Big Data and Development

Image Analysis

Lisa Chauvet

Semestre 1 2025 - CES Paris 1 Panthéon-Sorbonne

Lectures on Image analysis:

1. New data applied to development economics (3h, Lisa Chauvet)
    - Measuring poverty globally, regionally and at the individual level
    - Population and population dynamics
2. Image analysis (15h, Clément Gorin)
    - Image processing and GIS fundamentals
    - Predictive modelling
    - Neural networks
    - Convolutional networks
    - Image modelling applications (segmentation, localisation)

Evaluation based on a written in-class exam:

-One question on: Aiken, E, Bellue, S, Karlan, D, Udry, C, and Blumenstock, JE (2022). Machine Learning and Phone Data Can Improve the Targeting of Humanitarian Aid, Nature, 603: 864-870

https://www.nature.com/articles/s41586-022-04484-9.pdf

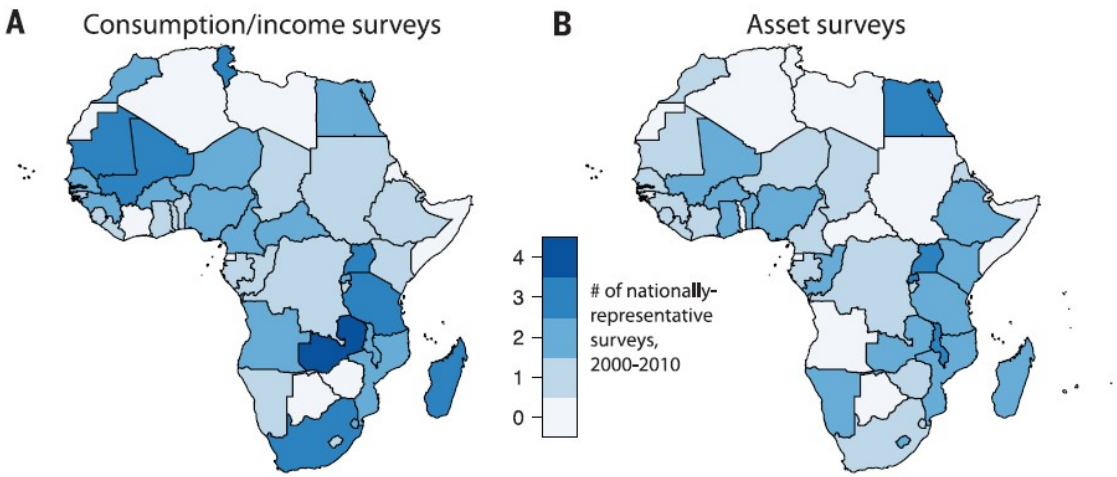
-Several questions on Clement Gorin's part

BIG DATA AND DEVELOPMENT

Image Analysis

Motivation

- Social science study increasingly dependent on real data
- Limitations of traditional way to obtain real data
  - survey data (sample size, social desirability bias)
  - census data (expensive, lack timeliness)



Motivation

- New data available (sensing devices, online platforms, etc)
  - reduce sparsity and increase sample size
  - require techniques in data mining and machine learning
- Three keywords:
  - Quantitative methods
  - Real data
  - Large-scale

Main reference: Gao, Zang, Zhou, 2019. Computational socioeconomics, *Physics Reports* 817, 1-104 (Gao et al., 2019).

Motivation

- Focus on three main sources of new data
  - Remote sensing (RS): nighttime light (NTL), daytime data, very-high resolution data (VHR)
  - Mobile phone
  - Online data: social media (FB, twitter) and other platforms (wiki, google trends, newspapers etc.)

  ⚠ Not exhaustive (e.g. trade data)
- Challenge: from description to causality

Plan

Nighttime light (NTL): Presentation

- DMPS-OLS (Defense Meteorological Satellite Program's Operational Linescan System)
- VIIRS (Visible Infrared Imaging Radiometer Suite of the Suomi National Polar-Orbiting Partnership)
- Gibson et al. (2020)

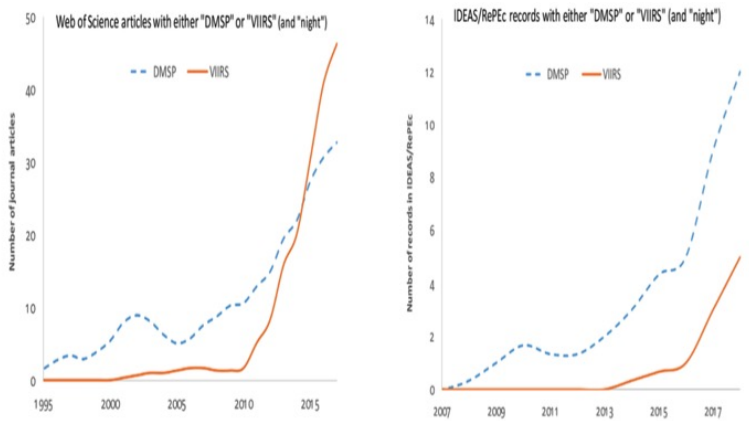


Figure 2. Scientists are Switching to VIIRS, Economists Less So. [Colour figure can be viewed at wileyonlinelibrary.com]

Nighttime light (NTL): Presentation

- DMPS-OLS
  - 1992-2013 on a yearly basis
  - **Low spatial acuracy: overglow/blurring issue**
  - Overstate cities by 77% (by 9% when deblurred). Abrahams et al. (2018) MATLAB code for deblurring overlooked by economists
  - Errors get smaller for larger areas

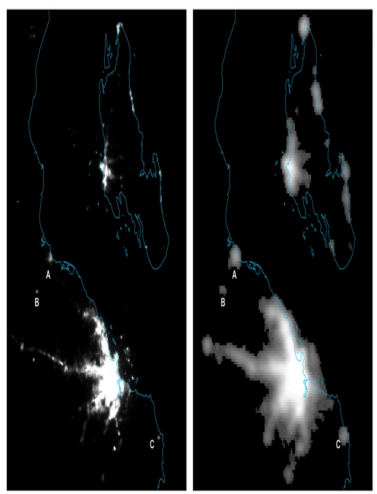


Figure 3. VIIRS Imagery (left) is Much Sharper than DMSP Imagery (right) for Dar es Salaam and Zanzibar, 2013. [Colour figure can be viewed at wileyonlinelibrary.com]

Source: Gibson et al. (2020)

Nighttime light (NTL): Presentation

- DMPS-OLS
  - **Lack of temporal comparability (Nordhaus and Chen, 2014)**
  - Average nb of nights used varies spatially and in time (Gibson et al., 2020)

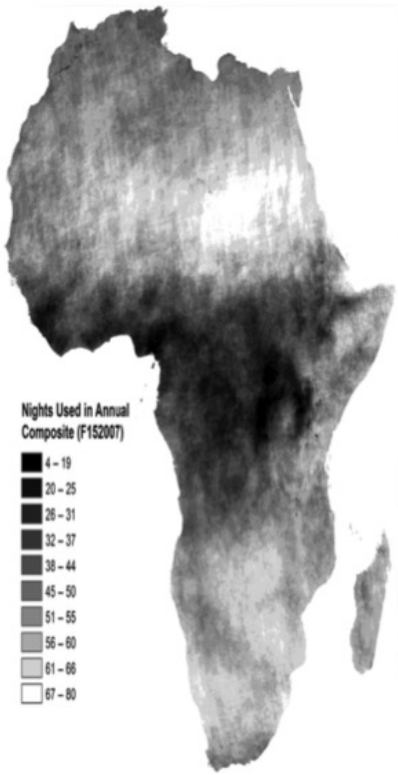


Figure 4. Variation Within Africa in Number of Nights Used for DMSP Annual Composite.

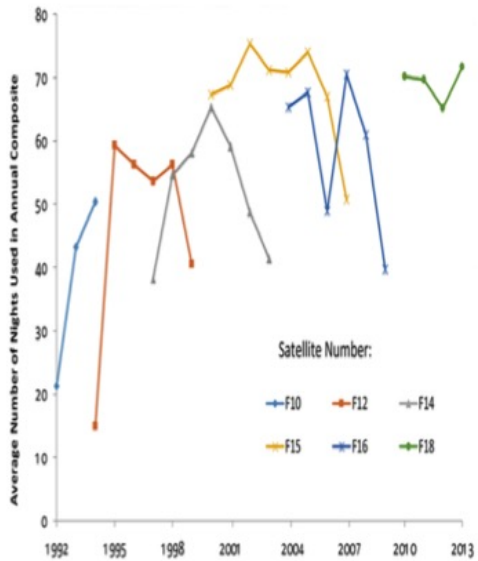


Figure 5. Variation by Satellite and Year in Average Number of Nights Used (Within Africa) for Annual Stable Lights Composite. [Colour figure can be viewed at wileyonlinelibrary.com]

Nighttime light (NTL): Presentation

- DMPS-OLS
  - **Saturation or "top-coding" of lights in urban areas (Bluhm and Krause, 2018)**
  - Limited dynamic range of DMSP (max DN = 63)
- DMPS-OLS and VIIRS: lights that can be detected are mainly for urban areas
- The correlation of annual average DMSP NL with GDP is weaker for countries with larger share of agricultural sector.

Nighttime light (NTL): Presentation

Table 1. Comparison of DMSP and VIIRS.

	DMSP	VIIRS
Original purpose	Detect moon-lit clouds, for Air Force weather forecasts	Earth observation for scientific research
Operational period	1970s–2013	October 2011 onward
Periodicity of processed data	Annual	Monthly and Annual
Time of nightly overpass	ca. 7.30 pm	ca. 1.30 am
Swath	3000 km (but only center half of each swath is processed)	3000 km
Geolocation errors	1.4–3.7 km (95% CI)	None
Spatial resolution of sensor	560 m x 560 m, smoothed to 5 x 5 blocks on-board, for 2.7 km x 2.7 km—at nadir	742 m x 742 m, across the entire swath
Spatial resolution of processed data	Allocated to grids of 30 arc seconds (≈930 m x 930 m at equator, or ≈930 m x 770 m at 35 degrees of latitude)	Allocated to grids of 15 arc seconds (≈465 m x 465 m at equator, or ≈465 m x 385 m at 35 degrees of latitude)
Other spectral bands	1 (thermal infrared)	21 during day, 11 at night
In-flight calibration	None	On-board solar diffuser
Saturation	In urban cores	None
Quantization	6-bit ($n = 64$)	14-bit ($n = 16{,}384$)
Dynamic range	Limited[a]	3×10^{-5} Watts m^{-2} sr^{-1} to 200 Watts m^{-2} sr^{-1} $L_{max}/L_{min} = 6{,}700{,}000$
Minimum detectable signal	5×10^{-5} Watts m^{-2} sr^{-1}	3×10^{-5} Watts m^{-2} sr^{-1}

[a]Figure 1 of Hsu *et al.* (2015) shows the radiance for DMSP (satellite F16) for the extremes of digital numbers (DN) 0 and 63 at different gain settings (amplification) has less than two orders of magnitude difference, compared with the almost seven orders of magnitude dynamic range for VIIRS shown by Shao *et al.* (2013).

Source: Gibson et al. (2020)

https://eogdata.mines.edu/products/vnl/
https://ncc.nesdis.noaa.gov/VIIRS/

Nighttime light (NTL): Presentation

- Li et al (2020) [A harmonized global nighttime light dataset 1992–2018 in Scientific Data]

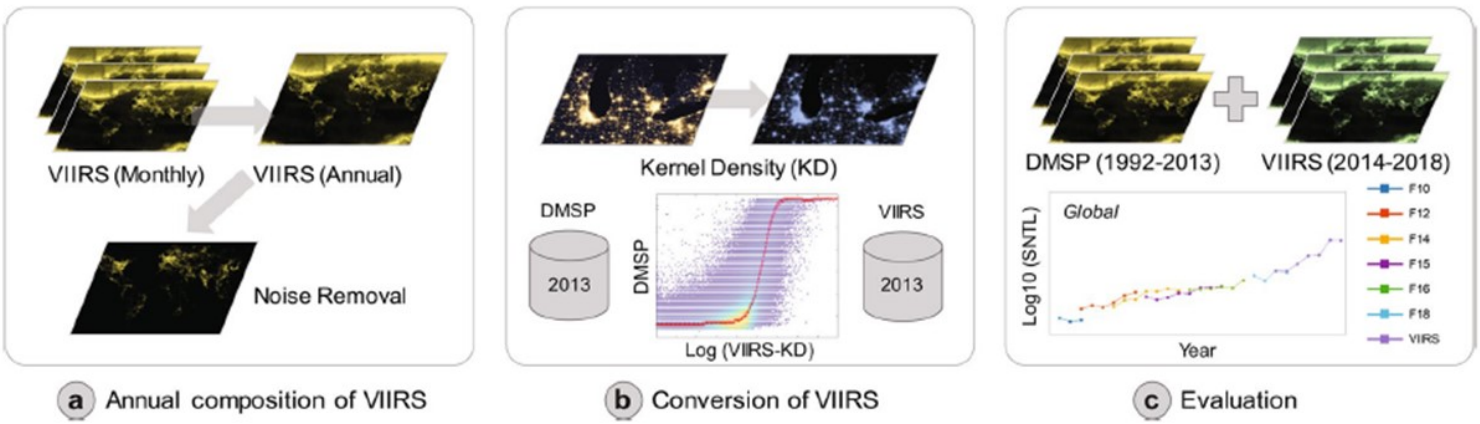


Fig. 1 The proposed framework of generating a consistent global NTL time series data through integration of DMSP and VIIRS. Annual composition of VIIRS data (a); conversion of VIIRS data (b); and evaluation of generated global NTL time series data (c).

Nighttime light (NTL): Presentation

- Li et al (2020) [A harmonized global nighttime light dataset 1992–2018 in Scientific Data]

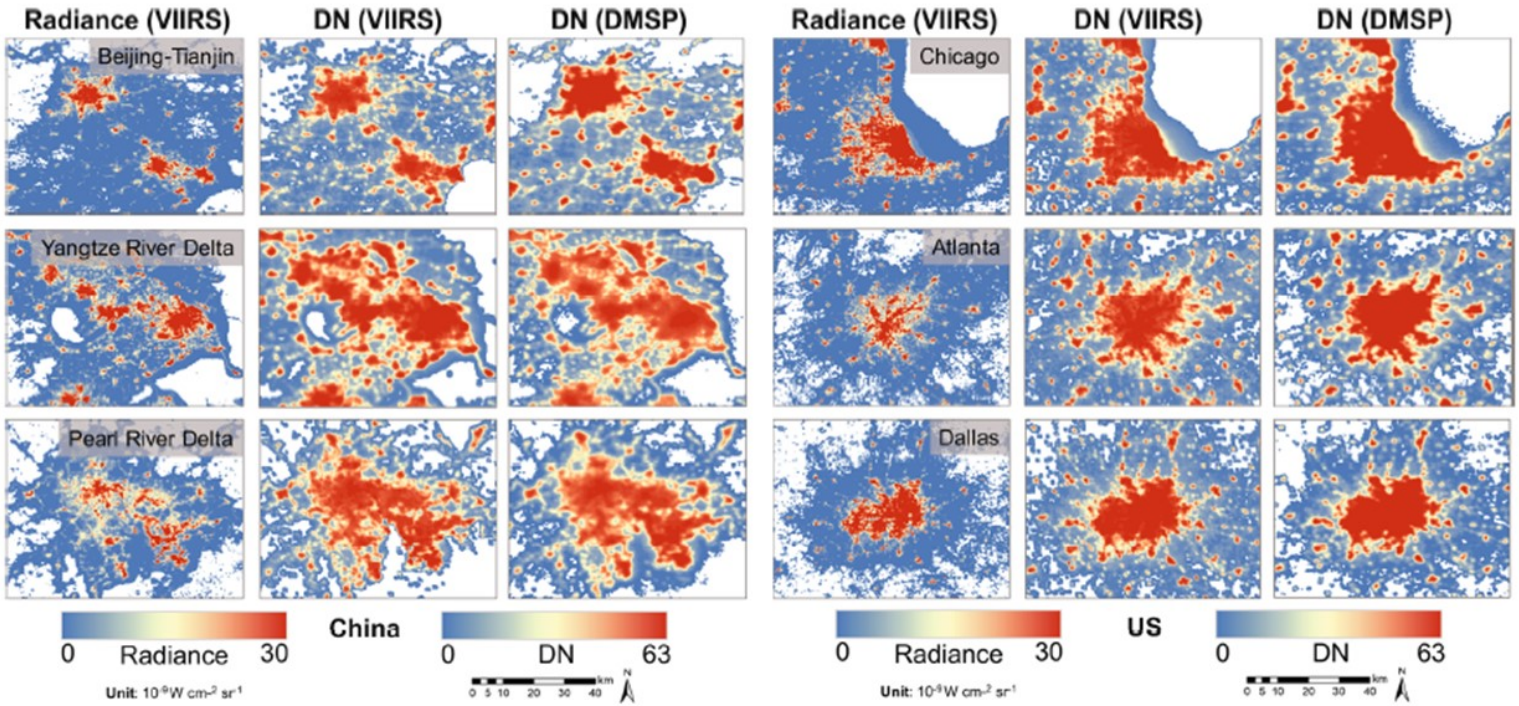


Fig. 6 Spatial patterns of the raw radiance, DMSP-like DNs from VIIRS, and DNs from the inter-calibrated DMSP in example metropolitan areas in China and US in 2013.

NTL reflect human activity: World development and poverty mapping

- Pioneer work: Elvidge et al. (1997)
- 21 countries, country-level correlations, 1994-95 NTL

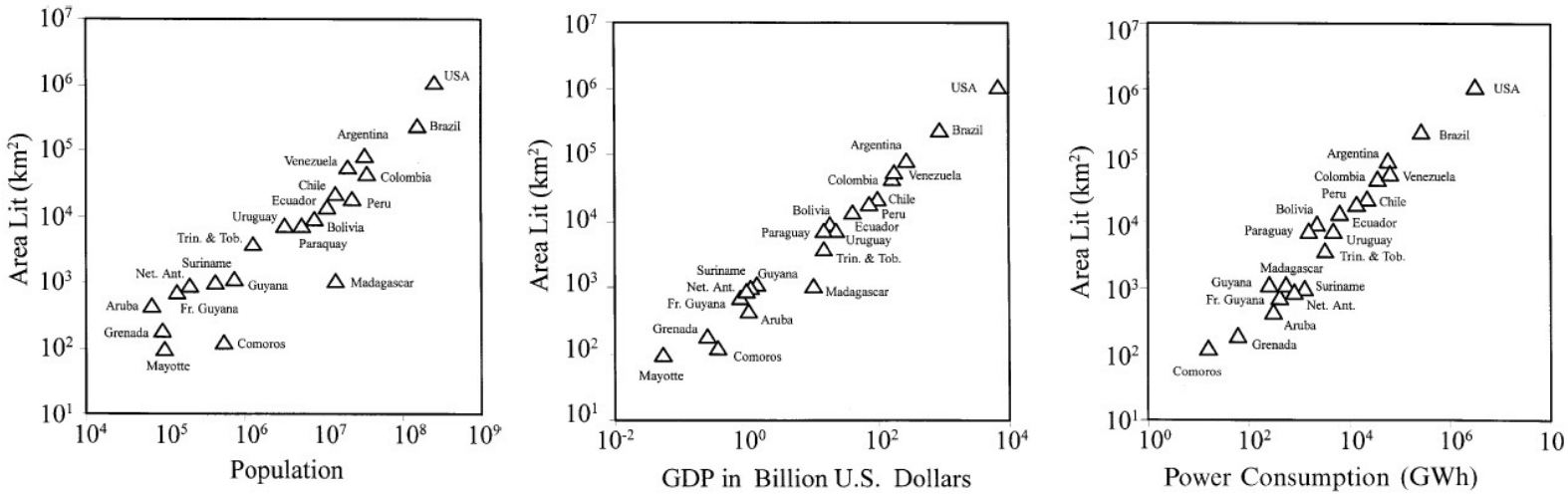
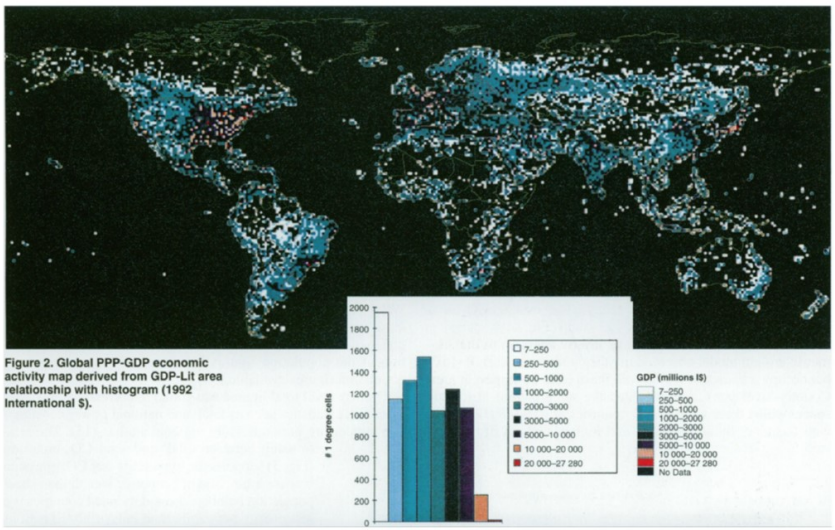


Table 3. Regression equations for estimating population, GDP and electric power consumption based on DMSP-OLS observed area lit.

Log of population = 3·097 + 0·920 (Log of area lit)	$R^2 = 0{\cdot}85$
Log of $GDP_{Billion\ USD}$ = −3·185 + 1·159 (Log of area lit)	$R^2 = 0{\cdot}97$
Log of GWH = −0·865 + 1·178 (Log of area lit)	$R^2 = 0{\cdot}96$

NTL reflect human activity: World development and poverty mapping

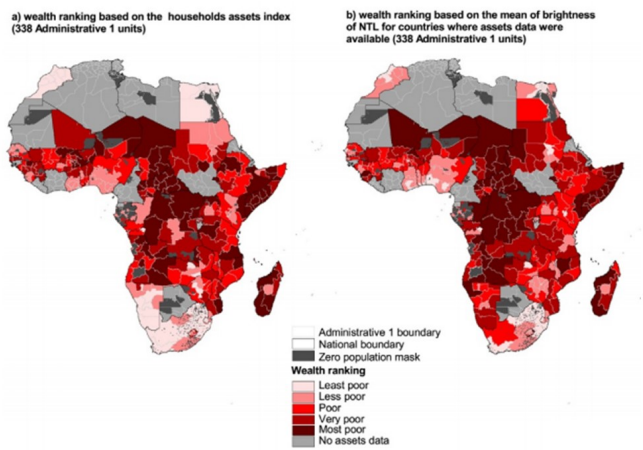
- Mapping: Doll et al. (2000)
- Country-level correlation of NTL with GDP and CO2 emissions for 46 countries.
- Sub-national extrapolation: mapping of GDP predicted by NTL



- Sutton and Costanza (2002): Country-level correlation bet. Light energy and GDP = 0.74. Build a 1km2 resolution map of economic activity

Combining NTL data with sub-national surveys and censuses to map development

- Ebener et al. (2005): Is it possible to generate sub-national income per capita series for countries where no sub-national figures are available?
- GDP figures at the first or second administrative level have been collected for 26 countries (653 units).
- Noor et al. (2008) Combination of MICS and DHS surveys to construct wealth indexes at the ADM1 level in 37 African countries (338 units). Pearson correlation with NTL = 0.64***



Using NTL to measure growth dynamics

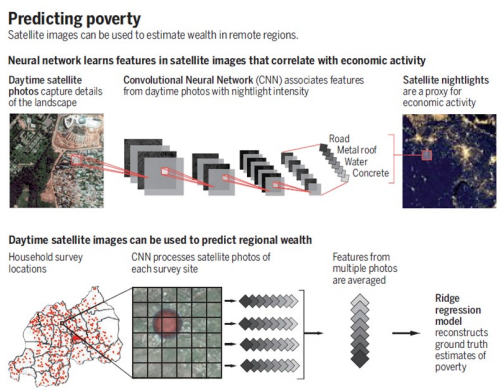
- Henderson et al. (2012): change in NTL as a proxy for GDP growth
- Issues with real GDP measures
  - Weak statistical capacity of many developing countries
  - Small fraction of economic activity is formal
  - Domestic price indices not always reliable and PPP issues
  - Very rarely sub-national information
- Alternative proxies in the literature: electricity consumption or household consumption (development economics), nb of letters mailed per capita or skeletal remains (economic history)
- NTL has informational value for countries with low-quality statistical systems (Chen and Nordhaus, 2011)

Combination of NTL and daytime imagery

- In poor areas, NTL very low with little variation $\Longrightarrow$ less useful to study living conditions of the very poor
- MP data may be an alternative, but not always and easily available
- Standard machine learning techniques: need million of labelled images
- Alternative: transfer learning: use only publicly available data and use NTL data to "train" the machine (Jean et al., 2016)
- https://www.youtube.com/watch?v=DafZSeIGLNE

Combination of NTL and daytime imagery

- Jean et al. (2016) use a convolutional neural network (CNN) to learn the relationship between daytime satellite images (rich in detail) and nighttime images
- Light areas are assumed to be wealthy → the network learns which features in the daytime imagery are indicative of wealth
- Enable the authors to accurately reconstruct survey-based indicators of regional poverty
- A model trained in one country can be used in another → useful for countries where no recent survey data exist



Source: Blumenstock (2016)

Using daytime imagery to predict growth

Khachiyan et al. (2022) (AER: insights)

- Use multispectral imagery from Landsat
- For labels, use household income and population for census blocks in the US Census and American Community Survey (ACS)
- Train a CNN using hundreds of thousands of images and training labels
- Predict levels and changes in income and population
- Model predictions achieve R2 values of greater than 0.85 in levels
- Outperforms NTL growth predictions

Poverty mapping using VHR data - Slums

- Remote sensing image analysis: increased availability of very-high-resolution (VHR) data and methodological advances (machine learning)

→ Production of information on the geography and dynamics of slums

- VHR data: spatial resolution of 5m or less (slums, refugee camps, informal settlements)

Table 1. Morphological features typical for slum areas (adapted from [41,54]).

Features	Slum Areas	Formal Built-Up Areas
Size	• Small (substandard) building sizes	• Generally larger building sizes
Density	• (Very) high roof coverage densities • Lack of public (green) spaces within or in the vicinity of slum areas	• Low to moderate density areas • Provision of public (green spaces) within or in vicinity of planned areas
Pattern	• Organic layout structure (no orderly road arrangement and noncompliance with set-back standards)	• Regular layout pattern (showing planned regular roads and compliance with set-back rules)
Site Characteristics	• Often at hazardous locations (e.g., flood prone, close to industrial areas, steep slope) • Proximity to infrastructure lines and livelihood opportunities	• Land has basic suitability for being built-up • (Basic) infrastructure is provided

Poverty mapping using VHR data - Slums

- Survey paper: Kuffer et al. (2016)
- Purposes of slum mapping using remote sensing
  - Where (urban fabric)
  - When (urban dynamics)
  - What (sociodemographic characteristics)

Table 7. Frequency of methods *versus* main focus for slum mapping using VHR imagery.

		Methods							Total Number (Percentage)
		Contour Model	Machine Learning	Object-Based Approach	Pixel-Based Approach	Statistical Model	Texture/ Morphology	Visual Image Interpretation	
FOCI	Analysis of types of informal/slum areas	0	1	1	0	1	1	2	6 (6.9%)
	Correlation with socioeconomic indicators	0	0	1	3	0	0	1	5 (5.7%)
	Identification of slum areas	0	8	15	3	2	9	11	48 (55.2%)
	Extractions of roofs/roads (objects)	4	0	7	0	0	1	1	13 (14.9%)
	Land use/cover mapping	0	2	4	5	1	3	0	15 (17.2%)
Total Number (Percentage)		**4 (4.6%)**	**11 (12.6%)**	**28 (32.2%)**	**11 (12.6%)**	**4 (4.6%)**	**14 (16.1%)**	**15 (17.2%)**	**87 (100%)**

Population mapping using RS

- Tobler et al. (1997) (Gridded Population of the World - GPW) gather information on 219 countries from various agencies and institutions
- Population figures were taken from the latest available census
- Problem: to convert the published population count for a given administrative unit into a population within the grid cells
- Grid assignment by a smoothing technique

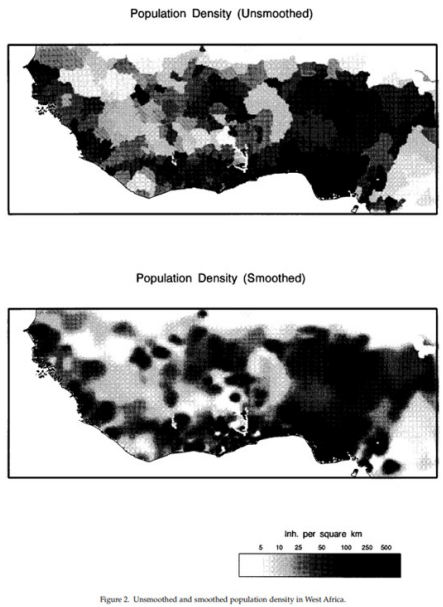


Figure 2. Unsmoothed and smoothed population density in West Africa.

https://sedac.ciesin.columbia.edu/

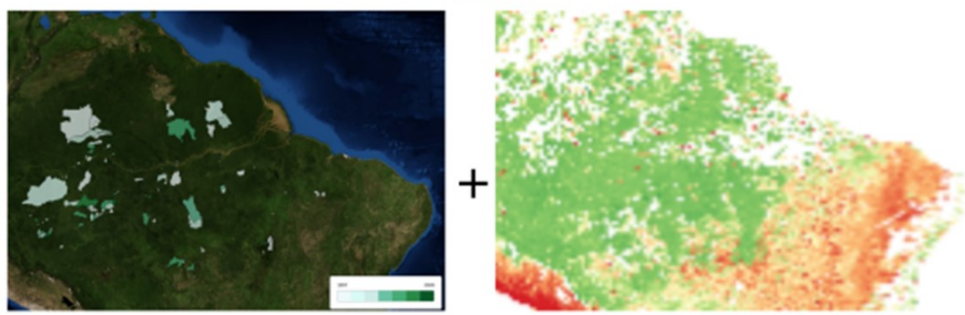
Population mapping using RS

- GRUMP project: 1kmx1km resolution
- Tatem et al. (2007) 100mx100m resolution
  - Combine land cover information and census data
- Cheriyadat et al. (2007): settlement maps
  - Feature extraction from high-resolution imagery (texture edge)
  - Semi-automated
- Population mobility: MP and SM data

From description to causality

- Geospatial impact evaluation
  https://www.aiddata.org/gie
- Quasi-experimental methods:
  - Strategic sub-sampling
  - Diff-in-diff
  - Spatial discontinuities

Figure 3: Joining Geocoded KFW Intervention Data With Remotely Sensed Deforestation Data in Brazil



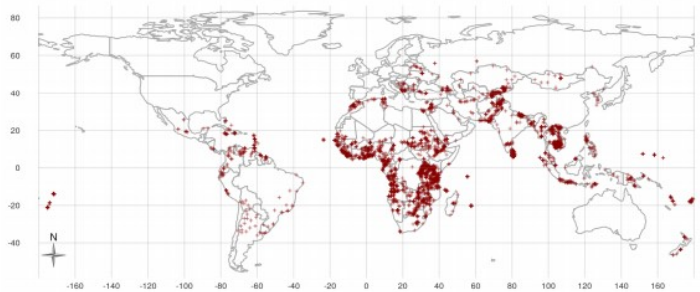
Georeferenced program locations Georeferenced outcome data

Source: BenYishay et al. (2017)

From description to causality

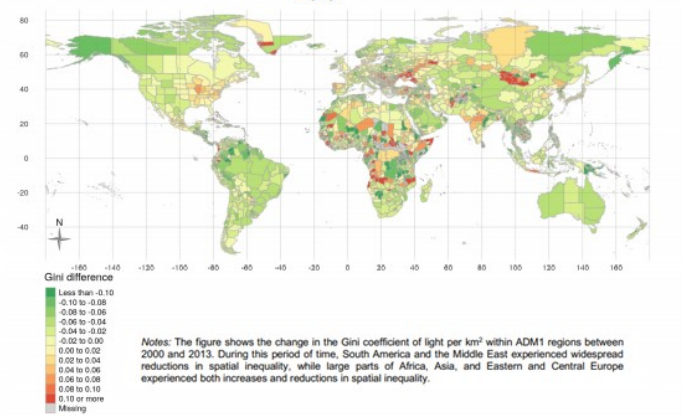
- Impact of Chinese investment in infrastructure on economic activity

Figure 1: Locations of Chinese Government-Financed Projects, 2000-2014



Notes: The figure shows all georeferenced Chinese Government-financed projects that reached the implementation or completion stage over the period 2000 to 2014.

Figure 4: Changes in Nighttime Light Inequality within ADM1 regions between 2000 and 2013



Notes: The figure shows the change in the Gini coefficient of light per km² within ADM1 regions between 2000 and 2013. During this period of time, South America and the Middle East experienced widespread reductions in spatial inequality, while large parts of Africa, Asia, and Eastern and Central Europe experienced both increases and reductions in spatial inequality.

Table 1: Chinese aid and within-region inequality, ADM1, 2002-2013, OLS & 2SLS

	Sector of projects			
	(1) All	(2) ODA	(3) OOF	(4) Transport
	Panel a) OLS estimates – Dependent variable: Gini			
$ChnAid_{t-2}$	-0.0008	-0.0018	-0.0013	-0.0112***
	(0.0017)	(0.0020)	(0.0021)	(0.0040)
Log population	-0.0198	-0.0198	-0.0198	-0.0199
	(0.0144)	(0.0144)	(0.0144)	(0.0144)
	Panel b) Reduced form estimates – Dependent variable: Gini			
$Steel_{t-3} \times \bar{p}$	-0.0267***	-0.0313***	-0.0371**	-0.0847***
	(0.0090)	(0.0109)	(0.0164)	(0.0240)
Log population	-0.0181	-0.0180	-0.0191	-0.0193
	(0.0141)	(0.0139)	(0.0144)	(0.0143)
	Panel c) 2SLS estimates – Dependent variable: Gini			
$ChnAid_{t-2}$	-0.0706***	-0.0977***	-0.0417**	-0.1038***
	(0.0249)	(0.0367)	(0.0189)	(0.0302)
Log population	-0.0180	-0.0178	-0.0190	-0.0203
	(0.0145)	(0.0139)	(0.0146)	(0.0145)
	Panel d) First-stage estimates – Dependent variable: $ChnAid_{t-2}$			
$Steel_{t-3} \times \bar{p}$	0.3783***	0.3200***	0.8897***	0.8158***
	(0.0594)	(0.0620)	(0.0697)	(0.1095)
Log population	0.0021	0.0024	0.0021	-0.0104
	(0.0288)	(0.0199)	(0.0198)	(0.0123)
Kleibergen-Paap F-statistic	40.60	26.66	162.90	55.52
Observations	29,881	29,881	29,881	29,881
Regions	2,674	2,674	2,674	2,674
Countries	158	158	158	158

Notes: The table reports OLS and 2SLS regression results at the regional (ADM1) level. The dependent variable is the Gini coefficient of light per km² within ADM1 regions in panels a-c and a binary indicator for the presence of Chinese aid projects in panel d. All specifications include ADM1-level fixed effects and country-year fixed effects. Standard errors clustered at the country level are in parentheses. $ChnAid_{t-2}$ refers to Chinese aid projects in the sector(s) indicated in the column header. The regional probability of receiving aid, $\bar{p}$, is also sector-specific and computed according the type indicated in the column title. Significant at: *** p<0.01, ** p<0.05, * p<0.1.

Source: Bluhm et al. (2018)

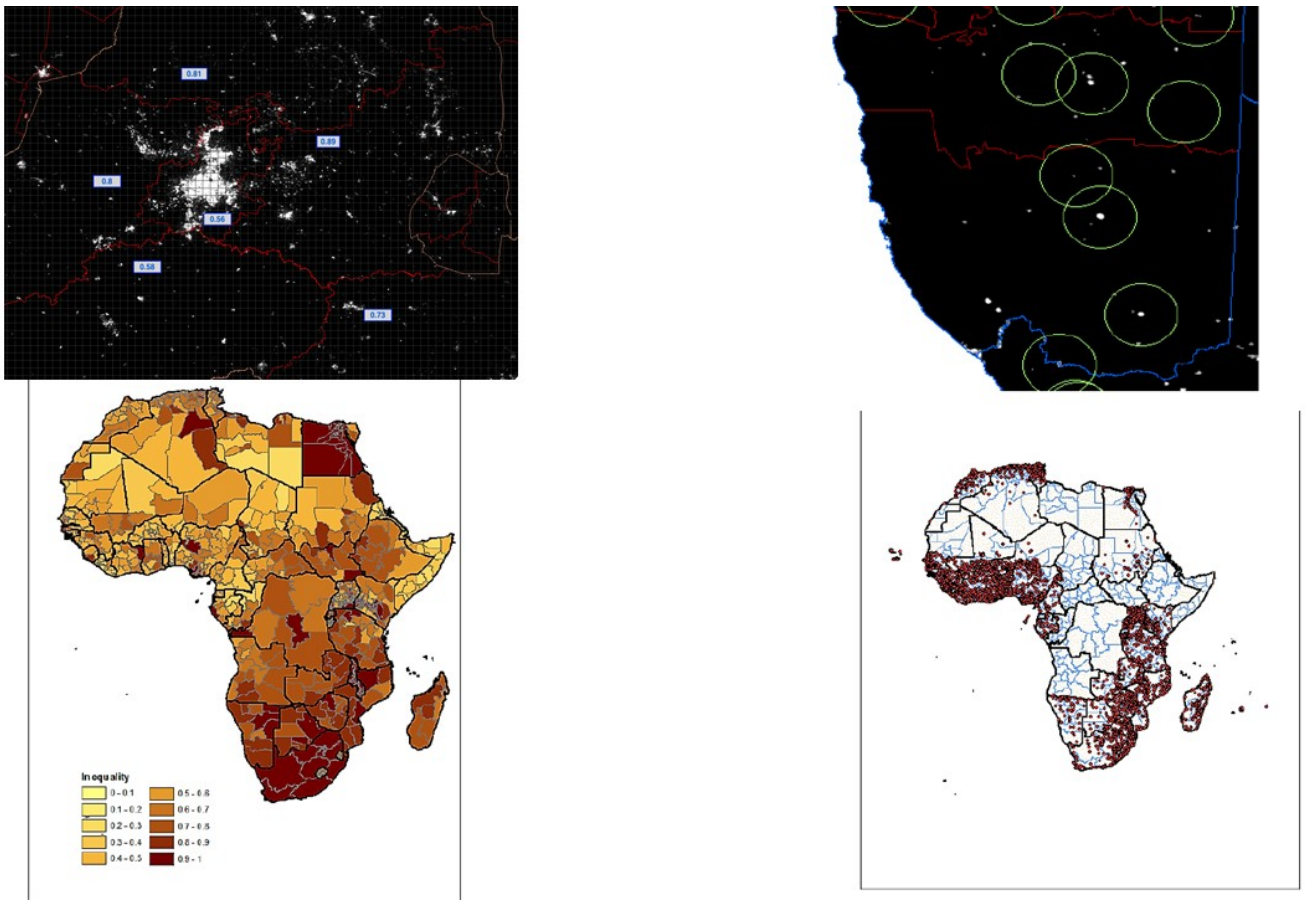
From description to causality

- Association of NTL data with geocoded data on:
  - Aid (Chinese, World Bank, country level e.g. UGA, NPL etc.)
  - Electoral results (CLEA dataset)
  - DHS, Afrobarometer, Conflict (PRIO, ACLED)

  $\rightarrow$ At various geographical levels, from ADM1 to cities, or even buffers

From description to causality

- Analysis at the buffer level



Source: Chauvet et al. (2020)

Plan

MP

- Cost-effective data at the individual level
  - Frequency and timing of communication events
  - Travelling patterns
  - Histories of consumption and expenditure
- Can be related to other data on public goods etc.

$\rightarrow$ with ML and regression analysis: socioeconomic status can be inferred at different levels

$\rightarrow$ Data Intensive Development Lab: https://didl.berkeley.edu/

Using phone network data to map development

- Eagle et al. (2010):
  - UK communication network data (2005): more than 90% of the mobile phones and greater than 99% of the residential and business landlines in the country
  - Used to construct community diversity measures
  - Combined with UK government's Index of Multiple Deprivation

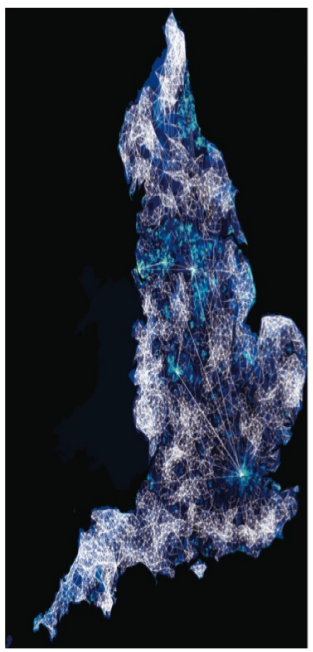


Fig. 1. An image of regional communication diversity and socioeconomic ranking for the UK. We find that communities with diverse communication patterns tend to rank higher (represented from light blue to dark blue) than the regions with more insular communication. This result implies that communication diversity is a key indicator of an economically healthy community. (C29) Crown copyright material is reproduced with the permission of the Controller of Her Majesty's Stationery Office]

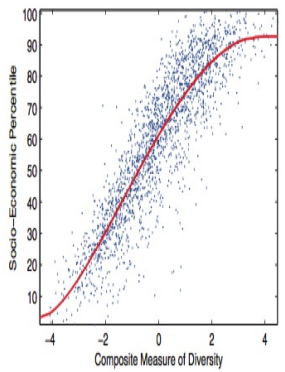


Fig. 2. The relation between social network diversity and socioeconomic rank. Diversity was constructed as a composite of Shannon entropy and Burt's measure of structural holes, by using principal component analysis. A fractional polynomial was fit to the data.

Using MP data to map wealth distribution

- Blumenstock et al. (2015): individuals' past history of mobile phone can be used to infer their socioeconomic status
- To reconstruct the distribution of wealth of an entire nation or to infer the asset distribution of microregions composed of just a few households
- Create high-resolution maps of the geographic distribution of wealth
- Using survey data on 856 individuals wealth with their phone records:
  - Show that individual's wealth can be predicted from his or her historical patterns of phone use
  - Generate out-of-sample prediction for 1.5 million MP users

Using MP data to map wealth distribution

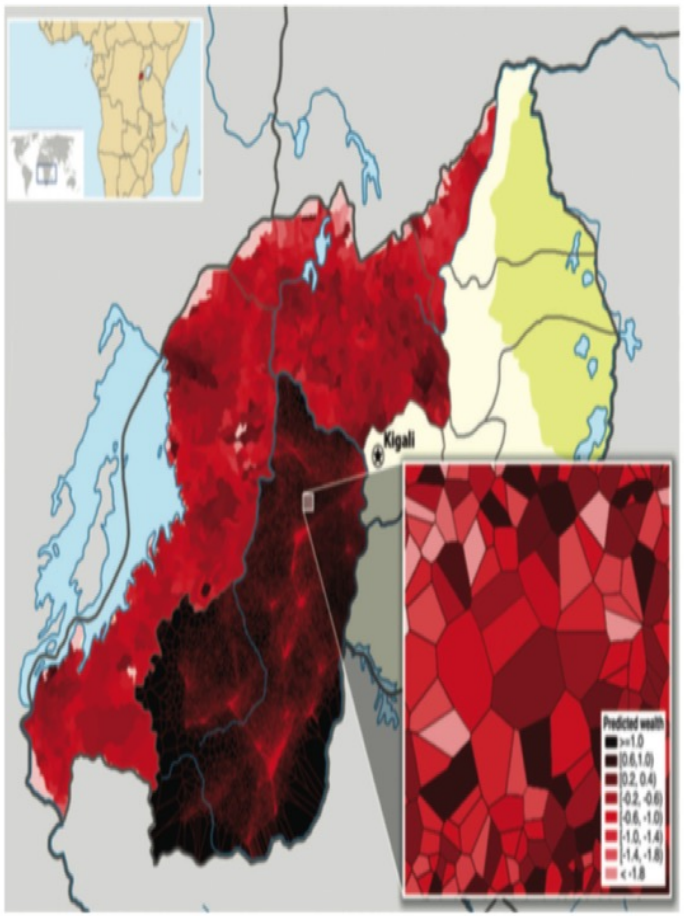


Fig. 2. Construction of high-resolution maps of poverty and wealth from call records. Information derived from the call records of 1.5 million subscribers is overlaid on a map of Rwanda. The northern and western provinces are divided into cells (the smallest administrative unit of the country), and the cell is shaded according to the average (predicted) wealth of all mobile subscribers in that cell. The southern province is overlaid with a Voronoi division that uses geographic identifiers in the call data to segment the region into several hundred thousand small partitions. (**Bottom right inset**) Enlargement of a 1-km^2 region near Kiyonza, with Voronoi cells shaded by the predicted wealth of small groups (5 to 15 subscribers) who live in each region.

Combining MP data with RS

- RS: Cost-effective information about physical properties of land, but coarse for urban areas
- MP: CDRs have high spatial resolution in urban areas, but insufficient coverage in rural areas (towers)
- Steele et al. (2017) combine the two: better predictive maps of socioeconomic status

Combining MP data with RS

poverty metric	model	r^2
whole country		
DHS WI	CDR–RS	0.76
	CDR	0.64
	RS	0.74
PPI	CDR–RS	0.25
	CDR	0.23
	RS	0.32
income	CDR–RS	0.27
	CDR	0.24
	RS	0.22

poverty metric	model	r^2
urban		
DHS WI	CDR–RS	0.78
	CDR	0.70
	RS	0.71
PPI	CDR–RS	0.00
	CDR	0.03
	RS	0.00
income	CDR–RS	0.15
	CDR	0.15
	RS	0.05

poverty metric	model	r^2
rural		
DHS WI	CDR–RS	0.66
	CDR	0.50
	RS	0.62
PPI	CDR–RS	0.18
	CDR	0.17
	RS	0.21
income	CDR–RS	0.14
	CDR	0.13
	RS	0.23

Measuring human mobility using MP data

- Location (Lat and Long) of the base transceiver stations (BTS) that the cell phone was connected to
- Rubio et al. (2010): In DCs, smaller average travelled distance, smaller geographical sparsity of social networks, etc.

Measuring human mobility using MP data

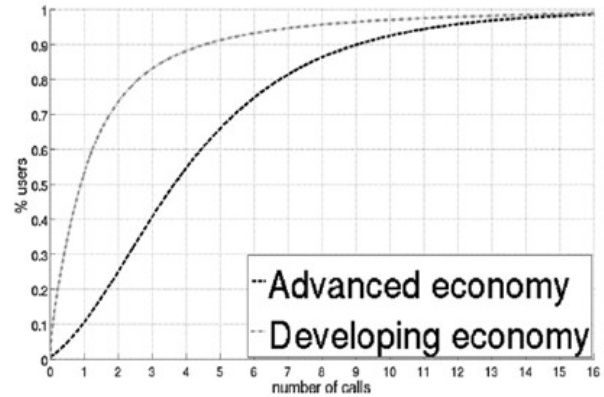
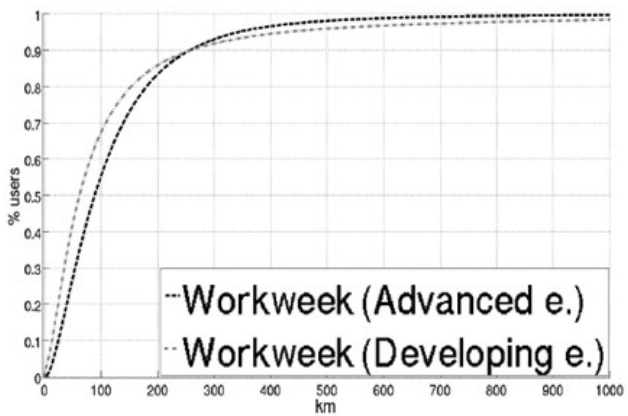
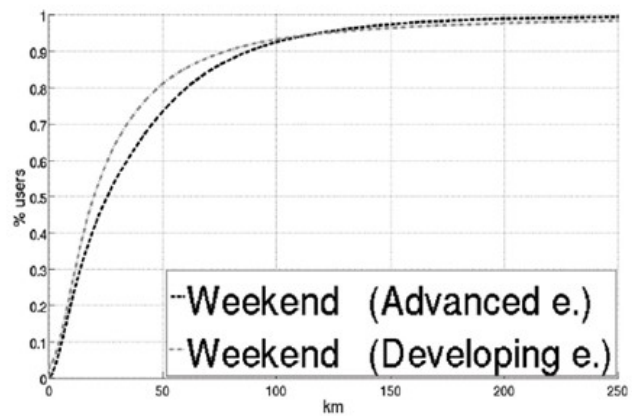


Figure 1: CDF of the average number of calls per day for a developing and an advanced economy.



(a) Workweeks



(b) Weekends

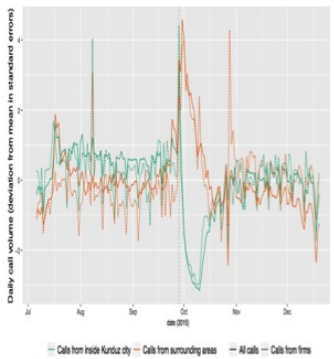
Figure 2: CDF of the average distance traveled during (a) workweeks and during (b) weekends.

From description to causal analyses

- Blumenstock et al. (2020): How the private sector responds to terrorist attacks in Afghanistan?
- Presence, entry, and exit of private firms measured at high frequency and spatial granularity using administrative records of corporate mobile phone activity
- Panel fixed-effects regressions: firm-by-district and month fixed effects, as well as linear and quadratic district-specific time trends
- Similarly, Lu et al. (2012) use MP data to analyse human mobility after 2010 Haiti earthquake

From description to causal analyses

Figure 1: Mobile Phone Activity and the Fall of Kunduz (September 2015)



Notes: Figure shows normalized mobile phone call volumes by corporate subscribers (dashed lines) and all subscribers (solid lines) in the Kunduz region in 2015. Green lines indicate calls from numbers within 10km of the city center; Orange lines indicate calls initiated from between 10km and 70km of the city center. Vertical dashed lines mark the initial date of the Taliban's attack on Kunduz city (September 28, 2015). See Appendix C1 for additional discussion.

Table 2: Firm District Activity After Major Terrorist Attacks

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Any calls made from district	Firm has employee who is active in district (=100)					
Major Terrorist Attack (1 lag)	13.058	-0.995**	-0.772*	-1.003**	-0.239***	-0.188***
	(10.588)	(0.477)	(0.446)	(0.497)	(0.068)	(0.063)
Mean Outcome	4.9935	4.9939	4.9939	4.9939	4.9939	4.9939
Beta/Mean	2.6151	-0.1994	-0.1547	-0.2009	-0.0479	-0.0376
Observations	15818428	15816179	15816179	15816179	15816179	15816179
Adj R2	0.0031	0.5802	0.5813	0.5817	0.5834	0.5835
Panel B: Employee based in district	Firm has employee whose primary tower is in district (=100)					
Major Terrorist Attack (1 lag)	13.234	-0.719	-0.657	-0.801	-0.147**	-0.107**
	(9.806)	(0.446)	(0.433)	(0.509)	(0.061)	(0.049)
Mean Outcome	1.7120	1.7122	1.7122	1.7122	1.7122	1.7122
Beta/Mean	7.7303	-0.4201	-0.3838	-0.4678	-0.0860	-0.0625
Observations	15818428	15816179	15816179	15816179	15816179	15816179
Adj R2	0.0091	0.6860	0.6861	0.6862	0.6878	0.6878
District-Firm FEs	No	Yes	Yes	Yes	Yes	Yes
Time FEs	No	No	Yes	Yes	Yes	Yes
District-Season FEs	No	No	No	Yes	Yes	Yes
District Lin Trends	No	No	No	No	Yes	Yes
District Quad Trends	No	No	No	No	No	Yes

Notes: Observation is a firm-district-month. Dependent variable in Panel A equals 1 if any call was made by that firm in that district-month, and 0 otherwise. Dependent variable in Panel B equals 1 if the modal calling tower for at least one of the firm's phones was in that district during that month, and 0 otherwise. Major Terrorist Attack equals 1 if previous month in top 1% of killings distribution, and 0 otherwise. Standard errors clustered at district level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Firm District Entry and Exit After Major Terrorist Attacks

	(1) Firm Active (=100)	(2) Firm Entry (=100)	(3) Firm Exit (=100)	(4) Modal Active (=100)	(5) Modal Entry (=100)	(6) Modal Exit (=100)
Major Terrorist Attack (1 lag)	-0.188***	-0.110***	0.079*	-0.107**	-0.041*	0.060*
	(0.063)	(0.034)	(0.042)	(0.049)	(0.021)	(0.034)
Mean Outcome	4.994	1.428	1.473	1.712	0.248	0.258
Beta/Mean	-0.0376	-0.0771	0.0538	-0.0625	-0.1666	0.2325
District-Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
District-Season FEs	Yes	Yes	Yes	Yes	Yes	Yes
District Lin Trends	Yes	Yes	Yes	Yes	Yes	Yes
District Quad Trends	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15816179	15417587	15417587	15816179	15417587	15417587
Adj R2	0.5835	0.0914	0.0924	0.6878	0.0685	0.0686

Notes: Observation is a firm-district-month. Firm Entry (Exit) equals 1 if firm is absent (present) for at least 1 prior month and then present (absent) for at least 1 month, where presence is measured by at least one call made by one of the firm's phones from that district in that month. Modal Entry (Exit) is defined analogously, but where presence is measured by the modal calling tower for at least one of the firm's phones being in that district during that month. Standard errors clustered at district level. *** p<0.01, ** p<0.05, * p<0.1.

https://www.youtube.com/watch?v=ga3y1QS9RiY

Text Analysis: SM and online data

to be continued...

Plan

Studies using NTLs

Table 3. Selected Uses of DMSP Night Lights Data: Studies With Lights Variables on the Left-Hand Side (most recent studies first).

Authors	Year	Spatial Unit	Dep Variable	Deblur?	Address Top-Coding?	Satellite FE?	Year FE?	Objective
Kocornik-Mina *et al.*	2020	Cities (n=1868 in 40 countries)	ln(DN), 2003–2008	No	No	No	Yes	Examine effects of floods on economic activity
Dreher *et al.*	2019a	Districts (n=5835) in 47 African countries	ln(DN+0.01 per capita), 2001–2012	No	No	No	Yes	Examine effects of Chinese aid projects on local economic development (proxied by lights)
Mamo *et al.*	2019	Districts (n=3635) in 42 African countries	ln(Σ DN per km^2 +0.01), 1992–2012	No	No	No	Yes	Examine effects of intensive and extensive margin of mining on local economic development
Heger and Neumayer	2019	Subdistricts (n=276) in Aceh, Indonesia	ln(sum of DN+0.001), 2003–2012	No	No	No	Yes	Examine effects of Boxing Day tsunami and reconstruction aid on subsequent economic growth
Prakash *et al.*	2019	State Assembly constituencies (n=2633) in India	Annual growth in Σ DN per km^2, 2004–2008	No	No	No	Yes	Examine effects of electing criminally accused politicians on rate of growth in night lights
Eberhard-Ruiz *et al.*	2019	Cities (n=180 in three African countries)	Annual Δln(DN), 1992–2013	No	No	No	Yes	Examine effect of RTA on city growth near internal borders

Source: Gibson et al. (2020)

Studies using NTLs

Table 3. *Continued.*

Authors	Year	Spatial Unit	Dep Variable	Deblur?	Address Top-Coding?	Satellite FE?	Year FE?	Objective
Düben and Krause	2019	Cities (n=13,844 in 194 countries)	Sum of Lights, 1992–2013	No	Yes (Pareto adjust)	Mostly not (except Table 6)	Yes	Examine Zipf.s law, in terms of lights and population
Lee	2018	1.9 km × 1.4 km grid, for N. Korea	ln(DN+0.01), 1992–2013	No	No	No	Yes	Examine within-North Korea regional impacts of sanctions
Smith and Wills	2018	10 km × 10 km grid, for 36 non-OECD countries	% in unlit rural area and sum of lights, 2000–2012	No	No	No	Yes	Examine how oil booms affect rural poverty (based on unlit area) and regional inequality
Mitnik *et al.*	2018	Communes (n=207) in Haiti	IHS Sum of lights 1994–2013[a]	Yes	Yes (use rad-cal data)	Yes	Yes	Examine economic impacts of road improvements
Castelló Climent *et al.*	2017	Districts (n=500) in 20 Indian states	ln(Σ DN per km^2), in 2006	No	Yes (use rad-cal data)	No	No	Examine how higher education affects economic development
Corral and Schling	2017	Beachs (n=23) in Barbados	ln(deblurred DN), 1992–2010	Yes	No	No	Synth control	Examine local economic impact of shoreline stabilization
Villa	2016	Districts (n=732 in Colombia)	Annual Δ(ΣDN), 1998–2004	No	No	No	Yes	Examine how social transfers affect local growth (in lights)
Storeygard	2016	Cities (n=289 in 15 African countries)	ln(Σ DN per city), 1992–2008	No	No[b]	Partly (average within sat-year)	Yes	Examine effects of transport costs on urban economic output

Source: Gibson et al. (2020)

Studies using NTLs

Table 3. *Continued.*

Authors	Year	Spatial Unit	Dep Variable	Deblur?	Address Top-Coding?	Satellite FE?	Year FE?	Objective
Baskaran *et al.*	2015	State Assembly Constituencies (n=3800) in India	ln(Σ DN+1 per capita),[c] 2001–2012	No	No	No	Yes	Examine political manipulation of electricity supply during special elections
Gibson *et al.*	2015	Cities (n=47) with pop>1 m in India	ln(urban lit area), 1992–2012	No	Yes (min DN criteria)	Yes	Yes	Examine urban area growth and change from prior land cover
Hodler and Raschky	2014	Districts (n=38,427) in 126 countries	ln (DN+0.01), 1992–2009	No	No	No	Yes	Examine how the birth regions of political leaders benefit (in terms of brighter lights)
Michalopoulos and Papaioannou	2014	Cross-border ethnic homeland partitions (n=507) in Africa[d]	ln (mean DN+0.01), 2007–2008	No	No	No	Yes	Examine the effect of national institutions on local economic development (lights), comparing across borders, within-ethnicity

[a]IHS is the inverse hyperbolic sine transformation.
[b]Uses a Tobit for bottom-coding.
[c]Also use annual growth rate of the sum of lights per capita, and share of villages with lights detected by DMSP.
[d]Also use a dummy for whether DMSP data indicate a pixel is lit or not, for aggregated pixels that are 1/64th of a decimal degree (ca 150 km^2 near the equator).

Source: Gibson et al. (2020)

Bibliographie I

Abrahams, A., C. Oram, and N. Lozano-Gracia (2018). Deblurring dmsp nighttime lights: A new method using gaussian filters and frequencies of illumination. *Remote Sensing of Environment 210*, 242–258.

BenYishay, A., D. Runfola, R. Trichler, C. Dolan, S. Goodman, B. Parks, J. Tanner, S. Heuser, G. Batra, and A. Anand (2017). A primer on geospatial impact evaluation methods, tools, and applications. In *AidData Working Paper# 44*. AidData at William & Mary Williamsburg, VA.

Bluhm, R., A. Dreher, A. Fuchs, B. Parks, A. Strange, and M. J. Tierney (2018). Connective financing: Chinese infrastructure projects and the diffusion of economic activity in developing countries.

Bibliographie II

Bluhm, R. and M. Krause (2018, November). Top Lights: Bright cities and their contribution to economic development. MERIT Working Papers 2018-041, United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT).

Blumenstock, J., G. Cadamuro, and R. On (2015). Predicting poverty and wealth from mobile phone metadata. *Science 350*(6264), 1073–1076.

Blumenstock, J., T. Ghani, S. Herskowitz, E. Kapstein, T. L. Scherer, and O. Toomet (2020). How do firms respond to insecurity? evidence from afghan corporate phone records. Technical report, mimeo.

Blumenstock, J. E. (2016). Fighting poverty with data. *Science 353*(6301), 753–754.

Bibliographie III

Chauvet, L., S. Eslami, M. Ferry, L. Pasquier-Doumer, et al. (2020). Inequality in public good provision and attitude towards taxation: Sub-national evidence from africa. Technical report.

Chen, X. and W. D. Nordhaus (2011). Using luminosity data as a proxy for economic statistics. *108*(21), 8589–8594.

Cheriyadat, A., E. Bright, D. Potere, and B. Bhaduri (2007). Mapping of settlements in high-resolution satellite imagery using high performance computing. *GeoJournal 69*(1-2), 119–129.

Doll, C. N., J.-P. Muller, and C. D. Elvidge (2000). Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. *Ambio*, 157–162.

Eagle, N., M. Macy, and R. Claxton (2010). Network diversity and economic development. *Science 328*(5981), 1029–1031.

Bibliographie IV

Ebener, S., C. Murray, A. Tandon, and C. C. Elvidge (2005). From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *international Journal of health geographics 4*(1), 1–17.

Elvidge, C. D., K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing 18*(6), 1373–1379.

Gao, J., Y.-C. Zhang, and T. Zhou (2019). Computational socioeconomics. *Physics Reports 817*, 1–104. Computational Socioeconomics.

Gibson, J., S. Olivia, and G. Boe-Gibson (2020). Night lights in economics: Sources and uses1. *Journal of Economic Surveys 34*(5), 955–980.

Bibliographie V

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American economic review 102*(2), 994–1028.

Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon (2016). Combining satellite imagery and machine learning to predict poverty. *Science 353*(6301), 790–794.

Khachiyan, A., A. Thomas, H. Zhou, G. Hanson, A. Cloninger, T. Rosing, and A. K. Khandelwal (2022, December). Using neural networks to predict microspatial economic growth. *American Economic Review: Insights 4*(4), 491–506.

Kuffer, M., K. Pfeffer, and R. Sliuzas (2016). Slums from space—15 years of slum mapping using remote sensing. *Remote Sensing 8*(6), 455.

Bibliographie VI

Lu, X., L. Bengtsson, and P. Holme (2012). Predictability of population displacement after the 2010 haiti earthquake. *Proceedings of the National Academy of Sciences 109*(29), 11576–11581.

Noor, A. M., V. A. Alegana, P. W. Gething, A. J. Tatem, and R. W. Snow (2008). Using remotely sensed night-time light as a proxy for poverty in africa. *Population Health Metrics 6*(1), 1–13.

Nordhaus, W. and X. Chen (2014, 05). A sharper image? Estimates of the precision of nighttime lights as a proxy for economic statistics1. *Journal of Economic Geography 15*(1), 217–246.

Rubio, A., V. Frias-Martinez, E. Frias-Martinez, and N. Oliver (2010). Human mobility in advanced and developing economies: A comparative analysis. In *2010 AAAI Spring Symposium Series*.

Bibliographie VII

Steele, J. E., P. R. Sundsøy, C. Pezzulo, V. A. Alegana, T. J. Bird, J. Blumenstock, J. Bjelland, K. Engø-Monsen, Y.-A. De Montjoye, A. M. Iqbal, et al. (2017). Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface 14*(127), 20160690.

Sutton, P. C. and R. Costanza (2002). Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecological Economics 41*(3), 509–527.

Tatem, A. J., A. M. Noor, C. Von Hagen, A. Di Gregorio, and S. I. Hay (2007). High resolution population maps for low income nations: combining land cover and census in east africa. *PloS one 2*(12), e1298.

Tobler, W., U. Deichmann, J. Gottsegen, and K. Maloy (1997). World population in a grid of spherical quadrilaterals. *International Journal of Population Geography 3*(3), 203–225.