Working Paper Series

SoDa LABS

Social [science] insights from alternative data

Est. 2018

Top Lights: Bright cities and their contribution to economic development

Richard Bluhm, Melanie Krause

SoDa Laboratories Working Paper Series

No. 2020-08

REF

Richard Bluhm, Melanie Krause (2020), SoDa Laboratories Working Paper Series No. 2020-08, Monash Business School, available at http://soda-wps.s3-website-ap-southeast-2.amazonaws.com/RePEc/ajr/sodwps/2020-08.pdf

PUBLISHED ONLINE

5 November 2020

SoDa Laboratories

https://www.monash.edu/business/soda-labs/

MONASH University | MONASH BUSINESS SCHOOL

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION AMBA ACCREDITED

ABN 12 377 614 012 CRICOS Provider Number: 00008C

Top Lights

Bright cities and their contribution to economic development

Richard Bluhm* Melanie Krause†

June 2020

Abstract

Tracking the development of cities in emerging economies is difficult with conventional data. Even the commonly-used satellite images of nighttime light intensity fail to capture the true brightness of larger cities. This paper shows that nighttime lights can be used as a reliable proxy for economic activity at the city level, provided they are first corrected for top-coding. We present a stylized model of urban luminosity and empirical evidence which both suggest that these 'top lights' can be characterized by a Pareto distribution. We then propose a correction procedure which recovers the full distribution of city lights. Our results show that the brightest cities account for nearly a third of global economic activity. Applying this approach to cities in Sub-Saharan Africa, we find that primate cities are outgrowing secondary cities but are changing from within. Poorer neighborhoods are developing and sub-centers are emerging, with the side effect that Africa's cities are also becoming increasingly fragmented.

JEL Classification: O10, O18, R11, R12

Keywords: Development, urban growth, night lights, top-coding, inequality

---

This paper has been presented at BrownU, UEA, ERSA, ECINEQ, EEA, EPCS, GGDC anniversary conference, IARIW, MonashU, RCEF, RES, SEM, UAE, UBrisbane, UGold Coast, UGroningen, UHamburg, UKiel, ULausanne, U St. Gallen, UUmeå, UVienna, VfS AEL, and VfS Annual Meeting. We have greatly benefited from discussions with the participants. We would like to thank Alexei Abrahams, Gordon Anderson, Gerda Asmus, Francesco Audrino, Benjamin Bechtel, Chris Elvidge, Xavier Gabaix, Oded Galor, Martin Gassebner, Roland Hodler, Robert Inklaar, Sebastian Kripfganz, Rafael Lalive, Christian Lessmann, Stelios Michalopoulos, Maxim Pinkovskiy, Stefan Pichler, Paul Raschky, Prasada Rao, Nicholas Rohde, Dominic Rohner, Paul Schaudt, André Seidel, Adam Storeygard, Eric Strobl, and David Weil for helpful comments and suggestions. We gratefully acknowledge financial support from the German Science Foundation (DFG). All remaining errors are ours.

*Corresponding author. Leibniz University Hannover, Institute of Macroeconomics, and University of California San Diego, Department of Political Science, e-mail: *bluhm@mak.uni-hannover.de*

†Hamburg University, Department of Economics, e-mail: *melanie.krause@uni-hamburg.de*

1 Introduction

Cities are hubs of economic activity and productivity. About 4.2 billion people, or 55% of the world's population, currently live in urban areas, and developing countries are urbanizing at a rapid pace (United Nations, 2018). The African continent alone might add up to a billion people to its urban population by 2050. Many important questions in development economics and macroeconomics are intimately linked to understanding the processes driving the concentration of people and economic activity in cities. The lion's share of the extant literature is, however, oriented towards cities in advanced economies, for which ample data are available (Glaeser and Henderson, 2017). Much less is known about the rising cities of the 21st century, who not only lack comparable data, but are undergoing such a fast-paced urbanization that existing sources quickly become dated.

Satellite images of Earth are transforming how economists and other social scientists are tracking human activity and its consequences.[1] Nighttime images of light emissions are now an established proxy for local economic activity (Chen and Nordhaus, 2011, Henderson et al., 2012, Donaldson and Storeygard, 2016) and have been used in a variety of innovative applications (e.g. Michalopoulos and Papaioannou, 2013, Hodler and Raschky, 2014, Alesina et al., 2016, Pinkovskiy and Sala-i Martin, 2016, Lessmann and Seidel, 2017, Pinkovskiy, 2017, Henderson et al., 2018). They have several advantages over conventional survey-based data. Night lights are measured uniformly around the globe, allowing us to bypass discussions over adjustments for exchange rates and regional price levels. Moreover, the most widely-used data are available as an annual panel from 1992 to 2013 at a resolution of less than a square kilometer.

Our primary objective in this paper is to establish how these data can be used to reliably track economic activity within and across cities. A serious drawback of the standard night lights data is that they are top-coded in larger cities. The Operation Linescan System (OLS)—a part of the US Defense Meteorological Satellite Program (DMSP)—was designed to pick up dim light sources, but the satellites have a limited on-board storage capacity and are based on outdated 1970s technology. They record light intensities as integerized digital numbers from 0 DN (dark) to 63 DN (bright) and truncate all observations above this limit to save space. The upper end of this scale is, however, easily reached by the light intensity emitted by a mid-sized city. As a result, the recorded signal "flatlines" when the satellites encounter bright city lights: the central business district and the outskirts of larger cities appear to be equally bright in the truncated data.[2]

[1] For reviews of the related literature see Donaldson and Storeygard (2016), who illustrate the advantages of remotely sensed data in general, and Michalopoulos and Papaioannou (2018), who focus on the night lights data in particular.

[2] The data are known to suffer from a variety of other problems, such as bottom-coding (Jean et al., 2016), overglow or blooming (Abrahams et al., 2018), and geolocation errors (Tuttle et al., 2013). The new Visible Infrared Imaging Radiometer Suite (VIIRS) has considerably improved sensors, including a

The scale of the truncation is sizable. In these so-called 'stable lights' data, the urban centers of large, busy cities, such as New York or London appear to emit as little light as smaller American or British towns. Our estimates instead suggest that they are more than an order of magnitude brighter than recorded in the original data. While top-coding tends to affect developed countries more than their less-developed counterparts, we show that it is pervasive and distorts the ranking of cities within and across countries. Nearly all primary cities in Africa and mid-sized cities in Asia hit the top-coding threshold. Large agglomerations in developing countries, such as Johannesburg or New Delhi, are affected particularly strongly.

In this paper, we analyze the global distribution of city lights at the pixel level, develop a new procedure to recover the details of within city activity, and then study the evolution of cities in Sub-Saharan Africa. We make three distinct contributions to the literature:

First, we argue that it is natural to characterize the distribution of the world's brightest lights, which we dub 'top lights', by a Pareto distribution. We provide theoretical and empirical evidence supporting this claim. In terms of theory, we present a stylized model of light emissions from cities, combining standard assumptions on the evolution of city sizes (e.g. Zipf's law) with regularities in urban scaling. The model gives rise to a power law in light emissions above a certain threshold. Our empirical tests based on auxiliary satellite data also strongly favor a heavy-tailed Pareto distribution in top lights with an inequality parameter comparable to top incomes or wealth in the US (e.g. see Atkinson et al., 2011).

Our second contribution is methodological. Building on the Pareto property of top lights, we develop a top-coding correction for the truncated data. To do so, we combine desirable features from the DMSP-OLS data set with observations from the less frequently available radiance-calibrated satellite data. Our correction procedure involves a geo-referenced ranking method at the pixel level. Based on this method, we present a new annual panel of nighttime lights over the entire period from 1992 to 2013.[3] After the discontinuation of the DMSP-OLS series in 2013, nighttime lights are now observed by the technically superior Visible Infrared Imaging Radiometer Suite (VIIRS) which does not suffer from top-coding. Two annual VIIRS cross-sections (2015 and 2016) have now been made available and this series allows researcher to accurately track the development of cities today and in the future. Our pixel-level panel for the 22-year period until 2013, on the other hand, should be prove to be useful to researchers interested in studying development of cities or other small geographies in the recent past.

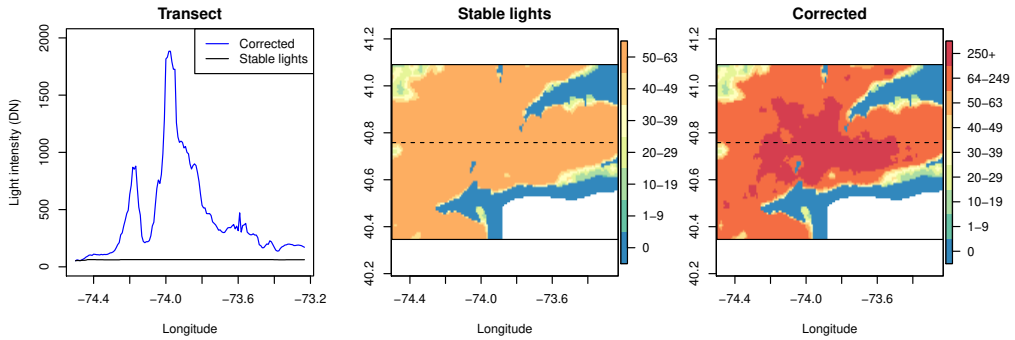
We find that our correction makes a substantial difference in virtually all major cities. The top 4% of pixels—the average share of pixels which we correct globally—account for

day-and-night band which records light intensity after midnight since late 2011. While these data will become more important in the future, the DMSP-OLS series is the only series covering the period from 1992 to 2013.

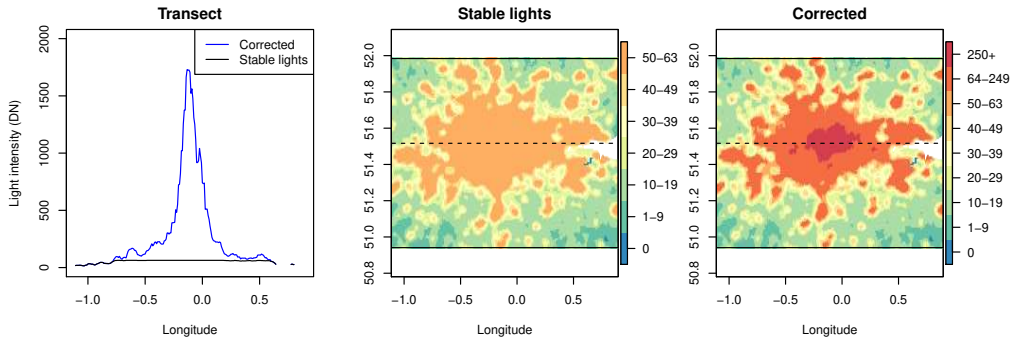
[3]Our corrected images can be downloaded from www.lightinequality.com.

Figure 1 – Selected cities in 1999, stable lights and corrected lights

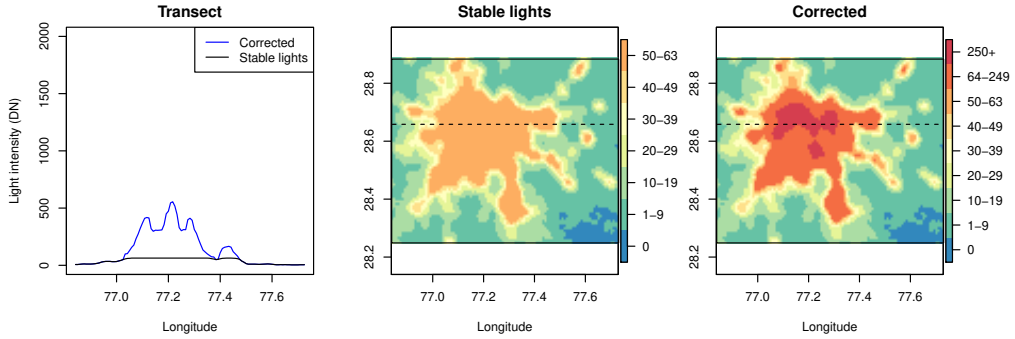
(a) New York City



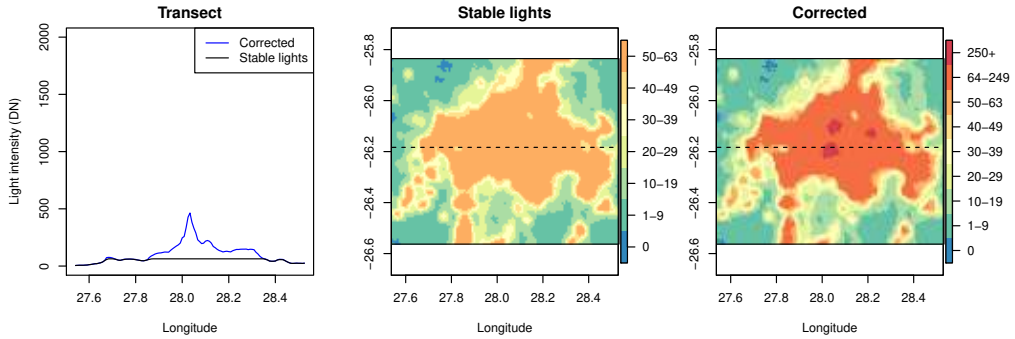
(b) London



(c) New Delhi



(d) Johannesburg



Notes: Comparison of the light intensities (DN) recorded by the stable lights and our corrected lights in four major cities. The left panel show the light intensity along a longitudinal transect through the brightest pixel in each city. The middle panel shows a map based on the stable lights data from satellite F121999. The right panel shows the same map using the corrected data presented in this paper. Both data have been binned and the color scales were adjusted so as to be comparable. Dashed lines indicate the transect path.

32% of the total brightness observed on Earth. This is nearly double their original share, underlining the contribution of big cities to global economic activity. As a result, the spatial distribution of economic activity turns out to be much less equal.

Figure 1 illustrates the problem and the effectiveness of our solution. It shows that the stable lights data are unable to differentiate among the light intensities originating from different locations within New York, London, New Delhi or Johannesburg. Across these four major cities, the average light intensity differs little and the sum of light in each city is largely determined by the urban extent. After our correction, we can clearly identify urban cores which are much brighter than the outskirts. This pattern carries over to other subnational aggregates. The corrected data also turn out to be a better predictor of regional economic activity in rich and densely populated countries, such as Germany.

Third, we use this new data to analyze the relative growth rates and internal structure of cities in Sub-Saharan Africa. The subcontinent is still the world's poorest region and urbanizing at a considerably lower level of development than other regions have in the past (Glaeser, 2014). Economic activity in Sub-Saharan Africa is heavily concentrated in primate and coastal cities due to a combination of factors, including high transport costs (Storeygard, 2016), outward-oriented colonial infrastructure (Jedwab and Moradi, 2016, Bonfatti and Poelhekke, 2017), urban bias (Lipton, 1977, Ades and Glaeser, 1995), the limited reach of national institutions (Michalopoulos and Papaioannou, 2014), climate change (Barrios et al., 2006), natural resource booms (Gollin et al., 2016), and the urban mortality transition (Jedwab and Vollrath, 2019). This structure is neither optimal nor static. On the one hand, excessive primacy in the city size distribution has been linked to the prevalence of slums and slow economic growth (Henderson, 2003, Castells-Quintana, 2017). On the other hand, recent studies suggest that Africa's secondary cities might be gaining ground vis-à-vis primate cities and playing an important role in poverty reduction (Henderson et al., 2012, Christiaensen and Todo, 2014, Christiaensen and Kanbur, 2017). Official numbers remain elusive though, so it remains an open question if secondary cities in Africa are, in fact, rising.

Our application shows that Africa's largest cities are maintaining their dominant position but are changing from within. Primary cities in Sub-Saharan Africa were growing faster than secondary cities over the period from 1992 to 2013, suggesting they are continuously absorbing growing populations in informal settlements (in line with Jedwab and Vollrath, 2019). However, as these cities grow, we observe two distinct developments: light inequality narrows and the within-city distribution of light becomes more fragmented. We interpret this finding as an indication that public services are being expanded throughout cities, while increasing fragmentation implies the formation of sub-centers with a continued lack of connectivity to other neighborhoods. Disconnectedness and long travel times, in turn, limit productivity and economies of scale (Lall et al., 2017,

Venables, 2017). Both of these findings would have been difficult to establish without the correction developed in this paper. In fact, the stable lights data indicate that secondary cities outperform primate cities during the period of study.

More broadly, this paper takes on the challenge of linking the pixel-level distribution of economic activity as observed from high resolution satellites to economic theory and empirical laws in urban economics. We argue that this approach leads to a better understanding of the features these data should have and at what scale they will be particularly useful. This perspective shows that the influence of top-coding is small when lights are aggregated to the country level but then rises steeply as the size of the unit of observation decreases. Moreover, as we demonstrate below, the impact of top-coding increases over time as countries and cities develop. Our application shows that even in Africa, which is still the poorest continent with the lowest electrification rates, top-coding makes a big difference. As the ultimate test of our argument, this suggests that top-coding will matter as least as much in other continents and regions. By providing a solution to this problem, we hope to further encourage researchers to use these historical data in innovative ways.

The remainder of this paper is organized as follows. Section 2 provides some background on the nighttime lights data and the extent of top-coding around the world. In Section 3 we present theoretical and empirical evidence in favor of a Pareto distribution in top lights. Section 4 outlines our correction procedure. Section 5 contains the application of our top-coding corrected data to African cities. Section 6 concludes. An Online Appendix contains the accompanying material, such as additional theoretical results, supplementary details on the data and summary statistics, a benchmarking exercise with German regional data, and a battery of robustness checks.

2 Top-coding around the world

The DMSP-OLS satellites circled the earth for several decades with the primary purpose of detecting clouds. As a byproduct, they measured night lights in the evening hours between 8:30 and 10:00 pm local time around the globe every day. The recorded data were pre-processed by the National Geophysical Data Center at the National Oceanic Administration Agency (NOAA) and averaged over cloud-free days.[4] The result are images of annual 'stable light' intensities from 1992 to 2013 for every 30 by 30 arc seconds pixel of the globe (about 0.86 square kilometers at the equator).[5]

We cannot use the truncated stable lights data to gauge the extent of top-coding.

[4]This was done to remove observations of cloudy days and sources of lights which are not man-made, such as auroral lights or forest fires. This process removed a lot of dim light sources. NOAA also made a series of unfiltered lights available, which we later use for the delineation of urban extents.

[5]In our analysis, we always exclude areas close to the polar zones (65 degrees south and 75 degrees north latitude) known to be influenced by ephemeral lights and remove areas affected by gas flaring.

Fortunately, for seven years, additional satellites were orbiting Earth with sensor settings that were less sensitive to light. NOAA generated a series of 'radiance-calibrated' lights by combining the stable lights data from normal flight operations with auxiliary data obtained from these low amplification sensors (Elvidge et al., 1999, Ziskin et al., 2010, Hsu et al., 2015). The resulting series is free of top-coding and has no theoretical upper bound.[6] The radiance-calibrated data can be used directly for cross-sectional analyses. Henderson et al. (2018), for example, use the 2010 version of this data in their study of the global spatial distribution of economic activity. Their measured values are, however, only of limited use for comparisons over time. Apart from their sparse temporal coverage, the radiance calibration process introduces a substantial amount of noise. For example, there are pixels whose radiance-calibrated values increase from a value of about 200 to 400 and back to 150 across the available years, without any underlying economic rationale. By contrast, the ranks of pixels in the radiance-calibrated data are considerably more stable over time than their absolute values, reaching correlations of 0.90-0.95 for adjacent years. For this reason, we rely on the shape of the distribution and the ranks of the radiance-calibrated pixels, rather than their precise values, in the pixel-level correction later on. This way, we combine desirable features from both the stable lights and the radiance-calibrated data. Online Appendix A compares summary statistics of the stable lights and radiance-calibrated data over time and discusses technical reasons for the large fluctuations we observe in the latter.

While the literature typically aggregates the pixel level data to some study area of interest (e.g. grid cell, city or region), we conduct our analysis at the native resolution of 30 arc seconds to avoid averaging over top-coded areas. On the global scale, this is a formidable task.[7] To ease the computational burden, we conduct large parts of the analysis with a spatial random sample of 10% of pixels within all countries that have a landmass larger than 500 km^2 (but later apply the correction to the full data). The sample contains more than two million pixels per year located in 194 countries or territories.

We still have to define where top-coding begins before we can assess its impact. Since the scale goes up to 63 DN, it seems natural to assume that this would be the appropriate threshold. There are, however, compelling technical and statistical arguments suggesting that the threshold should be much lower. Each value we observe in the annual data was averaged several times. In addition to averaging the high-resolution data on-board of the satellites,[8] the daily observations were averaged again in the process of generating an

[6]Note that in spite of calibration issues, Hsu et al. (2015, p. 1865) point out that, within the same year, the "DNs below saturation of the Stable Lights product and DN_EQs of merged fixed-gain imagery can be directly compared to each other."

[7]Every image contains more than 700 million pixels, about a third of which are on land and half of which are lit.

[8]Abrahams et al. (2018) provide a detailed explanation of how the DMSP-OLS satellites processed the data before transmitting them to Earth, including how this lead to geolocation errors, blurring, and top-coding. The satellite first truncated individual pixels at a much finer resolution and then aggregated

Table 1 – Summary statistics of global lights in 2010

	World	USA	Brazil	Israel	South Africa	China	Netherlands
	Panel a) Stable lights (from 0 to 63 DN)						
Mean	17.55	18.35	17.97	31.61	15.16	17.72	35.11
Standard deviation	15.35	16.90	15.45	21.49	15.29	15.50	16.09
Maximum	63.00	63.00	63.00	63.00	63.00	63.00	63.00
Spatial Gini	0.4258	0.4486	0.4165	0.3858	0.4604	0.4260	0.2637
Pixels	2,154,889	427,922	60,310	2,060	18,369	165,521	6,549
	Panel b) Radiance-calibrated lights (from 0 to ∞ DN)						
Mean	19.04	23.14	19.76	46.29	15.18	19.97	37.83
Standard deviation	44.35	53.42	41.76	82.70	29.99	46.71	44.08
Maximum	2109.67	1710.59	646.84	914.14	575.22	1862.04	435.63
Spatial Gini	0.6045	0.6613	0.5995	0.6505	0.5941	0.6093	0.4880
Pixels	2,154,889	427,922	60,310	2,060	18,369	165,521	6,549
	Panel c) Comparison of top-coded pixels						
Share = 63 DN	0.0176	0.0340	0.0228	0.1019	0.0206	0.0111	0.0299
Share $\geq$ 55 DN	0.0576	0.0847	0.0588	0.2447	0.0387	0.0597	0.1709
Radiance-cal. mean	142.05	153.77	143.36	135.70	113.27	143.96	107.92

Notes: The table reports summary statistics using the stable lights in Panel (a) and the radiance-calibrated lights in Panel (b). Panel (c) compares both sources at the pixel level. The top-coding shares are based on the stable lights data. The last line reports the radiance-calibrated mean of all pixels with saturated values $\geq$ 55 DN. All three panels are based on a 10% sample of all lit pixels, where each pixel is 30×30 arc seconds. The stable lights data are averaged across the whole year, while the radiance-calibrated data come from satellite F16, which recorded these data from January 11 to December 9, 2010.

annual image. Any pixel measured with a value of 63 on some, but not all, cloud-free days of a given year would have ended up with a yearly average below 63. Counting the share of pixels with the highest value, as is common practice, therefore only captures locations which are top-coded every could-free day of the year.

Our analysis shows that many pixels with DNs of 62, 61, down to the mid-50s, are subject to implicit top-coding and should be considerably brighter than they are recorded in the data (see Online Appendix B). NOAA even suggests that the first—albeit faint—influence of top-coding starts at much lower values.[9] Throughout this paper, we will conservatively set the top-coding threshold to 55 DN in order to not overstate its impact. Note that our correction approach will work with *any* sufficiently high top-coding threshold.

Table 1 compares the stable and radiance-calibrated lights across the globe in 2010—the latest year where both data sources are available. The first column already shows that

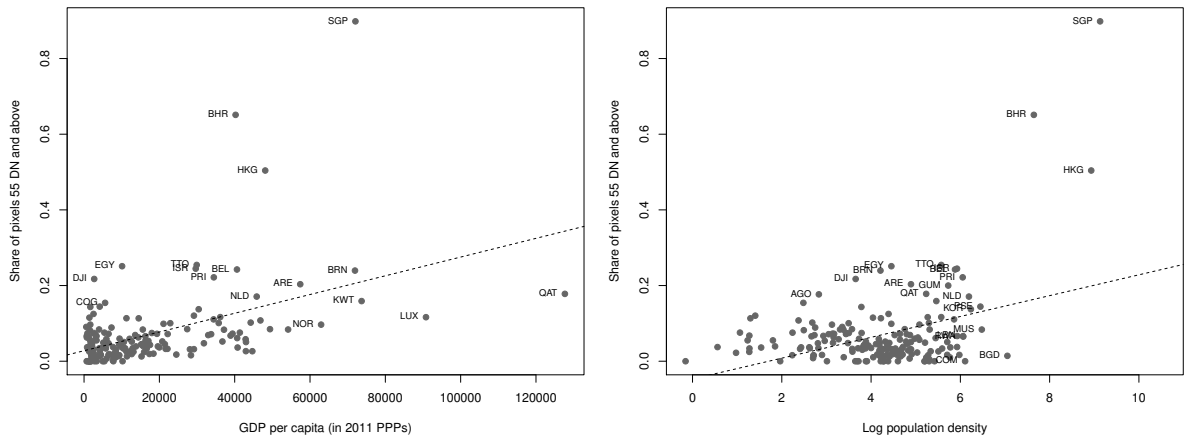
those to a coarser resolution. Each night the origin shifted a bit, recreating a finer resolution. Finally, these data were aggregated and averaged again at NOAA when the annual composites were created.

[9]Since "the OLS does onboard averaging to produce its global coverage data, saturation does not happen immediately when radiance reaches the maximum level. On the contrary, as the actual radiance grows, the observed DN value fails to follow the radiance growth linearly, causing a gradual transition into a plateau of full saturation" (Hsu et al., 2015, p. 1872).

the difference in scale is striking. The brightest radiance-calibrated pixel in our sample is more than 30 times brighter than the end of the stable lights scale. This is reflected in a standard deviation which is three times higher and a spatial Gini coefficient of inequality in lights which differs by 18 percentage points. In 2010, about six percent of all stable lights pixels are top-coded at 55 DN or higher, while their unsaturated counterparts are—on average—more than twice as bright.

Figure 2 – Share of top-coded pixels and country characteristics

(a) GDP per capita **(b)** Log population density



Notes: Illustration of the systematic bias introduced by top-coding. The data are a 10% representative sample of all non-zero lights in satellite F182010. GDP and population data are from the World Development Indicators.

Top-coding affects all countries but not uniformly. Countries which are *i)* highly developed, *ii)* small and *iii)* urbanized are more strongly affected by top-coding than others, but there is substantial heterogeneity. The remaining columns of Table 1 document this diversity in a selection of countries. Larger middle income countries, like China, Brazil or South Africa, have a top-coding share comparable to the world average. Mature economies of different sizes and population densities, such as the US, Israel or the Netherlands, have greater top-coding shares from 8% up to 25%. In Israel and the Netherlands, a high average light intensity coupled with a high incidence of top-coding in the stable lights data generate such an artificially low spatial Gini coefficient that it rises by more than 20 percentage points in the radiance-calibrated data. Figure 2 illustrates these patterns across all countries in the sample and highlights the exceptions. For example, the overwhelming majority of pixels in high income, high density city states, such as Singapore, Hong Kong, or Bahrain, are top-coded. Top-coding is also particularly pronounced in low income countries with a low average population density but large primate cities, such as Egypt. Top-coding is thus a complex function of the spatial equilibrium, that is, the size, density, number and location of cities in each country.

3 A Pareto distribution in top lights

Our correction approach rests on the claim that a Pareto distribution is a reasonable description of top lights. The Pareto distribution is an often-found empirical feature of data used in physics, biology and many other sciences (see Newman, 2005), including the distribution of top income or wealth (e.g. Piketty, 2003, Atkinson et al., 2011) and the size distribution of cities as delineated via night lights (Small et al., 2011). Yet, it has so far not been established for the upper tail of the distribution of night light intensities.

We approach this issue in two ways. We first present a tractable model of light intensities within and across cities, showing that the density of top lights at the pixel level is Pareto, or can at least be closely approximated by a Pareto density. The model is not meant to be taken literally but only to illustrate that a Pareto density can be derived from first principles. Certain assumptions, such as city monocentricity, make the model analytically tractable but are unrealistic. This is also why we do not rely on the model in the correction procedure. In the second part of this section, we make the empirical case in favor of a power law in top lights based on the radiance-calibrated data and subject it to a battery of tests.

3.1 A stylized model of city lights

When modeling the distribution of the lights around the world, our starting point is the size distribution of cities in terms of population. It is well known that combining Gibrat's (1931) law of homogeneous growth of cities with a lower bound of city sizes leads to a Pareto or Zipf distribution of large cities (Gabaix, 1999, Eeckhout, 2004). Zipf's law predicts that city ranks are inversely proportional to their size; for instance, the biggest city in the U.S. (New York) has twice the population of the second-ranked city (Los Angeles) and three times the population of the third-ranked city (Chicago).

Numerous studies have provided empirical evidence for this regularity based on U.S. cities or metropolitan areas (Gabaix and Ioannides, 2004, Rozenfeld et al., 2011, Ioannides and Skouras, 2013). Despite heterogeneity in the rank-size parameter in other countries (Rosen and Resnick, 1980, Soo, 2005), it is generally considered a good approximation of the size distribution of large cities (Luckstead and Devadoss, 2014) and evidence in its favor becomes stronger when cities are defined as "natural" agglomerations as opposed to administrative boundaries (e.g. see Small et al., 2011, Jiang and Jia, 2011).

Assumption 1. *The size distribution of big cities in terms of their population x is Zipf, i.e. a Pareto distribution with shape parameter $\eta = 1$ above some threshold x_c.*

The CDF of the number of cities with population x is

$$F(x) = 1 - \left(\frac{x_c}{x}\right)^{\eta} = \int_{x_c}^{x} \eta \frac{x_c^{\eta}}{x^{\eta+1}} \mathrm{d}x = \int_{x_c}^{x} \frac{x_c}{x^2} \mathrm{d}x. \tag{1}$$

Cities have a certain area extent in which they house their population. The urban allometry literature (Stewart, 1947, Jones, 1975), which deals with the scaling of human-made structures within cities, links the two quantities in the following way:

Assumption 2. *The population x and the area s of a city are proportional, i.e. $x \sim s^{\phi}$.*

It is typically assumed that $1 < \phi < 2$, so that larger cities not only spread out on the plane by converting the surrounding agricultural land but also grow in terms of height (Batty and Longley, 1994). Bettencourt (2013), for example, motivates scaling laws based on a network theory of human interactions using the parameter value $\phi = 1.5$.

To derive the distribution of individual pixels y within cities, we yo model the shape of cities. The standard assumption in the workhorse model of urban economics is monocentricity (Mills, 1967, Amson, 1972, Desmet and Rossi-Hansberg, 2013).

Assumption 3. *Cities are monocentric and of circular shape. They consist of rings of unit width from the center to the outskirts.*

Depending on its area s and therefore its population x, each city has $r = \pi^{-1/2} x^{1/(2\phi)}$ rings.[10] Integration by substitution yields the CDF of the number of rings per city[11]

$$F(r) = 2\phi x_c \pi^{-\phi} \int_{\tilde{r}}^{r} r^{-2\phi-1} \mathrm{d}r = \begin{cases} 0 \text{ for } r < \tilde{r} \\ 1 - x_c \pi^{-\phi} r^{-2\phi} \text{ for } r >= \tilde{r} \end{cases} \tag{2}$$

with $\tilde{r} = \pi^{-1/2} x_c^{1/(2\phi)}$. Its density

$$f(r) = x_c \pi^{-\phi} r^{1-2\phi} \text{ for } r >= \tilde{r} \tag{3}$$

follows a power law with a shape parameter $2\phi - 1 > \eta = 1$ for $\phi > 1$. The higher shape parameter implies that the distribution of rings has fewer extreme values than the distribution of city sizes in terms of their population.

Each city with r rings consists of πr^2 rectangular pixels of unit size. The pixels are located at distance d from their respective city center. There are two opposing effects governing how the global number of pixels depends on this distance: *i)* within a given city, the number of pixels increases linearly with d because rings farther from the center contain more pixels, and *ii)* the larger the distance d from the city center, the fewer cities of that size are left, namely only the cities with $r \geq d$ rings.

Dividing the absolute amount of pixels at each distance d by the total number of

[10] For simplicity of exposition, we suppress the proportionality constant in $x \sim s^{\phi}$ and assume $x = s^{\phi}$.
[11] Online Appendix C derives this result.

pixels yields their global density

$$f(d) = \begin{cases} 2\pi \frac{\phi-1}{\phi} x_c^{-1/\phi} d \text{ for } d < \tilde{d} \\ 2\pi^{1-\phi} \frac{\phi-1}{\phi} x_c^{1-1/\phi} d^{1-2\phi} \text{ for } d \geq \tilde{d} \end{cases} \tag{4}$$

with $\tilde{d} = \pi^{-1/2} x_c^{1/(2\phi)}$.[12]

To derive the global density of luminosity $f(l)$, we require a last assumption on how light intensity within a city varies with the distance to the center d. As the within-city light distribution has not yet been modeled explicitly, we resort to standard models of population density. Light intensity is the product of population and income per capita. Differences in light intensity within countries are mostly driven by variations in population density (Henderson et al., 2018), which we take as a justification to focus on the population density gradient.

A popular choice in the literature is the negative exponential function, that is, $p(d) = P_0 \exp(-\gamma d)$ where $p(d)$ is the population density at distance d from the city center, $P_0 \geq p$ is the density at the center, and $\gamma > 0$ is a decay parameter so that the city periphery is more sparsely populated or, in our case, lit. The negative exponential can be motivated on the basis of the standard Alonso-Muth-Mills model (Brueckner, 1982). Empirical studies have found $\gamma \approx 0.15$ on average for large world cities (Bertaud and Malpezzi, 2014). An alternative to the exponential distribution is the inverse power function, which was originally proposed for gravity models of traffic flow (Smeed, 1961, Coleman, 1964, Batty and Longley, 1994). It is defined as $p(d) = P_0 d^{-a}$ with $d > 0$ and shape parameter $a > 0$.

Both functions are qualitatively similar for intermediate distances, with the inverse power function slightly below the negative exponential. They differ from each other in the city center, where the inverse power function has higher and more sharply declining values, as well as in the outskirts, where it predicts a higher density. A greater concentration at the center makes the inverse power function particularly suitable for business floor-space models (as recommended by Zielinski, 1980). At the same time, its tail is known to fit the urban fringe well (Parr, 1985). Since we are interested in lights rather than population density, we would like to capture both the very bright central business districts in most cities and the more dimly lit but significant suburban sprawl in the outskirts, which typically features prominently in the footprint of city lights (e.g. see Small et al., 2011).

Modeling light densities within cities with an inverse power function better reflects the light gradient and turns out to be analytically appealing. Hence, it serves as our baseline case. Note that using the negative exponential function also leads to an expression for the distribution of top lights which (under certain conditions) can be approximated by a

[12] Online Appendix C derives this result.

Pareto density.[13]

Assumption 4. *Within cities, the light density $l(d)$ follows an inverse power function $l(d) = L_0 d^{-a}$ for $d > 0$ with $L_0 > l$ as the maximum luminosity at the center and $a > 0$ as the decay parameter.*

Applying the variable transformation from the inverse power function to the pixel density in eq. (4) yields

$$f(l) = \begin{cases} 2\pi^{1-\phi}\frac{\phi-1}{\phi}x_c^{1-1/\phi}\Big(L_0/l\Big)^{(1-2\phi)/a} & \text{for } 0 < l \leq \tilde{l} \\ \underbrace{2\pi\frac{\phi-1}{\phi}x_c^{-1/\phi}}_{c}\Big(L_0/l\Big)^{1/a} & \text{for } \tilde{l} < l \leq L_0, \end{cases} \tag{5}$$

where $\tilde{l} = L_0\pi^{a/2}x_c^{-a/(2\phi)}$.

Restricting our attention to the upper part of the light distribution from $\tilde{l}$ onwards, we can establish our key result:

Result 1. *Based on assumptions 1–4, top lights above threshold $\tilde{l}$ follow the Pareto distribution $f(l) = c\Big(L_0/l\Big)^{1/a}$ with shape parameter $1/a$.*

Using a grid-search, we find that for $a = 0.7$ the inverse power function comes closest to the negative exponential function of the empirical parameter $\gamma = 0.15$.[14] A back-of-the-envelope calculation thus implies a Pareto shape parameter around one and a half. Note that we estimate this parameter using auxiliary data in the next subsection.

In sum, a limited set of four assumptions allows us to analytically derive a Pareto distribution in top lights, with shape parameter $1/a$, maximum luminosity L_0, and a multiplicative constant. We believe that this is an important insight, even though the model is very stylized. Note that the correction we propose later on only relies on the Pareto property and does *not* require any of the simplifying assumptions, such as monocentricity or a particular population density gradient. In fact, part of our application below investigates how sub-centers have been forming in African cities.

There are alternative ways to characterizing the distribution of top lights. We could, for instance, consider the truncation purely as a statistical issue of tail probabilities and use extreme value theory to derive their distribution. Reassuringly, this approach also points towards a Paretian distribution of top lights (see Online Appendix D).

[13] Online Appendix C derives this result.

[14] The curve-fitting was conducted on the domain for d from 1 to 50, minimizing the squared error between the negative exponential function with $\gamma = 0.15$ and the inverse power function with $a \in [0.05, 2]$.

3.2 Empirical tests

Empirical tests for a Pareto tail were popularized in economics by the literature on top incomes (e.g. Piketty, 2003, Atkinson et al., 2011) and city sizes (e.g. Rosen and Resnick, 1980, Gabaix, 1999, Gabaix and Ioannides, 2004, Rozenfeld et al., 2011). These tests usually exploit particular properties of the Pareto distribution.

Recall that data y which is Pareto distributed above a certain threshold y_c has the probability density function, $f(y) = \alpha y_c^{\alpha} y^{-\alpha-1}$, where α is the relevant shape parameter and only takes on positive values. The survival function, $1 - F(y) = (y_c/y)^{\alpha}$, gives the probability that the random variable Y is larger than the given value y. This maps directly to our model from the previous section, only that we now denote the light intensity at the pixel level by y and the shape parameter by $\alpha = 1/a$ to simplify the exposition.

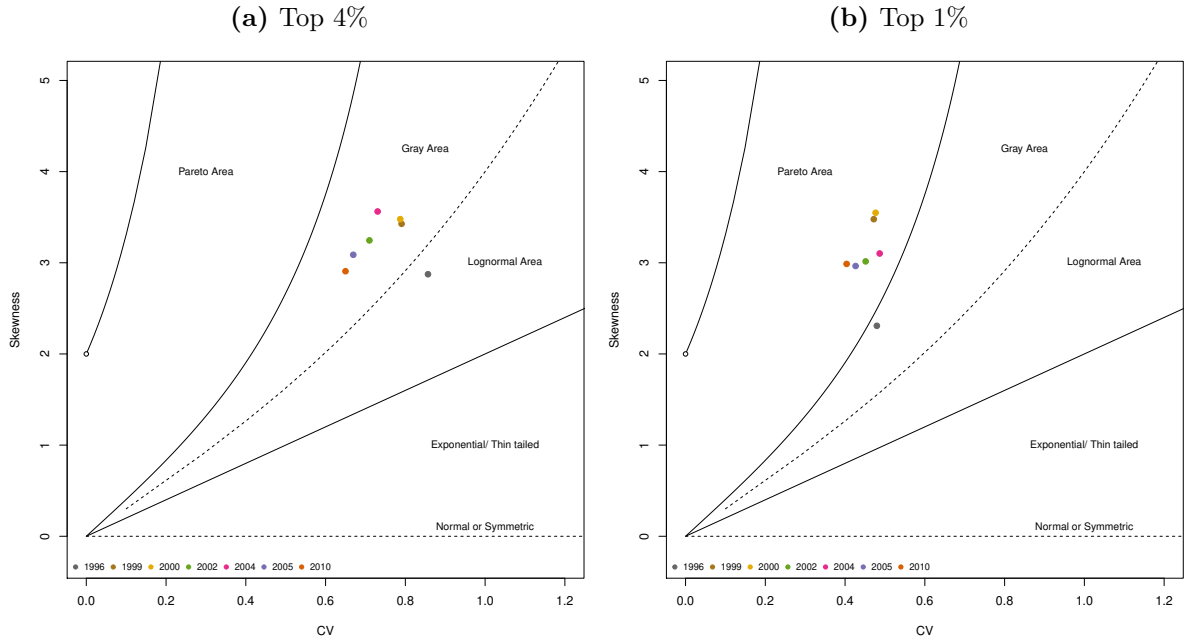
Visual inspection: Following Cirillo (2013), we first visually check whether our data are Pareto distributed, before estimating shape parameters that are only meaningful if this condition is fulfilled. We use the seven radiance-calibrated satellites to analyze the shape of the missing tail in the stable lights data. Figure 3 shows a discriminant moment ratio plot with the coordinate pair of the coefficient of variation (i.e., standard deviation divided by the mean) on the x-axis and skewness on the y-axis (Cirillo, 2013). Each parametric distribution has its particular curve of feasible coordinates, so that the relevant part of the plane can be divided into a Pareto area, a lognormal area, and a gray area possibly belonging to both. This type of plot provides a more reliable indication of the Paretian nature of the data than more traditional graphical devices (such as Zipf plots shown in Online Appendix E).

The visual evidence in favor of a Pareto distribution is strongest for higher thresholds. For the top 4% of pixels in panel (a), all but one satellite are located in the area of indeterminacy.[15] This ambiguity mirrors similar findings in the city-size literature, where both the Pareto and the log-normal with a large standard deviation generate tails which are virtually indistinguishable (Eeckhout, 2004, 2009).[16] For smaller percentages, such as the top 1% in panel (b), the evidence in favor of a Pareto distribution becomes much stronger. All but one satellite are in the Pareto area, far from the lognormal area and

[15] The only exception is satellite F16 in 1996 which is based on considerably fewer cloud-free overpasses than later years. It is the earliest and dimmest of the radiance-calibrated products. Its highest values are far below those in subsequent years and the data contain many ties. Online Appendix A contains the relevant summary statistics for the radiance-calibrated data.

[16] It is well-established in the literature on city sizes and top incomes that the Pareto distribution is often only a good representation for the very top of data, with a moderately decreasing fit as the threshold decreases. Our data suggest that a Pareto distribution of top lights remains a good approximation down to the top 10% of the data. This is more than sufficient for our purposes, since we only consider 3% to 5% of all pixels to be top-coded in any given year. Indeterminacy vis-à-vis the lognormal occurs is common in this literature as well, since "the tail of a lognormal is indistinguishable from the Pareto under certain circumstances, [so that] the researcher who is interested in the tail properties of a size distribution can choose which one to use" (Eeckhout, 2009).

Figure 3 – Discriminant moment ratio plots



Notes: The panels show discriminant moment ratio plots (Cirillo, 2013). The input data are a 10% representative sample of all non-zero lights in the radiance-calibrated data at the pixel level, where each pixel is 30×30 arc seconds.

those of other thin-tailed distributions.

We take a closer look at the issue of lognormal versus Pareto in Online Appendix E. We conduct several goodness-of-fit tests, where we compare the empirical distribution of the radiance-calibrated data to the best-fitting theoretical distribution. The findings are unambiguous. The Pareto distribution mimics the distribution of top lights much better than the lognormal (with R^2s close to unity). Hence, we conclude that our data are best characterized by a heavy-tailed Pareto distribution.

Log-rank regressions: Log-rank regressions are a popular approach to estimating the Pareto shape parameter and firmly rooted in urban economics. They are based on the following approximation. For Pareto-distributed observations y_i, $i = 1, ...N$, with the survival function given above, we have $\text{rank}(y_i) \approx N y_c^{\alpha} y_i^{-\alpha}$, or, in logarithms $\log \text{rank}(y_i) - \log N \approx \alpha \log y_c - \alpha \log y_i$. Gabaix and Ibragimov (2011) show that directly estimating this relationship systematically underestimates the true coefficient and standard error. However, once the ranks are shifted by minus one-half and the standard errors are adjusted, rank regressions consistently estimate the parameters of interest and turn out to be relatively robust to deviations from power laws.

Table 2 reports the corresponding results. We separately estimate the Pareto shape parameter for each of the radiance-calibrated satellites and for the top 5% to 1% of the data. The point estimates are relatively stable for the top 3% to 5% of lights—

Table 2 – OLS rank regressions

Year	1996	1999	2000	2003	2004	2006	2010	Average
				Panel a) Top 5%				
Pareto $\hat{\alpha}$	1.3122	1.2993	1.2471	1.4482	1.4484	1.4790	1.4453	1.3828
	(0.0060)	(0.0054)	(0.0054)	(0.0065)	(0.0063)	(0.0066)	(0.0062)	[0.0932]
Observations	96,685	116,858	106,914	100,095	106,899	99,487	107,745	–
				Panel b) Top 4%				
Pareto $\hat{\alpha}$	1.3514	1.4000	1.3594	1.5890	1.5837	1.6250	1.5963	1.5007
	(0.0069)	(0.0065)	(0.0066)	(0.0079)	(0.0077)	(0.0081)	(0.0077)	[0.1236]
Observations	77,348	93,484	85,482	80,075	85,489	79,590	86,196	–
				Panel c) Top 3%				
Pareto $\hat{\alpha}$	1.4270	1.5665	1.5444	1.7667	1.7876	1.8423	1.8330	1.6811
	(0.0084)	(0.0084)	(0.0086)	(0.0102)	(0.0100)	(0.0107)	(0.0102)	[0.1654]
Observations	58,011	70,115	64,111	60,058	64,134	59,692	64,647	–
				Panel d) Top 2%				
Pareto $\hat{\alpha}$	1.6714	1.8614	1.9314	2.0737	2.0976	2.1751	2.2160	2.0038
	(0.0120)	(0.0122)	(0.0132)	(0.0147)	(0.0143)	(0.0154)	(0.0151)	[0.1932]
Observations	38,673	46,742	42,740	40,039	42,756	39,794	43,097	–
				Panel e) Top 1%				
Pareto $\hat{\alpha}$	2.2300	2.4350	2.4470	2.5914	2.5046	2.7075	2.8641	2.5399
	(0.0227)	(0.0225)	(0.0237)	(0.0259)	(0.0242)	(0.0271)	(0.0276)	[0.2052]
Observations	19,337	23,373	21,372	20,020	21,377	19,898	21,551	–

Notes: The table reports the results of OLS rank regressions with $\log(\text{rank}(y_i) - 1/2) - \log N$ as the dependent variable. Asymptotic standard errors computed as $(2/N)^{1/2}\hat{\alpha}$ are reported in parentheses (see Gabaix and Ibragimov, 2011). The data are a 10% representative sample of all non-zero lights in the radiance-calibrated data at the pixel level, where each pixel is 30×30 arc seconds. The last column reports the point average of the seven satellites and its standard deviation in brackets.

approximately the range of shares we will later replace each year—and then rise as smaller percentages are considered.[17] The last column reports the simple average of the estimated coefficients. Somewhat remarkably, the average Pareto shape parameter is about one and a half for the top 4% of lights—not far from our model-based guesstimate. Moreover, averaging all 21 coefficients in panels (a) to (c) also yields a central estimate of about one and a half. Note that there is only limited variation over time at the lower thresholds, apart perhaps from a small discontinuity in the early 2000s. Our range of parameter estimates implies inequality comparable to the top tail of the U.S. income or wealth distribution (Piketty, 2003, Atkinson et al., 2011). Top lights are considerably more equally distributed than city sizes, just as predicted by the model in the previous section.

We provide an array of additional robustness checks in Online Appendix E and F. Our main findings are robust to *i)* using the Hill estimator instead of OLS rank regressions,

[17] In theory, with Pareto-distributed data, conducting the estimation with the portion of the distribution above a higher threshold $y_c^h > y_c$ should lead to the same estimated α. In practice, this will often not hold exactly and the results for tail regressions are known to depend on the precise threshold used (see for instance Rosen and Resnick, 1980).

ii) estimating unrestricted rank regressions, which are another way of testing for a Pareto distribution, and *iii)* using top-coding free data from the new VIIRS satellites, whose first available year is 2015. The VIIRS satellites are technologically superior and recorded at a much finer resolution. Reassuringly, they are either as indicative of a Pareto tail in top lights as the radiance-calibrated data, or provide even stronger evidence in its favor, such as linear Zipf plots.[18]

4 Correcting for top-coding

We propose a simple correction procedure inspired by the methods used in the top incomes literature (Piketty, 2003, Atkinson et al., 2011). All lights below the top-coding threshold are unaltered, while those above are replaced by a Pareto-distributed counterpart. An appealing feature of this approach is that it keeps the overwhelming majority of the data intact and replaces only a small but highly influential fraction of pixels.[19]

Our theoretical arguments and empirical tests suggest a Pareto parameter around *one and a half*. Recall that a plausible parameterization of our model also implies $\alpha = 1/0.7 \approx 1.43$ and our empirical estimates inferred from the radiance-calibrated data are centered on one and a half. We use this fixed value as a rule-of-thumb parameter. Of course, our procedure is not predicated on a particular value, nor does it require the parameter to be constant over time. However, we do not detect an unambiguous trend indicating that the distribution of top lights has become more equal over time. As will become clear shortly, assuming a constant parameter value still allows for significant variation over time as particular pixels *i)* cross the top-coding threshold and *ii)* achieve a higher rank vis-à-vis other pixels. Cities can thus become brighter in absolute terms and grow relative to other cities after the correction.[20]

[18]The estimated shape parameters are a bit higher for top shares around 3% to 5% but then also appear to be more stable in the upper tail. Since the VIIRS data are five years after the most recent radiance-calibrated image and have a different overpass time, it is difficult to identify the source of these slight discrepancies.

[19]There have been a number of attempts to correct the stable lights data in the remote sensing literature. One type of corrections is based on cubic pixel-level regressions (see e.g. Letu et al., 2012), which is a heavily parametric approach and usually assumes that light emissions were constant over some area or period. An alternative approach uses land cover and vegetation data to non-linearly transformed night lights. Zhang et al. (2013), for example, define an index which combines night lights with a Normalized Difference Vegetation Index (NDVI). While this recovers some of the city-wide variation in light, the conflation of lights with NDVI data is not ideal for a number of reasons: *i)* the original scale of the stable lights data is lost, *ii)* all values—not just the top-coded data points—are being adjusted, and *iii)* researchers may wish to use the NVDI and other land cover data as explanatory or dependent variables and could reasonably worry about mechanical correlations introduced by this approach. All of these studies agree that the radiance-calibration method of Hsu et al. (2015) is ideal in terms of its spectral response and use it as a benchmark. None of these studies directly combine the two data sources.

[20]In additional analyses, we assigned the seven satellite-specific estimates to the adjacent stable light satellite-years. This approach gives very similar results overall, albeit with more jumps. Country-specific estimates also lead to similar results, although comparatively darker and poorer countries experience larger corrections. By using the same parameter for the whole world, we deliberately follow an agnostic

Our preferred approach is a pixel-level replacement in which we directly substitute the top-coded lights by their corresponding quantile from the theoretical Pareto distribution. Correcting the raw data has the advantage that they can then be flexibly aggregated to the unit of interest. Another option is to analytically correct the relevant summary statistics, such as mean lights and spatial Gini coefficients. We provide closed-form solutions that are particularly useful for back-of-the envelope calculations in Online Appendix G. Exclusively correcting the summary statistics, however, limits the potential applications and is not sufficient when the actual locations of the top-coded pixels matter.

A remaining challenge is to geo-reference the Pareto quantiles so that the brightest pixels actually end up in the centers of dense urban agglomerations. It turns out that there is a straightforward solution. Since we know the exact location of all pixels, we can rank them according to their radiance-calibrated values from the nearest year and distribute the highest values from the Pareto distribution in the same manner. Working with ranks avoids importing the artificial variability of the radiance-calibrated satellites but preserves a crucial part of the data structure.[21] Our algorithm for replacing top-coded pixels with their quantile counterpart from Pareto distribution works as follows:

(1) For each of the 34 satellite-years t of stable lights data, calculate the number N_t of pixels ≥ 55 DN to be replaced.

(2) Produce a ranking of these N_t pixels based on the radiance-calibrated data associated with the same satellite-year or the data from the closest year.[22]

(3) Generate N_t theoretical values from a truncated Pareto distribution with the rule-of-thumb $\alpha = 1.5$, the top-coding threshold $y_c = 55$, and upper bound $H = 2000$.

(4) Replace the N_t stable lights pixels ≥ 55 so that the stable lights pixel with the i-th highest rank from (2) is replaced by the i-th highest theoretical value.

In other words, our procedure combines the desirable features of the stable lights (annual availability and measurement for all non-top-coded values) with those of the radiance calibrated lights (distributional shape of the top as well as stability of the ranks). We apply this procedure to every stable lights satellite image over the entire period from 1992 to 2013. This is the data which we use in the subsequent application and which

approach. We do not want to subject some countries to specifically stronger corrections than others without an a priori reason, but technically, our procedure would allow for that. Additional results are available on request.

[21]The ranks of the pixels are much more stable over time than the values of the radiance-calibrated data. The rank correlation of maximum city lights typically ranges from 0.90–0.95 for adjacent radiance-calibrated satellites (see Online Appendix A).

[22]Whenever the radiance-calibrated data creates ties, we first try to break those ties using the ranks of the stable lights data and then use the ranks of neighboring radiance-calibrated satellites. This produces a near unique ranking each time. We also experimented with other rankings. The results are very similar and available on request.

underlies Figure 1. Note that the support of the Pareto distribution is unbounded, so that its highest quantiles yield a handful of values far exceeding the natural limit of man-made light intensity. To generate a realistic depiction of city lights we impose an upper bound of 2000 DN, which approximately corresponds to the average maximum observed in the world's brightest cities (see Online Appendix A). Note that this barely affects the overall results but only a few extreme values.[23]

The correction has a significant impact on how we understand the global distribution of economic activity. On average, 3.7% pixels above the top-coding threshold account for 17.7% of the total sum of lights observed in the stable lights data. This share *almost doubles* to 31.94% after the correction—a statement which is approximately true in every individual satellite-year, although later shares are bigger in absolute terms before and after the correction. In other words, about four percent of all lit pixels account for about a third of all visible economic activity. The worldwide spatial Gini coefficient rises by nine percentage points, on average, after the correction. Note that the annual variation in the global Gini coefficient is only a few percentage points, so that it is swamped by the size of the top-coding correction. These findings are not particularly sensitive to the choice of the Pareto shape parameter, although the size of the top-coding correction varies somewhat. Online Appendix H provides a variety of summary statistics and sensitivity checks using different Pareto shape parameters. The size of the correction in both the country averages of light intensities and the country-wide spatial Gini coefficients varies systematically with GDP per capita, country size and population density, in line with the occurrence of top-coding illustrated in Figure 2. Moreover, we provide comparison checks between the corrected data and the radiance-calibrated data for those 7 years when both are available, finding correlations of 0.94-0.96.

We conduct two benchmarking exercises to assess the properties of the corrected data and better understand at which scale of aggregation top-coding influences the conclusions we are likely to draw when using night lights. We only briefly summarize the results here and relegate a full discussion to Online Appendix I. The first exercise estimates national light-output elasticities as in Henderson et al. (2012). Even at this high level of aggregation, the top-coding corrected series performs marginally better. For the most part, however, the estimates using the corrected data are not just economically but also numerically very close to the original results. Our second benchmark goes from the national to the regional level using German as a case study. This exercise allows us to confirm the performance of our correction procedure in a developed economy with high-

[23] Online Appendix A reports and discusses the observed city-wide maxima obtained by the radiance-calibrated satellites which motivate this upper bound in man-made luminosity. The truncated Pareto has the CDF $F(y) = \left(1 - \left(\frac{y_c}{y}\right)^{\alpha}\right) / \left(1 - \left(\frac{y_c}{H}\right)^{\alpha}\right)$ for $y_c \leq y \leq H$, where an upper bound of $H = 2000$ does not affect the overwhelming majority of stable lights pixels ≥ 55 to be replaced, but does ensure realistic values at top, say, 0.01% of the data. Our results in the subsequent sections are very similar, no matter if we use the simple Pareto or the truncated Pareto.

quality regional accounts. Here we obtain two notable results. First, the light-output elasticity rises considerably after the top-coding correction. Only the corrected data allow us to recover estimates of the light-output elasticity comparable to the national-level in a cross-section and panel of German regions. Second, light intensities in the corrected series are approximately linear in population density and diverge from the stable lights data after about one thousand people per square kilometer, that is, in denser urban areas. The corrected data provides a realistic ranking of larger German cities, while the standard data paints a distorted picture. Both findings suggest that the corrected data better capture the light-output gradient in developed economies with mature urban structures.

5 Application: Cities in Sub-Saharan Africa

Armed with this new data, we now return to the question of whether primate cities are outgrowing secondary cities in Sub-Saharan Africa and how city structures are adjusting to continuously growing populations. In most African countries, economic activity is concentrated in the biggest city. Dar es Salaam, Kinshasa and Lagos already are mega cities with more than 10 million inhabitants, or will attain that status in a few years (United Nations, 2018). Strong spatial concentration is often a feature of countries with poor infrastructure and a low level of development (Krugman, 1991, Puga, 1998, Jedwab and Vollrath, 2019), although Africa's urbanization differs from the historical experience of industrialized economies for a variety of reasons.

The new millennium marked a turning point for most African economies. Sustained consumption growth and pro-poor distributional change brought about falling poverty rates (Bluhm et al., 2018). However, as countries are developing, it is theoretically unclear whether secondary cities will catch up (Duranton, 2008). There is some empirical evidence suggesting that this is the case. Henderson et al. (2012), for example, estimate that the African hinterland was growing about 2.3% faster than primate cities over the period from 1992 to 2008.[24] Moreover, many in the World Bank view secondary city development as key to sustained poverty reduction (Christiaensen and Todo, 2014, Christiaensen and Kanbur, 2017). Yet, manufacturing is heavily concentrated in primate and coastal cities with greater access to world markets. The oil price boom of the 2000s, for example, hurt remote secondary cities more than primate cities (Storeygard, 2016). Secondary cities in Africa have also been characterized as "consumption cities", catering to the agricultural hinterland rather than the modern sector (Gollin et al., 2016).

Whether or not primate cities are driving productivity growth depends—at least in part—on their internal structure. Informal settlements in large cities can more easily

[24]Note that Henderson et al. (2012) also suggest that a "detailed study would be required to explain the result" (p. 1024). In India, Gibson et al. (2017) find that secondary cities matter more for rural poverty reduction when studying the spillover of urban growth at the intensive and extensive margin.

absorb growing populations (Jedwab and Vollrath, 2019), but if these neighborhoods remain badly connected to the center, then cities will be crowded, fragmented and less productive (Lall et al., 2017). An adverse urban form implies that companies and inhabitants are faced with high transport costs and long commuting times, which limits interactions and positive spillovers across the city (Rosenthal and Strange, 2004, Harari, 2020). Ultimately, this may prevent the industries located in primate cities from reaping increasing returns and diversifying into tradables, creating more urbanization without industrialization (Venables, 2017, Gollin et al., 2016).

Our application tackles both of these aspects. The first part uses the top-coding corrected data to show that primate cities have grown at least as quickly as secondary cities and towns over the period from 1992 to 2013. The second part provides a new perspective on within-city activity which is made possible by recovering the full distribution of light intensities within cities. We show that the concentration of light decreases over time as poorer neighborhoods develop. At the same time, we find evidence for increasing fragmentation, suggesting that newly formed sub-centers are not well-connected with the city as a whole.

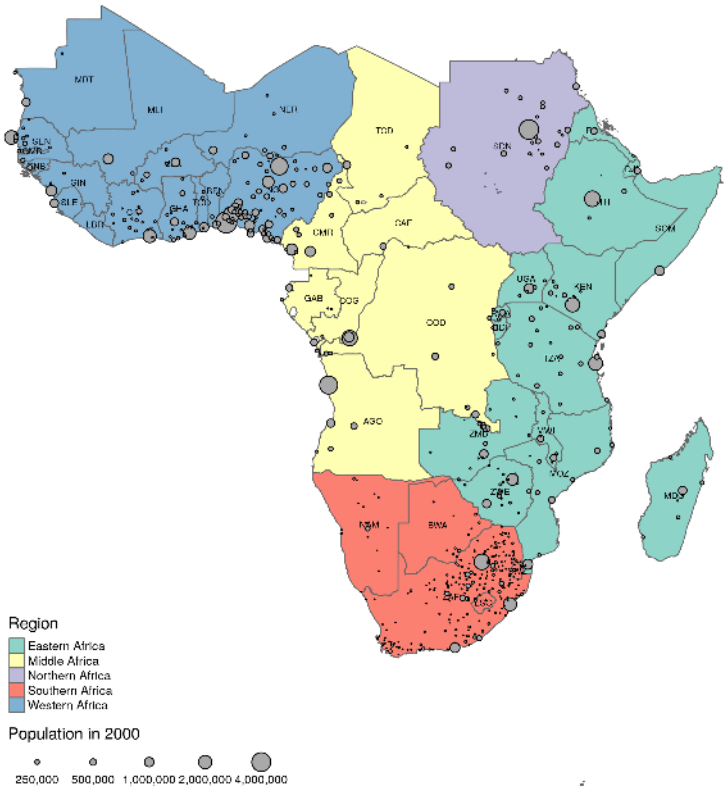
City boundaries: Urban areas are often delineated using night light by defining them as contiguously lit clusters above some fixed luminosity threshold (Small et al., 2011, Storeygard, 2016, Harari, 2020). Satellite-derived footprints are particularly useful in Africa, where administrative boundaries quickly become outdated. Relying on nighttime lights is appealing, since they offer a time-varying measure of urban expansion, but the thresholding approach suffers from a well-known problem: no single threshold works well for all cities. Thresholding overestimates the urban extent of larger cities and penalizes other cities at the same time (Small et al., 2011, Abrahams et al., 2018). We address these issues using a method developed by Abrahams et al. (2018) to resolve exactly this issue. Their de-blurring algorithm reduces the non-linear "overglow" in the lights data and considerably improves the accuracy of the identified urban extents.[25]

To capture changes at the extensive margin and minimize measurement errors, we define cities as contiguously illuminated pixels in the de-blurred lights, provided that a light source is detected in at least two satellites over a period of three years.[26] The period from 1992 to 1994 marks the initial boundaries and the period from 2011 to

[25]The de-blurring approach is based on two insights into the data generating process: *i)* the original light sources are blurred by a symmetric Gaussian point-spread function, and *ii)*, pixels in which light sources are located must be local maxima in the so-called percent frequency of light detection image (for more detail see Abrahams et al., 2018). Note that we have used light intensity thresholds for the detection of city boundaries in previous version of this paper which yields qualitatively similar results.

[26]Two images of the same place in the same year may indicate different urban footprints due to a lack of on-board calibration. Considering a window of three years and requiring multiple detection points effectively cancels out most of this artificial variation.

Figure 4 – Primate and secondary cities in Sub-Saharan Africa



Notes: Illustration of the location of our panel of 41 primate and 527 secondary cities in SSA using the urban footprint detection algorithm outlined in the text. Note that we dropped Equatorial Guinea due to gas flaring on the capital island of Malabo and consider South Sudan as part of Sudan for the entire sample.

2013 represents the latest available boundaries.[27] We then identify city locations by overlaying these urban areas with all settlement points from the Global Rural-Urban Mapping Project (GRUMP) within three kilometers of the urban perimeter.[28] An urban area receives the name and attributes of the most populous settlement located within the expanded urban perimeter.[29]

We create two data sets based on these urban areas. The first relies only on the initial boundaries and allows us to track how economic activity of a city develops over time at the intensive margin. The second takes the union of the initial and final period boundaries to identify the "envelope" or maximum urban extent of the cities in our sample. This data allows us to study the structure of cities as they grow and separate

[27] Online Appendix J shows both the initial and latest boundaries together with daytime images of Lagos, Luanda, and Johannesburg taken at the end of both periods.

[28] We manually extend this data to include all coordinates of cities which at some point over the period from 1992 to 2013 were designated the administrative capital of a province or state. We first identify the administrative capitals of subnational regions using `www.statoids.com` and then geocode each city using multiple online gazetteers.

[29] When population estimates are unavailable, we use the name and attributes of the settlement point closest to the polygon centroid.

the intensive margin from extensive growth in the fringe (defined here as the envelope minus the city boundaries from the initial period). Since several smaller cities merge over time or become absorbed by a larger city in their vicinity, we aggregate all parts of a city that will eventually become a single agglomeration, no matter if we are analyzing their initial footprint or the envelope. Finally, we identify the *primate city* in each country as the city with largest population and define all other cities and towns as secondary. Not all small settlements qualify as secondary cities but there is no consensus on their definition (minimum sizes typically range from 100,000 to 500,000 people). We also lack accurate population data for the smaller settlements. Hence, for our main results, we work with an area threshold of 9 km^2—about 3×3 pixels—and show that our results are robust to variations of this threshold in Online Appendix J.

Figure 4 shows a map of our universe of 568 cities and their population in 2000 (if available). 41 cities, one per country, are primate cities, while 527 are secondary cities.[30] The largest urban agglomeration in the sample is Johannesburg, followed by Lagos and Cape Town.

City growth: Cities grow at the intensive margin, as existing city quarters develop and become brighter, and at the extensive margin, as they expand and absorb surrounding areas. Table 3 focuses on the intensive margin, that is, growth within the initial urban boundaries. It shows summary statistics for the commonly used stable lights data and our top-coding-free counterpart in 1992 and 2013. Panel (a) shows that the initial sum of light originating from primary cities is, on average, more than ten times as large as that of secondary cities. Panel (b) neutralizes the size difference and examines average light densities. The initial light density in primate cities is about one and a half to twice as large as in secondary cities. These stylized facts hold regardless of whether we use the stable lights or the corrected data.

Substantial differences appear once we focus on city growth or compare the two groups towards the end of the sample. The raw data shows that *i)* top-coding affects primary cities much more than secondary cities, and *ii)* the size of the correction becomes substantially larger over time. The correction adds 4.7% to the average light density of primate cities in 1992, but increases it by 23.7% in 2013. The summary statistics also show why this is happening. By 2013, the average light intensity of primate cities grew to almost 62 DN according to the stable lights data, suggesting that nearly all pixels in these cities are top-coded (regardless of where exactly we draw the line) and their light intensity cannot continue to grow.

The correction has important implications for our understanding of agglomeration

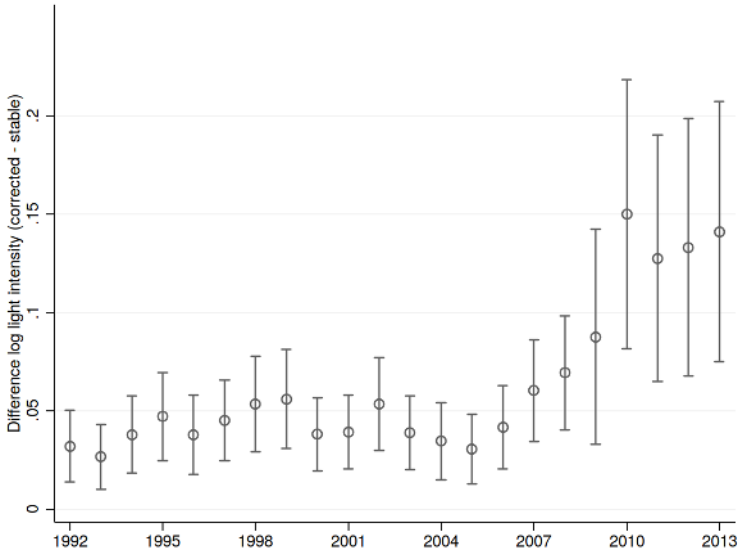
[30]This lines up well with census-based counts of medium-sized secondary and primate cities in the early 1990s. The 2018 update of the Africapolis database, for example, includes 551 cities which have over 50,000 inhabitants in 1990.

Table 3 – Summary statistics for African cities, 1992–2013, initial boundaries

	Stable Lights		*Corrected Lights*	
	Primate	Secondary	Primate	Secondary
	Panel a) Sum of lights			
Sum in 1992	20321.93	1761.87	22777.48	1865.97
Sum in 2013	27485.29	2446.96	37126.00	2768.77
Annualized growth rate	0.0143	0.0156	0.0233	0.0188
	Panel b) Lights per km^2			
Average in 1992	39.23	22.89	40.88	23.03
Average in 2013	61.61	38.42	76.21	39.31
Annualized growth rate	0.0215	0.0247	0.0297	0.0255

Notes: The table reports a selection of summary statistics for African cities based on their initial boundaries. Annualized growth rates are computed as $\frac{1}{21}(\ln x_{2013} - \ln x_{1992})$, where x refers to the data per group reported in the table.

Figure 5 – Difference between primate and secondary cities in the two data sources



Notes: Illustration of how the component of the cross-sectional difference between primate and secondary cities which can be fully attributed to the top-coding correction evolves over time. The figure plots the coefficients, β_p, and the corresponding 95% confidence intervals obtained from the following regression: $\ln \Delta_{ijt} = \sum_{p=1}^{T} \beta_p (s_p \times P_{ij}) + s_t + \epsilon_{ijt}$ where $\ln \Delta_{ijt}$ is the log difference between the corrected data and the stable lights data, P_{ij} is an indicator for primate cities and s_p or s_t are year fixed effects. Standard errors are clustered at the city level.

economies and city growth in Sub-Saharan Africa. The stable lights data show that secondary cities are catching up with primate cities—an artificial byproduct of more and more pixels in primate cities becoming top-coded. This picture is reversed by the top-coding correction. Primary cities outgrew secondary cities by about half a percentage point, while the growth rate of secondary cities is hardly affected. This is reflected in

the growth rates of individual cities as well. Top-coding plays virtually no role in small primate cities, such as Bissau, while the annualized growth rates of larger cities, like Kinshasa, Johannesburg, Luanda, Dakar or Khartum, increase by about a percentage point or more.[31] Figure 5 adds another piece of evidence along these lines. It shows that the component of the difference between primate and secondary cities which can be fully attributed to top-coding (and is net of annual measurement error) starts to rise strongly after 2005. As a result, both cross-sectional differences and relative growth rates are increasingly distorted in the later years of the stable lights data.

We underpin these descriptive results with panel growth regressions focusing on growth at the intensive margin. In each case, we regress the log of lights per km^2 recorded for city j in country i at time t, or $\ln \text{LIGHTS}_{ijt}$, on a linear time trend, an interaction of the linear time trend with an indicator for primate cities, P_{ij}, and a set of fixed effects that varies across specifications. We typically include city fixed effects and then progressively add satellite, year or country-year dummies to purge systematic measurement errors across satellites.[32]

Table 4 – Growth regressions for African cities, intensive margin

	Dependent variable: Log lights in the initial footprint					
	Stable lights data			*Corrected data*		
	(1)	(2)	(3)	(4)	(5)	(6)
Linear trend	0.842			0.831		
	(0.126)***			(0.126)***		
	[0.543]			[0.555]		
Primate × Linear trend	0.275	0.275	0.249	0.738	0.737	0.821
	(0.185)	(0.185)	(0.340)	(0.255)***	(0.255)***	(0.313)***
	[0.198]	[0.195]	[0.343]	[0.260]***	[0.258]***	[0.314]***
City FE	✓	✓	✓	✓	✓	✓
Satellite FE	✓	–	–	✓	–	–
Year FE	–	✓	–	–	✓	–
Country-Year FE	–	–	✓	–	–	✓
Observations	12487	12487	12487	12487	12487	12487
Cities	568	568	568	568	568	568

Notes: The table reports the results of city-level panel regressions using the stable lights and top-coding corrected data. All coefficients are scaled by 100 for readability. The specifications are variants of $\ln \text{LIGHTS}_{ijt} = \beta_1 t + \beta_2 (t \times P_{ij}) + c_{ij} + s_{jt} + \epsilon_{ijt}$ where t is a linear time trend, P_{ij} is an indicator for primate cities, c_{ij} is a city fixed effect and s_{jt} contains a varying set of fixed effects (satellite, year, or country-year). Standard errors clustered at the city level are reported in parentheses. Conley errors with a spatial cutoff of 1,000 km and a time-series HAC with a lag cutoff of 1,000 years are reported in brackets. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[31] Online Appendix J presents data for each country, listing the primate city, as well as the growth rates of mean lights in both the primary and all secondary cities.

[32] Our vector of satellite dummies accounts for all the possible combinations of satellites, e.g. the fact that the value for 1993 are averages of the satellites with flight numbers F10 and F12.

Table 4 confirms that primate cities outgrew secondary cities by a significant margin. Columns (1) and (4) show that the average growth rate of secondary cities was slightly more than 0.8% per annum. The trend is statistically significant when errors are clustered at the city level but is not robust to allowing for spatial dependence across cities. Annual fluctuations in the light sensitivity of the satellites imply that this baseline growth rate might not be identified, which is why we include year or country-year fixed effects in the subsequent columns and focus on the difference in growth rates. The differential between primate and secondary cities can still be estimated in those specifications and is net of this specific measurement error. We would draw very different conclusions about their relative growth rates in either of these two data sets. The stable lights data in columns (1) to (3) suggest no significant differences in growth across city types. On the contrary, the corrected data in columns (3) to (5) indicate that primate cities grew 0.7-0.8 percentage points faster than secondary cities. The estimated coefficient on the interaction term directly corresponds to a difference in means test. Hence, we can reject the null hypothesis that both trends are the same at conventional significance levels. The growth rates of secondary cities are virtually unaffected by the correction. If we take these estimates at face value, then primate cities grew nearly *twice as fast* as secondary cities—a difference of almost 19% over the entire period of 21 years.

Online Appendix J presents several robustness checks, which we only briefly summarize here. The differential remains intact when we *i)* vary the area threshold for the minimum secondary city size, *ii)* omit any one region of Sub-Saharan Africa, or *iii)* use the radiance-calibrated data in the seven available years. In fact, the radiance-calibrated data deliver estimates of the interaction term that are close to those presented here. However, the average light intensity recorded by the radiance-calibrated data exhibits such large fluctuations that they produce overall trends which are negative—a result which cannot be supported by any other data or an appeal to fundamentals. Although the difference in growth rates between secondary and primate cities is less affected by these common trends, we take this as further indication that these data are not a workable alternative to our data in a panel setting.

So far we have focused exclusively on the intensive margin. When the city footprint is kept fixed, we can interpret growth in lights directly as increases in population density and economic activity per square kilometer. Population density, in turn, is strongly correlated with living standards and public good provision in developing countries (Gollin et al., 2017). Estimates from developed countries suggest that a doubling of population density raises productivity by about 5% (Rosenthal and Strange, 2004). Dense city centers are also the places where top-coding is most pronounced. New developments at the fringe of a city are usually dimmer. However, there too, primate cities have been growing substantially faster than secondary cities (according to both data sources).[33] Taken

[33] Online Appendix J contains the corresponding regression results.

together, this set of results could be good news for African economies. If these increases go hand-in-hand with improvements in infrastructure and greater housing density, then they signify rising welfare in primate cities.

City structure: Next, we study how the interior structure of cities is transforming over time. Our aim is to better understand whether neighborhoods within African cities are becoming better connected or whether they increasingly resemble loose clusters of disconnected informal settlements. For this part of the analysis, we only use the corrected data and focus on the "envelope" of the city, that is, the maximum urban extent observed in both the initial and final boundaries. The rationale for these two choices is straightforward. First, Figure 1 clearly illustrates that any measure of dispersion would be severely biased if calculated on the original data where all top-coded pixels have similar light intensities. New York, London, New Delhi or Johannesburg would appear as homogenous cities within which most locations look alike. Second, focusing on the maximum urban extent allows cities to sprawl and become less connected over time (Harari, 2020), while physically constrained cities or cities in which slums are replaced with formal housing can increase in density and compactness.[34]

We compute two proxies for the variation of urban population density or within-city fragmentation, both of which have been previously used in the literature on urban forms (e.g. see Tsai, 2005). Our first measure is the coefficient of variation of lights per km^2. The coefficient of variation is a simple inequality measure capturing the variation of light intensities across an entire city. It is defined as the ratio of the standard deviation to the mean. A high (low) value indicates large (small) within-city differences in the dispersion of light. The index is not bounded from above.

Our second measure of fragmentation is Moran's I (Moran, 1950). Moran's I takes the precise location of each pixel within a city into account and indicates whether similar light intensities cluster together in space. It is defined as

$$I = \frac{N}{S_0} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where N is the number of pixels in the initial footprint of the city, w_{ij} are elements of an $N \times N$ inverse distance weight matrix, S_0 is the sum of all w_{ij}, x_i or x_j is the pixel-level light intensity, and $\bar{x}$ is mean luminosity.[35]

Positive values of Moran's I indicate that pixels are surrounded by others of similar luminosity or population density (positive autocorrelation), while negative values reflect

[34] We still focus on agglomerations which are now defined as all sub-cities which will eventually merge into a single metropolitan area. All results presented here are robust to using the initial footprint only.

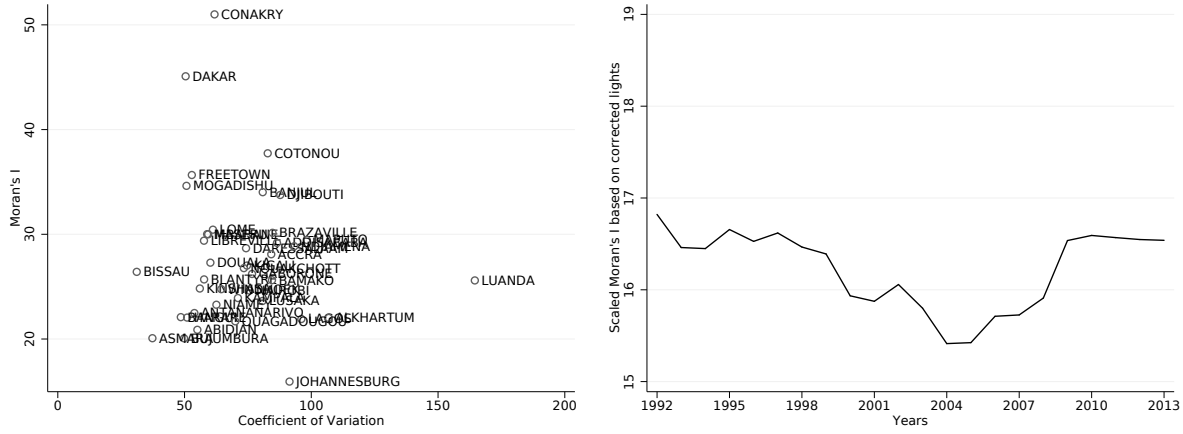
[35] We work with a scaled version of Moran's I to make cities consisting of different numbers of pixels comparable, that is, we subtract its expected value under the null hypothesis of no spatial correlation: $I^* = I - \mathbb{E}[I] = I - (-1/(N-1))$.

a checkerboard pattern (negative autocorrelation). The index ranges from minus one to one. Light intensities within cities are positively spatially correlated but there is a clear ranking. The index continuously falls as we move from monocentric cities over polycentric cities to decentralized urban sprawl. A monocentric city in which luminosity slowly and gradually decreases from the densely populated center to the sparsely populated outskirts will have a higher Moran's I than a checkered city in which dense and sparsely populated areas take turns. We scale both indices by 100 for a better exposition.

Panel (a) of Figure 6 illustrates the heterogeneity of urban structures on the subcontinent and shows that our light-based measures capture meaningful variation. Consider, for example, cities with a high Moran's I and relatively low coefficients of variation, such as Conakry, Dakar, and Cotounou. This combination indicates a regular structure with a bright center surrounded by similarly bright areas with a slow decay towards darker outskirts. Other cities with the same coefficient of variation have a much lower Moran's I. Their spatial distribution is considerably more fragmented, matching other accounts. A large part of Abidjan's population, for example, lives in slums characterized by illegal land tenure, buildings made out of non-permanent materials, and little or next to no infrastructure (UN-Habitat, 2003).

Figure 6 – Varying structures of cities in Sub-Saharan Africa

(a) Fragmentation in primate cities, 2000 **(b)** Moran's I in Johannesburg



Notes: Illustration of urban structures in Sub-Saharan Africa. Panel (a) shows a cross-sectional scatter plot of the estimated coefficient of variation (CV) and Moran's I in 2000. Panel (b) displays the evolution of Moran's I in Johannesburg over the period from 1992 to 2013.

Johannesburg is a particularly interesting case in terms of fragmentation. In 2000, it has one of the highest coefficients of variation and the lowest Moran's I in our sample of primate cities. Owing to a legacy of racial segregation during Apartheid, Johannesburg consists of alternating poor and rich neighborhoods which do not form a single integrated city. There is some limited evidence that this pattern is changing. Panel (b) of Figure 6 shows that we observe a moderate increase in Moran's I since the mid-2000s. The

coefficient of variation is decreasing at the same time. This suggests that the different neighborhoods are integrating, although the overall levels of inequality and fragmentation remain very high when compared with other cities in our sample.[36] Just as before, we have to be careful when interpreting raw trends based on these data, as they include substantial measurement error.

We employ the same approach as before to analyze these data in a more structured manner and focus on the differential between city types—i.e., we regress one of the measures of city fragmentation, F_{ijt}, on a linear time trend, an interaction of a linear time trend with an indicator for primate cities, P_{ij}, the log of lights per km^2 in the city, $\ln \text{LIGHTS}_{ijt}$, and a set of fixed effects that vary across specifications. We include the city-wide average light intensity to analyze their changing structure net of scale effects.[37]

Table 5 – Trends in fragmentation of African cities, envelopes

	Dependent variable:					
	Coefficient of Variation			*Moran's I*		
	(1)	(2)	(3)	(4)	(5)	(6)
Linear trend	-0.436			-0.523		
	(0.050)***			(0.030)***		
	[0.109]***			[0.060]***		
Primate × Linear trend	-0.796	-0.797	-0.657	-0.019	-0.020	-0.004
	(0.128)***	(0.128)***	(0.146)***	(0.027)	(0.027)	(0.036)
	[0.132]***	[0.132]***	[0.148]***	[0.030]	[0.030]	[0.036]
Lights per km^2	-18.759	-18.732	-17.334	-1.343	-1.232	-1.456
	(1.787)***	(1.871)***	(1.997)***	(0.467)***	(0.496)**	(0.617)**
	[1.980]***	[2.060]***	[2.224]***	[0.477]***	[0.504]**	[0.605]**
City FE	✓	✓	✓	✓	✓	✓
Satellite FE	✓	–	–	✓	–	–
Year FE	–	✓	–	–	✓	–
Country-Year FE	–	–	✓	–	–	✓
Observations	12356	12356	12356	12356	12356	12356
Cities	562	562	562	562	562	562

Notes: The table reports results of city-level panel regressions using the top-coding corrected data. The specifications are variants of $F_{ijt} = \beta_1 t + \beta_2 (t \times P_{ij}) + \beta_3 \ln \text{LIGHTS}_{ijt} + c_{ij} + s_{jt} + \epsilon_{ijt}$, where F_{ijt} is either the coefficient of variation or Moran's I, t is a linear time trend, P_{ij} is an indicator for primate cities, c_{ij} is a city fixed effect and s_{jt} contains are varying set of fixed effects (satellite, year, or country-year). Standard errors clustered at the city level are reported in parentheses. Conley errors with a spatial cutoff of 1,000 km and a time-series HAC with a lag cutoff of 1,000 years are reported in brackets. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5 reveals two new insights. First, we observe a decrease in the dispersion of lights over time which differs strongly across the two city types. Column (1) shows that

[36] The evidence is stronger if we focus on the initial footprint and ignore that Johannesburg is sprawling. The rise towards the end of the sample is steeper and exceeds the values of the 1990s.

[37] Note that this does not alter our conclusions in any important manner.

the coefficient of variation has been decreasing steadily over the period from 1992 to 2013. Columns (2) and (3) add that this trend is two or almost three times faster in primate cities relative to the decline secondary cities. Second, this decrease in the concentration of light is accompanied by an increasingly less regular structure of cities, as is implied by decreasing Moran's I. However, we observe few differences between primary and secondary cities on this index. Our results suggest that African cities as a whole are sprawling and becoming less compact. This trend is compounded by increases in light density with the envelope of a city. Growing cities experiences further reductions in the coefficient of variation as well as as Moran's I.[38]

Our preferred interpretation of these findings is that Africa's biggest cities are at a crossroads. They are growing rapidly at the intensive and extensive margin, while the distribution of economic activity and people is starting to equalize across the city. The simultaneous decrease in clustering, which is also present when we only examine the initial footprint, tells us that subcenters are forming but these may or may not ultimately coalesce into a cohesive whole.[39] Although poorer neighborhoods are becoming denser and brighter relative to the center, a lack of connectivity to other neighborhoods remains a major obstacle and constrains economies of scale. This limits the ability of African cities to facilitate matching, sharing and learning (Duranton and Puga, 2004). Moreover, the lack of a compact shape itself induces large welfare losses for urban consumers in developing countries (Harari, 2020).

It is important to emphasize that there is nothing tautological about these results in our correction; if anything, the procedure provides a lower bound of primate city growth and fragmentation. Top-coded pixels which become brighter over time move up in the global ranking and thus receive higher theoretical values. Our correction does not explicitly consider whether these pixels are located in primary or secondary cities, or if they are surrounded by other bright pixels. This is precisely how Johannesburg is able to buck the trend of other large cities in Sub-Saharan Africa.

6 Concluding remarks

While satellite data of nighttime lights are an increasingly popular proxy for economic activity, they suffer from top-coding and severely underestimate the brightness of most cities. The key contribution of this paper is to provide a solution to this problem and establish new findings about the economic performance of cities in Sub-Saharan Africa.

[38] We also present tentative evidence on whether an increasing fragmentation of cities inhibits their subsequent growth in Online Appendix J. Light inequality seems to lower future growth, while the effect of fragmentation appears to be ambiguous.

[39] Simulations with a fixed city size show that the emergence of sub-centers goes in line with an increase in light density, a decrease in light inequality and an increase in fragmentation, just as we observe in the African data.

Our solution rests on the claim that top lights can be characterized by a Pareto distribution. We support this conjecture in two ways. First, a model of luminosity emitted by large cities suggests that plausible assumptions directly lead to a power law in light emissions. Second, a battery of empirical tests indicates that a Pareto distribution is a sound representation of the data. Other parametric or non-parametric approaches are possible, but we find it appealing to directly link the distribution of bright lights to both Zipf's law of cities and the standard tail extrapolation problem. On this basis, we develop a geo-referenced ranking procedure to replace the top-coded pixels with their theoretical counterparts and present a new global panel of light intensities over the period from 1992 to 2013.

The new data lends itself to numerous applications and performs well in several benchmarking exercises. In this paper, we focus on city growth and city structure in Sub-Saharan Africa. Our main finding is that primary cities have maintained their dominant position but are becoming more fragmented internally. This limits economies of scale and their ability to break into world markets. Institutional features certainly play a role in this development and warrant more research. If public services are improving, and public infrastructure connects striving neighborhoods, then this will have wide-ranging benefits extending well beyond the city. Finally, by focusing on cities in Sub-Saharan Africa, we submit our data to a conservative test. Given that the top-coding correction makes a substantial difference in a setting where electrification rates and urban building densities are low, it will certainly play a larger role in other parts of the world.

References

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.

Ades, A. F. and E. L. Glaeser (1995). Trade and circuses: Explaining urban giants. *Quarterly Journal of Economics 110*(1), 195–227.

Alesina, A., S. Michalopoulos, and E. Papaioannou (2016). Ethnic inequality. *Journal of Political Economy 124*(2), 428–488.

Amson, J. (1972). Equilibrium models of cities: 1. An axiomatic theory. *Environment and Planning A 4*(4), 429–444.

Atkinson, A. B., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of Economic Literature 49*(1), 3–71.

Barrios, S., L. Bertinelli, and E. Strobl (2006). Climatic change and rural–urban migration: The case of Sub-Saharan Africa. *Journal of Urban Economics 60*(3), 357–371.

Batty, M. and P. Longley (1994). *Fractal Cities: A Geometry of Form and Function*. San Diego, CA and London: Academic Press.

Bertaud, A. and S. Malpezzi (2014). The spatial distribution of population in 57 world cities: The role of markets, planning, and topography. Unpuplished, University of Wisconsin-Madison.

Bettencourt, L. M. A. (2013). The origins of scaling in cities. *Science 340*(6139), 1438–1441.

Bluhm, R., D. de Crombrugghe, and A. Szirmai (2018). Poverty accounting. *European Economic Review 104*(C), 237–255.

Bonfatti, R. and S. Poelhekke (2017). From mine to coast: Transport infrastructure and the direction of trade in developing countries. *Journal of Development Economics 127*, 91–108.

Brueckner, J. K. (1982). A note on sufficient conditions for negative exponential population densities. *Journal of Regional Science 22*(3), 353–359.

Castells-Quintana, D. (2017). Malthus living in a slum: Urban concentration, infrastructure and economic growth. *Journal of Urban Economics 98*, 158–173.

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

Christiaensen, L. and R. Kanbur (2017). Secondary towns and poverty reduction: Refocusing the urbanization agenda. *Annual Review of Resource Economics 9*, 405–419.

Christiaensen, L. and Y. Todo (2014). Poverty reduction during the rural-urban transformation: The role of the missing middle. *World Development 63*(C), 43–58.

Cirillo, P. (2013). Are your data really Pareto distributed? *Physica A: Statistical Mechanics and its Applications 392*(23), 5947–5962.

Coleman, J. S. (1964). *Introduction to Mathematical Sociology*. New York, NY: The Free Press.

Desmet, K. and E. Rossi-Hansberg (2013). Urban accounting and welfare. *American Economic Review 103*(6), 2296–2327.

Donaldson, D. and A. Storeygard (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives 30*(4), 171–198.

Duranton, G. (2008). Viewpoint: From cities to productivity and growth in developing countries. *Canadian Journal of Economics 41*(3), 689–736.

Duranton, G. and D. Puga (2004). Micro-Foundations of Urban Agglomeration Economies. In J. V. Henderson and J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4 of *Handbook of Regional and Urban Economics*, Chapter 48, pp. 2063–2117. Elsevier.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review 94*(5), 1429–1451.

Eeckhout, J. (2009). Gibrat's law for (all) cities: Reply. *American Economic Review 99*(4), 1676–1683.

Elvidge, C. D., K. E. Baugh, J. B. Dietz, T. Bland, P. C. Sutton, and H. W. Kroehl (1999). Radiance calibration of DMSP-OLS low-light imaging data of human settlements. *Remote Sensing of Environment 68*(1), 77–88.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *Quarterly Journal of Economics 114*(3), 739–767.

Gabaix, X. and R. Ibragimov (2011). Rank - 1/2: A simple way to improve the ols estimation of tail exponents. *Journal of Business & Economic Statistics 29*(1), 24–39.

Gabaix, X. and Y. Ioannides (2004). The evolution of city size distribution. In J. V. Henderson and

J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (1 ed.), Volume 4, Chapter 53, pp. 2341–2378. Elsevier.

Gibrat, R. (1931). *Les Inegalités Économiques: Applications: Aux Inegalités des Richesses, à la Concentration des Entreprises, aux Populations des Villes, aux Statistiques des Familles, etc. D'une Loi Nouvelle, la Loi de l'Effet Proportionnel*. Paris: Librairie du Recueil Sirey.

Gibson, J., G. Datt, R. Murgai, and M. Ravallion (2017). For India's rural poor, growing towns matter more than growing cities. *World Development 98*, 413–429.

Glaeser, E. and J. V. Henderson (2017). Urban economics for the developing world: An introduction. *Journal of Urban Economics 98*, 1–5.

Glaeser, E. L. (2014). A world of cities: The causes and consequences of urbanization in poorer countries. *Journal of the European Economic Association 12*(5), 1154–1199.

Gollin, D., R. Jedwab, and D. Vollrath (2016). Urbanization with and without industrialization. *Journal of Economic Growth 21*(1), 35–70.

Gollin, D., M. Kirchberger, and D. Lagakos (2017). In search of a spatial equilibrium in the developing world. Working Paper 23916, National Bureau of Economic Research.

Harari, M. (2020). Cities in bad shape: Urban geometry in India. *American Economic Review* (forthcoming).

Henderson, J. V. (2003). The urbanization process and economic growth: The so-what question. *Journal of Economic Growth 8*(1), 47–71.

Henderson, J. V., T. Squires, A. Storeygard, and D. Weil (2018). The global distribution of economic activity: Nature, history, and the role of trade. *Quarterly Journal of Economics 133*(1), 357–406.

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

Hodler, R. and P. A. Raschky (2014). Regional favoritism. *Quarterly Journal of Economics 129*(2), 995–1033.

Hsu, F.-C., K. E. Baugh, T. Ghosh, M. Zhizhin, and C. D. Elvidge (2015). DMSP-OLS radiance calibrated nighttime lights time series with intercalibration. *Remote Sensing 7*(2), 1855–1876.

Ioannides, Y. and S. Skouras (2013). US city size distribution: Robustly Pareto, but only in the tail. *Journal of Urban Economics 73*(1), 18–29.

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.

Jedwab, R. and A. Moradi (2016). The permanent effects of transportation revolutions in poor countries: Evidence from Africa. *Review of Economics and Statistics 98*(2), 268–284.

Jedwab, R. and D. Vollrath (2019). The urban mortality transition and poor country urbanization. *American Economic Journal: Macroeconomics 11*(1), 223–75.

Jiang, B. and T. Jia (2011). Zipf's law for all the natural cities in the United States: A geospatial perspective. *International Journal of Geographical Information Science 25*(8), 1269–1281.

Jones, A. (1975). Density-size rule, a further note. *Urban Studies 12*(2), 225–228.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy 99*(3), 483–499.

Lall, S. V., J. V. Henderson, and A. J. Venables (2017). *Africa's Cities: Opening Doors to the World*. Washington, DC: World Bank.

Lessmann, C. and A. Seidel (2017). Regional inequality, convergence, and its determinants: A view from outer space. *European Economic Review 92*(B), 110–132.

Letu, H., M. Hara, G. Tana, and F. Nishio (2012). A saturated light correction method for DMSP/OLS nighttime satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing 50*, 389–396.

Lipton, M. (1977). *Why Poor People Stay Poor: Urban Bias in World Development*. Cambridge, MA: Harvard University Press.

Luckstead, J. and S. Devadoss (2014). Do the world's largest cities follow Zipf's and Gibrat's laws? *Economics Letters 125*(2), 182–186.

Michalopoulos, S. and E. Papaioannou (2013). Pre-colonial ethnic institutions and contemporary African development. *Econometrica 81*(1), 113–152.

Michalopoulos, S. and E. Papaioannou (2014). National institutions and subnational development in Africa. *Quarterly Journal of Economics 129*(1), 151–213.

Michalopoulos, S. and E. Papaioannou (2018). Spatial patterns of development: A meso approach. *Annual Review of Economics 10*(1), 383–410.

Mills, E. (1967). An aggregative model of resource allocation in a metropolitan area. *American Economic Review 57*(2), 197–210.

Moran, A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika 37*(1/2), 17–23.

Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics 46*(5), 323–351.

Parr, J. (1985). A population-density approach to regional spatial structure. *Urban Studies 22*(4), 289–303.

Piketty, T. (2003). Income inequality in France, 1901–1998. *Journal of Political Economy 111*(5), 1004–1042.

Pinkovskiy, M. (2017). Growth discontinuities at borders. *Journal of Economic Growth 22*(2), 145–192.

Pinkovskiy, M. and X. Sala-i Martin (2016). Lights, camera ... income! illuminating the national accounts-household surveys debate. *Quarterly Journal of Economics 131*(2), 579–631.

Puga, D. (1998). Urbanization patterns: European versus less developed countries. *Journal of Regional Science 38*(2), 231–252.

Rosen, K. T. and M. Resnick (1980). The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics 8*(2), 165–186.

Rosenthal, S. and W. Strange (2004). Evidence on the nature and sources of agglomeration economics. In J. Henderson and J. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4, pp. 2119–2171. Elsevier North Holland.

Rozenfeld, H., D. Rybski, X. Gabaix, and H. Makse (2011). The area and population of cities: New insights from a different perspective on cities. *American Economic Review 101*(5), 2205–2225.

Small, C., C. Elvidge, D. Balk, and M. Montgomery (2011). Spatial scaling of stable night lights. *Remote Sensing of Environment 115*(2), 269–280.

Smeed, R. (1961). *The Traffic Problem in Towns*. Manchester Statistical Society Papers.

Soo, K. T. (2005). Zipf's law for cities: A cross-country investigation. *Regional Science and Urban Economics 35*(3), 239–263.

Stewart, J. Q. (1947). Suggested principles of "social physics". *Science 106*(2748), 179–180.

Storeygard, A. (2016). Farther on down the road: Transport costs, trade and urban growth in sub-saharan africa. *Review of Economic Studies 83*(3), 1263–1295.

Tsai, Y.-H. (2005). Quantifying urban form: Compactness versus 'sprawl'. *Urban Studies 42*(1), 141–161.

Tuttle, B., S. Anderson, P. Sutton, C. Elvidge, and K. Baugh (2013). It used to be dark here. *Photogrammetric Engineering and Remote Sensing 79*(3), 287–297.

UN-Habitat (2003). *The Challenge of Slums: Global Report on Human Settlements, 2003*. Earthscan Publications.

United Nations (2018). *World Urban Prospects, the 2018 Revision*. United Nations New York.

Venables, A. J. (2017). Breaking into tradables: Urban form and urban function in a developing city. *Journal of Urban Economics 98*, 88–97.

Zhang, Q., C. Schaaf, and K. C. Seto (2013). The vegetation adjusted NTL urban index: A new approach to reduce saturation and increase variation in nighttime luminosity. *Remote Sensing of Environment 129*, 32–41.

Zielinski, K. (1980). The modelling of urban population density: A survey. *Environment and Planning A 12*(2), 135–154.

Ziskin, D., K. Baugh, F. C. Hsu, T. Ghosh, and C. Elvidge (2010). Methods used for the 2006 radiance lights. *Proceedings of the Asia-Pacific Advanced Network 30*, 131–142.

Online appendix

A Additional summary statistics

In this section we provide further details on the construction of the stable lights and radiance-calibrated lights, and compare their characteristics.

The main advantage of the stable lights series is that they are available as an annual panel. Moreover, for several years more than one satellite orbited Earth, resulting in a total of 34 satellite-years over the period of 1992 to 2013. The radiance-calibrated data, by contrast, are based on rarely occurring additional flights of satellites which were about to be decommissioned and could be operated with a different gain setting (lower or higher amplification settings). These auxiliary data are only available for seven years over the entire period from 1996 to 2010. NOAA blended the stable lights data from normal flight operations with these auxiliary satellite data to obtain the radiance-calibrated series (Elvidge et al., 1999, Ziskin et al., 2010, Hsu et al., 2015). The resulting night light intensities are free of top-coding and have no upper bound (at least in theory).

Several technical issues and measurement errors, occurring when the different fixed gain images were merged at NOAA, produced a lot of variability in the radiance-calibrated data: *i)* the low amplification data are based on considerably fewer orbits than the stable lights series (often covering only small parts of a year), *ii)* they were generated by blending different parts of the frequency spectrum which are deemed reliable, *iii)* higher light intensities are supported by fewer and fewer fixed-gain images[1], and *iv)* fires or stray lights were not fully removed from the auxiliary data. All this contributes to the high variance across different radiance-calibrated satellite-years.[2] Because of this instability, together with the fact that they are only available for seven out of 22 years, we only rely on the radiance-calibrated data to infer the shape of the distribution at the top. The relative ranks of pixels are consistently measured across the different satellites and less prone to be affected by measurement errors.

Table A-1 reports summary statistics for the 34 stable light satellite-years and the seven radiance-calibrated years. Between 2.7% and 5.9% of all pixels in the stable lights images reach the top of the scale (i.e., 55 DN to 63 DN), more so in later years. As the radiance-calibrated lights do not suffer from top-coding, their mean, standard deviation and Gini in lights are much higher. Rather than being capped at 63 DN, they reach maximum values from 2000 to 5000 DN. The fluctuations across satellites are reflected in the overall mean light intensity, but are most apparent at the top. The maximum

[1]Consider the 2010 radiance-calibrated product for example, the maximum number of cloud-free images is 134, the suburbs of Paris are informed by about 50–60 cloud-free images, but the city core only by 10–20 images. This pattern repeats itself throughout all major cities.

[2]Measurement errors are also present in the stable lights data and affect their reliability in the time series dimension but to a much lesser extent. The sensors of the satellites deteriorated over their lifetime and had to be replaced every couple of years, which implies that later recordings of any particular satellite tend to be the brightest (although this is not a hard rule). In panel regressions, economists usually resort to a combination of satellite and time fixed effects to partially address this issue.

light intensity doubles within three years and then decreases again by a similar amount (whereas the mean increases and decreases by about 27% over the same period).

Table A-2 confirms that these fluctuations are not driven by a few outliers. Instead of examining overall maxima, we now report various percentiles for the seven radiance-calibrated satellites and the means above these percentiles. For example, the top 2% begin at 147.01 DN in the 1996 data, at 214.59 DN in 2003, and again at 150.90 DN in 2010. The means above the various percentiles vary similarly over time. The differences are largest in absolute values at the very top but remain sizable throughout the distribution. This variation cannot be explained economically.

Table A-3 shows the maximum values attained by the seven radiance-calibrated satellites in 30 selected cities. Despite considerable variability over time, the relative ranking is in line with our expectations. The light intensity of the brightest pixel in New York City, for example, is about ten times greater than that of the brightest pixel in Nairobi. Note that the average maximum light intensity hardly exceeds 2000 DN, no matter if we compute it for London, New York, or Shanghai. This is why we restrict the maximum light intensities generated by our pixel-level correction to 2000 DN.

Table A-4 illustrates that not all differences between the stable lights and radiance-calibrated data can be attributed to top-coding. It regresses all pixels below 55 DN of the stable lights on the radiance-calibrated lights, where top-coding is supposed to not play a role. We find a regression coefficient around one-half rather than equivalence. This absence of a one-to-one correspondence is owed to the lack of on-board calibration, blooming (Abrahams et al., 2018), the presence of stray light (Hsu et al., 2015), and geo-location errors (Tuttle et al., 2013).

Table A-5 reports the maximum light intensities recorded within 25 kilometers of the city center in 988 world cities with more than 500,000 inhabitants. Table A-6 adds the rank-correlations. The latter are much higher and typically around 0.90–0.95 for adjacent radiance-calibrated years, which supports our preference for pixel ranks over their actual values.

Table A-1 – Summary statistics of the stable lights and radiance-calibrated data

	Stable lights					*Radiance-calibrated*			
Year	Flight No.	Mean	Std. Dev.	Gini	% ≥ 55	Mean	Std. Dev.	Gini	Max
1992	F10	13.83	13.51	0.44	3.81				
1993	F10	11.96	12.81	0.46	3.12				
1994	F10	12.02	13.31	0.48	3.49				
	F12	14.65	13.93	0.44	4.20				
1995	F12	13.09	13.57	0.46	3.76				
1996	F12	12.69	13.36	0.46	3.51	19.42	55.63	0.65	2064
1997	F12	13.45	13.74	0.45	3.94				
	F14	10.98	12.87	0.49	3.16				
1998	F12	13.89	13.89	0.45	4.18				
	F14	10.94	12.78	0.49	3.05				
1999	F12	14.74	14.34	0.44	4.67	19.53	56.93	0.64	4698
	F14	10.15	12.31	0.49	2.78				
2000	F14	11.34	12.99	0.49	3.18	22.88	65.84	0.63	5552
	F15	13.25	13.34	0.44	3.70				
2001	F14	11.64	13.32	0.49	3.50				
	F15	12.93	13.26	0.45	3.54				
2002	F14	12.14	13.70	0.49	3.77				
	F15	13.18	13.44	0.45	3.72				
2003	F14	11.96	13.72	0.49	3.82	24.83	67.57	0.65	4186
	F15	10.28	12.45	0.50	2.70				
2004	F15	10.08	12.52	0.51	2.76	24.07	65.94	0.66	4357
	F16	11.82	13.04	0.46	3.40				
2005	F15	10.44	12.73	0.51	2.79				
	F16	10.44	12.54	0.49	2.85				
2006	F15	10.56	12.91	0.51	2.93	20.63	50.93	0.63	3333
	F16	12.26	13.37	0.47	3.48				
2007	F15	10.74	12.82	0.50	2.79				
	F16	13.05	13.79	0.46	4.03				
2008	F16	12.97	13.84	0.47	3.95				
2009	F16	13.50	14.12	0.47	4.17				
2010	F18	17.55	15.35	0.43	5.91	19.04	44.35	0.60	2110
2011	F18	14.78	14.68	0.46	4.94				
2012	F18	16.44	15.20	0.44	5.76				
2013	F18	16.23	15.20	0.44	5.78				

Notes: The table reports summary statistics using a 10% sample of the stable lights and radiance-calibrated data at the pixel level, where each pixel is 30 × 30 arc seconds. There are several years when two DMSP satellites were concurrently recording data for the stable lights series, so that there are 34 satellite-years between 1992 and 2013. The radiance-calibrated data are only available for the following periods: 16 Mar 96 – 12 Feb 97 (1996), 19 Jan 99 – 11 Dec 99 (1999), 03 Jan 00 – 29 Dec 00 (2000), 30 Dec 02 – 11 Nov 2003 (2003), 18 Jan 04 – 16 Dec 04 (2004), 28 Nov 05 – 24 Dec 06 (2006), and 11 Jan 10 – 9 Dec 10 (2010), although the actual coverage in terms of days often refers to a much smaller period.

Table A-2 – Summary statistics of the radiance-calibrated data (top shares)

Year	1996	1999	2000	2003	2004	2006	2010
				Panel a) Top 5%			
Percentile (x)	62.87	66.74	73.75	94.60	90.97	74.97	64.84
Mean above x	186.04	197.42	228.61	245.59	236.80	189.34	166.90
				Panel b) Top 4%			
Percentile (x)	76.30	84.79	95.62	119.27	114.26	94.03	81.98
Mean above x	215.29	228.01	264.84	280.40	270.51	215.70	190.42
				Panel c) Top 3%			
Percentile (x)	98.42	114.12	131.13	154.40	149.82	122.72	108.27
Mean above x	258.23	271.33	315.97	328.70	317.06	251.82	222.49
				Panel d) Top 2%			
Percentile (x)	147.01	166.33	198.77	214.59	207.97	168.84	150.90
Mean above x	327.60	338.23	393.22	402.32	387.54	305.83	269.84
				Panel e) Top 1%			
Percentile (x)	259.04	275.41	318.53	331.88	314.53	255.44	229.79
Mean above x	460.17	463.60	534.81	538.85	519.98	404.80	354.36
				Panel f) Top 0.1%			
Percentile (x)	729.41	716.94	815.16	822.00	805.43	605.13	511.62
Mean above x	979.91	960.86	1117.96	1110.62	1111.93	806.63	687.53
				Panel g) Top 0.01%			
Percentile (x)	1355.38	1279.48	1528.71	1491.25	1516.16	1085.71	936.22
Mean above x	1551.16	1652.31	1893.03	1828.03	1914.32	1316.93	1137.76

Notes: The table shows summary statistics of the radiance-calibrated data at the various percentiles. The input data are a 10% representative sample of all non-zero lights in the radiance-calibrated data above the defined threshold at the pixel level, where each pixel is 30×30 arc seconds.

Table A-3 – Maximum light intensities in 30 selected cities over time

City	1996	1999	2000	2003	2004	2006	2010	Average
Beijing	2265.07	3160.86	977.94	2979.75	2911.25	1575.00	1262.30	2161.74
Berlin	757.45	518.67	430.80	490.48	556.83	418.00	375.34	506.79
Bogota	661.73	774.02	416.75	602.80	828.82	622.18	489.89	628.02
Brussels	1334.32	1765.47	1632.47	1561.80	1767.69	1058.33	674.38	1399.21
Cairo	543.47	622.09	527.60	730.67	753.50	550.00	414.87	591.74
Calgary	1084.76	2077.92	1669.85	822.00	1520.70	731.15	721.22	1232.51
Casablanca	919.44	769.98	729.69	1214.73	1075.77	708.33	620.97	862.70
Damascus	800.90	724.38	470.05	472.74	893.66	820.00	675.15	693.84
Dhaka	1026.01	1410.89	882.44	935.03	840.66	920.00	465.11	925.73
Dubai	329.21	323.17	269.86	457.58	607.57	1280.34	420.12	526.84
Edinburgh	811.04	537.69	453.18	973.48	767.20	425.24	518.75	640.94
Foshan	715.98	1410.36	537.46	1499.66	1625.73	1142.86	1164.98	1156.72
Istanbul	579.83	551.57	413.14	653.84	543.84	421.08	327.46	498.68
Jakarta	683.82	664.56	1100.27	1381.62	788.43	805.95	632.81	865.35
Johannesburg	349.23	406.95	271.40	343.70	510.57	314.39	304.96	357.31
London	3342.95	2145.71	1664.97	1575.50	2123.50	1815.38	1366.45	2004.92
Los Angeles	1741.22	1807.50	1519.95	1757.50	1794.70	1288.89	1087.33	1571.01
Manila	1117.10	768.36	390.44	969.96	1260.40	692.31	551.60	821.45
Moscow	1011.28	1308.20	1270.86	1282.32	1142.01	945.45	655.45	1087.94
Mosul	225.95	232.32	211.38	370.24	284.30	321.64	319.33	280.74
Mumbai	1456.54	1790.01	1515.98	1775.52	1963.82	1322.22	1842.57	1666.67
Nairobi	180.27	188.66	211.83	173.54	191.45	174.02	164.13	183.41
New York	2299.18	2090.67	1971.10	2283.33	2877.00	1592.86	1399.78	2073.42
Paris	1827.80	2444.32	1177.72	1430.28	1794.70	1425.00	874.55	1567.77
Rio de Janeiro	926.51	917.27	748.92	699.31	708.83	484.08	461.57	706.64
Seoul	629.42	695.65	629.67	808.30	810.81	580.00	513.82	666.81
Shanghai	1965.24	1906.01	1123.89	2931.80	3982.13	2307.14	1926.59	2306.12
Sydney	1482.57	1470.49	1006.67	1923.48	1600.94	1275.00	751.58	1358.68
Tel Aviv	1284.19	1679.72	997.83	1397.40	1446.72	1188.24	1099.83	1299.13
Tokyo	1709.40	1768.79	1785.16	1876.90	2013.90	1273.33	940.31	1623.97

Notes: The table report the maximum light intensity in DN recorded within 25 km radius of the city center in selection of cities. The input data are the radiance-calibrated lights. City locations are obtained from the Natural Earth point data of major populated places.

Table A-4 – Regression of stable lights on radiance-calibrated data

Year	1996	1999	2000	2003	2004	2006	2010
Stable lights	0.5557	0.5502	0.4241	0.4357	0.3473	0.4874	0.7468
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0004)
Constant	4.6422	5.5218	3.9172	3.6007	3.2069	2.9392	6.4094
	(0.0045)	(0.0054)	(0.0044)	(0.0043)	(0.0036)	(0.0037)	(0.0066)
R^2	0.7440	0.7013	0.7115	0.7709	0.7873	0.8011	0.6319

Notes: The table reports OLS estimates of a regression of all pixels smaller than 55 DN of the stable lights on their radiance-calibrated counterpart in all those years for which both data sources are available. Standard errors are in parentheses. The data are a 10% random sample of lights at the pixel level, where each pixel is 30×30 arc seconds.

Table A-5 – Correlation matrix of maximum city lights

Years	1996	1999	2000	2003	2004	2006	2010
1996	1.0000						
1999	0.9142	1.0000					
2000	0.8615	0.8561	1.0000				
2003	0.8588	0.8619	0.7715	1.0000			
2004	0.8733	0.9002	0.8073	0.9181	1.0000		
2006	0.8737	0.8974	0.8109	0.9307	0.9379	1.0000	
2010	0.7831	0.7872	0.7428	0.8529	0.8525	0.8955	1.0000

Notes: The table reports correlations between the maximum light intensities recorded within 25 km radius of the city center of 988 world cities with more than 500,000 inhabitants.

Table A-6 – Rank correlation matrix of maximum city lights

Years	1996	1999	2000	2003	2004	2006	2010
1996	1.0000						
1999	0.9557	1.0000					
2000	0.9167	0.9122	1.0000				
2003	0.9048	0.9162	0.8451	1.0000			
2004	0.9129	0.9331	0.8447	0.9536	1.0000		
2006	0.9063	0.9256	0.8611	0.9482	0.9549	1.0000	
2010	0.8495	0.8651	0.8108	0.8964	0.8973	0.9270	1.0000

Notes: The table reports rank correlations between the maximum light intensities recorded within 25 km radius of the city center of 988 world cities with more than 500,000 inhabitants.

B The top-coding threshold

The influence of top-coding in the DMSP-OLS satellite data has been underestimated in part because much of the literature assumes it only affects pixels with the highest recorded value. However, even though the scale of stable lights goes up to 63, we have good reason to assume that many pixels with DNs of 62, 61, down to the mid-50s, are subject to top-coding and should be brighter than they are recorded in the data.

The rationale behind this conjecture is straightforward. The stable lights data were already averaged at least twice during the data construction. First, the DMSP satellites averaged several higher resolution pixels on-board to reduce the amount of information that needs to be transmitted down to Earth. The OLS system recorded images at a nominal resolution of 0.56 km, which was averaged on-board into 5×5 blocks to create a 2.77 km (smooth) resolution and then reprojected onto a 30 arc second grid.[3] Second, the data providers at NOAA processed the daily images into a single annual composite. As a result, many pixels suffering from top-coding in at least one of the underlying fine resolution data points or smooth resolution daily images would have ended up with an average value of less than 63. Hsu et al. (2015) suggest that this subtle type of top-coding may even start at a DN as low as 35. Since "the OLS does onboard averaging to produce its global coverage data, saturation does not happen immediately when radiance reaches the maximum level. On the contrary, as the actual radiance grows, the observed DN value fails to follow the radiance growth linearly, causing a gradual transition into a plateau of full saturation" (Hsu et al., 2015, p. 1872).

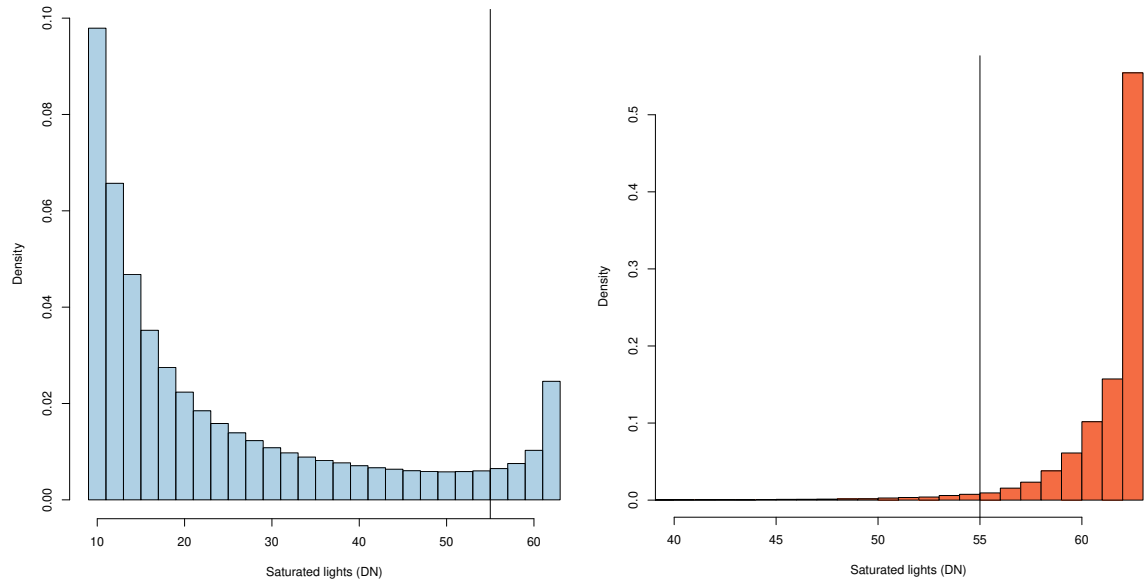
We explore the location of the top-coding threshold with a statistical approach. If only the stable lights at 63 DN were subject to top-coding, we would expect the histogram in panel (a) of Figure B-1 to show a decreasing shape ending in a spike only at 63 DN. Instead, we observe an increase in the number of pixels from 55 onwards (e.g. a bathtub shape), signaling that these values are top-coded as well. Further evidence along these lines is provided by panel (b) of Figure B-1. It shows a histogram of the light intensity of the stable lights DNs associated with high radiance-calibrated values (above 160 DN). There are a large number of pixels with DNs down to the mid-50s which correspond to very high radiance-calibrated values, but the density falls rapidly below the mid-50s. Other years show very similar patterns.

Table B-1 list the percentile values of the radiance-calibrated lights corresponding to stable lights at 55 DN, 56 DN and so on. The stable lights at 63 DN have the highest radiance-calibrated values (50% of them are higher than 390 DN). But there is also a significant share of 55 DN lights corresponding to high radiance-calibrated values, for instance, 25% are recorded with 140 DN or brighter.

[3] See https://directory.eoportal.org/web/eoportal/satellite-missions/d/dmsp-block-5d or Abrahams et al. (2018) for a detailed description of the sensors and on-board processing.

Figure B-1 – Histograms of stable lights in 1999

(a) If stable DN > 9 **(b)** If radiance-calibrated DN > 160



Notes: Illustration of the location of the top-coding threshold in the stable lights. Panel a) shows a histogram of the F12 satellite in 1999 for all pixels with a DN greater 9. Panel b) shows a histogram of the same satellite only for pixels where the radiance-calibrated light intensity is greater 160 DN. The input data are a 10% representative sample of all non-zero lights in the stable lights and radiance-calibrated data at the pixel level (see Elvidge et al., 2009, Hsu et al., 2015).

Table B-1 – Percentiles of radiance-calibrated values at given stable lights values in 2000

Stable lights	Radiance-calibrated percentiles					
DN	5%	25%	50%	75%	95%	99%
55	53.20	74.94	99.41	140.85	232.90	328.86
56	56.15	79.99	108.20	153.92	250.93	344.05
57	60.14	84.99	115.11	164.63	262.18	357.60
58	64.13	92.81	125.35	179.57	277.59	392.33
59	70.32	101.97	141.92	203.17	306.77	423.28
60	79.16	116.64	163.92	231.91	344.57	497.25
61	89.33	137.89	196.68	268.21	410.91	625.30
62	109.03	176.36	246.66	331.46	524.18	762.63
63	160.91	276.92	390.08	560.28	952.14	1494.85

Notes: The table reports values from the cumulative distribution function of the radiance-calibrated lights which are associated with a given stable lights value (from 55 to 63). For instance, 25% of the radiance-calibrated values associated with a stable lights value of 61 DN, are below 122.06. The data are a representative 10% sample for the year 2000.

C Proofs and extensions of the model

In this section we provide additional proofs and extensions of the model presented in the main text.

The CDF of the number of rings: Note that $r = \pi^{-1/2} x^{1/(2\phi)}$ implies $x = \pi^{\phi} r^{2\phi}$ and $\mathrm{d}x = 2\phi\pi^{\phi} r^{2\phi-1}\mathrm{d}r$. Substituting these definitions into eq. (1) and integrating yields the CDF of the number of rings per city as presented in eq. (2) of the main text

$$\begin{aligned} F(r) &= 2\phi x_c \pi^{-\phi} \int_{\tilde{r}}^{r} r^{-2\phi-1}\mathrm{d}r = 2\phi x_c \pi^{-\phi} \left[-\frac{1}{2\phi} r^{-2\phi}\right]_{\tilde{r}}^{r} \\ &= \begin{cases} 0 \text{ for } r < \tilde{r} = \pi^{-1/2} x_c^{1/(2\phi)} \\ 1 - y_c \pi^{-\phi} r^{-2\phi} \text{ for } r >= \tilde{r} = \pi^{-1/2} x_c^{1/(2\phi)}. \end{cases} \end{aligned} \tag{C-1}$$

The density of pixels: Start with the distribution of the number of pixels. At distances $d < \tilde{d}$, the amount of pixels increases linearly in d as rings farther away from the center contain more pixels: $\frac{\mathrm{d}}{\mathrm{d}d}\pi d^2 = 2\pi d$. Beyond $\tilde{d}$, the effect within each city has to be multiplied by the survival function $1 - F(r)$ from eq. (2), as there are fewer and fewer cities of such size. Denoting the number of cities as M, the absolute amount of pixels N as a function of d is

$$P(d) = \begin{cases} 2\pi d M \text{ for } d < \tilde{d} = \pi^{-1/2} x_c^{1/(2\phi)} \\ 2\pi^{1-\phi} M x_c d^{1-2\phi} \text{ for } d \geq \tilde{d} = \pi^{-1/2} x_c^{1/(2\phi)}. \end{cases} \tag{C-2}$$

The total number of pixels, N, can be obtained by integration

$$\begin{aligned} N &= \int_0^{\tilde{d}} 2\pi d M \mathrm{d}d + \int_{\tilde{d}}^{\infty} 2\pi^{1-\phi} M x_c d^{1-2\phi}\mathrm{d}d = 2\pi M \left[\frac{1}{2}d^2\right]_0^{\tilde{d}} + 2\pi^{1-\phi} M x_c \left[\frac{1}{2-2\phi} d^{2-2\phi}\right]_{\tilde{d}}^{\infty} \\ &= \pi M \frac{y_c^{1/\phi}}{\pi} + \frac{\pi^{1-\phi} M y_c}{\phi - 1}\left(\frac{y_c^{1/\phi}}{\pi}\right)^{1-\phi} = M x_c^{1/\phi} + \frac{1}{\phi-1} M x_c^{1/\phi} = \frac{\phi}{\phi-1} M x_c^{1/\phi}. \end{aligned} \tag{C-3}$$

Dividing eq. (C-2) by N yields the density, $f(d)$, shown in eq. (4):

$$f(d) = \begin{cases} 2\pi \frac{\phi-1}{\phi} x_c^{-1/\phi} d \text{ for } d < \tilde{d} \\ 2\pi^{1-\phi} \frac{\phi-1}{\phi} x_c^{1-1/\phi} d^{1-2\phi} \text{ for } d \geq \tilde{d} \end{cases} \tag{C-4}$$

with $\tilde{d} = \pi^{-1/2} x_c^{1/(2\phi)}$.

The density is illustrated in panel (a) of Figure C-1.

An exponential distribution of light in cities: Next, we derive the distribution of top lights when lights within a city follow a negative exponential function instead of an inverse power function. The following replaces assumption 4 from the main text.

Assumption 5. *Within cities, the light density $l(d)$ follows an exponential function $l(d) = L_0 \exp(-\gamma d)$, where $L_0 \geq l$ is the density at the center and $\gamma > 0$ is a decay parameter.*

Panel (b) of Figure C-1 illustrates the negative exponential distribution (with $\gamma = 0.15$) and inverse power distribution (with $a = 0.7$). Both functions exhibit a comparable decay from the city center to the outskirts of the city. The main difference is that the negative exponential function attains values which are not as high in the center but decreases more quickly towards zero at the outskirts, whereas the inverse power function has a longer tail.

Contrary to our baseline case this altered setting does not directly generate a Pareto distribution in lights. The distribution now depends on L_0, the maximum luminosity of the center of each city. We consider three cases. In each case, we focus on the light density $f(l)$ conditional on L_0. Using the variable transformation of eq. (4) together with $d = \frac{1}{\gamma} \ln(L_0/l)$ yields

$$f(l \mid L_0) = \begin{cases} 2\pi^{1-\phi} \frac{\phi-1}{\phi} x_c^{1-1/\phi} \left[\frac{1}{\gamma} \ln(L_0/l)\right]^{1-2\phi} & \text{for } l \leq \tilde{l} \\ \underbrace{2\pi \frac{\phi-1}{\phi} x_c^{-1/\phi} \frac{1}{\gamma}}_{c} \ln(L_0/l) & \text{for } \tilde{l} < l < L_0, \end{cases} \tag{C-5}$$

where $\tilde{l} = L_0 \exp(-\gamma \tilde{d})$.

This conditional density increases for dim luminosities (at the fringes of the largest cities) and decreases from $\tilde{l}$ onwards. The turning point and maximum, $\tilde{l}$, corresponds to the minimum size, $\tilde{d}$, of each city in terms of distance from center. At higher luminosities, there are fewer and fewer pixels as these are the ones located in ever smaller rings closer to the center.

To derive the marginal density $f(l)$, we have to make assumptions about the distribution of maximum luminosities L_0 across cities.

Case 1: L_0 is a constant, so that all cities large and small are equally bright in the center. This is unrealistic, but mathematically simple. The marginal density of lights $f(l)$ equals the conditional density $f(l \mid L_0)$. Analyzing the top end of the distribution $c \ln(L_0/l)$ for $\tilde{l} < l < L_0$ we observe a near linear decrease with a slope of $c \frac{\mathrm{d}}{\mathrm{d}l} \ln(L_0/l) \propto -\frac{1}{l}$ for l close to L_0. There is no power law. The assumption of a constant L_0 across all cities generates too many pixels with the highest luminosities.

Case 2: L_0 follows a Pareto distribution across cities so that some city centers are much brighter than others. Empirical evidence points in this direction (see Table A-3).

We are only interested in the upper part of the density in eq. (C-5).[4] If L_0 follows the Pareto density with $\eta = 1$ so that $f(L_0) = L_{min}/L_0^2$, with L_{min} as the minimum center luminosity, we have the joint density of l and L_0 as

$$f(l, L_0) = f(L_0)f(l \mid L_0) = c\frac{L_{min}}{L_0^2}\ln(L_0/l) \text{ for } \tilde{l}_{max} < l < L_{max} \text{ and } L_0 > l. \tag{C-6}$$

The marginal density, $f(l)$, is found by integrating L_0 out of the above

$$\begin{aligned} f(l) &= cL_{min}\int_l^{L_{max}} \frac{1}{L_0^2}\ln(L_0/l)dL_0 = cL_{min}\left[\frac{\ln l - \ln L_0 - 1}{L_0}\right]_l^{L_{max}} \\ &= cL_{min}\left[\frac{1}{l} - \frac{1}{L_{max}}(\ln(L_{max}/l) + 1)\right], \end{aligned} \tag{C-7}$$

which holds for $\tilde{l}_{max} < l < L_{max}$.

Case 3: Alternatively, an intermediate case for L_0 is a uniform distribution between values L_{min} and L_{max} with density $f(L_0) = (L_{max} - L_{min})^{-1}$. Cities (big and small) differ in their maximum luminosity, but all maximum luminosities are equally likely across cities. Following the same steps as in the second case, the marginal density is

$$\begin{aligned} f(l) &= \frac{c}{L_{max} - L_{min}}\int_l^{L_{max}} \ln(L_0/l)dL_0 = \frac{c}{L_{max} - L_{min}}\Big[L_0(\ln L_0 - \ln l - 1)\Big]_l^{L_{max}} \\ &= \frac{c}{L_{max} - L_{min}}\Big[l + L_{max}(\ln(L_{max}/l) - 1)\Big], \end{aligned} \tag{C-8}$$

which holds for $\tilde{l}_{max} < l < L_{max}$.

Cases 2 and 3 are heavy-tailed distributions which differ mathematically from the simple Pareto. But given their heavy tails, they may be approximated by a Pareto. As shown in Figure C-2, this works particularly well for case 2 with a Pareto distribution of $\alpha = 1.5$, while for case 3 the Pareto distributions with $\alpha = 1.2$ works reasonably well.

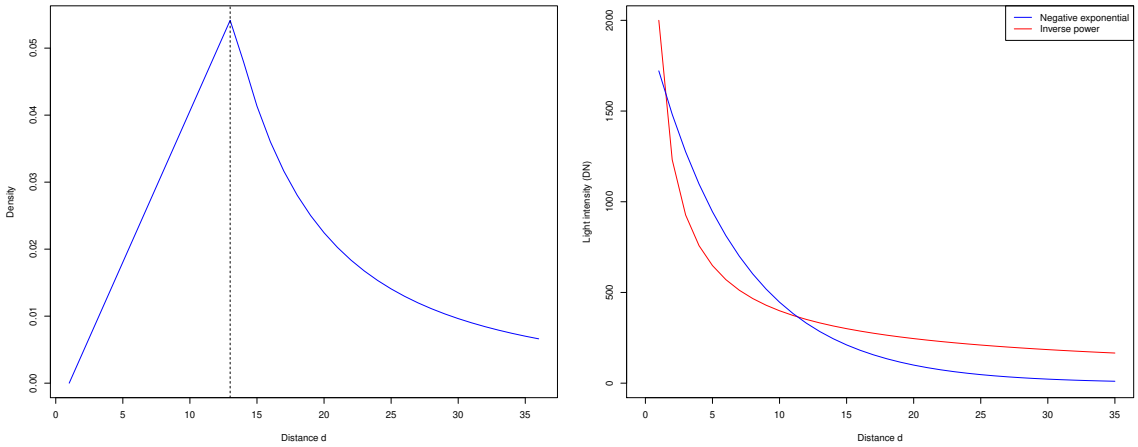
Result 2. *Based on Assumptions 1–3 and 5, as well as a sufficient variation in maximum light intensity across cities, it follows that top lights above a threshold $\tilde{l}$ can be approximated by a Pareto distribution.*

In sum, when lights within cities follow a negative exponential function, a Pareto distribution of top lights does not arise analytically in the three cases considered here. However, depending on the exact assumptions made about differences in the maximum brightness across cities, the resulting distribution is heavy-tailed and can be approximated by a Pareto distribution.

[4]Note that the threshold $\tilde{l}$ depends on the random variable L_0. Hence, we restrict our analysis to the area of the density starting at the highest possible threshold value $\tilde{l}_{max}$, corresponding to $L_0 = L_{max}$.

Figure C-1 – Illustration of distributions

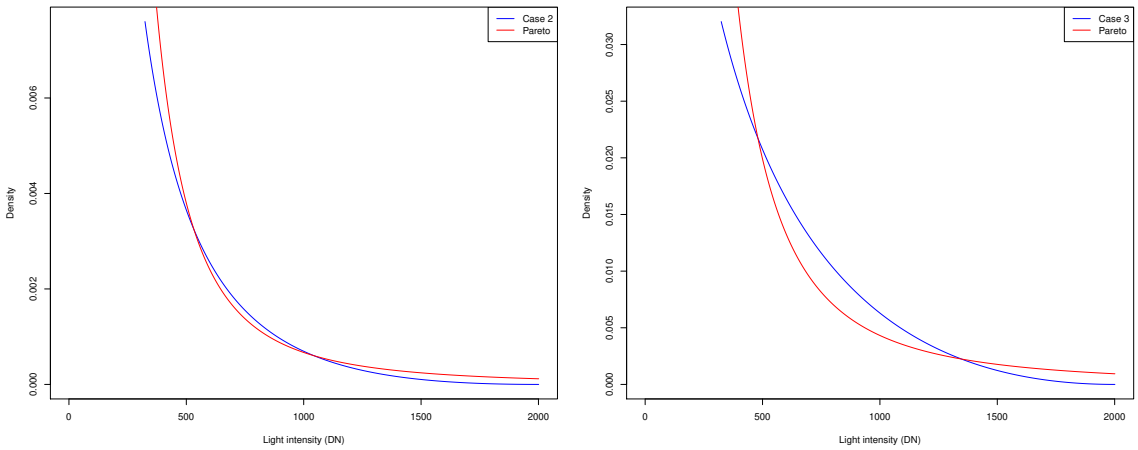
(a) Pixel density **(b)** Light gradient



Notes: The left panel shows the pixel density from eq. (4) with $x_c = 10000$ and $\phi = 1.5$. The right panel shows the negative exponential distribution with $\gamma = 0.15$ as well as the inverse power distribution with $a = 0.7$, both start at $P_0 = 2000$.

Figure C-2 – Approximating the theoretical densities with Pareto distributions

(a) Case 2 and Pareto with $\alpha = 1.5$ **(b)** Case 3 and Pareto with $\alpha = 1.2$



Notes: The left panel shows the pixel density from eq. (C-7) with $x_c = 10000$, $\phi = 1.5$ and $\gamma = 1.5$. The Pareto distribution with $\alpha = 1.5$ is scaled to fit. The right panel shows the pixel density from eq. (C-8) with $x_c = 10000$, $\phi = 1.5$ and $\gamma = 1.5$. The Pareto distribution with $\alpha = 1.2$ is scaled to fit.

D Extreme value theory

As an alternative to our stylized urban economics model, we can also motivate a Pareto distribution in top lights purely on statistical grounds using extreme value theory (EVT). EVT deals with the probability distributions of sparse observations such as threshold exceedances. A key result of this theory is that these quantities observe a Generalized Pareto distribution (Coles, 2001).

More precisely, let $X_1, X_2, \ldots$ be a sequence of independent random variables—such as light—with common but unknown distribution function F, and let $M_n = \max\{X_1, \ldots, X_n\}$. If F satisfies the *extremal types theorem* (Coles, 2001), so that for large n, $\mathbb{P}[M_n > z] \approx G(z)$ with $G(z)$ as the Generalized Extreme Value distribution, then, for a high enough threshold u, the distribution of the threshold exceedance $\mathbb{P}[(X - u) > y | X > u]$ is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}, \tag{D-1}$$

no matter which regular distribution X was drawn from.

This means that we will observe a Generalized Pareto distribution with parameters ξ and σ for all lights values above a specified threshold. With $\xi = 0$, this reduces to the exponential distribution and with $\xi > 0$ the distribution is Pareto. There is strong evidence that the latter case holds for the lights data.

Table D-1 shows the results of fitting the Generalized Pareto distribution to various top shares of the light distribution of the seven radiance-calibrated satellites. The fit is very good and the estimated ξ parameters are always significantly positive. This clearly points towards a Pareto distribution.

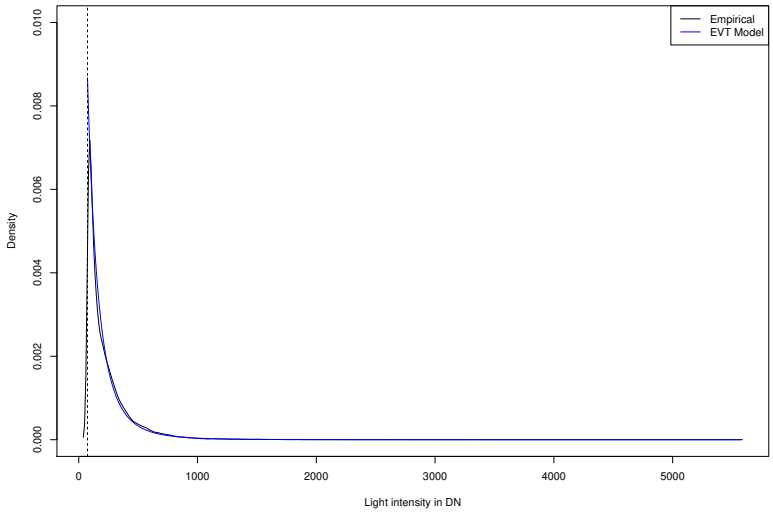
Figure D-1 plots the Generalized Pareto distribution against the empirical distribution function of the radiance-calibrated data from 2010. It visualizes the close fit and confirms the results from the previous regression.

Table D-1 – Fitted Generalized Pareto distributions, varying thresholds

Year	1996	1999	2000	2003	2004	2006	2010	Average
				Panel a) Top 5%				
$\ln \sigma$	4.2422	4.5197	4.7304	4.7560	4.7136	4.5239	4.4477	4.5619
	(0.0062)	(0.0051)	(0.0053)	(0.0052)	(0.0050)	(0.0051)	(0.0049)	[0.1860]
ξ	0.4917	0.3136	0.2792	0.2354	0.2398	0.1972	0.1652	0.2746
	(0.0056)	(0.0043)	(0.0044)	(0.0042)	(0.0040)	(0.0040)	(0.0038)	[0.1075]
Threshold	63	67	74	95	91	75	65	–
Observations	96,685	116,858	106,914	100,094	106,899	99,486	107,745	–
				Panel b) Top 4%				
$\ln \sigma$	4.5136	4.6989	4.9099	4.8569	4.8241	4.6278	4.5574	4.7127
	(0.0066)	(0.0055)	(0.0055)	(0.0057)	(0.0054)	(0.0055)	(0.0052)	[0.1545]
ξ	0.3720	0.2401	0.2020	0.2051	0.2051	0.1605	0.1212	0.2152
	(0.0057)	(0.0044)	(0.0044)	(0.0045)	(0.0042)	(0.0043)	(0.0039)	[0.0790]
Threshold	76	85	96	119	114	94	82	–
Observations	77,348	93,484	85,481	80,075	85,489	79,589	86,195	–
				Panel c) Top 3%				
$\ln \sigma$	4.8260	4.8674	5.0702	4.9841	4.9127	4.7153	4.6387	4.8592
	(0.0069)	(0.0060)	(0.0060)	(0.0063)	(0.0061)	(0.0062)	(0.0058)	[0.1491]
ξ	0.2266	0.1753	0.1387	0.1629	0.1873	0.1356	0.0944	0.1601
	(0.0056)	(0.0047)	(0.0045)	(0.0049)	(0.0047)	(0.0047)	(0.0042)	[0.0424]
Threshold	98	114	131	154	150	123	108	–
Observations	58,010	70,112	64,110	60,057	64,133	59,691	64,646	–
				Panel d) Top 2%				
$\ln \sigma$	5.0520	5.0042	5.1151	5.0807	4.9771	4.7847	4.6836	4.9568
	(0.0079)	(0.0070)	(0.0073)	(0.0076)	(0.0075)	(0.0075)	(0.0070)	[0.1614]
ξ	0.1355	0.1332	0.1438	0.1432	0.1933	0.1266	0.0903	0.1380
	(0.0060)	(0.0053)	(0.0055)	(0.0057)	(0.0058)	(0.0056)	(0.0050)	[0.0304]
Threshold	147	166	199	215	208	169	151	–
Observations	38,673	46,742	42,740	40,039	42,755	39,795	43,097	–
				Panel e) Top 1%				
$\ln \sigma$	5.2035	5.1025	5.2287	5.1729	5.1013	4.8650	4.7009	5.0535
	(0.0109)	(0.0099)	(0.0103)	(0.0108)	(0.0109)	(0.0108)	(0.0100)	[0.1936]
ξ	0.0961	0.1262	0.1374	0.1483	0.2037	0.1324	0.1163	0.1372
	(0.0082)	(0.0074)	(0.0078)	(0.0082)	(0.0086)	(0.0082)	(0.0074)	[0.0337]
Threshold	259	275	319	332	315	255	230	–
Observations	19,337	23,371	21,370	20,019	21,378	19,897	21,548	–

Notes: The table reports parameter estimates from fitted the Generalized Pareto distribution shown in eq. (D-1). The input data are a 10% representative sample of all non-zero lights in the radiance-calibrated data above the defined threshold at the pixel level, where each pixel is 30×30 arc seconds. The last column reports the point average of the seven satellites and its standard deviation in brackets.

Figure D-1 – Generalized Pareto CDF versus EDF, radiance-calibrated data in 2010



Notes: Illustration of Generalized Pareto CDF fitted to the data and the empirical distribution function (EDF). The EDF and Generalized Pareto CDF are fitted to the top 4% of stable lights in 2010. The input data are a 10% representative sample of all non-zero lights of the radiance-calibrated data at the pixel level, where each pixel is 30×30 arc seconds.

E Additional results using the radiance-calibrated data

This section complements the analysis in the paper by proving additional robustness checks of our Pareto hypothesis using the seven radiance-calibrated satellites.

Visual Inspection: Panel (a) of Figure E-1 shows Zipf plots for the top 2% of lights for each of the seven radiance-calibrated satellites. A Zipf plot is a visualization of the Pareto survival function in logs. A linear Zipf plot is usually considered evidence in favor of the Pareto distribution, but its practical relevance is being contested (Cirillo, 2013). Our plots for the lights data are qualitatively similar to those of the top incomes literature, in that they display linear sections together with some initial curvature and outliers at the end.[5] It is well-known that Zipf plots often deviate from linearity at the very top since fewer and fewer values are observed at the extremes. Sometimes this is addressed by removing the very top. We use logarithmic bins so that the size of the bins increases by a multiplicative factor (Newman, 2005). The sensitivity of Zipf plots to outliers is compounded by instability and measurement errors afflicting the radiance-calibrated satellites. While we conclude that the Zipf plot using the radiance-calibrated data is ambiguous, we obtain a near-linear Zipf plot using the superior VIIRS data (see the next section).

Panel (b) of Figure E-1 provides another graphical test for the Pareto distribution based on 'Van der Wijk's Law'. The Pareto distribution is unique in that the average above some level y is proportional to y at all points in the tail, with a factor of proportionality equal to $\frac{\alpha}{\alpha-1} > 1$. The graph plots, for each DN on the x-axis, the average luminosity of all pixels brighter than this value on the y-axis. As expected, we observe a linear relationship with a slope above unity.

Tests against the lognormal distribution: As a robustness check, we pit the Pareto distribution against other plausible candidates. We pay particular attention to the lognormal distribution, since it is commonly used to describe the complete distribution of incomes or city sizes.

Table E-1 shows the results from separate regressions of the empirical distribution function on the Pareto CDF and the lognormal CDF based on the top 4% of the data. The estimated coefficient for the Pareto CDF is closer to unity and the R^2 is substantially larger than in the lognormal counterpart (0.98 vs. 0.83).

[5] Working with any top share, from the top 5% to the top 1% gives qualitatively similar results, even if the case for a Pareto distribution tends to be stronger the higher we set the threshold. This is in line with the empirical literature on Pareto applications in other fields.

Figure E-2 visualizes this difference in fit for the year 2010. The lognormal CDF fits the data poorly, while the Pareto CDF is always closer to the empirical distribution.

Unrestricted rank regressions: Recall that for Pareto-distributed observations y_i, $i = 1, ...N$, we have $\text{rank}(y_i) \approx N y_c^{\alpha} y_i^{-\alpha}$, or, in logarithms $\log \text{rank}(y_i) - \log N \approx \alpha \log y_c - \alpha \log y_i$. Hence, in the regression

$$\log\Big(\text{rank}(y_i) - \frac{1}{2}\Big) - \log N = \alpha_1 \log y_c + \alpha_2 \log y_i + \epsilon \tag{E-1}$$

only the Pareto distribution satisfies the null hypothesis that $-\alpha_1 = \alpha_2$ with $\alpha_2 < 0$. As before, we follow Gabaix and Ibragimov (2011) and subtract one half from the rank to improve the OLS estimation of the tail exponent in the rank regression.

Table E-2 reports the OLS rank regression results of eq. (E-1) for all seven satellites at various different thresholds, i.e. the top 5% to top 1%. The two coefficients are usually very close and the R^2s are high (0.96–0.99).[6]

The Hill estimator: If the null hypothesis $-\alpha_1 = \alpha_2 = \alpha$ is enforced in eq. (E-1), one can directly obtain the parameter estimate for the Pareto α. In the main text we estimate this parameter using OLS rank regressions. As a robustness check, we now use the Hill estimator (Hill, 1975), $\hat{\alpha}_{Hill} = (N-1)\left(\sum_{i=1}^{N-1} \log y_i - \log y_c\right)^{-1}$, for the restricted rank regression

$$\log \text{rank}(y_i) - \log N \approx \alpha \log y_c - \alpha \log y_i. \tag{E-2}$$

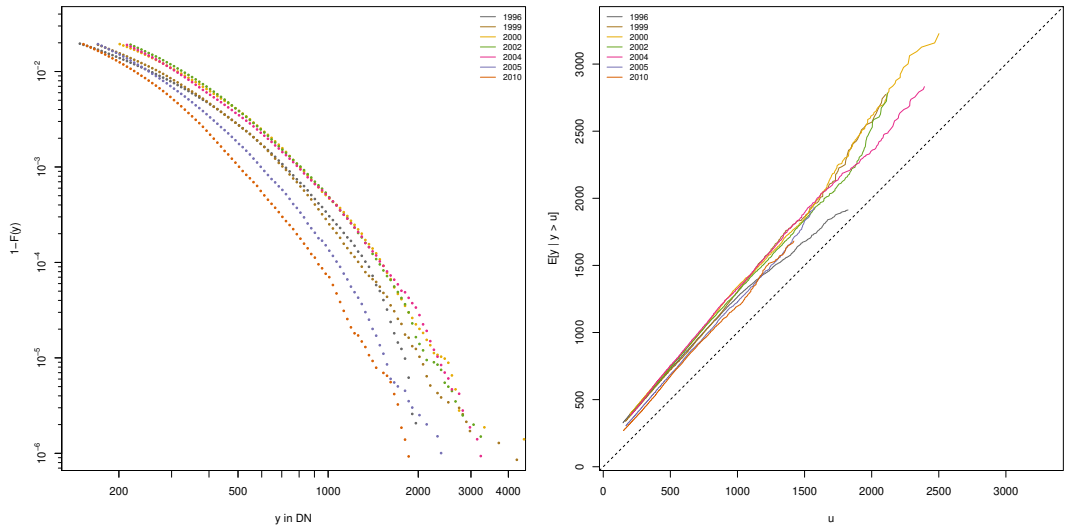
Under the assumption of a Pareto distribution, the Hill estimator equals the efficient maximum likelihood estimator and is known for its superior properties for fitting the tail of the Pareto distribution (Soo, 2005, Eeckhout, 2009). The standard errors are given by $\hat{\alpha}_{Hill}/\sqrt{N-3}$ (see Gabaix, 2009).

Table E-3 report the results for all seven satellites at various different thresholds, i.e. the top 5% to top 1%. The Pareto parameters obtained using the Hill estimator are very similar to the OLS estimates in the main text. For the top 3-4%, the values are between 1.3 and 1.6 for the seven satellites, very close to the OLS average parameter estimate of 1.5. For higher thresholds, we observe also the same increase in the parameter estimate that we observe in the OLS results.

[6]Note that formal statistical tests, e.g. tests of coefficient equality or Kolmogorov-Smirnoff tests, do not make much sense in huge samples such as ours. Gabaix and Ioannides (2004, p. 2350) capture this nicely: "with an infinitely large dataset one can reject any non-tautological theory." The extremely small standard errors lead to overrejections of the null hypothesis unless the empirical value equals exactly the theoretical value.

Figure E-1 – Zipf plot and Van der Wijk's plot

(a) Zipf **(b)** Van der Wijk



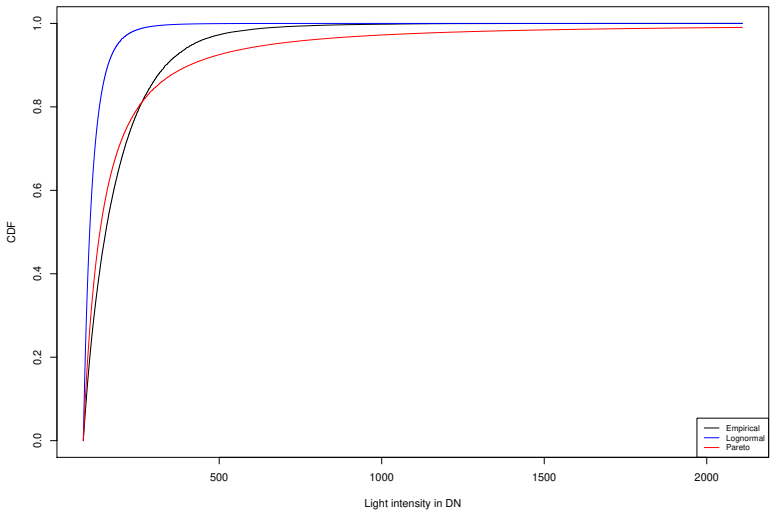
Notes: Popular graphical tests for an approximate Pareto distribution in top lights. Panel (a) shows the Zipf plot for the top 2% of all pixels. The figure uses logarithmic binning to reduce noise and sampling errors in the right tail of the distribution (see Newman, 2005). There are about 100 bins in the tail, where the exact number depends on the range of the input data. Panel (b) demonstrates Van der Wijk's law, which states that the average light above some value u is proportional to u, this is $\mathrm{E}[y|y > u] \propto u$. Here, too, the data is the top 2% of all pixels. The input data are a 10% representative sample of all non-zero lights in the radiance-calibrated data at the pixel level.

Table E-1 – Regression of the EDF on theoretical CDFs, top 4%

Year	1996	1999	2000	2003	2004	2006	2010	Average
	Panel a) Pareto CDF on RHS							
Slope	1.0108	1.0551	1.0616	1.0575	1.0746	1.0722	1.0796	1.0588
	(0.0003)	(0.0004)	(0.0005)	(0.0004)	(0.0004)	(0.0005)	(0.0005)	[0.0231]
Constant	-0.0320	-0.0668	-0.0746	-0.0666	-0.0787	-0.0784	-0.0858	-0.0690
	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	[0.0177]
R^2	0.9914	0.9866	0.9802	0.9884	0.9869	0.9860	0.9831	–
	Panel b) Lognormal CDF on RHS							
Slope	0.9004	0.9265	0.9181	0.9387	0.9520	0.9472	0.9488	0.9331
	(0.0014)	(0.0016)	(0.0018)	(0.0014)	(0.0014)	(0.0014)	(0.0015)	[0.0190]
Constant	-0.1653	-0.2186	-0.2238	-0.1954	-0.2088	-0.2031	-0.2179	-0.2047
	(0.0011)	(0.0013)	(0.0015)	(0.0011)	(0.0011)	(0.0011)	(0.0012)	[0.0200]
R^2	0.8496	0.7913	0.7626	0.8508	0.8467	0.8480	0.8268	–

Notes: The table reports results of a regression of the empirical distribution function (EDF) on the Pareto or lognormal CDF, using the top 4% of the data. The data are a 10% representative sample of all non-zero lights in the radiance-calibrated data at the pixel level, where each pixel is 30×30 arc seconds. The last column reports the point average of the seven satellites and its standard deviation in brackets.

Figure E-2 – Pareto and lognormal CDF versus EDF, radiance-calibrated lights in 2010



Notes: Illustration of the difference between the Pareto and lognormal CDFs fitted to the data and the empirical distribution function (EDF). Note that the lognormal distribution was fitted to the whole distribution rather than the tail because of its unimodal shape, while the Pareto distribution is estimated only on the tail. For comparison, we adjust the CDFs so that they all start at the top 4% of radiance-calibrated lights in 2010. The input data are a 10% representative sample of all non-zero lights in the radiance-calibrated data at the pixel level, where each pixel is 30×30 arc seconds.

Table E-2 – Unrestricted rank regressions

Year	1996	1999	2000	2003	2004	2006	2010	Average
				Panel a) Top 5%				
y_i	-1.4334	-1.4996	-1.4630	-1.6903	-1.6933	-1.7388	-1.7170	-1.6050
	(0.0012)	(0.0012)	(0.0013)	(0.0013)	(0.0013)	(0.0014)	(0.0015)	[0.1330]
y_c	1.4736	1.5632	1.5318	1.7539	1.7582	1.8087	1.7936	1.6690
	(0.0014)	(0.0015)	(0.0016)	(0.0016)	(0.0015)	(0.0017)	(0.0018)	[0.1405]
R^2	0.9694	0.9651	0.9600	0.9701	0.9721	0.9677	0.9620	0.9666
Observations	96,685	116,858	106,914	100,095	106,899	99,487	107,745	104,955
				Panel b) Top 4%				
y_i	-1.5130	-1.6328	-1.6165	-1.8403	-1.8513	-1.9101	-1.9056	-1.7528
	(0.0015)	(0.0014)	(0.0016)	(0.0016)	(0.0014)	(0.0017)	(0.0018)	[0.1612]
y_c	1.5618	1.6974	1.6870	1.8978	1.9132	1.9767	1.9804	1.8163
	(0.0017)	(0.0017)	(0.0019)	(0.0018)	(0.0016)	(0.0019)	(0.0020)	[0.1655]
R^2	0.9662	0.9661	0.9623	0.9725	0.9759	0.9711	0.9658	0.9685
Observations	77,348	93,484	85,482	80,075	85,489	79,590	86,196	83,952
				Panel c) Top 3%				
y_i	-1.6609	-1.8385	-1.8624	-2.0491	-2.0633	-2.1470	-2.1746	-1.9708
	(0.0019)	(0.0018)	(0.0019)	(0.0019)	(0.0016)	(0.0020)	(0.0021)	[0.1882]
y_c	1.7225	1.9017	1.9340	2.1044	2.1174	2.2068	2.2429	2.0328
	(0.0022)	(0.0020)	(0.0022)	(0.0021)	(0.0018)	(0.0022)	(0.0024)	[0.1872]
R^2	0.9646	0.9695	0.9691	0.9761	0.9811	0.9762	0.9721	0.9727
Observations	58,011	70,115	64,111	60,058	64,134	59,692	64,647	62,967
				Panel d) Top 2%				
y_i	-1.9711	-2.1628	-2.2315	-2.3687	-2.3478	-2.4809	-2.5663	-2.3042
	(0.0025)	(0.0023)	(0.0022)	(0.0023)	(0.0018)	(0.0024)	(0.0025)	[0.2009]
y_c	2.0329	2.2180	2.2831	2.4156	2.3880	2.5295	2.6215	2.3555
	(0.0029)	(0.0025)	(0.0025)	(0.0025)	(0.0020)	(0.0026)	(0.0027)	[0.1974]
R^2	0.9698	0.9757	0.9798	0.9825	0.9871	0.9826	0.9807	0.9797
Observations	38,673	46,742	42,740	40,039	42,756	39,794	43,097	41,977
				Panel e) Top 1%				
y_i	-2.5471	-2.7216	-2.7241	-2.8508	-2.7006	-2.9769	-3.1652	-2.8123
	(0.0039)	(0.0031)	(0.0031)	(0.0030)	(0.0027)	(0.0032)	(0.0029)	[0.2049]
y_c	2.5922	2.7593	2.7596	2.8823	2.7258	3.0097	3.2005	2.8471
	(0.0043)	(0.0034)	(0.0034)	(0.0033)	(0.0029)	(0.0035)	(0.0031)	[0.2031]
R^2	0.9781	0.9849	0.9864	0.9889	0.9895	0.9886	0.9911	0.9868
Observations	19,337	23,373	21,372	20,020	21,377	19,898	21,551	20,990

Notes: The table reports OLS results obtained from the unrestricted rank regressions eq. (E-1) at various relative thresholds. The input data are a 10% representative sample of all non-zero lights in the radiance-calibrated data above the defined threshold at the pixel level, where each pixel is 30×30 arc seconds. Standard errors are in parentheses. The last column reports the point average of the seven satellites and its standard deviation in brackets.

Table E-3 – Parameter estimates from rank regressions (Hill estimator)

Year	1996	1999	2000	2003	2004	2006	2010	Average
				Panel a) Top 5%				
Pareto $\hat{\alpha}$	1.2286	1.1833	1.1289	1.3112	1.3100	1.3356	1.3012	1.2570
	(0.0040)	(0.0035)	(0.0035)	(0.0041)	(0.0040)	(0.0042)	(0.0040)	[0.0780]
Observations	96,685	116,858	106,914	100,095	106,899	99,487	107,745	–
				Panel b) Top 4%				
Pareto $\hat{\alpha}$	1.2487	1.2689	1.2233	1.4431	1.4315	1.4666	1.4333	1.3593
	(0.0045)	(0.0042)	(0.0042)	(0.0051)	(0.0049)	(0.0052)	(0.0049)	[0.1065]
Observations	77,348	93,484	85,482	80,075	85,489	79,590	86,196	–
				Panel c) Top 3%				
Pareto $\hat{\alpha}$	1.2948	1.4152	1.3805	1.6023	1.6234	1.6672	1.6478	1.5188
	(0.0054)	(0.0053)	(0.0055)	(0.0065)	(0.0064)	(0.0068)	(0.0065)	[0.1509]
Observations	58,011	70,115	64,111	60,058	64,134	59,692	64,647	–
				Panel d) Top 2%				
Pareto $\hat{\alpha}$	1.5068	1.6869	1.7536	1.8920	1.9325	1.9860	2.0095	1.8239
	(0.0077)	(0.0078)	(0.0085)	(0.0095)	(0.0093)	(0.0100)	(0.0097)	[0.1832]
Observations	38,673	46,742	42,740	40,039	42,756	39,794	43,097	–
				Panel e) Top 1%				
Pareto $\hat{\alpha}$	2.0363	2.2458	2.2613	2.4101	2.3582	2.5190	2.6558	2.3552
	(0.0146)	(0.0147)	(0.0155)	(0.0170)	(0.0161)	(0.0179)	(0.0181)	[0.2011]
Observations	19,337	23,373	21,372	20,020	21,377	19,898	21,551	–

Notes: The table reports the results of the restricted rank regression eq. (E-2) using the Hill estimator. The data are a 10% representative sample of all non-zero lights in the radiance-calibrated data at the pixel level, where each pixel is 30×30 arc seconds. The last column reports the point average of the seven satellites and its standard deviation in brackets.

F Additional results using the VIIRS data

Since October 2011, the first satellite of the Suomi National Polar Partnership Visible Infrared Imagining Radiometer Suite (NPP-VIIRS) has been in orbit. The VIIRS day-night-band (DNB) on-board sensors have a much higher native resolution of 15 arc seconds, are radiometrically calibrated, do not suffer from top-coding, and record a physical quantity (radiance). This section complements the analysis in the paper by proving additional robustness checks of our Pareto hypothesis using this new data.

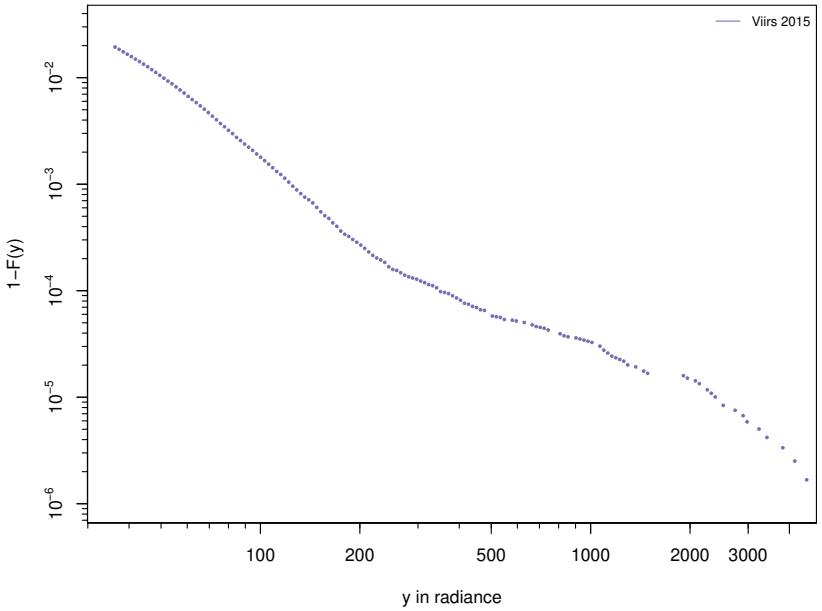
Although the new system is undoubtedly superior in many respects, comparability with the previous series is limited for at least two reasons: *i)* the first annual VIIRS composite made available by NOAA refers to the year 2015, so that there is no temporal overlap with the 1992-2013 DMSP-OLS series, *ii)* the VIIRS satellites have an overpass time around midnight, in contrast to the evening hours of the DMSP-OLS satellites, so that it is not entirely clear what kind of production and consumption activity they capture (Elvidge et al., 2014, Nordhaus and Chen, 2015). While we do not rely on the VIIRS data for our replacement procedure, we use the first VIIRS cross-section from 2015 as another robustness check for whether the Pareto distribution holds. The VIIRS data are particularly insightful in this respect because of their superior quality.

To compare the higher resolution VIIRS image to the DMSP data, we resample the raster to the DMSP resolution and then extract radiances of each pixel at the locations of the 10% sample that we have been using thus far. Naturally, there are considerable differences in the scale since the VIIRS-DNB records radiance. Note that radiance is measured in nano watt per steradian per square centimeter ($10^{-9} W cm^{-2} sr^{-1}$). The difference in scale is reflected in the summary statistics of the VIIRS data. The mean is 3.98, the standard deviation is 18.65, and the maximum is 6567.42. The spatial Gini is much higher using the VIIRS data than in the radiance-calibrated data (0.79 vs. 0.60-0.65) which is owed to their improved sensors and finer resolution. Nevertheless, the top tail of the light distribution essentially exhibits the same properties.

Figure F-1 shows the Zipf plot for the VIIRS data. The shape is nearly linear, even high up in the tail and displays less curvature than the corresponding plot for the radiance-calibrated data. This also suggests that the radiance-calibration process introduces noise and understates the Paretian nature of night lights.

Table F-1 replicates the results of the rank regressions from the previous section using the VIIRS data. The results are qualitatively similar to those obtained with the radiance-calibrated data, but some small differences are notable. In particular, the estimated shape parameters are a bit higher for top shares around 3% to 5% but then also appear to be more stable in the upper tail. Since the VIIRS data are from five years after the most recent radiance-calibrated image and have a different overpass time, it is difficult to identify the source of these discrepancies.

Figure F-1 – Zipf plot using the top 2% of pixels in the VIIRS data



Notes: The figure shows a Zipf plot for the top 2% of all pixels of the VIIRs data, after resampling the data to the DMSP-OLS grid and resolution. The figure uses logarithmic binning to reduce noise and sampling errors in the right tail of the distribution (see Newman, 2005). There are about 140 bins in the tail, where the exact number depends on the range of the input data. The VIIRs pixels correspond to the same 10% representative sample of all non-zero lights in the radiance-calibrated data at the pixel level obtained from Hsu et al. (2015) and used in the rest of the paper.

Table F-1 – Rank regressions based on the VIIRS data in 2015

Unrestricted regressions			Hill estimates		Restricted regressions
		Panel a) Top 5%			
y_i	-1.9603	α	1.5876	α	1.7331
	(0.0013)		(0.0065)		(0.0100)
y_c	2.0405				
	(0.0015)				
R^2	0.9879				
Observations	59,633		59,633		59,633
		Panel b) Top 4%			
y_i	-2.0831	α	1.7331	α	1.8747
	(0.0013)		(0.0079)		(0.0121)
y_c	2.1479				
	(0.0015)				
R^2	0.9909				
Observations	47,705		47,705		47,705
		Panel c) Top 3%			
y_i	-2.2150	α	1.9144	α	2.0419
	(0.0014)		(0.0101)		(0.0153)
y_c	2.2622				
	0.0016				
R^2	0.9933				
Observations	35,780		35,780		35,780
		Panel b) Top 2%			
y_i	-2.3438	α	2.1671	α	2.2481
	(0.0017)		(0.0140)		(0.0206)
y_c	2.3665				
	(0.0019)				
R^2	0.9939				
Observations	23,854		23,854		23,854
		Panel e) Top 1%			
y_i	-2.3778	α	2.4716	α	2.4235
	(0.0033)		(0.0226)		(0.0314)
y_c	2.3682				
	(0.0036)				
R^2	0.9888				
Observations	11,927		11,927		11927

Notes: The table uses the VIIRS data to repeat three regressions which were conducted with the radiance-calibrated data before: the unrestricted OLS rank regression eq. (E-1) and the restricted regression eq. (E-2) using both the OLS and the Hill estimator. Standard errors are reported in paretheses. For the OLS restricted rank regression, these are the asymptotic standard errors computed as $(2/N)^{1/2}$. The data are a 10% representative sample of all non-zero lights in the radiance-calibrated data at the pixel level, where each pixel is 30×30 arc seconds.

G An analytical top-coding correction

Researchers are often interested in aggregate measures, such as average luminosity or light inequality in a region or a country. Here we present simple formulas to correct these summary statistics for top-coding. These corrections work with arbitrary thresholds and Pareto shape parameters.

Mean luminosity: The top-coding corrected mean luminosity μ of a country or region is simply the weighted average of the bottom and top means μ_B and μ_T. If the latter is the mean of a Pareto distribution starting at y_c, we have

$$\mu = \omega_B \mu_B + (1-\omega_B)\mu_T = \omega_B \mu_B + (1-\omega_B)\frac{\alpha}{\alpha-1} y_c \tag{G-1}$$

where ω_B and $\omega_T = 1-\omega_B$ are the shares of pixels below and above the threshold. A simple numerical illustration shows how correcting for top-coding drives up the mean luminosity. If top-coding starts at $y_c = 55$, affects 5% of the study area of interest, α is 1.5 and mean luminosity in the non-top-coded pixels is $\mu_B = 10$, then the corrected mean luminosity is 17.75 rather than 12.25.

Spatial Gini coefficients: The overall Gini coefficient can be written as the weighted sum of the bottom-share and top-share Ginis (i.e., the within-group Gini) as well as the difference between the top share of total lights minus the top share of pixels (i.e., the between-group Gini), such that

$$G = \omega_B \phi_B G_B + \omega_T \phi_T G_T + [\phi_T - \omega_T], \tag{G-2}$$

where the shares of all light accruing to the top and bottom groups are $\phi_B = \omega_B \mu_B / \mu$ and $\phi_T = \omega_T \mu_T / \mu$, and $G_T = 1/(2\alpha - 1)$. A greater share of top-coded pixels ω_T, brighter top-coded pixels ϕ_T, and a greater spread in the distribution of the top-coded data G_T all increase the size of the correction.

The above decomposition of the Gini coefficient can be derived by defining the Gini coefficient over multiple groups as in Mookherjee and Shorrocks (1982)

$$G = \frac{1}{2N^2\mu} \sum_i \sum_j |y_i - y_j| \tag{G-3}$$

$$= \frac{1}{2N^2\mu} \sum_k \left(\sum_{i \in N_k} \sum_{j \in N_k} |y_i - y_j| + \sum_{i \in N_k} \sum_{j \notin N_k} |y_i - y_j| \right) \tag{G-4}$$

$$= \sum_k \left(\frac{N_k}{N} \right)^2 \frac{\mu_k}{\mu} G_k + \frac{1}{2N^2\mu} \sum_k \sum_{i \in N_k} \sum_{j \notin N_k} |y_i - y_j| \,. \tag{G-5}$$

where G_K is the within group Gini coefficient of group k. The second term is a measure of group overlap including their between group differences.

Perfect separation (no overlap between groups) implies $\sum_{i \in N_k} \sum_{j \in N_h} |y_i - y_j| = N_k N_h |\mu_k - \mu_h|$. Hence, we can simplify equation eq. (G-5) to

$$G = \sum_k \left(\frac{N_k}{N}\right)^2 \frac{\mu_k}{\mu} G_k + \sum_k \sum_h \frac{N_k N_h}{2N^2 \mu} |\mu_k - \mu_h| \,. \tag{G-6}$$

With two bottom and top groups $k, h \in \{B, T\}$ (where $\mu_T > \mu_B$) and some algebra, this becomes

$$G = \left(\frac{N_B}{N}\right)^2 \frac{\mu_B}{\mu} G_B + \left(\frac{N_T}{N}\right)^2 \frac{\mu_T}{\mu} G_T + \left[\left(\frac{N_T}{N}\right)^2 \frac{\mu_T}{\mu} - \frac{N_T}{N}\right]. \tag{G-7}$$

Now define the pixel shares below and above the threshold as ω_B and ω_T, where $\omega_T = 1 - \omega_B$ and the group's share of all income (light) as $\phi_B = \omega_B \frac{\mu_B}{\mu}$ and $\phi_T = \omega_T \frac{\mu_T}{\mu}$ to obtain eq. (G-2) above.

H Characteristics of the corrected data

In this section we compare the back-on-the-envelop analytical corrections from the previous section with our corrected data at the pixel level, examine the correlations between our corrected data and the radiance calibrated data, and discuss the size of the top-coding correction around the world.

Table H-1 reports the mean luminosities and global Gini coefficients before and after the correction for each satellite, using both the analytic formulas and the data corrected at the pixel level. Our geo-referenced pixel-level replacement comes close to the analytic solutions but is generally more conservative (due to the fixed upper bound). Mean luminosity increases on average from 12.7 DN to 15.3–16.6 DN and inequality in lights from 0.47 to 0.56–0.59.

Table H-2 reports mean luminosity and the Gini coefficient of inequality for 2010, using a wider range of parameters as robustness checks in the analytical correction eq. (G-2). Working with a smaller (larger) parameter than our benchmark $\alpha = 1.5$ implied more (less) inequality in the tail of the light distribution. The corrections are consequently larger (smaller). We can also see that parameter values of 1.4–1.6 only lead to very small differences in the magnitude of the correction. Also, using a higher α does not change the magnitude of the correction as much as using a smaller α, as the comparison of the extreme values of 1.2 and 1.8 shows.

Figure H-1 plots the time series graph of global inequality in lights from 1992 to 2013, both before and after the top-coding correction based on eq. (G-2). Parameter values of 1.4 and 1.6 serve as comparison bands for the benchmark case of 1.5. The global distribution of lights became slightly more unequal over the 1990s, remained flat in the first decade of the new millennium and then became temporarily more equal in the aftermath of the global financial crises and great recession. However, this year-to-year variation is completely swamped by the size of the top-coding correction.

Table H-3 provides another comparison check for our pixel-level corrected data. It shows the correlations between the corrected lights and the radiance-calibrated data for the seven years where both are available. These figures refer to the whole distribution, not just the top. Remember that in the correction procedure, we only rely on the ranks for the top, not the precise radiance-calibrated values, and infer values from the Pareto distribution. It is all the more remarkable that our corrected values turn out to be strongly correlated to the radiance-calibrated values, with correlations of 0.94–0.96 for all seven available years.

The global summary statistics presented so far hide a lot of between-country heterogeneity. Figure H-2 illustrates the size of the correction in different countries with various scatter plots. As expected, the same characteristics that drive the number of top-coded pixels (see Section 2 in the paper) turn out to be predictive of the size of

the correction in terms of country-wide mean luminosity and inequality in light. The correction is strongly increasing in GDP per capita, weakly in country size and moderately in population density. Numerous developing countries experience sizable corrections (such as Egypt, Paraguay or Mexico). City states, such as Singapore, have large top-coding corrections, as do smaller countries, like Israel and Estonia. Nevertheless, even large countries like the US experience a sizable increase in both mean luminosity (plus 7 DN) and the Gini coefficient (plus 14 percentage points). No single factor captures all the relevant heterogeneity.

Table H-1 – Satellite level statistics of the top-coding correction

Satellite	Top share (pixels)	*Top share (light)* Unadj	Adj	*Mean luminosity* Unadj	Form Adj	Pixel Adj	*Gini coefficient* Unadj	Form Adj	Pixel Adj
F101992	0.0381	0.1663	0.3096	13.83	17.81	16.70	0.4390	0.5626	0.5334
F101993	0.0312	0.1568	0.2938	11.96	15.23	14.28	0.4593	0.5737	0.5456
F101994	0.0349	0.1754	0.3207	12.02	15.67	14.60	0.4783	0.5980	0.5684
F121994	0.0420	0.1733	0.3176	14.65	19.04	17.74	0.4353	0.5634	0.5316
F121995	0.0376	0.1733	0.3174	13.09	17.02	15.85	0.4580	0.5813	0.5505
F121996	0.0351	0.1670	0.3068	12.69	16.37	15.25	0.4607	0.5801	0.5494
F121997	0.0394	0.1766	0.3210	13.45	17.57	16.31	0.4540	0.5799	0.5477
F121998	0.0418	0.1816	0.3276	13.89	18.25	16.90	0.4474	0.5774	0.5436
F121999	0.0467	0.1915	0.3412	14.74	19.62	18.08	0.4447	0.5802	0.5449
F141997	0.0316	0.1739	0.3169	10.98	14.29	13.28	0.4876	0.6047	0.5747
F141998	0.0305	0.1680	0.3067	10.94	14.13	13.13	0.4883	0.6023	0.5720
F141999	0.0278	0.1648	0.3011	10.15	13.06	12.13	0.4895	0.6017	0.5714
F142000	0.0318	0.1689	0.3062	11.34	14.67	13.59	0.4852	0.6003	0.5687
F142001	0.0350	0.1817	0.3276	11.64	15.30	14.16	0.4856	0.6069	0.5754
F142002	0.0377	0.1872	0.3375	12.14	16.08	14.90	0.4896	0.6126	0.5818
F142003	0.0382	0.1930	0.3409	11.96	15.96	14.65	0.4928	0.6177	0.5836
F152000	0.0370	0.1685	0.3063	13.25	17.13	15.89	0.4399	0.5647	0.5308
F152001	0.0354	0.1645	0.3011	12.93	16.64	15.46	0.4463	0.5679	0.5351
F152002	0.0372	0.1700	0.3085	13.18	17.08	15.82	0.4465	0.5710	0.5370
F152003	0.0270	0.1582	0.2894	10.28	13.11	12.17	0.4982	0.6055	0.5751
F152004	0.0276	0.1642	0.2979	10.08	12.97	12.00	0.5080	0.6163	0.5853
F152005	0.0279	0.1604	0.2953	10.44	13.36	12.43	0.5115	0.6171	0.5886
F152006	0.0293	0.1666	0.2988	10.56	13.63	12.55	0.5135	0.6217	0.5892
F152007	0.0279	0.1547	0.2844	10.74	13.68	12.69	0.5049	0.6099	0.5795
F162004	0.0340	0.1734	0.3129	11.82	15.38	14.23	0.4641	0.5863	0.5528
F162005	0.0285	0.1642	0.2993	10.44	13.43	12.46	0.4926	0.6040	0.5732
F162006	0.0348	0.1707	0.3041	12.26	15.91	14.61	0.4714	0.5908	0.5546
F162007	0.0403	0.1861	0.3302	13.05	17.28	15.86	0.4624	0.5916	0.5554
F162008	0.0395	0.1832	0.3285	12.97	17.11	15.78	0.4702	0.5961	0.5622
F162009	0.0417	0.1862	0.3356	13.50	17.87	16.54	0.4694	0.5966	0.5644
F182010	0.0591	0.2033	0.3614	17.55	23.73	21.89	0.4258	0.5720	0.5361
F182011	0.0494	0.2020	0.3595	14.78	19.95	18.41	0.4552	0.5936	0.5598
F182012	0.0576	0.2118	0.3734	16.44	22.45	20.68	0.4361	0.5838	0.5481
F182013	0.0578	0.2151	0.3800	16.23	22.28	20.55	0.4389	0.5876	0.5530
Average	**0.0374**	**0.1765**	**0.3194**	**12.65**	**16.56**	**15.34**	**0.4691**	**0.5917**	**0.5595**

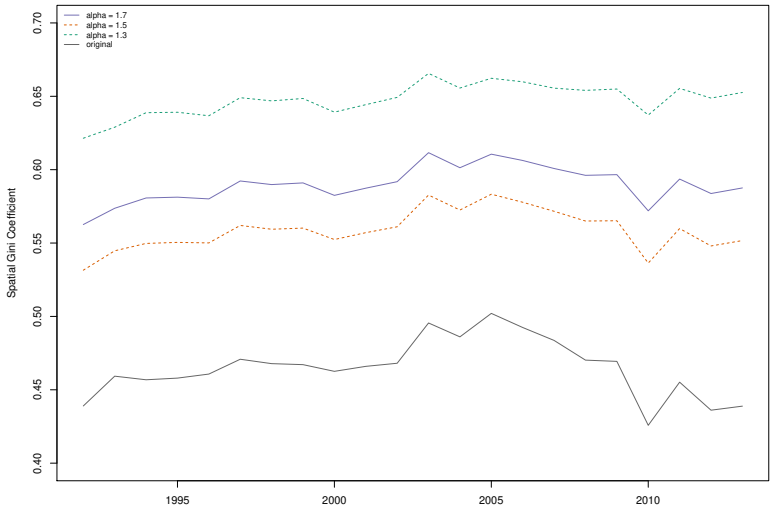
Notes: The table reports summary statistics of the global lights data before the top-coding correction and after the analytical, formula-based correction at the aggregate level (eq. (G-1) and eq. (G-2)) as well as the pixel-level correction from the paper. Column 1 reports the share of pixels above 55 DN, Column 2 and 3 the share of lights emitted by these top pixels respectively in the unadjusted and adjusted data set. Columns 4-6 and 7-9 report the mean luminosity and Gini coefficient, respectively for the unadjusted data, the analytical, formula-based correction and the pixel-level corrected data. All corrections use $\alpha = 1.5$ and $y_c = 55$ for the Pareto tail.

Table H-2 – Correction of global mean and Gini coefficient in 2010, different parameters

	Unadj.	*Pareto parameter* $\alpha =$						
		1.2	1.3	1.4	**1.5**	1.6	1.7	1.8
Mean luminosity	17.55	33.49	28.07	25.36	**23.73**	22.65	21.88	21.30
Spatial Gini	0.4258	0.6954	0.6372	0.5990	**0.5720**	0.5519	0.5363	0.5240

Notes: The table computes the top-coding corrected mean and Gini coefficient of global inequality in lights for the year 2010 with different α parameters based on eq. (G-1) and eq. (G-2) with $y_c = 55$. The input data are a representative 10% sample of non-zero lights from satellite F182010.

Figure H-1 – Global Gini coefficient in lights before and after the correction



Notes: Illustration of the global top-coding correction. The figure shows global inequality in lights calculated by eq. (G-2) using the specified Pareto shape parameters. The input data are a representative 10% sample of non-zero lights. For years when more than two satellites flew concurrently, the values were averaged.

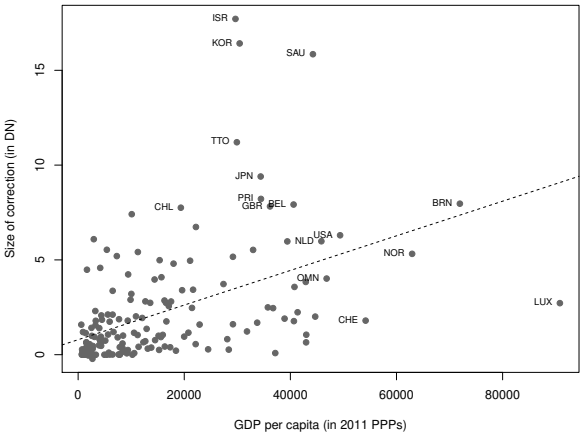
Table H-3 – Correlations between corrected data and radiance-calibrated data

	Radiance-calibrated						
Corrected	1996	1999	2000	2003	2004	2006	2010
1996	0.9546						
1999		0.9543					
2000			0.9587				
2003				0.9454			
2004					0.9592		
2006						0.9614	
2010							0.9530

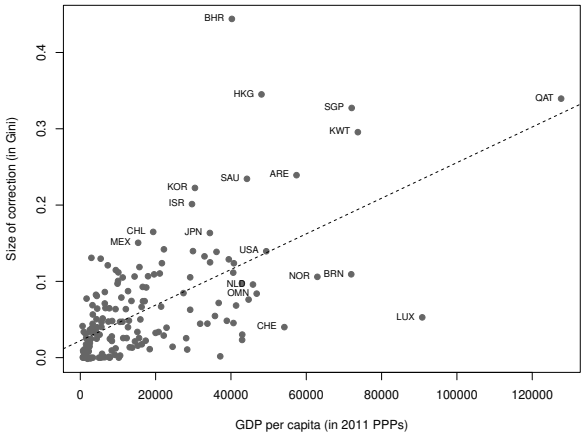
Notes: The table reports correlations between the corrected lights and the radiance calibrated lights for the seven years where the radiance-calibrated data are available. The corrections are based on satellites F121996, F141999, F152000, F152003, F162004, F162006 and F182010.

Figure H-2 – Size of the correction and country characteristics

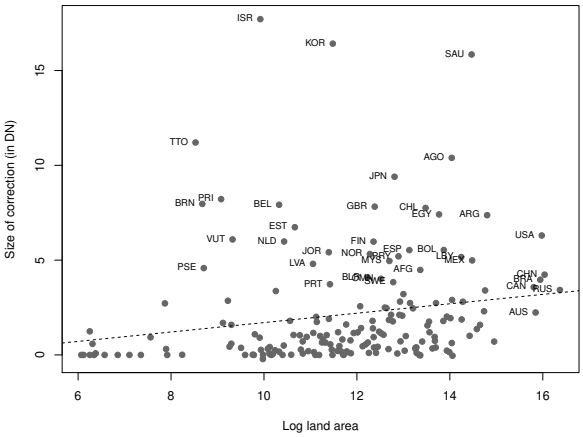
(a) Mean correction vs. GDP per capita



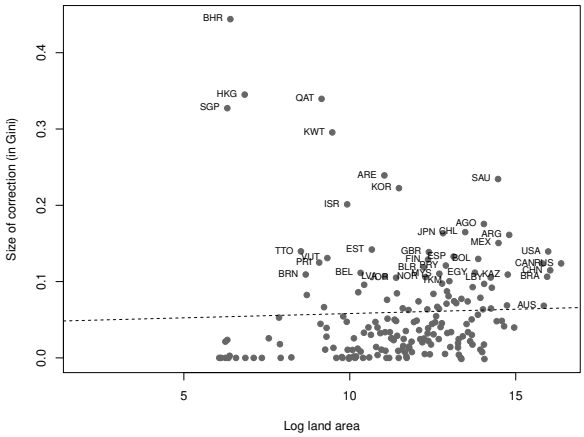
(b) Gini correction vs. GDP per capita



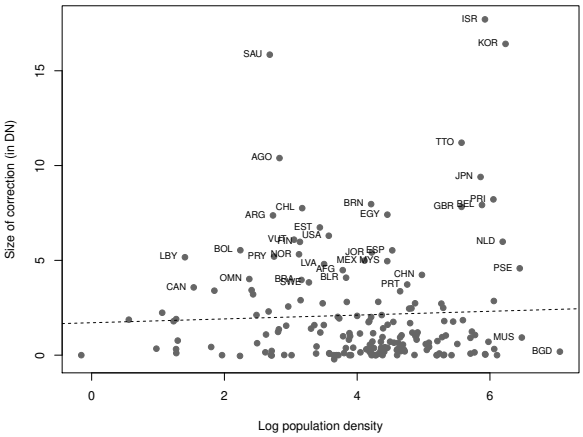
(c) Mean correction vs. log area



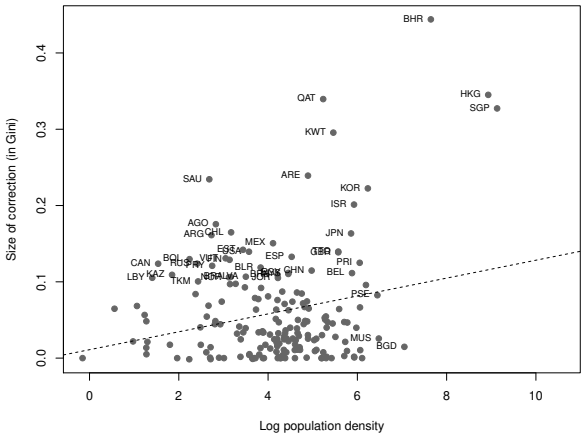
(d) Gini correction vs. log area



(e) Mean correction vs. log density



(f) Gini correction vs. log density



Notes: Illustration of how the correction of means and Gini coefficients, based on eq. (G-1) and eq. (G-2), correlate with log GDP per capita (PPP), log land area, and log population density. The data are a 10% representative sample of all non-zero lights in satellite F182010. GDP and population data are from the World Development Indicators. For display purposes, the left panels exclude countries with a correction of mean luminosity larger than 20 DN, these are Singapore, Hong Kong, Qatar, Bahrain, Kuwait, and the UAE. These countries are included in all panels on the right.

I Benchmarking exercises

Light-output elasticities at the national level: To validate our corrected data, we estimate light-output elasticities at the national level in the spirit of Henderson et al. (2012). Henderson et al. (2012) run fixed effects regressions of log GDP on log lights per square kilometer. They report an income elasticity of lights around 0.28. We replicate these results using a matched sample of the stable lights data, our top-coding corrected lights and the radiance-calibrated data for the seven years which all three data sources have in common over the period from 1996 to 2010.

Table I-1 reports the corresponding estimates and shows that—even at the highly aggregated country level—our top-coding correction leads to marginal improvements. The corrected data always yield the highest within-R^2, no matter if we use lights per square kilometer or control for population density (which implies a per capita interpretation[7]). As expected, the results are not materially different at the national level, so that for an analysis of the light-output relationship at such a high level of aggregation either data can be used without explicitly considering the role of top-coding. Note that it is not clear whether we should expect the corrected data to deliver regression coefficients which are closer to the radiance-calibrated data. The spectral mixing process of Hsu et al. (2015) created a lot of noise in areas which are not top-coded and provide the overwhelming share of the variation analyzed by these regressions.

Light-output elasticities at the subnational level: Lights are particularly useful as a proxy for economic activity at the regional level or other smaller geographies. For this benchmarking exercise, we use Germany as an example of a rich and decentralized economy. Germany publishes regional accounts down to the district level.[8] We compile annual data on GDP and population for 401 districts from 2000 until 2013. We use these data to analyze the cross-sectional and panel relationship between output and light intensity in Germany, with a particular emphasis on whether light intensity contains any information on economic activity beyond population density.

Table I-2 compares the light-output elasticities obtained from cross-sectional regressions using the stable lights data (columns 1 to 3) to the corrected data (columns 4 to 6). In both sets of estimates, we first introduce state and then so-called 'government region' fixed effects. This purges some of the confounding variation, such as persistent differences between West and East Germany.[9] Columns (1) to (3) suggest that there is

[7]To see this, note that $\ln(y_{it}/pop_{it}) = \beta \ln(lights_{it}/pop_{it}) + \gamma \ln(pop_{it})$ is equivalent to $\ln y_{it} = \beta \ln lights_{it} - (1-\beta-\gamma) \ln pop_{it}$, regardless of whether this is specified in levels or in terms of densities.

[8]Germany is administratively divided into states (Bundesländer), government regions (Regierungsbezirke), districts (Kreise) and municipalities (Gemeinden). These correspond to the European Nomenclature of Territorial Units for Statistics (NUTS) levels 1, 2, 3 and 4. Regional accounts are published until NUTS level 3.

[9]The East has benefited substantially from public investments in unified Germany and also adopted

little value added in using night lights to predict local output once population density is accounted for. This broadly reflects the conventional wisdom about the utility of nighttime lights in highly development economies. The picture changes completely in columns (4) to (6) where we use the corrected data. The sub-national light-output elasticity rises substantially and remains highly significant even in the presence of stricter fixed effects and a dummy for urban areas. The corrected data are strongly correlated with local output, on top of differences in population density, and the size of the light-output elasticity reaches levels comparable to national level estimates.

Table I-3 shows that this result carries over to changes over time, albeit with some qualifications. We now use the full panel from 2000 to 2013. The first two columns for each data source include state or government region fixed effects and year dummies, while the last column only uses within district variation. The stable lights data suggest that the light-output elasticity is relatively small. The corrected data recover results comparable to the cross-sectional estimates in columns (4) and (5). However, once we include district fixed effects, both the estimates become similar and their size decreases substantially. The within-district variation is relatively small to begin with, making it difficult to empirically trace out the light-output relationship (the R^2s decline in tandem). More importantly, the urban structure of Germany is very mature, so that our correction primarily affects the cross-sectional ordering of districts, rather than their ranking over time. In any case, the correction either substantially improves estimates of the light-output relationship at the local level or, at the minimum, provides comparable answers.

Figure I-1 plots the average light intensity per square kilometer indicated by both data sources over population density in order to illustrate the differences in the implied ranking of districts. The two series begin to diverge strongly after a density of about one thousand people per square kilometer. The stable lights series displays an asymptotic movement towards the top-coding threshold of slightly above 100 DN whereas the corrected series is approximately *linear* in population density.[10] Another notable feature is that the two out of the three brightest cities according to the stable lights data are medium-sized cities with populations well below 200,000 people (Herne and Gelsenkirchen), whereas our data correctly identifies the three largest and most populated economic centers as the brightest (Frankfurt, Munich and Berlin).

In sum, top-coding is a major issue for estimating urban-rural differences, precisely because there the cross-sectional comparison matters most. Using Germany as an example, we can establish two findings which are likely to hold in other settings as well. First, top-coding rises with urbanization. Second, the economic ranking of cities distorted in the stable lights data but a sensible ranking emerges after the correction.

energy-saving lights differently than the West.

[10]Note that lights per square kilometer are not bounded by 63 DN because of the division by area. Most pixels in Germany are smaller than 1 km^2 but bigger than 500 m^2, hence the theoretical upper bound is slightly above 100 DN, eg. 63 DN/0.5 km^2 = 126 DN/km^2.

Table I-1 – Light-output elasticity, country-level, 1996–2010

	Dependent variable: GDP in 2005 PPPs					
	(1) Stable	(2) Corrected	(3) Radiance	(4) Stable	(5) Corrected	(6) Radiance
Lights per km^2	0.275***	0.278***	0.233***	0.254***	0.257***	0.214***
	(0.067)	(0.064)	(0.054)	(0.064)	(0.062)	(0.051)
Population per km^2				0.382***	0.364***	0.397***
				(0.118)	(0.116)	(0.122)
Country FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Within-R^2	0.721	0.725	0.712	0.732	0.735	0.725
Observations	1288	1288	1288	1288	1288	1288
Countries	186	186	186	186	186	186

Note(s): The table reports panel FE estimates. Lights per km^2 and population per km^2 are measured in logs. The specifications are variants of $\ln y_{it} = \beta \ln \textsc{Lights}_{it} + \mathbf{x}'_{it}\boldsymbol{\gamma} + c_i + s_t + \epsilon_{it}$ where $\mathbf{x}_{it}$ is a vector of controls, c_i is a country fixed effect, and s_t are year dummies. Country-clustered standard errors are provided in parentheses. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table I-2 – Light-output elasticity, Germany at district level, cross-section

	Dependent variable: GDP in LCU					
	Stable lights data			*Corrected data*		
	(1)	(2)	(3)	(4)	(5)	(6)
Lights per km^2	0.119	0.145	0.173	0.271**	0.314**	0.336***
	(0.099)	(0.115)	(0.115)	(0.111)	(0.124)	(0.125)
Population per km^2	1.154***	1.151***	1.151***	1.044***	1.029***	1.033***
	(0.062)	(0.070)	(0.070)	(0.073)	(0.081)	(0.082)
Urban			-0.079**			-0.083**
			(0.036)			(0.036)
State FE	✓	–	–	✓	–	–
Gov't region FE	–	✓	✓	–	✓	✓
Adjusted-R^2	0.969	0.969	0.970	0.970	0.970	0.970
Districts	401	401	401	401	401	401

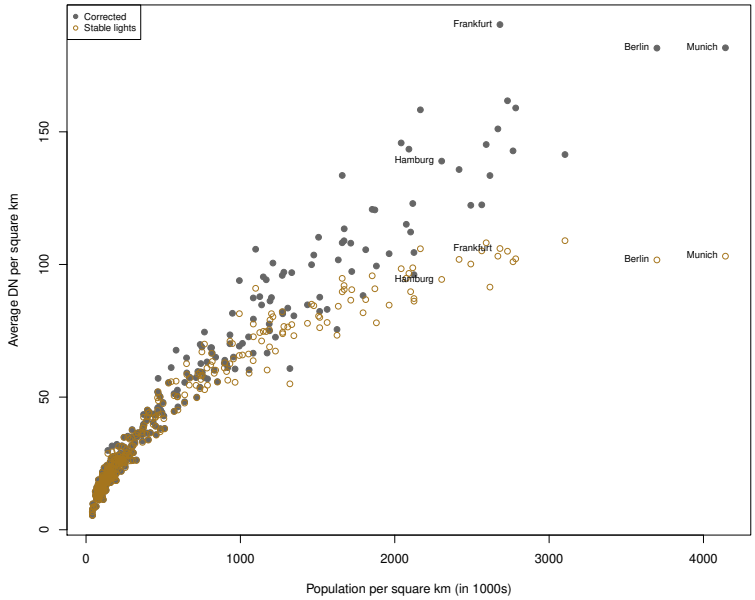
Notes: The table reports cross-sectional OLS estimates. All columns include a constant (not shown). Lights per km^2 and population per km^2 are measured in logs. Urban is a binary variable based on the official district classification. The specifications are variants of $\ln y_i = \beta \ln \textsc{Lights}_i + \mathbf{x}'_i\boldsymbol{\gamma} + \text{FE}_i + \epsilon_i$, where $\mathbf{x}_i$ is a vector of controls and c_i is a state or government region fixed effect. The data have been averaged over the period from 2000 to 2013. Robust standard errors are provided in parentheses. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table I-3 – Light-output elasticity, Germany at district level, 2000–2013

	Dependent variable: GDP in LCU					
	Stable lights data			*Corrected data*		
	(1)	(2)	(3)	(4)	(5)	(6)
Lights per km^2	0.116*	0.148**	0.063***	0.212***	0.248***	0.042***
	(0.064)	(0.067)	(0.009)	(0.074)	(0.077)	(0.008)
Population per km^2	1.155***	1.165***	0.427***	1.082***	1.089***	0.412***
	(0.041)	(0.044)	(0.078)	(0.049)	(0.052)	(0.078)
Urban		-0.076**			-0.077**	
		(0.035)			(0.034)	
Time FE	✓	✓	✓	✓	✓	✓
State FE	✓	–	–	✓	–	–
Gov't region FE	–	✓	–	–	✓	–
District FE	–	–	✓	–	–	✓
Within-R^2	0.961	0.960	0.060	0.962	0.961	0.052
Observations	5614	5614	5614	5614	5614	5614
Districts	401	401	401	401	401	401

Notes: The table reports panel fixed effects estimates. Lights per km^2 and population per km^2 are measured in logs. Urban is a binary variable based on the official district classification. The specifications are variants of $\ln y_{it} = \beta \ln \textsc{Lights}_{it} + \mathbf{x}'_{it}\boldsymbol{\gamma} + c_i + s_t + \epsilon_{it}$ where $\mathbf{x}_{it}$ is a vector of controls, c_i is a state, government region, or district fixed effect, and s_t are time dummies. District-clustered standard errors are in parentheses. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure I-1 – Light intensity versus population density in German districts



Notes: Illustration of the cross-sectional differences implied by the stable lights and corrected data. The data are averages of lights per km^2 and population per km^2 in 401 German districts over the period from 2000 to 2013.

J Additional results for African cities

Here we present additional results for the application on cities in Sub-Saharan Africa.

Figure J-1 shows the urban extents of selected cities and compares them with Google Earth images at the end of the periods we use to delineate urban areas, i.e., 12/1994 and 12/2013. The urban footprint detected by our algorithm coincides well with built-up structures. Abrahams et al. (2018) report similar results based on a more systematic comparison of urban areas derived in this manner to those obtained with remotely-sensed built-up grids.

Table J-1 gives an overview of the countries included in our study. The table reports the names of the primary city, the number of secondary cities, and the annualized growth rates for the stable lights and corrected data. The corrections are larger in primate than in secondary cities for each individual country.

Table J-2 varies the minimum city size to account for the uncertainty in classifying secondary cities. It only reports the strictest specification with country-year fixed effects and contrasts the results in both data sets. The results are very robust to larger thresholds and decrease only moderately in size and significance when smaller cities are included. Including smaller settlements increases noise, as their light intensity is based on fewer and fewer data points per year—no matter if they should be classified as secondary cities or not.

Table J-3 shows that our results are robust to excluding each of the African regions in turn. Regardless of whether we exclude Northern, Middle, Western or Southern Africa, the difference the interactions of the linear trend and the primacy dummy is only significant when the corrected data is used. When we exclude East Africa, the coefficient marginally loses significance but remains within a standard error of the main result. This suggests that cities in East and Southern Africa are particularly top-coded.

Table J-4 examines sprawl, that is, growth in lights outside of the initial footprint of the city but within the envelope of both the initial and final period boundaries. Primary cities are sprawling more rapidly than secondary cities. As expected, the top-coding correction does not have a noticeable impact at the urban fringe. Urban sprawl is typically suburban and density only increases later when these areas become part of the city proper. Together with our main finding, this underlines that primate cities are consolidating their dominant position in Sub-Saharan Africa.

Table J-5 provides some preliminary evidence on whether urban form has an effect on city growth. To study this question, we regress log mean lights on the coefficient of variation or Moran's I in the previous year, a linear time trend, an interaction with primacy, and a combination of city and satellite-year fixed effects. Spatial inequality appears to have a robust negative effect on growth, but the impact of fragmentation is small and not robust. City structure and growth are endogenous, so that we consider

this an interesting partial correlation which could be explored in future research.

Table J-6 shows city-level growth regressions using the radiance-calibrated data side-by-side with our main results to facilitate the comparison. Primary cities are growing significantly faster than secondary cities according to both data sets. Moreover, the interaction of primacy with the time trend in column (6) yields estimates which are numerically close to our main results. The regressions also show that the radiance-calibrated data imply a negative time trend in the light density of secondary cities (which is likely to be measurement error, made worse by the absence of coincident satellites and only a handful of years).

Figure J-2 illustrates the annual variation in all three data sets. The radiance-calibrated data exhibit sizable fluctuations across the limited time span for which they are available. These differences are not due to differences in top-coding, but the presence of background noise, a complicated process of merging the frequency spectrum from multiple satellites and, most importantly, a lack of comparability when this process was employed in different years. This is precisely why we only use them to infer the top-coding threshold and obtain the rank of top-coded pixels. The time series of the stable lights and corrected data fluctuate considerably less and exhibit a positive trend over the entire period.

Finally, Table J-7 repeats the fragmentation regressions from the main text but computes the coefficient of variation and Moran's I using the radiance-calibrated data. The results confirm that light intensity within primate cities has become considerably less concentrated while there is no such trend in fragmentation.

Figure J-1 – Urban extents of selected cities

(a) Lagos, 12/1994

(b) Lagos, 12/2013

(c) Luanda, 12/1994

(d) Luanda, 12/2013

(e) Johannesburg, 12/1994

(f) Johannesburg, 12/2013

Notes: Illustration of the urban extents detection algorithm presented in the text. Note the differences in map scale. Comparison of 1992-1994 urban footprint with December 1994 and December 2013 Landsat/ Copernicus images obtained via Google Earth Pro. Google Earth images are used as part of their "fair use" policy. All rights to the underlying maps belong to Google.

Table J-1 – Annualized growth rates of cities in Africa, 1992–2013, intensive margin

Country	Primate city	Primate city growth: Stable lights	Primate city growth: Corrected	Secondary cities	Secondary city growth: Stable lights	Secondary city growth: Corrected
Angola	Luanda	0.0282	0.0834	5	0.2034	0.2115
Burundi	Bujumbura	0.0114	0.0123	0		
Benin	Cotonou	0.0291	0.0317	3	0.1827	0.1823
Burkina Faso	Ouagadougou	0.0265	0.0325	4	0.1842	0.1846
Botswana	Gaborone	0.0328	0.0406	10	0.1775	0.1776
Central African Republic	Bangui	-0.0076	-0.0076	0		
Côte d'Ivoire	Abidjan	0.0128	0.0229	18	0.1817	0.1819
Cameroon	Douala	0.0193	0.0250	8	0.1642	0.1671
Congo (Dem.Rep.)	Brazzaville	0.0074	0.0181	13	0.1640	0.1651
Congo	Kinshasa	0.0124	0.0217	2	0.1897	0.1923
Djibouti	Djibouti	0.0205	0.0247	0		
Eritrea	Asmara	0.0217	0.0221	1	0.0627	0.0627
Ethiopia	Addis Abbeba	0.0227	0.0248	4	0.1771	0.1771
Gabon	Libreville	0.0138	0.0182	4	0.1700	0.1718
Ghana	Accra	0.0217	0.0284	18	0.1688	0.1694
Guinea	Conakry	0.0268	0.0273	2	0.1212	0.1212
Gambia, The	Banjul	0.0300	0.0314	0		
Guinea-Bissau	Bissau	0.0154	0.0154	0		
Kenya	Nairobi	0.0194	0.0213	12	0.1517	0.1525
Lesotho	Maseru	0.0390	0.0409	1	0.0517	0.0517
Madagascar	Antananarivo	0.0288	0.0309	5	0.1569	0.1569
Mali	Bamako	0.0233	0.0303	1	0.1787	0.1787
Mozambique	Maputo	0.0341	0.0475	11	0.1870	0.1876
Mauritania	Nouakchott	0.0294	0.0365	2	0.1928	0.1925
Malawi	Blantyre	0.0167	0.0206	5	0.1526	0.1537
Namibia	Windhoek	0.0130	0.0198	14	0.1735	0.1744
Niger	Niamey	0.0188	0.0181	5	0.1767	0.1767
Nigeria	Lagos	0.0175	0.0244	65	0.1643	0.1655
Rwanda	Kigali	0.0212	0.0218	1	0.1366	0.1366
Sudan	Khartum	0.0210	0.0317	20	0.1873	0.1879
Senegal	Dakar	0.0213	0.0329	11	0.1765	0.1776
Sierra Leone	Freetown	0.0270	0.0270	0		
Somalia	Mogadishu	0.0635	0.0635	0		
Swaziland	Mbabane	0.0288	0.0303	6	0.1340	0.1340
Tchad	Ndjamena	0.0300	0.0324	3	0.1560	0.1560
Togo	Sokode	0.0200	0.0228	2	0.1429	0.1429
Tanzania	Dar es Salaam	0.0231	0.0245	18	0.1594	0.1595
Uganda	Kampala	0.0318	0.0367	4	0.1528	0.1528
South Africa	Johannesburg	0.0100	0.0186	211	0.1697	0.1711
Zambia	Lusaka	0.0185	0.0293	19	0.1769	0.1776
Zimbabwe	Harare	0.0011	0.0014	19	0.1403	0.1403

Notes: The table reports summary statistics for the primary city and secondary cities in each country. The annualized growth rates are based on average light intensity and computed as $\frac{1}{21}(\ln x_{2013} - \ln x_{1992})$. The growth rate of secondary cities is an average across all such cities in the country.

Table J-2 – Growth regressions varying the minimum secondary city size, intensive margin

	Dependent variable: Log lights in the initial footprint							
	Stable lights	Corrected	Stable lights	Corrected	Stable lights	Corrected	Stable lights	Corrected
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primate × linear trend	-0.063	0.501	0.249	0.821	0.648	1.240	0.022	0.663
	(0.309)	(0.283)*	(0.340)	(0.313)***	(0.382)*	(0.359)***	(0.237)	(0.208)***
	[0.309]	[0.280]*	[0.343]	[0.314]***	[0.377]*	[0.352]***	[0.225]	[0.193]***
Minimum size (in km^2)	6	6	9	9	15	15	25	25
City FE	✓	✓	✓	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	16061	16061	12487	12487	8800	8800	6006	6006
Cities	731	731	568	568	400	400	273	273

Notes: The table reports the results of city-level panel regressions using the stable lights and top-coding corrected data with varying minimum size thresholds for secondary cities. All coefficients are scaled by 100 for readability. The specifications are variants of $\ln \text{LIGHTS}_{ijt} = \beta_1 t + \beta_2 (t \times P_{ij}) + c_{ij} + s_{jt} + \epsilon_{ijt}$ where t is a linear time trend, P_{ij} is an indicator for primate cities, c_{ij} is a city fixed effect and s_{jt} contains a varying set of fixed effects (satellite, year, or country-year). Standard errors clustered at the city level are reported in parentheses. Conley errors with a spatial cutoff of 1,000 km and a time-series HAC with a lag cutoff of 1,000 years are reported in brackets. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table J-3 – Growth regressions excluding major regions, intensive margin

	Dependent variable: Log lights in the initial footprint							
	Stable lights	Corrected	Stable lights	Corrected	Stable lights	Corrected	Stable lights	Corrected
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primate × linear trend	0.548	1.152	0.489	0.912	-0.130	0.568	-0.067	0.529
	(0.464)	(0.424)***	(0.378)	(0.357)**	(0.404)	(0.379)	(0.299)	(0.263)**
	[0.462]	[0.419]***	[0.380]	[0.359]**	[0.413]	[0.384]	[0.296]	[0.257]**
Excluded region	West Africa		Middle Africa		East Africa		Southern Africa	
City FE	✓	✓	✓	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	9363	9363	11631	11631	10005	10005	7054	7054
Cities	426	426	529	529	455	455	321	321

Notes: The table reports the results of city-level panel regressions using the stable lights and top-coding corrected data omitting major regions of Sub-Saharan Africa. All coefficients are scaled by 100 for readability. The specifications are variants of $\ln \text{LIGHTS}_{ijt} = \beta_1 t + \beta_2 (t \times P_{ij}) + c_{ij} + s_{jt} + \epsilon_{ijt}$ where t is a linear time trend, P_{ij} is an indicator for primate cities, c_{ij} is a city fixed effect and s_{jt} contains a varying set of fixed effects (satellite, year, or country-year). Standard errors clustered at the city level are reported in parentheses. Conley errors with a spatial cutoff of 1,000 km and a time-series HAC with a lag cutoff of 1,000 years are reported in brackets. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table J-4 – Growth regressions for African cities, outside of initial boundaries

	Dependent variable: Log lights in the fringe					
	Stable lights data			*Corrected data*		
	(1)	(2)	(3)	(4)	(5)	(6)
Linear trend	2.575			2.576		
	(0.161)***			(0.161)***		
	[0.682]***			[0.683]***		
Primate × Linear trend	2.959	2.955	2.830	2.996	2.992	2.883
	(0.433)***	(0.433)***	(0.304)***	(0.442)***	(0.442)***	(0.309)***
	[0.442]***	[0.433]***	[0.303]***	[0.451]***	[0.442]***	[0.306]***
City FE	✓	✓	✓	✓	✓	✓
Satellite FE	✓	–	–	✓	–	–
Year FE	–	✓	–	–	✓	–
Country-Year FE	–	–	✓	–	–	✓
Observations	11410	11410	11410	11410	11410	11410
Cities	519	519	519	519	519	519

Notes: The table reports the results of city-level panel regressions using the stable lights and top-coding corrected data outside of the initial boundaries. The fringe is defined as the difference between the envelope of a city, i.e., the maximum urban extend, and the initial boundary. This leads to slightly fewer cities as some are not observed in the final period. Areas with zero or few lights are adjusted following the procedure in Storeygard (2016) where the minimal detectable light intensity (in DN) for a given area is added to the total before dividing by the area of the fringe. All coefficients are scaled by 100 for readability. The specifications are variants of $\ln \textsc{Lights}_{ijt} = \beta_1 t + \beta_2 (t \times P_{ij}) + c_{ij} + s_{jt} + \epsilon_{ijt}$ where t is a linear time trend, P_{ij} is an indicator for primate cities, c_{ij} is a city fixed effect and s_{jt} contains a varying set of fixed effects (satellite, year, or country-year). Standard errors clustered at the city level are reported in parentheses. Conley errors with a spatial cutoff of 1,000 km and a time-series HAC with a lag cutoff of 1,000 years are reported in brackets. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table J-5 – Impact of fragmentation on city growth, envelopes

	Dependent variable: Log lights in the envelope					
	(1)	(2)	(3)	(4)	(5)	(6)
Linear trend	0.228			0.810		
	(0.144)			(0.129)***		
	[0.558]			[0.555]		
Primate × Linear trend	0.739	0.724	0.602	1.269	1.274	0.921
	(0.293)**	(0.292)**	(0.216)***	(0.351)***	(0.351)***	(0.239)***
	[0.303]**	[0.301]**	[0.220]***	[0.351]***	[0.350]***	[0.239]***
Lagged CV	-0.549	-0.564	-0.403			
	(0.088)***	(0.090)***	(0.071)***			
	[0.100]***	[0.102]***	[0.077]***			
Lagged Moran				-0.396	-0.269	-0.217
				(0.167)**	(0.174)	(0.153)
				[0.172]**	[0.177]	[0.150]
City FE	✓	✓	✓	✓	✓	✓
Satellite FE	✓	–	–	✓	–	–
Year FE	–	✓	–	–	✓	–
Country-Year FE	–	–	✓	–	–	✓
Observations	11791	11791	11791	11791	11791	11791
Cities	562	562	562	562	562	562

Notes: The table reports the results of city-level panel regressions using the top-coding corrected data, where either the lagged coefficient of variation or Moran's I are used as regressors ($F_{ij,t-1}$). All coefficients are scaled by 100 for readability. The specifications are variants of $\ln \text{LIGHTS}_{ijt} = \beta_1 t + \beta_2(t \times P_{ij}) + \beta_3 F_{ij,t-1} + c_{ij} + s_{jt} + \epsilon_{ijt}$ where t is a linear time trend, P_{ij} is an indicator for primate cities, c_{ij} is a city fixed effect and s_{jt} contains a varying set of fixed effects (satellite, year, or country-year). Standard errors clustered at the city level are reported in parentheses. Conley errors with a spatial cutoff of 1,000 km and a time-series HAC with a lag cutoff of 1,000 years are reported in brackets. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table J-6 – Growth regressions for African cities, radiance-calibrated data

	Dependent variable: Log lights in the initial footprint					
	Corrected data			*Radiance-calibrated data*		
	(1)	(2)	(3)	(4)	(5)	(6)
Linear trend	0.834			-0.515		
	(0.126)***			(0.226)**		
	[0.554]			[0.934]		
Primate × linear trend	0.742	0.742	0.826	1.062	1.062	0.828
	(0.255)***	(0.255)***	(0.314)**	(0.399)***	(0.399)***	(0.339)**
	[0.260]***	[0.258]***	[0.314]*	[0.409]***	[0.407]***	[0.386]**
City FE	✓	✓	✓	✓	✓	✓
Satellite FE	✓	–	–	✓	–	–
Year FE	–	✓	–	–	✓	–
Country-Year FE	–	–	✓	–	–	✓
Observations	12465	12465	12465	3966	3966	3966
Cities	567	567	567	567	567	567

Notes: The table reports the results of city-level panel regressions using the top-coding corrected data and radiance-calibrated data. All coefficients are scaled by 100 for readability. The specifications are variants of $\ln \text{LIGHTS}_{ijt} = \beta_1 t + \beta_2 (t \times P_{ij}) + c_{ij} + s_{jt} + \epsilon_{ijt}$ where t is a linear time trend, P_{ij} is an indicator for primate cities, c_{ij} is a city fixed effect and s_{jt} contains a varying set of fixed effects (satellite, year, or country-year). Standard errors clustered at the city level are reported in parentheses. Conley errors with a spatial cutoff of 1,000 km and a time-series HAC with a lag cutoff of 1,000 years are reported in brackets. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure J-2 – Trends in light emitted by African cities according to different data sets

Notes: Illustration of the average sum of lights over time in the sample of African cities.

Table J-7 – Trends in fragmentation of African cities, envelopes, radiance-calibrated data

	Dependent variable:					
	Coefficient of Variation			*Moran's I*		
	(1)	(2)	(3)	(4)	(5)	(6)
Linear trend	0.500			0.075		
	(0.114)***			(0.052)		
	[0.211]**			[0.073]		
Primate × linear trend	-0.960	-0.935	-0.992	-0.095	-0.092	-0.018
	(0.220)***	(0.217)***	(0.257)***	(0.052)*	(0.052)*	(0.059)
	[0.219]***	[0.215]***	[0.255]***	[0.061]	[0.061]	[0.063]
Log light intensity	-3.951	-5.630	-1.699	-0.681	-0.902	-1.169
	(4.131)	(4.267)	(3.388)	(0.545)	(0.574)	(0.715)
	[4.250]	[4.346]	[3.781]	[0.560]	[0.587]	[0.695]*
City FE	✓	✓	✓	✓	✓	✓
Satellite FE	✓	–	–	✓	–	–
Year FE	–	✓	–	–	✓	–
Country-Year FE	–	–	✓	–	–	✓
Observations	3932	3932	3932	3932	3932	3932
Cities	562	562	562	562	562	562

Notes: The table reports results of city-level panel regressions using the radiance-calibrated data for the available years. The specifications are variants of $F_{ijt} = \beta_1 t + \beta_2(t \times P_{ij}) + \beta_3 \ln \text{LIGHTS}_{ijt} + c_{ij} + s_{jt} + \epsilon_{ijt}$, where F_{ijt} is either the coefficient of variation or Moran's I, t is a linear time trend, P_{ij} is an indicator for primate cities, c_{ij} is a city fixed effect and s_{jt} contains are varying set of fixed effects (satellite, year, or country-year). Standard errors clustered at the city level are reported in parentheses. Conley errors with a spatial cutoff of 1,000 km and a time-series HAC with a lag cutoff of 1,000 years are reported in brackets. Significant at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Additional references

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.

Cirillo, P. (2013). Are your data really Pareto distributed? *Physica A: Statistical Mechanics and its Applications 392*(23), 5947–5962.

Coles, S. (2001). *An Introduction to the Statistical Modeling of Extreme Values*. Springer Series in Statistics.

Eeckhout, J. (2009). Gibrat's law for (all) cities: Reply. *American Economic Review 99*(4), 1676–1683.

Elvidge, C. D., K. E. Baugh, J. B. Dietz, T. Bland, P. C. Sutton, and H. W. Kroehl (1999). Radiance calibration of DMSP-OLS low-light imaging data of human settlements. *Remote Sensing of Environment 68*(1), 77–88.

Elvidge, C. D., F.-C. Hsu, K. E. Baugh, and T. Ghosh (2014). National trends in satellite-observed lighting. In Q. Weng (Ed.), *Global Urban Monitoring and Assessment through Earth Observation*, Taylor & Francis Series in Remote Sensing Applications, Chapter 6, pp. 97–119. CRC Press.

Elvidge, C. D., D. Ziskin, K. E. Baugh, B. T. Tuttle, T. Ghosh, D. W. Pack, E. H. Erwin, and M. Zhizhin (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies 2*(3), 595–622.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics 1*(1), 255–294.

Gabaix, X. and R. Ibragimov (2011). Rank - 1/2: A simple way to improve the ols estimation of tail exponents. *Journal of Business & Economic Statistics 29*(1), 24–39.

Gabaix, X. and Y. Ioannides (2004). The evolution of city size distribution. In J. V. Henderson and J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (1 ed.), Volume 4, Chapter 53, pp. 2341–2378. Elsevier.

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

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annuals of Statistics 3*(5), 1163–1174.

Hsu, F.-C., K. E. Baugh, T. Ghosh, M. Zhizhin, and C. D. Elvidge (2015). DMSP-OLS radiance calibrated nighttime lights time series with intercalibration. *Remote Sensing 7*(2), 1855–1876.

Mookherjee, D. and A. Shorrocks (1982). A decomposition analysis of the trend in UK income inequality. *Economic Journal 92*(368), 886–902.

Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics 46*(5), 323–351.

Nordhaus, W. and X. Chen (2015). A sharper image? Estimates of the precision of nighttime lights as a proxy for economic statistics. *Journal of Economic Geography 15*(1), 217–246.

Soo, K. T. (2005). Zipf's law for cities: A cross-country investigation. *Regional Science and Urban Economics 35*(3), 239–263.

Storeygard, A. (2016). Farther on down the road: Transport costs, trade and urban growth in sub-saharan africa. *Review of Economic Studies 83*(3), 1263–1295.

Tuttle, B., S. Anderson, P. Sutton, C. Elvidge, and K. Baugh (2013). It used to be dark here. *Photogrammetric Engineering and Remote Sensing 79*(3), 287–297.

Ziskin, D., K. Baugh, F. C. Hsu, T. Ghosh, and C. Elvidge (2010). Methods used for the 2006 radiance lights. *Proceedings of the Asia-Pacific Advanced Network 30*, 131–142.