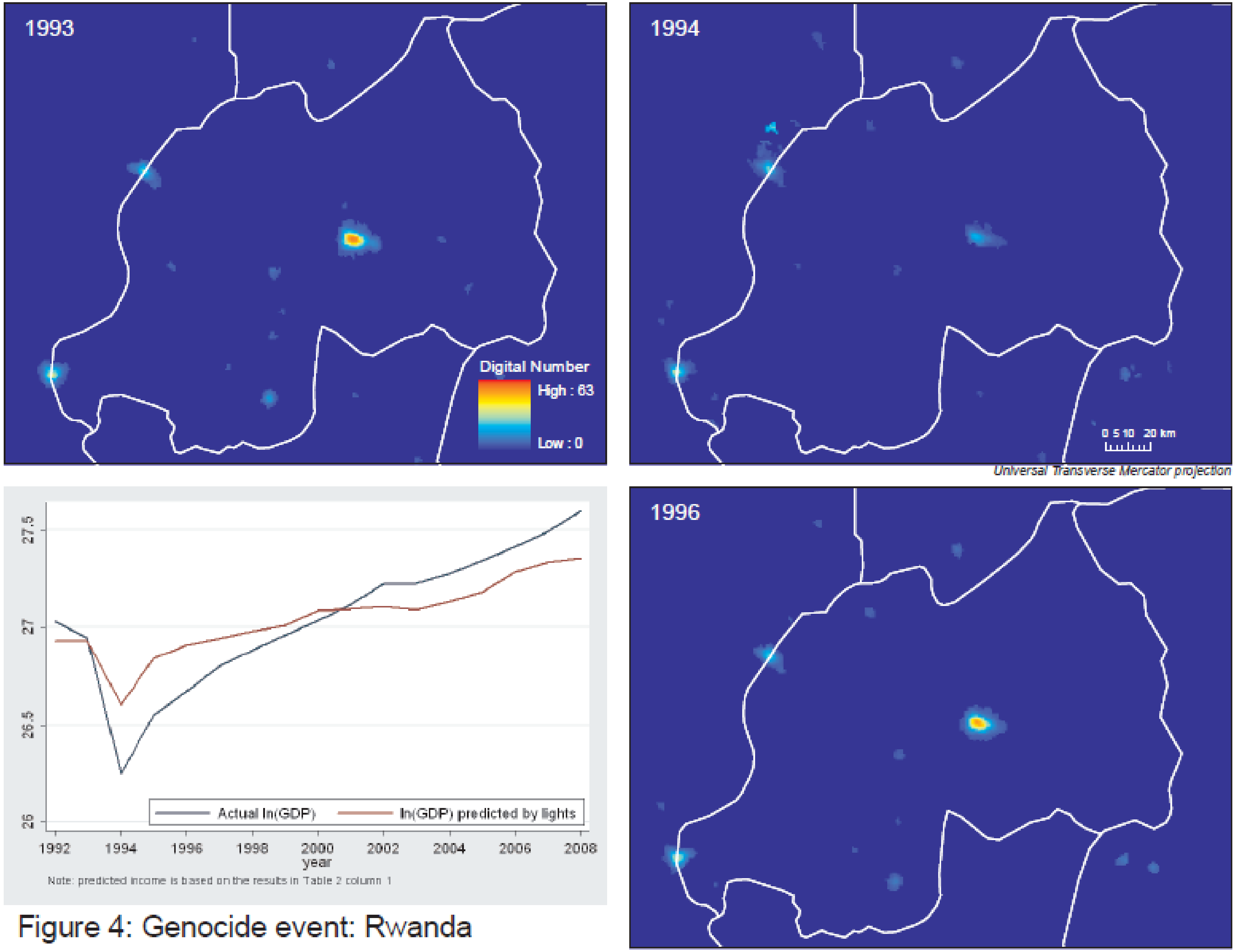


Figure 4: Genocide event: Rwanda

MASSIVE AMOUNTS OF MEASUREMENT ERROR

- Light emitted into space is an imperfect measure of economic activity.
  - Las Vegas vs. Salt Lake City.
  - Oil refinery vs. office where people write software.
- Light registered by satellite is an imperfect measure of light emitted into space.
  - Variation in cloud cover, haze, whether tall building block view, etc.

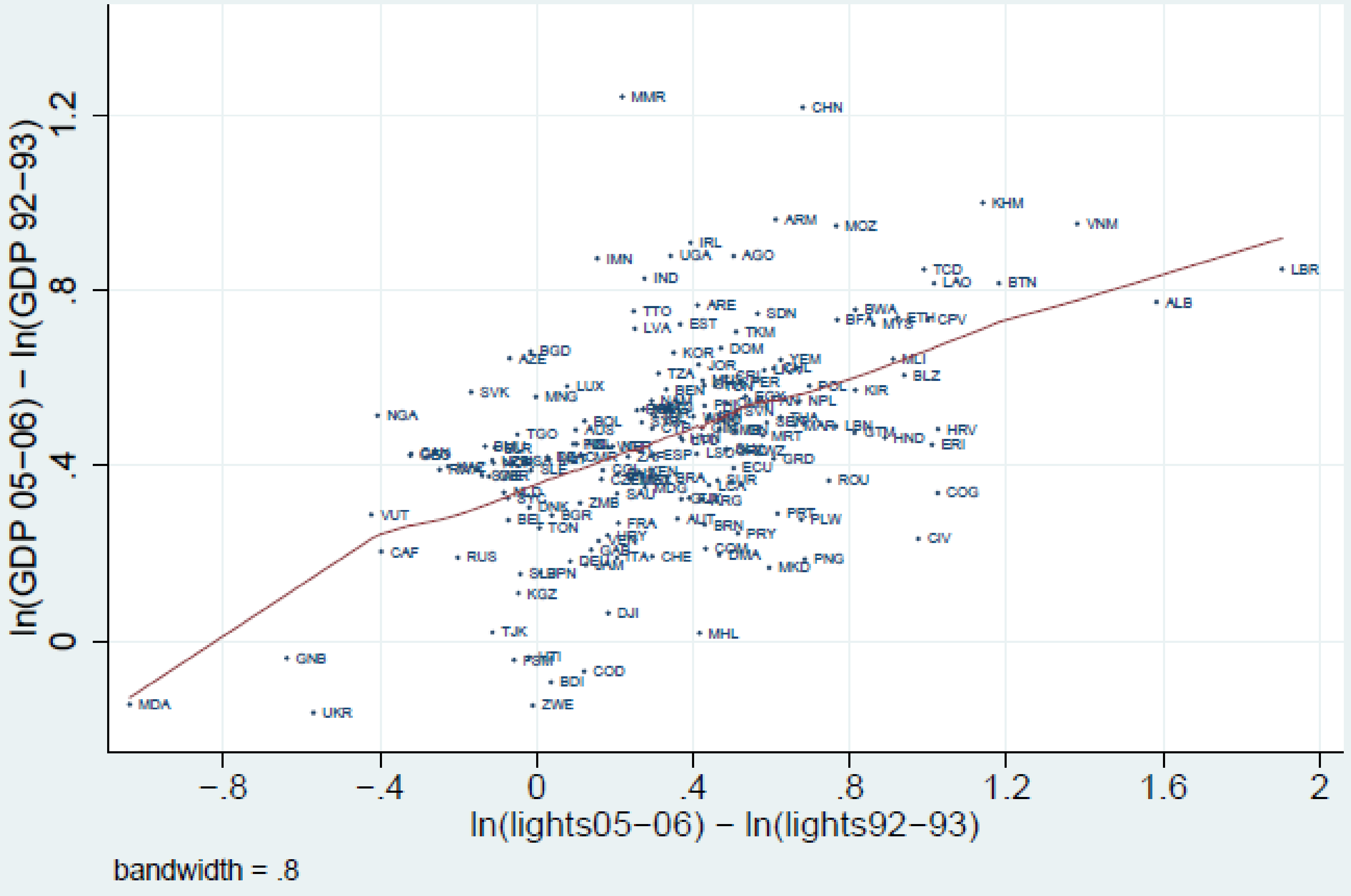
- Because of measurement error (in my view)
  - It is *somewhat* dangerous to look at cross-sectional comparisons. Much more reliable to look at changes in one place over time.
  - It is *very* dangerous to look at high-frequency variation in lights (quarterly, annual). Much better to look at changes over very long periods, like decades.
    - Also: much light growth results from build-out of network electricity. You might expect a "ratchet" effect whereby income growth would raise lights more than income contraction would lower lights.

“Measuring Economic Growth from Outer Space” (Henderson, Storeygard, and Weil, 2012)

- GDP growth is poorly measured, especially in poor countries.
- Change in lights provides an alternative measure of GDP growth.
  - Two bad measures that are uncorrelated are better than one bad measure.
- Paper goes through the math of how to
  - Estimate GDP growth from observed lights growth.
  - Form an optimal composite of this estimate and officially reported GDP growth.

- Analogy of measuring windows…..

Figure 6c. GDP versus lights: long differences



"Measuring Economic Growth from Outer Space" (Henderson, Storeygard, and Weil, 2012)

- GDP growth is poorly measured, especially in poor countries.
- Change in lights provides an alternative measure of GDP growth.
  - Two bad measures that are uncorrelated are better than one bad measure.
- Paper goes through the math of how to
  - Estimate GDP growth from observed lights growth.
  - Form an optimal composite of this estimate and officially reported GDP growth.
- Results:
  - $\Delta \ln(GDP) \approx 0.3\Delta \ln(lights)$
  - Weight on lights estimate $\approx$ .5 for countries with "bad data"

Table 6. Average annual growth rates in true income, for bad data countries (1992/93-2005/06)

Country	ISO code	WDI (LCU)	fitted lights	optimal combination of WDI and fitted lights	difference
Myanmar	MMR	10.02%	3.25%	6.47%	-3.22%
Angola	AGO	6.99%	3.88%	5.37%	-1.51%
Nigeria	NGA	4.04%	1.90%	2.93%	-1.07%
Sudan	SDN	5.92%	4.01%	4.93%	-0.94%
Vietnam	VNM	7.60%	5.82%	6.67%	-0.86%
Burkina Faso	BFA	5.80%	4.46%	5.10%	-0.66%
Benin	BEN	4.52%	3.50%	3.99%	-0.50%
Ghana	GHA	4.60%	3.71%	4.14%	-0.44%
Rwanda	RWA	3.06%	2.24%	2.63%	-0.41%
Algeria	DZA	3.29%	2.84%	3.06%	-0.22%
Oman	OMN	4.28%	3.84%	4.05%	-0.22%
Mali	MLI	5.08%	4.77%	4.92%	-0.15%
Sierra Leone	SLE	3.04%	2.75%	2.89%	-0.15%
Cameroon	CMR	3.29%	2.99%	3.14%	-0.15%
Iran, Islamic Rep.	IRN	4.03%	3.74%	3.88%	-0.15%
Niger	NER	3.48%	3.20%	3.34%	-0.14%
Gambia, The	GMB	3.80%	3.80%	3.80%	0.00%
Liberia	LBR	6.75%	6.99%	6.87%	0.12%
Central African Republic	CAF	1.59%	1.92%	1.76%	0.17%
Mauritania	MRT	3.68%	4.04%	3.86%	0.18%
Swaziland	SWZ	3.42%	3.94%	3.69%	0.26%
Lebanon	LBN	3.85%	4.45%	4.16%	0.30%
Madagascar	MDG	2.74%	3.38%	3.07%	0.32%
Eritrea	ERI	3.51%	4.99%	4.27%	0.74%
Guinea-Bissau	GNB	-0.29%	1.41%	0.58%	0.88%
Congo, Rep.	COG	2.63%	5.02%	3.86%	1.20%
Haiti	HTI	-0.28%	2.74%	1.27%	1.55%
Côte d'Ivoire	CIV	1.82%	4.91%	3.40%	1.56%
Congo, Dem. Rep.	COD	-0.52%	3.04%	1.30%	1.83%
Burundi	BDI	-0.71%	2.86%	1.12%	1.84%

Combining Several Indicators

- “Measuring Economic Growth With A Fully Identified Three-Signal Model” Working Paper (2024) Civelli, Gaduh, and Yousuf.
- Measure are
  - Night lights
    - Combined DMSP and VIIRS using methodology of Nechaev et al (2021)
  - Nitrogen oxides (produced by combustion)
    - Note: Nitrogen oxides are better than lights for high frequency variation. Reflect current activity. See pictures.
  - Urban land cover

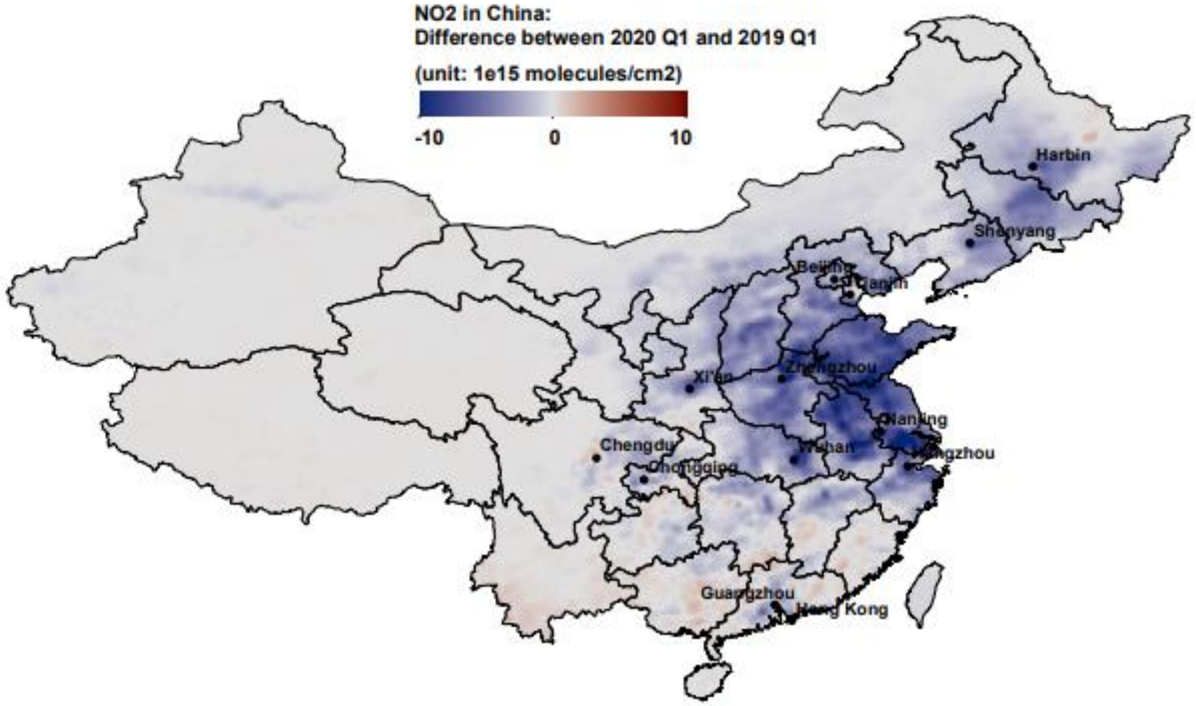


Figure 3: China: Business Cycle Frequency Effects of COVID-19 Shutdown on NO_2.

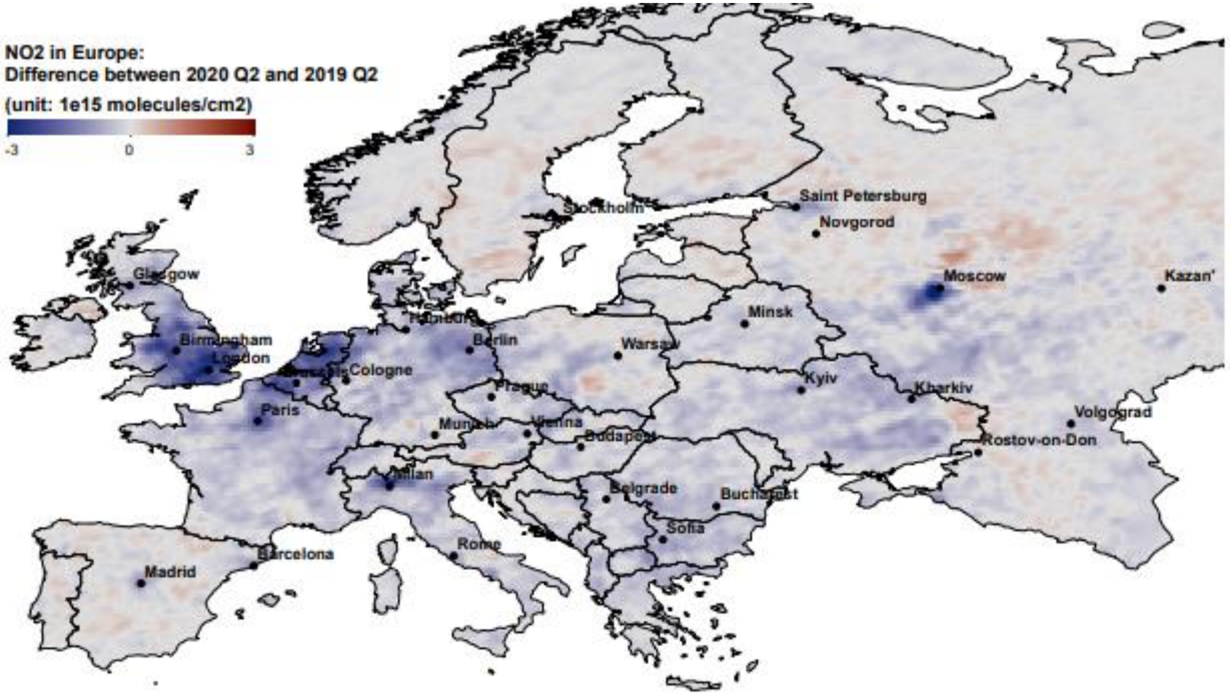
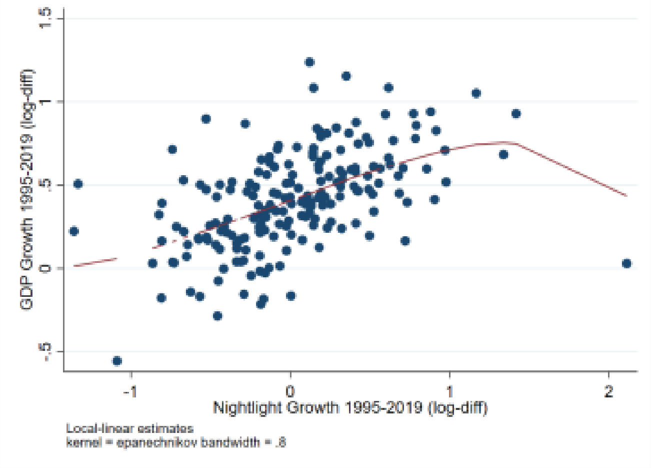
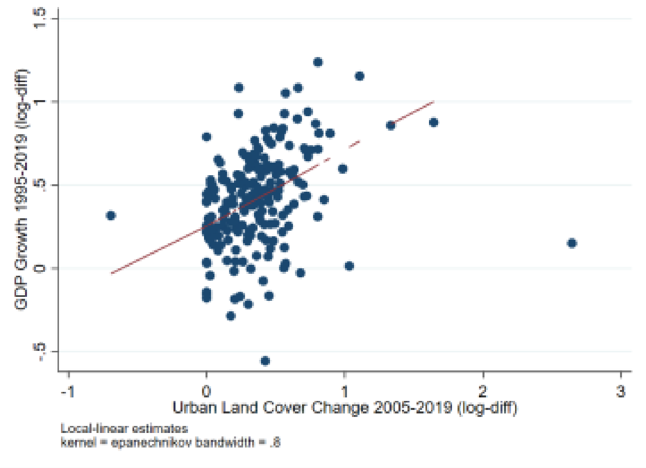


Figure 4: Europe: Business Cycle Frequency Effects of COVID-19 Shutdown on NO_2.

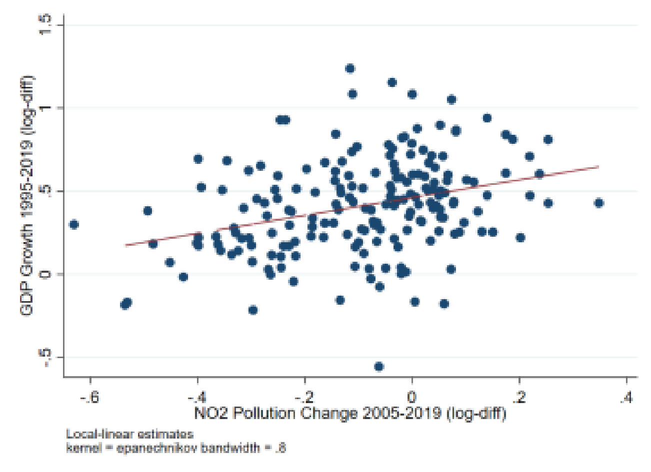
Ezran *et al*., “Measuring Global Economic Activity Using Air Pollution” Working Paper 2023.



(a) Nightlights



(b) Urban land cover



(c) NO_2

Note: Long-term comparison of the growth rates of official GDP and: (a) nightlights luminosity; (b) urban land cover; and (c) NO_2 emission for all countries between 2005 and 2019. Growth rates are calculated as log-differences over the sample, taking the mean of the observations of the first and last two years of the sample as initial and final values. The red lines correspond to a non-parametric, local-linear fit estimated with an Epanechnikov kernel (bandwidth = .8).

Figure C.2: Official GDP Growth v. Signals Growth

Table C.2: Official GDP Growth (x_1) Predictive Regressions, All Countries, 2005-2019

	NO_2			$urban$		
	(1)	(2)	(3)	(4)	(5)	(6)
$nighlights$ (x_2)	0.254	0.284		0.218	0.284	
	(0.058)	(0.058)		(0.063)	(0.058)	
NO_2 (x_3)	0.367		0.531			
	(0.105)		(0.099)			
$urban$ (x_3)				0.311		0.467
				(0.080)		(0.070)
N	189	189	189	189	189	189
adj. R^2	0.280	0.236	0.100	0.304	0.236	0.187
$Period$	2005-2019	2005-2019	2005-2019	2005-2019	2005-2019	2005-2019

Note: Estimation of the predictive regression model (8) for the 2005-2019 period for all countries. $urban$ and NO_2 respectively indicate the growth rates of urban land cover and nitrogen dioxide pollution, two alternative x_3 signals. Growth rates are calculated as log-differences over the sample, taking the mean of the observations of the first and last two years of the sample as initial and final values. Columns (2) and (5) report the estimates of the regression in the corresponding HSW two-signal version of the model in equation (3). Columns (3) and (6) correspond to the two-signal model with the urban land cover signal/NO_2 instead of nightlights. Robust standard errors in parentheses.

Some Applications of Night Lights to Proxy Growth

- Hodler and Raschky (2014)
  - More economic activity happens in leader's home region, especially in dictatorships, especially in Africa and Asia (see picture).
- Pinkovsky and Sala i Martin (2016)
  - Compare developing country GDP growth as measured in national income accounts (NIPA) vs. as measured from mean income in household surveys.
  - Find that NIPA matches lights growth much better.
- Martinez (2022)
  - Gap between reported GDP growth and growth measured by lights is higher on average in autocracies.

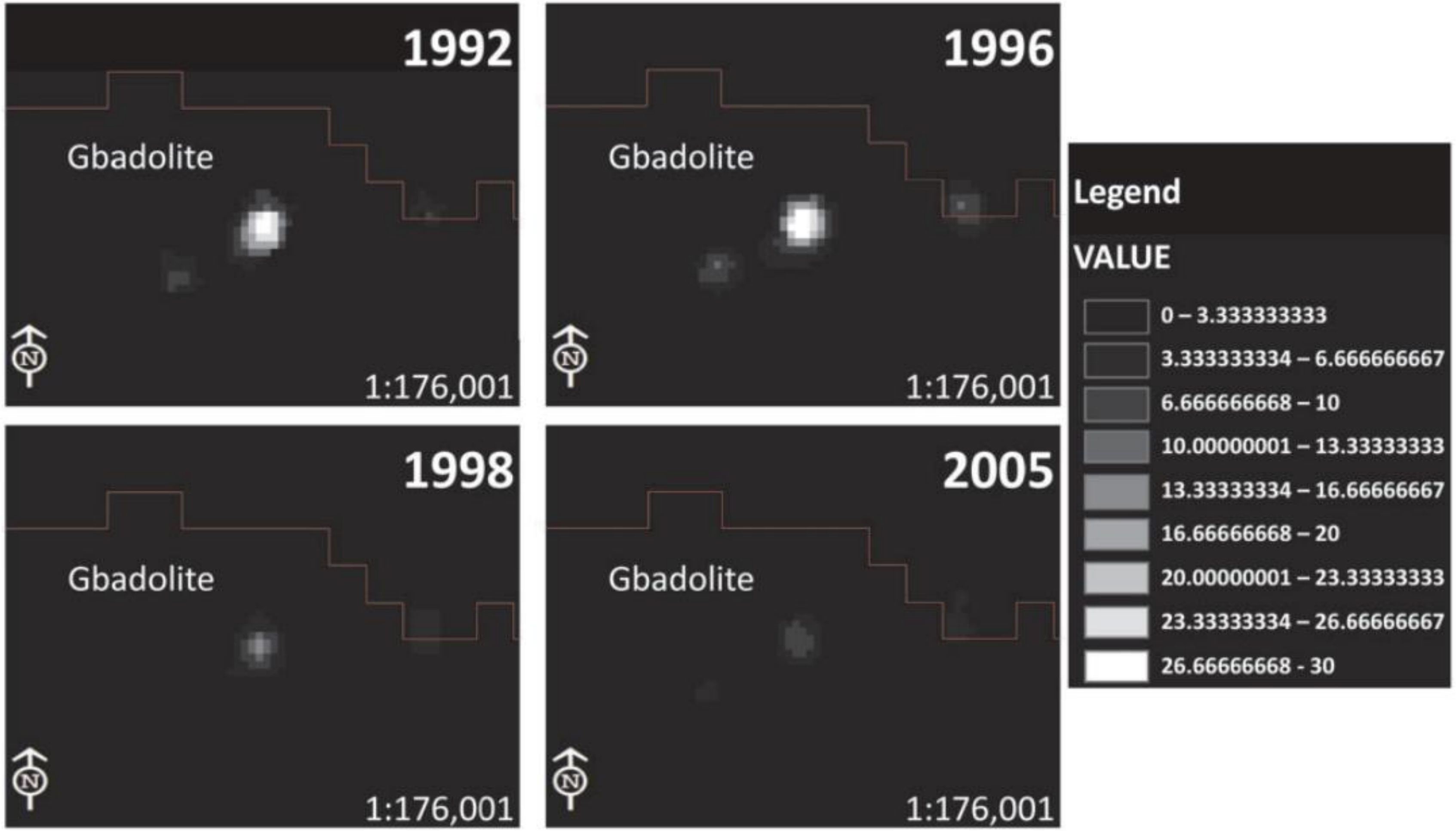


FIGURE I

Nighttime Light Intensity in Gbadolite in 1992, 1996, 1998, and 2005

Mobutu Sese Seko was president of Zaire until 1997.

Adam Storeygard, Tufts University

Proxying Sub-National Income *Growth*

- HSW methodology can produce a proxy for income *growth* for sub-national regions, where no alternative measure is available.
- Regions can be anything that you can draw on a map.
  - HSW find that inland regions in Africa grew 1/3 percentage point per year faster than coastal over the period 1992-2008
  - Can also apply to ethnic homelands, biomes, etc.
  - You can even create arbitrary "virtual countries" as a useful counterfactual (Michalopoulos, 2012).

Proxying for Sub-National Income *Level*

- You can similarly do this, but it runs into my earlier “somewhat dangerous” warning.
- There may be differences in light levels due to culture, industrial structure, geography, climate, etc.
- Also, the structure of the function that converts income and population density to light observed is unclear.
- These are differenced out when you look at lights growth, but remain present when you look at levels.
- Still, people do look at levels.

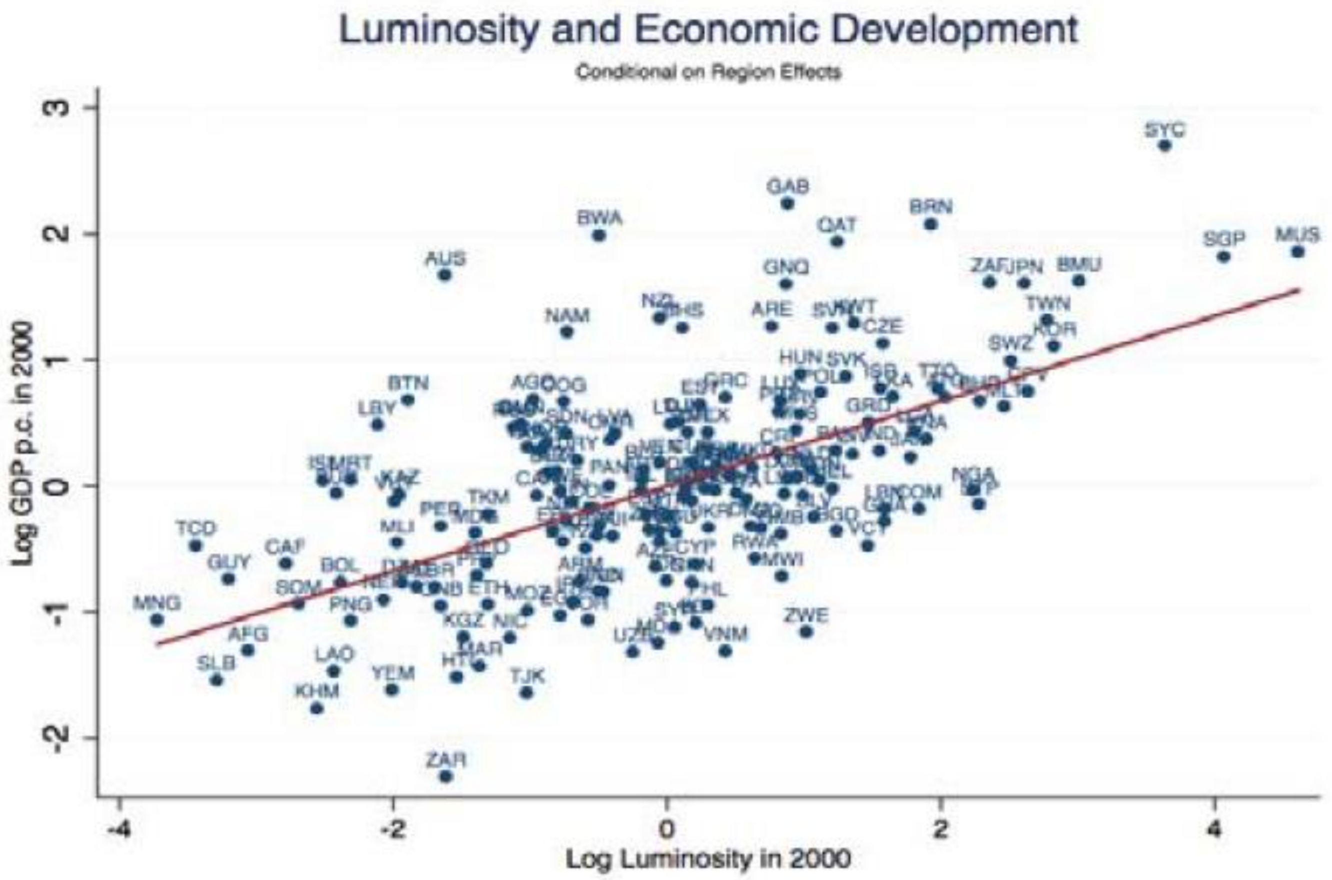


Figure 2

“Luminosity” is light per square kilometer (!)

Source: Michalopoulos and Papaioannou “Spatial Patterns of Development: A Meso Approach,” *Annual Review of Economics*, 2018 (10): 383-410.

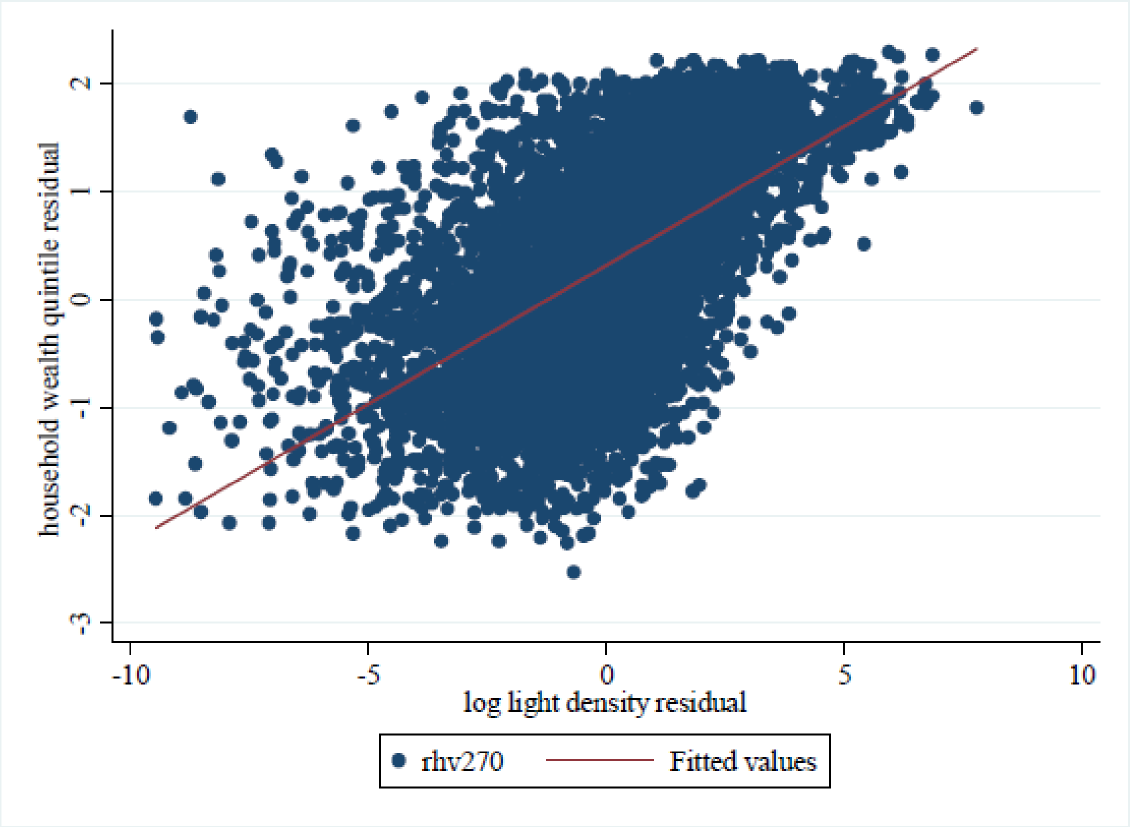


Figure 4a

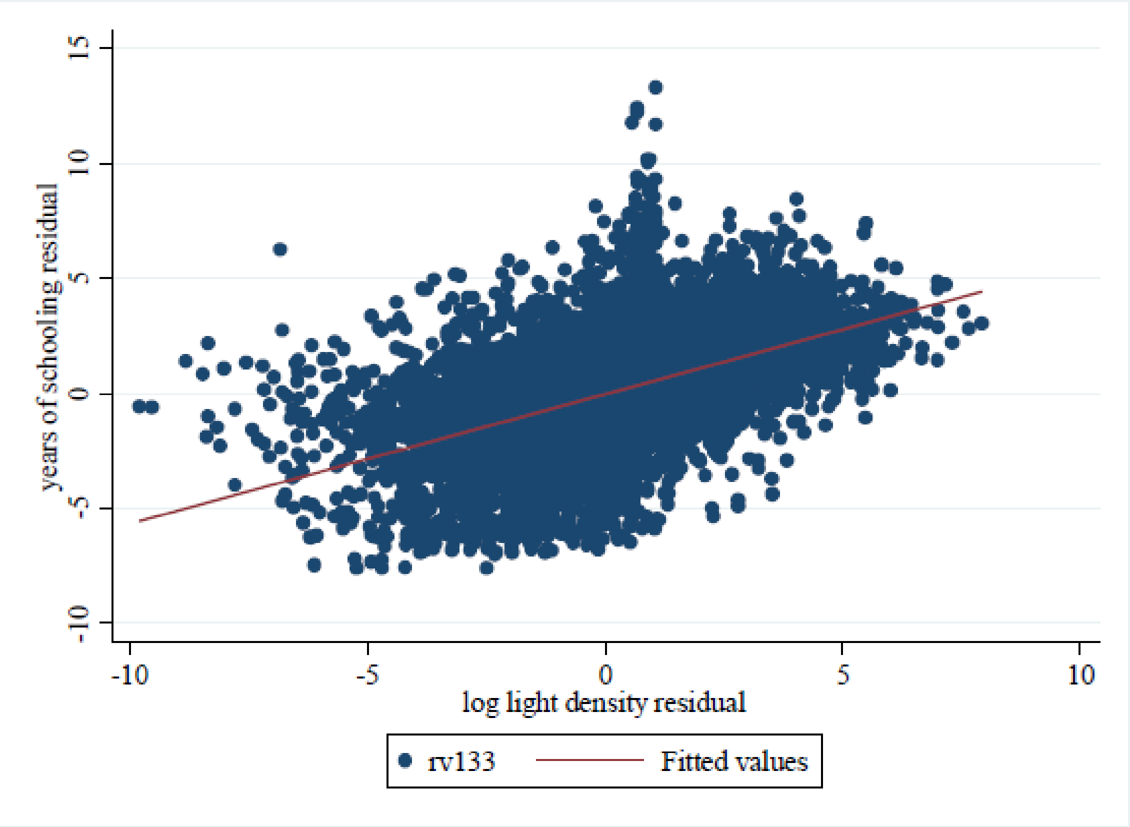


Figure 4b

Household wealth and schooling from 74 DHS surveys that cover close to 9,000 third-level administrative units from 21 African countries. Source: Michalopoulos and Papaioannou "Spatial Patterns of Development: A Meso Approach," *Annual Review of Economics*, 2018 (10): 383-410.

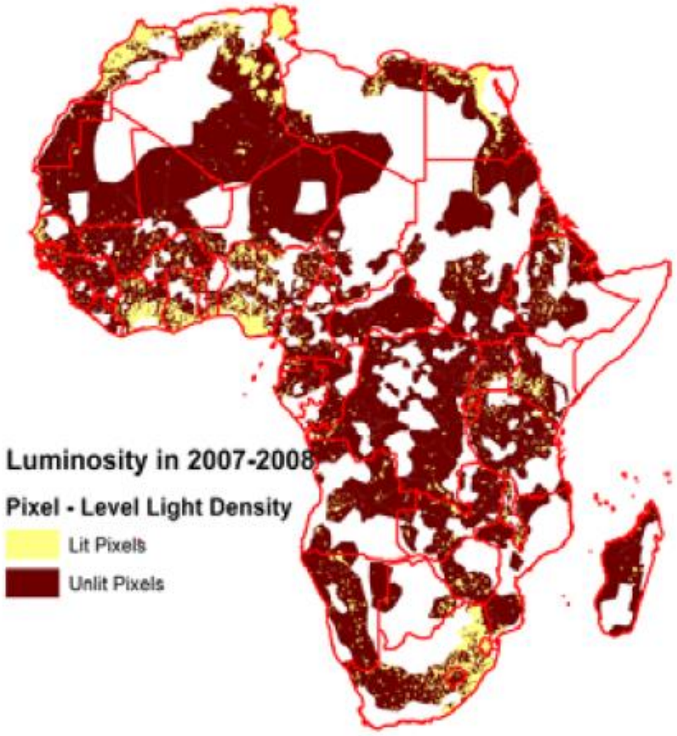


Figure 9*a*

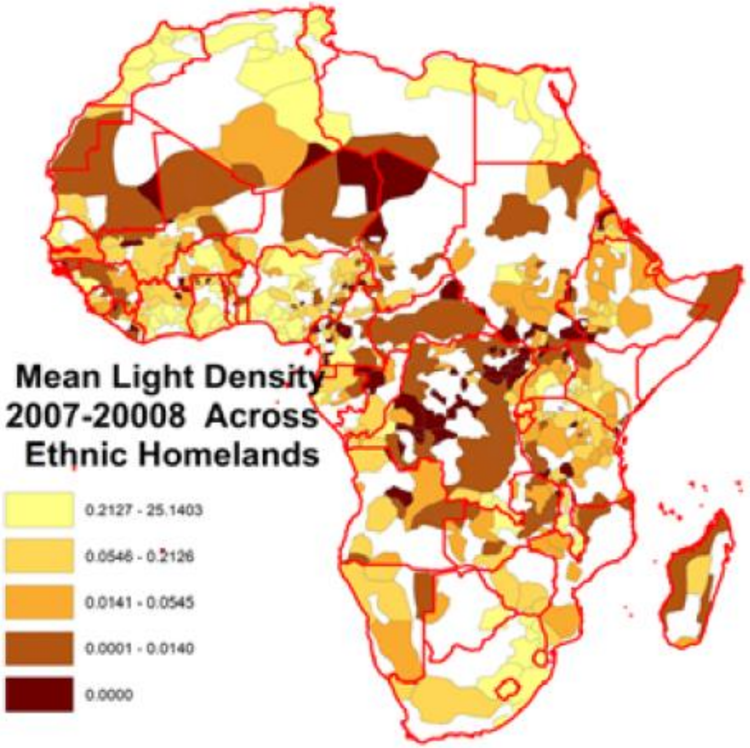


Figure 9*b*

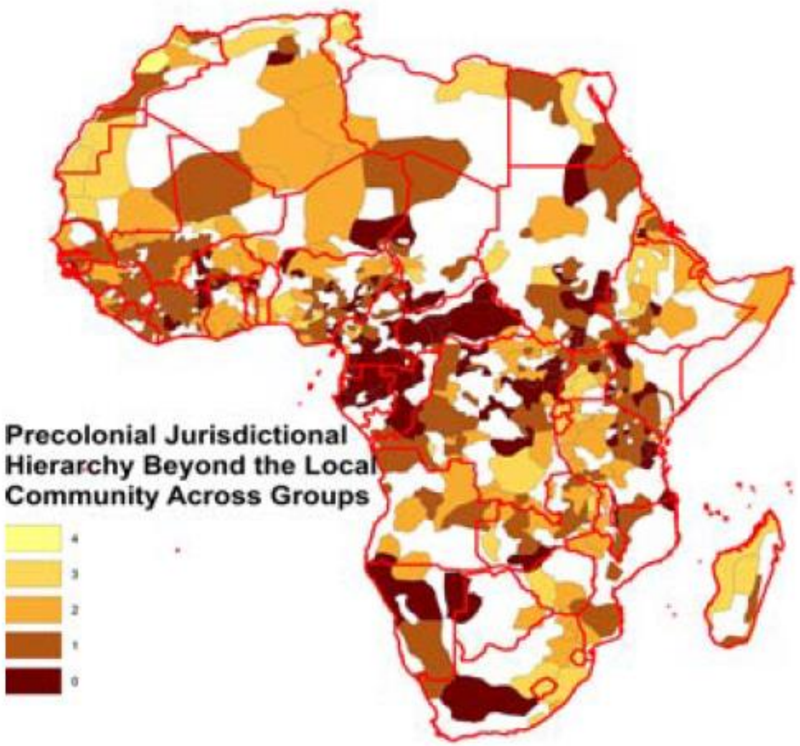
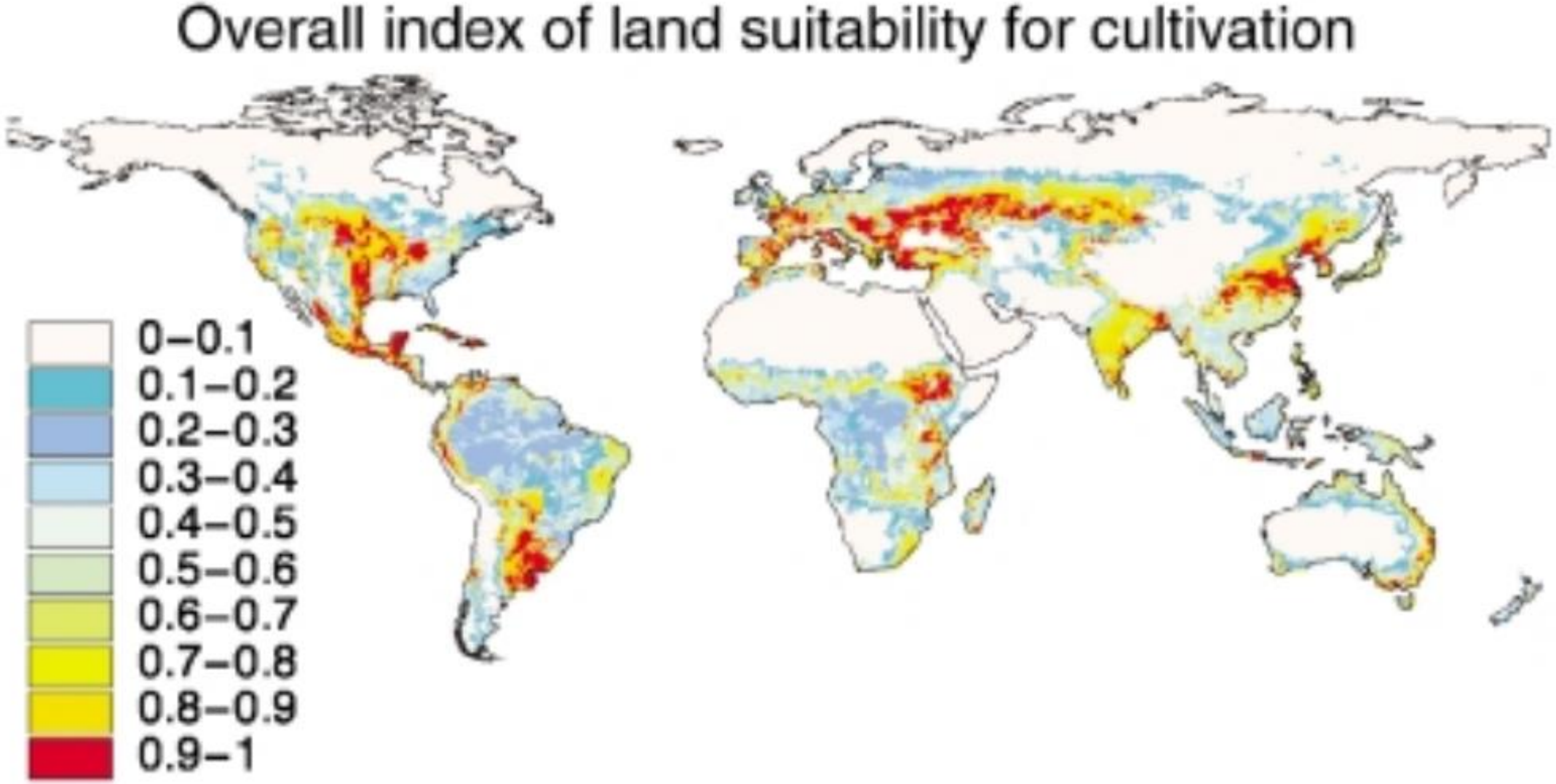


Figure 9*c*

Michalopoulos and Papaioannou "Spatial Patterns of Development: A Meso Approach," *Annual Review of Economics*, 2018 (10): 383-410.

Geographic Data

- Data on physical features
  - Soil fertility and other characteristics, crop suitability
  - Terrain roughness
  - Many dimensions of climate (temperature, precipitation, seasonality, etc.)
- Much of this is remote sensed, but also includes on-the-ground measures (and some modeling).



Ramankutty *et al*., "The global distribution of cultivable lands: current patterns and sensitivity to possible climate change," 2002.

Population Data

- Also available at a fine scale.
- Combination of government collected data by regions and satellite observation used to put people in the right places.

Population and Geography

- "Land Quality" Henderson, Storeygard and Weil (2023) [on my website]
- See references there for details of geographic and population data.

- Data on quarter-degree grid squares (237,000)
  - Population density
  - Vector of geographic characteristics
    - Elevation, latitude, ruggedness, malaria index, coastal dummy, dummies for within 25 km of a natural harbour, ocean-navigable major river, or large lake, distance to a coast (from Henderson et al, 2018)
    - 44 variables from Global Agro Ecological Zone (GAEZ): thermal and moisture regime, growing period, etc. plus suitability indices for 11 major crops.

Land Quality

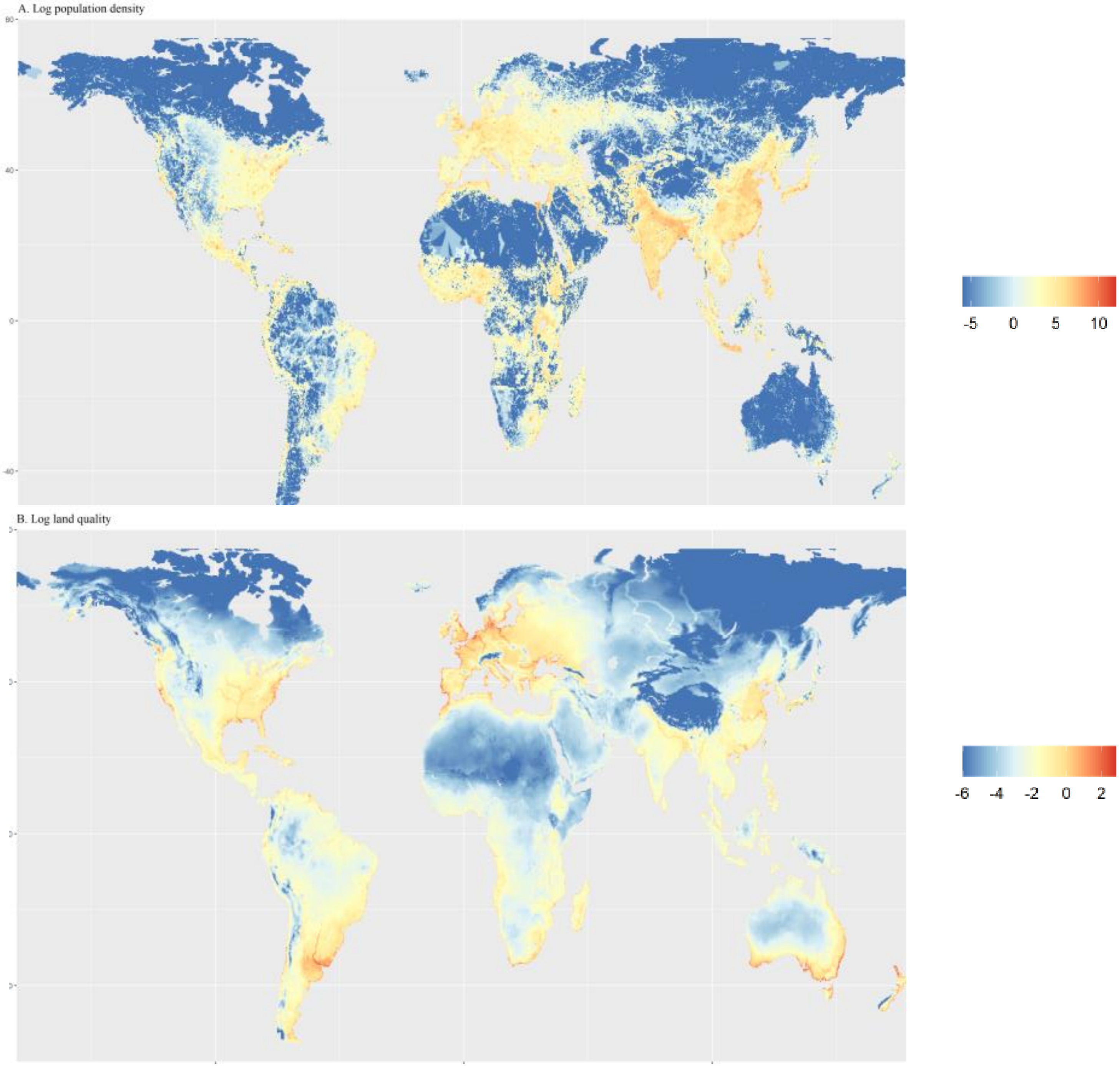
Poisson Specification (c is country, i is grid cell):

$$\frac{population_{i,c}}{area_{i,c}} = \frac{L_{i,c}}{Z_{i,c}} = \exp(C_c + X'_{i,c}\beta)$$

We define *Quality* of a grid square as fitted value, suppressing country fixed effects:

$$Quality_{i,c} = \exp(X'_{i,c}\widehat{\beta})$$

Figure 1. Population Density and Land Quality



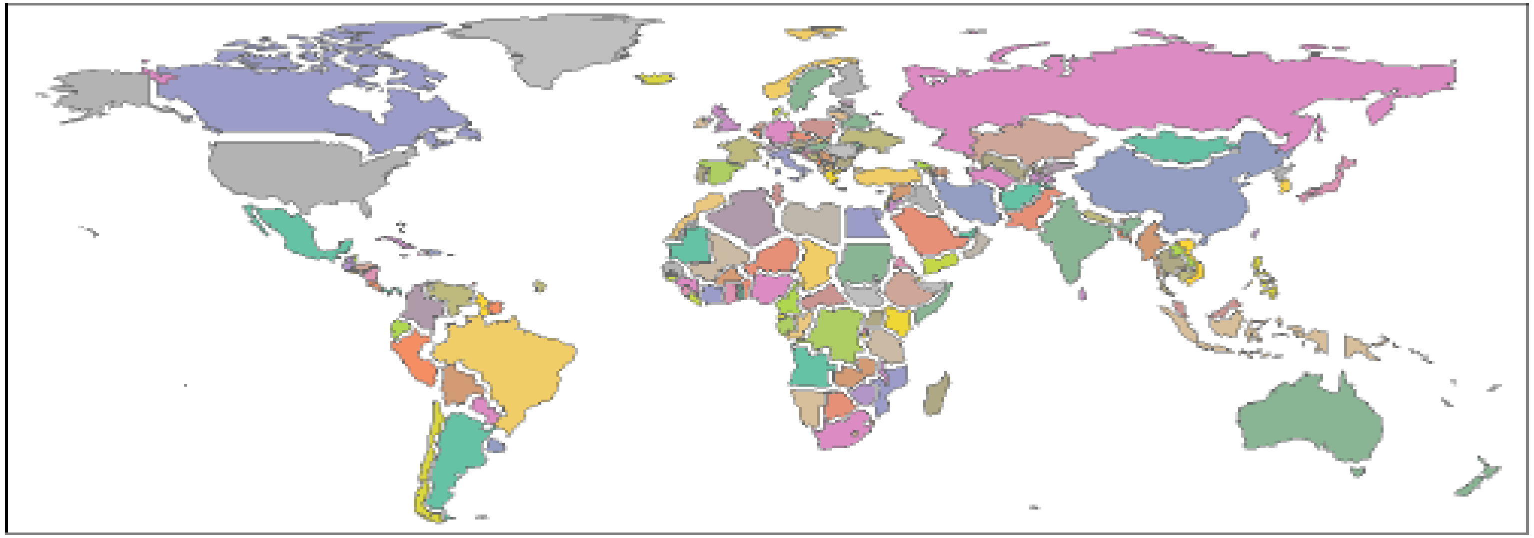
Quality Adjusted Area

$$QAA_c = \sum_{i \in c} \exp\left(X'_{i,c}\,\widehat{\beta}\,\right) Z_{i,c}$$

- **Key Point:** Because we estimate $\hat{\beta}$ with country fixed effects:
- An individual country being more populous does not affect our estimate of its QAA.
- Countries with particular *average* characteristics being more populous does not affect our estimate of their QAA.
- We normalize this so that the sum across countries is equal to the actual land area of the world.

Figure 2. Country Level Quality Adjusted Area

A. Countries by Land Area



B. Countries by Quality Adjusted Area

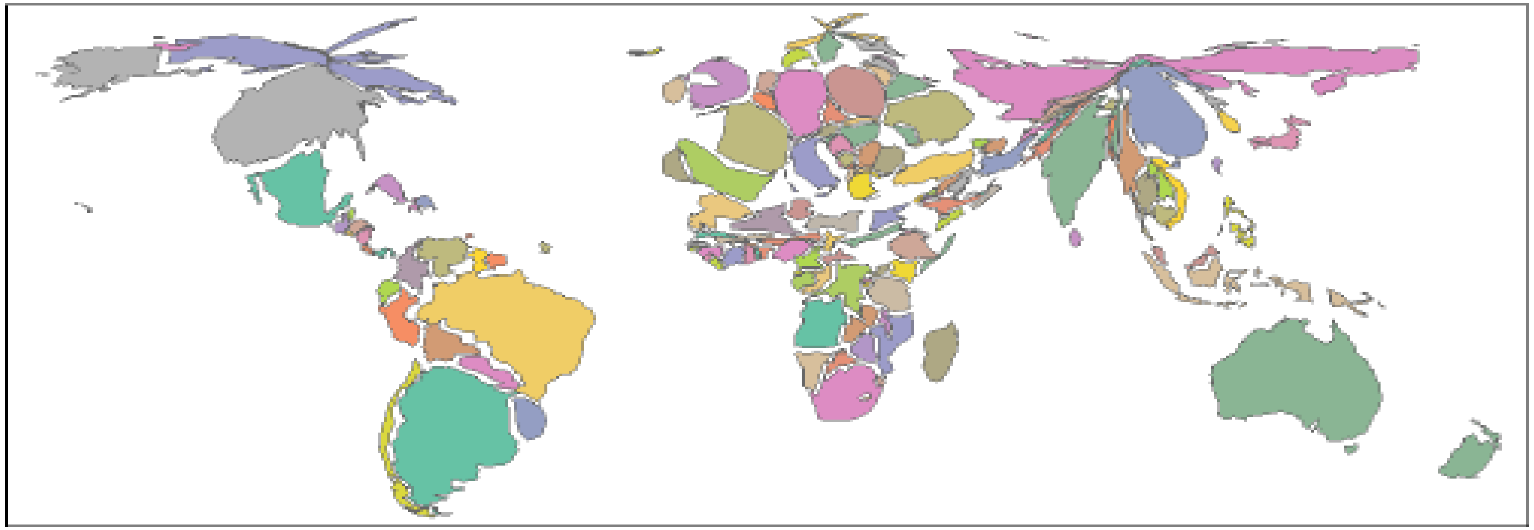
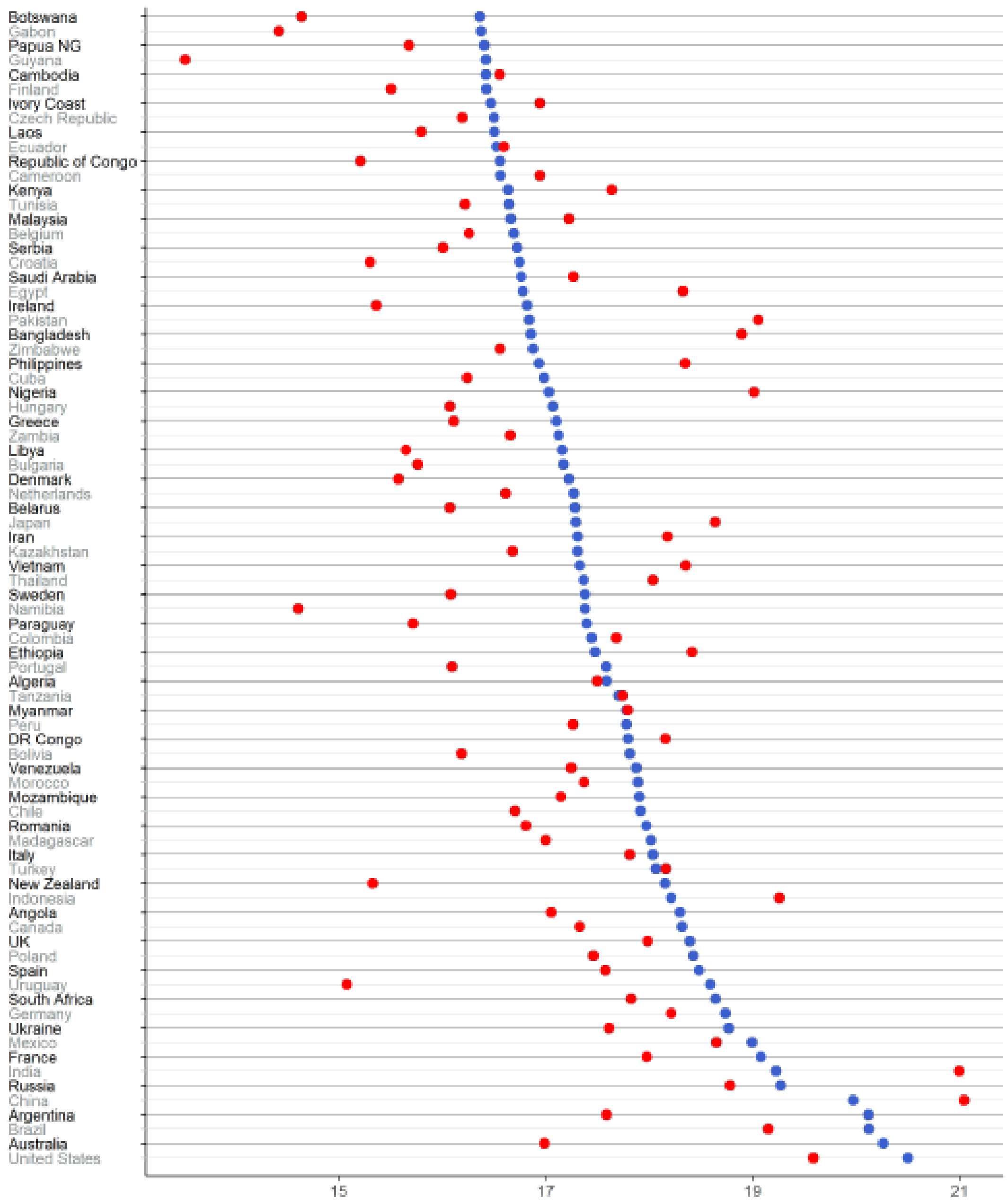


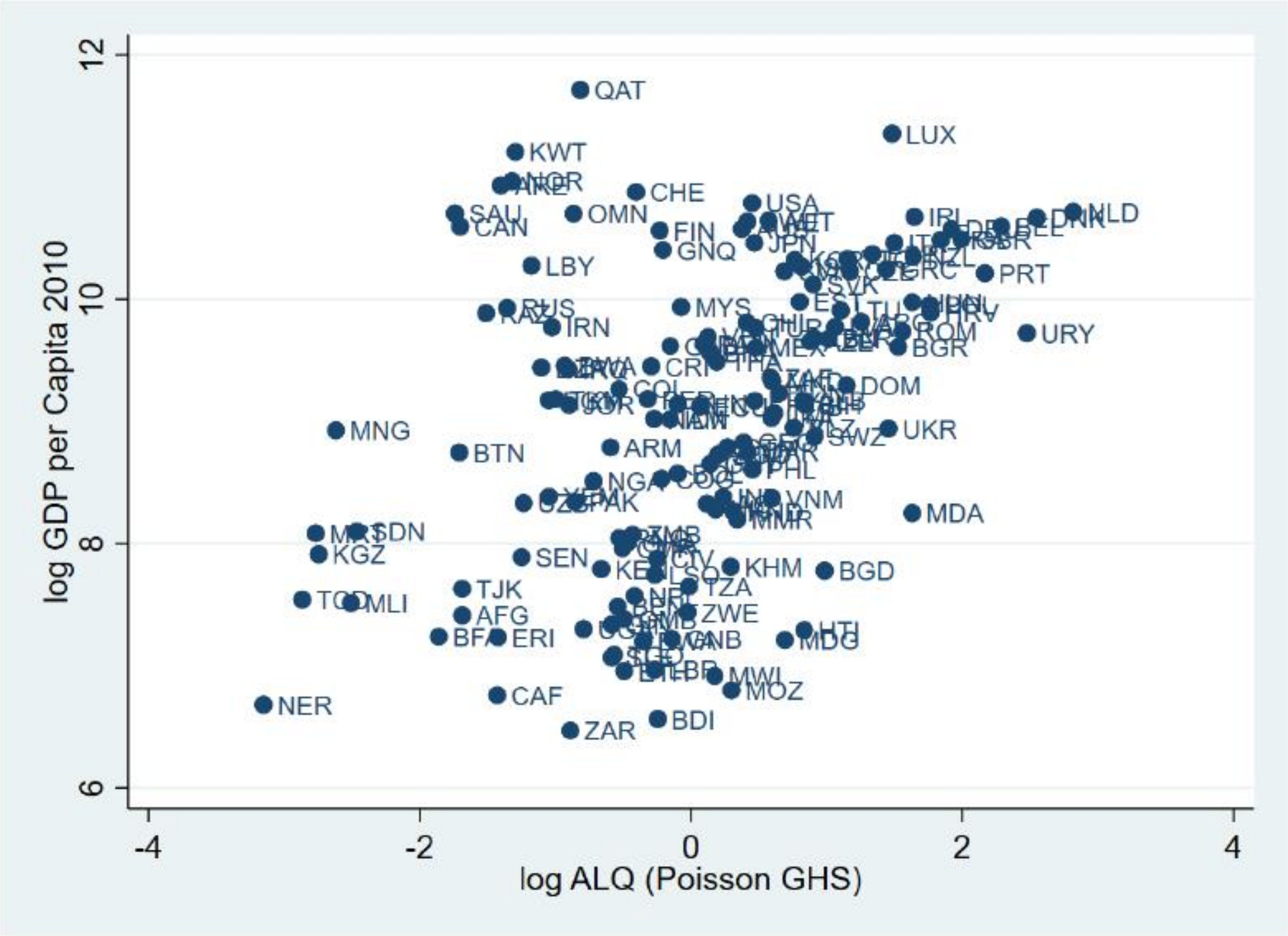
Figure 3. Top 80 Countries by Fitted Population



Blue dots depict log fitted population; red dots depict log actual population.

Population reallocated according to share of world QAA	
Country	Gains/losses (millions)
Australia	605
Argentina	502
USA	479
Brazil	341
France	129
India	-1009
China	-903
Pakistan	-167
Nigeria	-156
Indonesia	-149

Land Quality and Income



Quality Adjusted Population Density

$$QAPD = \frac{Population}{QAA}$$

➔ Normalized so that for the world as a whole, $QAPD$ is 56.6 people per square km, which is the same as conventional population density.

$QAPD$ Country Values (all in appendix)

Top 5 (excluding city states and pop<1 million)

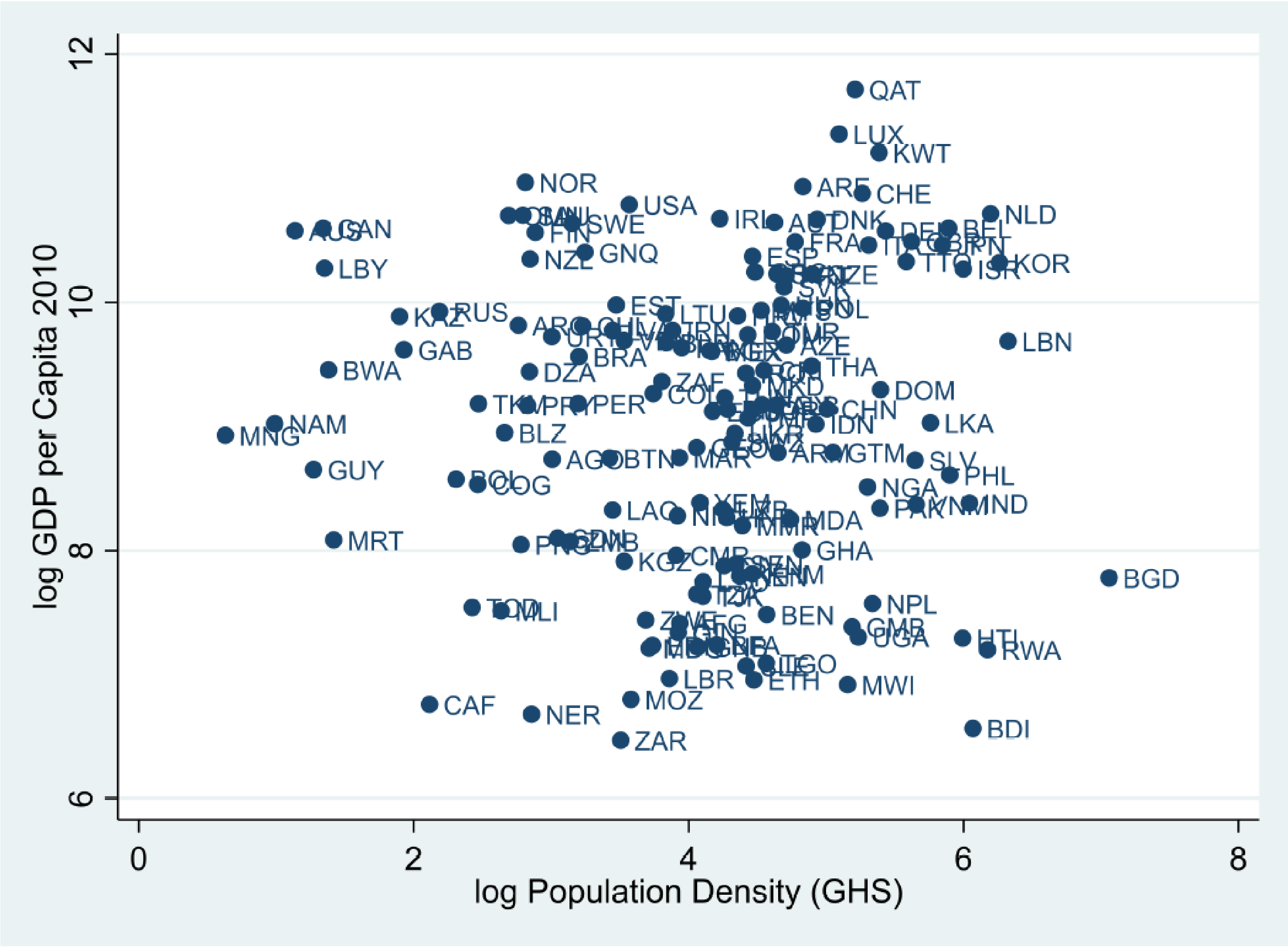
Bottom 5

Country	$QAPD$
Kuwait	797
Rwanda	683
Burundi	553
Kyrgyzstan	533
Pakistan	517
Argentina	4.51
Namibia	3.54
New Zealand	3.35
Australia	2.14
Uruguay	1.70

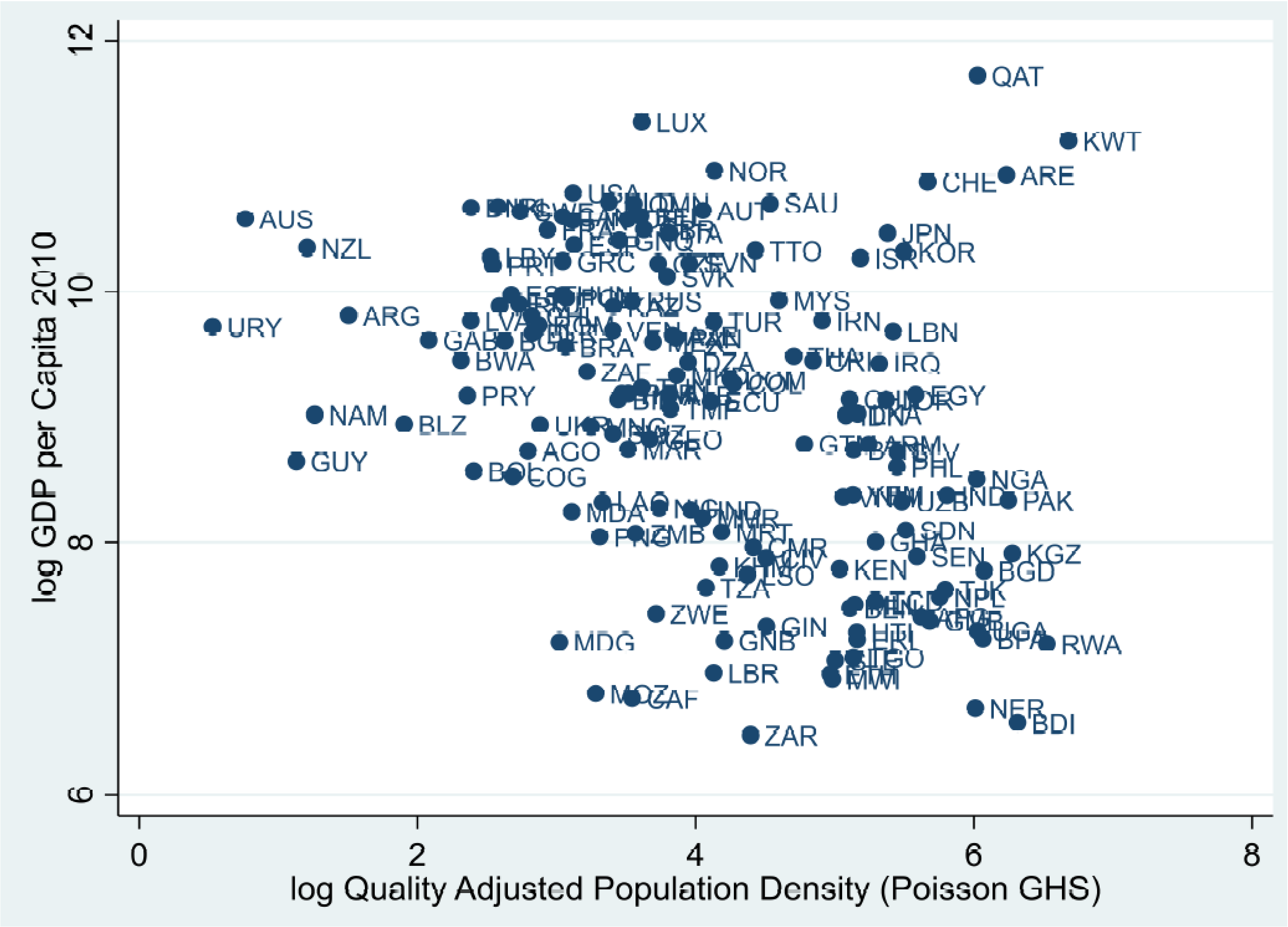
Other Interesting Countries

Country	$QAPD$
Bangladesh	434
Nigeria	412
India	332
Japan	217
Ghana	199
China	165
Indonesia	161
Turkey	62.2
UK	37.6
Germany	33.6
USA	22.6
Brazil	21.4
France	18.8
Ireland	13.2

Conventional Density vs. GDP per Capita



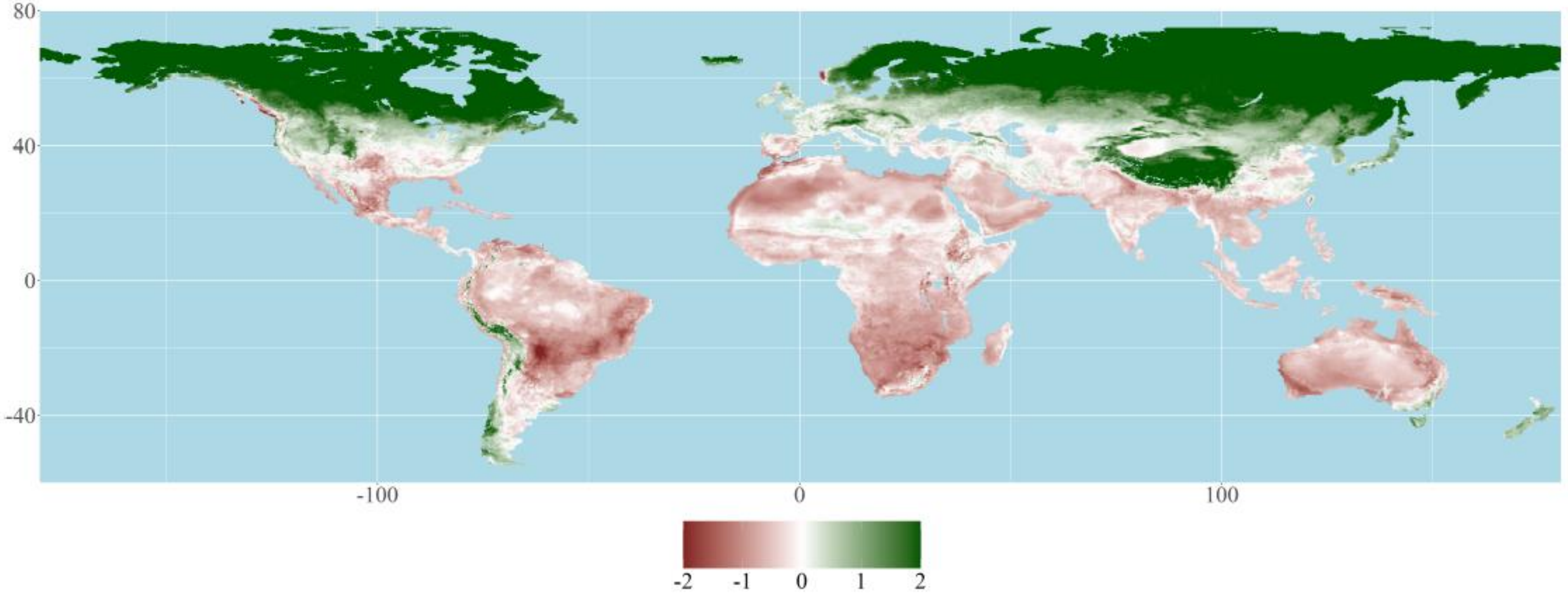
$QAPD$ vs. GDP per capita



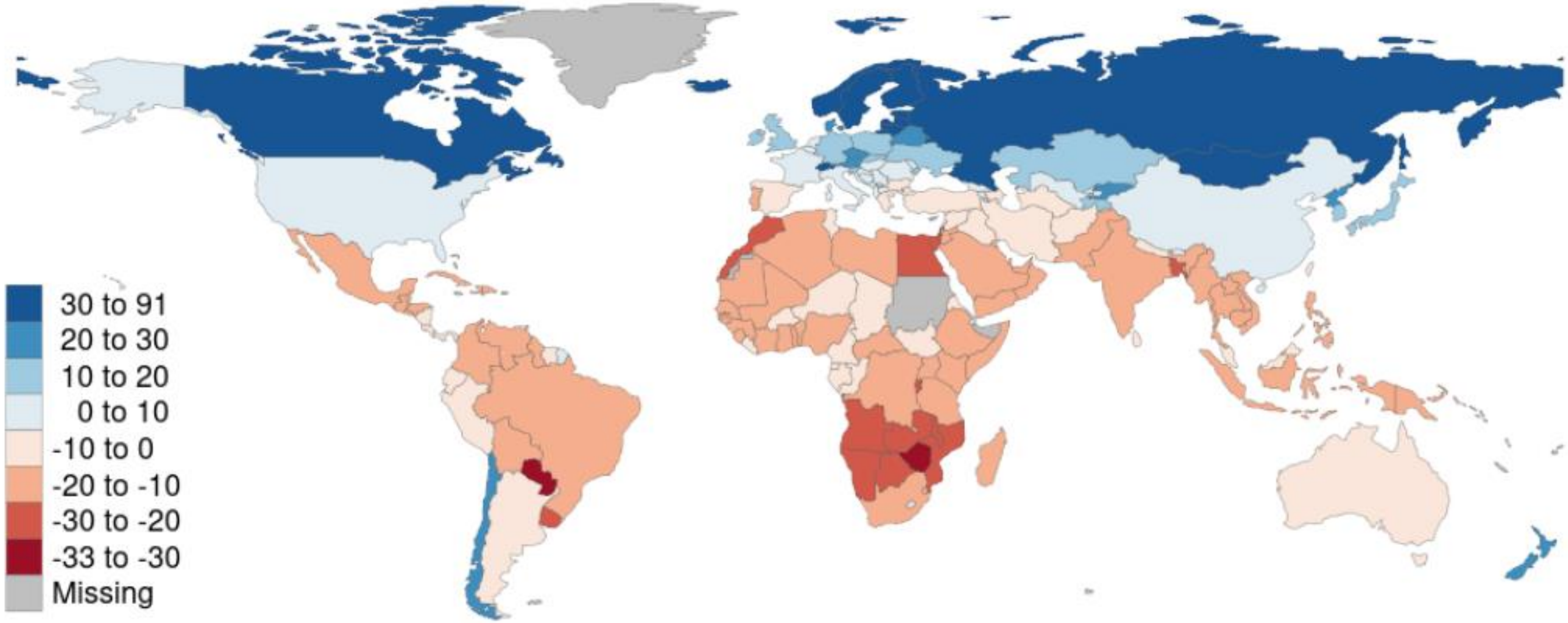
One More Application: Climate Change

- Henderson, Jang, Storeygard, and Weil (2024) “Climate Change, Population Pressure, and Population Growth” [on my website]
- Estimate land quality as in previous paper.
- Look at predicted changes in climate 2010-2100 to see how this will affect land quality.
- Economic model maps this into changes in income per capita.

Change in Log Location Quality 2010-2100 under RCP 8.5



Country-Level Impacts from Climate Change with Mobil Labor



Other Results

- As of 2100, projected climate change under a high-emissions scenario
  - Lowers total world GDP in 2100 by 2-3%.
  - Lowers average GDP in countries below median by 10%.
  - Increases world Gini coefficient by about 0.05.
- For most poor countries, the projected effect of population growth on per capita income is larger than the projected effect of climate change.