Illuminating Africa?*

Giorgio Chiovelli†
Universidad de Montevideo

Stelios Michalopoulos‡
Brown University, CEPR and NBER

Elias Papaioannou§
London Business School, CEPR

Tanner Regan¶
George Washington University

November 23, 2023

Abstract

Satellite images of nighttime lights are commonly used to proxy local economic conditions. Despite their popularity, there are concerns about how accurately they capture local development in low-income settings and different scales. We compile a yearly series of comparable nighttime lights for Africa from 1992 to 2020, considering key factors that affect accuracy and comparability over time: sensor quality, top coding, blooming, and, importantly, variations in satellite systems (DMPS and VIIRS) using an ensemble, machine learning, approach. The harmonized luminosity series outperforms the unadjusted series as a stronger predictor of local development, particularly over time and at higher spatial resolutions.

KEYWORDS: Night Lights, Economic Development, Measurement, Africa.

JEL CLASSIFICATION: O1, R1, E01, I32

*We thank Gonzalo Ferrés and Federico Ferro for providing superb research assistance.

†Giorgio Chiovelli. Universidad de Montevideo, Department of Economics, Prudencio de Pena 2440, Montevideo, 11600, Uruguay; gchiovelli@um.edu.uy. Web: https://sites.google.com/view/giorgiochiovelli/

‡Stelios Michalopoulos. Brown University, Department of Economics, 64 Waterman Street, Robinson Hall, Providence RI, 02912, United States; smichalo@brown.edu. Web: https://sites.google.com/site/steliosecon/

§Elias Papaioannou. London Business School, Economics Department, Regent's Park. London NW1 4SA. United Kingdom; eliasp@london.edu. Web: https://sites.google.com/site/papaioannouelias/home

¶Tanner Regan. George Washington University, Department of Economics, Monroe Hall 2115 G Street NW, Washington DC, 20052. United States; tanner_regan@ gwu.edu. Web: https://sites.google.com/site/tannerregan

1 Introduction

A considerable literature in economics, political science, and remote sensing employs satellite nighttime lights to proxy development (Donaldson and Storeygard, 2016; Levin *et al.*, 2020).[1] The use of satellite data appears *a priori* helpful for low and middle-income countries with weak state capacity and recurrent conflict. Earlier research reveals that luminosity is a valuable proxy for cross-country GDP (Henderson *et al.*, 2012; Chen and Nordhaus, 2011); hence, researchers use luminosity to correct for inconsistencies and noise in country-level statistics stemming from challenges measuring output in economies specializing in agriculture, with a large informal economy, and underfunded statistical agencies (Pinkovskiy and Sala-i Martin, 2016). Luminosity appears helpful in correcting inflated output statistics that non-democratic governments often produce (Martinez, 2022) and quantifying profit shifting by multinationals (Bilicka and Seidel, 2022). Applied research used luminosity to measure regional development, as data unavailability and error-in-variables are more serious at the local level. Many works use luminosity to proxy development across administrative units (Hodler and Raschky, 2014; Alesina *et al.*, 2016), historical ethnic homelands (Michalopoulos and Papaioannou, 2013, 2014), and pixels (Henderson *et al.*, 2018).[2]

However, there are still open issues regarding how accurately luminosity predicts development. First, while most studies conduct validation exercises that luminosity correlates with local development, there are concerns about the strength of the association. Second, researchers have leeway on the validation, for example, selecting the development proxy and the spatial unit, which sometimes are large areas and, in other cases, small pixels. Third, while some works use unadjusted series (e.g., Michalopoulos and Papaioannou, 2013), others adjust for top coding and the tendency of light to spill to neighboring areas (e.g., Henderson *et al.*, 2018). Fourth, mapping the lower resolution and more coarse pre-2013 series (DMSP-OLS) with the post-2013 (VIIRS) series is not straightforward. Most works abstain from dealing with this issue using the pre or post-2013 series. Consequently, despite the many uses of luminosity in various fields of economics and political science, there are ambiguities about its ability to capture development in low-income settings, at which scale, location, and periods.

Recent works paint a somewhat conflicting picture. On the one hand, India-based tabulations suggest that lights proxy well regional economic conditions in levels and changes (Asher *et al.*, 2021). On the other hand, studies from Indonesia, China, and South Africa suggest that variation of GDP does not correlate strongly with lights outside cities, at least for the early period,

[1] This literature comprises work in Africa (Hjort and Poulsen, 2019; Storeygard, 2016; Michalopoulos and Papaioannou, 2014; Dreher *et al.*, 2019; Henderson *et al.*, 2017), Asia (Harari, 2020; Chodorow-Reich *et al.*, 2020; Baum-Snow *et al.*, 2017), Europe (Gibson, 2021), North America (Bleakley and Lin, 2012), and worldwide (Ch *et al.*, 2021; Henderson *et al.*, 2018; Pinkovskiy, 2017; Martinez, 2022).

[2] A related research stream combines light density with daytime imagery to better proxy local activity (Jean *et al.*, 2016; Yeh *et al.*, 2020; Khachiyan *et al.*, 2022). However, the daytime data is proprietary, and the methods are prohibitive for typical economics research.

1992 − 2013 (Gibson *et al.*, 2021). Besides, Bluhm and McCord (2022) comprehensive validation of nightlights locally in the US, Spain, Germany, Brazil, and Italy demonstrate a non-constant lights-GDP elasticity in changes.

Here, we zoom in on Africa, where the output data quality is poor, and there is relatively limited regional data on economic activity, especially at granular levels. As research has moved from cross-country designs to *meso* approaches that exploit variation within-country across regions (administrative units, ethnic homelands, and pixels), there are applications at spatial units of various sizes. Our first contribution is to create a standardized pixel-level panel of nightlights over three decades (1992 − 2020), integrating the pre-2013 (DMSP) and the post-2013 (VIIRS) series after various adjustments to reduce measurement error. While the VIIRS data series offers various improvements (Gibson *et al.*, 2021), research often requires the longest possible time-series, for example, to study the long-run impact of democratic transitions in Africa, China shock, and trade liberalization. Adjusting and integrating the old luminosity series with the newer ones can allow exploring a wider range of questions.[3]

Our second contribution is to validate the newly compiled luminosity series as a proxy for African local development. We compare the performance of the unadjusted luminosity series to our adjusted and harmonized one, examining their correlation with education, household wealth, electrification, and other development measures using 139 georeferenced DHS surveys from 34 African countries and all Mozambican censuses (spanning both the DMSP and the VIIRS period). Our analysis reveals three main takeaways. First, the new luminosity series correlates strongly with local development in the cross-section and over time. The correlation is, however, far from perfect. Second, the correlation is stronger, especially in panel estimation, with the adjusted and harmonized series, telling of a reduction in measurement error. Third, the luminosity-development correlation retains significance across gridcells and administrative units of various sizes; in rural and urban areas; even when exploiting very localized variation.

Structure Section 2 details our ML methodology for correcting and merging the DMSP-OLS with the VIIRS series. Sections 3, 4, and 5, focus on the cross-sectional and over time association of luminosity with development across countries, regions/gridcells of various sizes, and Mozambican administrative units, respectively. Section 6 concludes, discussing avenues for future research.

2 Methods and New Series

This Section presents the data and our approach to compiling a new time series of nighttime lights across Africa for $1km$ by $1km$ pixels from 1992 to 2020. First, we discuss the adjustments in the

[3] We intend to make the code and data publicly available and update the series annually.

DMSP series. Second, we present the machine learning (ML) approach that converts and merges the new VIIRS series to the adjusted DMSP.

2.1 Adjusted DMSP Series

The DMSP data from the Earth Observation Group (EOG) for $1992 - 2013$ has three main deficiencies: cross-sensor calibration, top coding, and blooming. As earlier studies have addressed these issues, we discuss them briefly.

Cross-sensor Accuracy As the DMSP-OLS series comes from six satellites, luminosity readings vary. Sometimes, values differ even for the same satellite due to the sensor's (eyes) degradation. We use the DMSP series from Li *et al.* (2020), which integrates the cross-sensor correction from Li and Zhou (2017). Using data from sensors with overlapping or nearby years, these studies estimate a second-order polynomial to map values across sensors.

Top Coding The DMSP data are 8-bit integers, Digital Numbers (DN), ranging from 0 to 63. This limits stored information. In addition, as sensors are calibrated to detect clouds, they miss brighter lights. DN of 63 corresponds to a range of actual radiance. Pixels with DNs in the mid-50s also suffer from implicit top coding, as they are averages of multiple inputs, which could be top-coded. A 'radiance-calibrated' (RC) vintage of DMSP, which is not top coded, is available for seven years. Most pixels in Africa are unlit (98.4% in 1992, 97.4% in 2002, and 96.8% in 2012). Even among lit pixels the share of top-coded ones ($DN = 63$) is tiny (0.98% in 1992, 1.4% in 2002, and 1.7% in 2012). The share of lit pixels with values close to top-coding (DN$\geq$55) is 2.8% in 1992, 5.6% in 2002, and 10% in 2012. We correct the series by combining the DMSP with the RC data with the approach of Bluhm and Krause (2020). First, for each year, we identify pixels (number N_t) with DN$\geq$ 55 for replacement. Second, we rank the N_t pixels using the RC series from the nearest year. Third, generate "structural values" from a truncated Pareto distribution, i.e., $f(x) = \frac{\alpha L^{\alpha} x^{-\alpha-1}}{1-(L/H)^{\alpha}}$.[4] Fourth, we replace the N_t top coded pixels so that the pixel with the i-th highest rank is replaced by the i-th highest "structural value".

Blooming The DMSP data suffer from blooming (also called 'blurring' or 'bleeding') due to weak spatial accuracy.[5] The sensor records a window where central pixels cover less space on the ground than pixels along the edges, 'stretching out' the edge pixels (Gibson *et al.*, 2021). The sensor can also be displaced up to $3km$. We follow Cao *et al.* (2019), who model blooming as spatial

[4] We use the parameter values from Bluhm and Krause (2020): $\alpha = 1.5$, lower bound threshold $L = 55$, and upper bound threshold $H = 2000$.

[5] Some works suggested that blooming may be more important close to the sea and lakes, but Gibson *et al.* (2021) show that this is not the case. Nevertheless, we allow the blooming correction to differ flexibly in coastal areas and across broad African regions.

spillovers and remove them. The background light is identified based on all lit pixels that neighbor at least one unlit pixel; *pseudo-light pixels* (PLPs). The method works as follows: First, identify PLPs, lit pixels ($DN > 0$) with one of its neighboring pixels dark (DN= 0). Second, for each PLP take the inverse squared distance weighted sum of light in a surrounding 7×7 window, excluding pixels with less light.[6] Third, run an OLS regression of the PLP light on the sum of its neighbors' light: $DN_p = \alpha + \beta \sum_q \frac{DN_q}{d_{q,p}^2}$. [We estimate the spatial decay separately for broad African regions (north, west, east, central, south, and coastal), although this does not affect the estimates much.] Fourth, for lit pixels, remove blooming by subtracting the model predicted lights; replace pixels' light with $DN'_p = DN_p - \widehat{\alpha} - \widehat{\beta} \sum_q \frac{DN_q}{d_{q,p}^2}$. Fifth, smooth each cell's value with the mean of it and its eight nearest neighbors and then set all negative values to zero. The blooming correction increases the share of unlit pixels, which is already large, as it removes light that spills over from nearby highly lit pixels. The share of unlit pixels in 1992 rises from 98.4% to 99.1% and in 2012 from 96.8% to 97.9%. However, as shown below, the adjusted series correlates more strongly with local development despite the higher share of unlit pixels.

2.2 Harmonizing the DMSP and VIIRS Series

We use the VIIRS VNL V2 series (Elvidge *et al.*, 2021). While it does not suffer from top coding, blooming, and sensor degradation, VIIRS becomes available only after 2013. Besides, VIIRS' units, range, variance, and spatial resolution are not comparable to the DMSP. First, VIIRS records 14-bit DN, allowing for a wider range and more distinct values than DMSP. Second, VIIRS is recorded at a finer spatial resolution than DMSP ($500m$ vs. $1km$ at the equator). Third, the quality of the sensors is (much) better. Consequently, almost all studies rely on one of the two series. We develop a method to fuse the VIIRS with the DMSP series, creating an uninterrupted panel of comparable nightlights over three decades. We merge VIIRS to both the unadjusted and the adjusted DMSP series that account for top coding, sensor degradation, and blooming. Since the VIIRS data does not suffer from these issues, our preferred series downgrades VIIRS to the adjusted DMSP series. Our validation exercises below corroborate this choice.

There is a very recent parallel literature to our work that "downgrades" the VIIRS series to make it comparable to previous data (Li *et al.*, 2020; Nechaev *et al.*, 2021). Li *et al.* (2020) use a sigmoid function calibrated in overlapping DMSP-VIIRS years. But, as the authors acknowledge, the new data are comparable *only* for high light values, which is particularly problematic for Africa's low luminosity. Compared to the remote sensing literature, our innovation is to take a machine learning (ML) approach agnostic to the functional form of the two series' mapping. The concordance of the two series is a problem well-suited to an ML design, which discovers complex

[6] The 7×7 window follows the approach of Cao *et al.* (2019) and works from the assumption that blooming does not 'stretch out' from an origin pixel across more than ≈3 adjacent pixels (i.e., ≈3km).

structures not specified in advance (Mullainathan and Spiess, 2017). In parallel work, Nechaev *et al.* (2021) use a Convolutional Neural Network (CNN) to downgrade the VIIRS series. While their approach is comparable to ours, they only apply it to the 'unadjusted' DMSP data, which suffers from blooming, top-coding, and sensor quality. As shown below, the unadjusted series does a poorer job of explaining local development.

2.2.1 Method

To implement the downgrading of VIIRS and its merging with the DMSP series, we use an *ensemble method*, considered as state-of-the-art in machine learning (Sagi and Rokach, 2018).[7] Ensemble methods combine different models to improve out-of-sample performance over a single model (Athey and Imbens, 2019). We use an 'extremely randomized trees,' (Geurts *et al.*, 2006), an averaging ensemble method that combines many decision trees. The driving principle behind averaging methods is to build multiple independent models and then average their predictions. Random forests combine decision trees, each built from a random sample of observations and features (covariates/predictors); the decision trees use the best splits from the respective samples (Breiman, 2001). Extremely randomized trees take an extra step: instead of picking the 'best' thresholds from the sample of observations and features, pick them randomly, as doing so improves accuracy and computational efficiency (Geurts *et al.*, 2006).

We work with a 30-arc second pixel (roughly $1km$ at the equator), as this matches the DMSP resolution and the following features: (i) pixel statistics, mean, median, min, and max values; (ii) statistics outside the pixel, mean, and variance at windows of varying widths in pixels (3, 4, 7, 9, 11, 13, 17, 21); and (iii) indicators for six broad African regions, North, West, Central, East, South, and sea off-coast. The decision trees allow for all complex interactions between features. The extremely randomized tree has a set of (regularization) parameters that must be calibrated.[8] We implement a randomized search across parameter values evaluated by cross-fold validation to avoid misjudgment; we use the *'scikit learn'* library in *python* that picks parameters that maximize the out-of-sample prediction, as measured by the out-of-sample R^2.

2.2.2 Performance

The downgrading of VIIRS to the adjusted DMSP performs well both in absolute terms and compared to the recent efforts of Li *et al.* (2020) and Nechaev *et al.* (2021). Figure 1 reports the results

[7] Mullainathan and Spiess (2017) write that *'while it may be unsurprising that such ensembles perform well on average... it may be more surprising that they come on top in virtually every prediction competition.'*

[8] The full set of parameters is: n_estimators (number of trees), min_samples_split (the minimum number of observations required to split an internal node), min_samples_leaf (the minimum number of samples required at a leaf node), max_features (the number of features to consider when looking for the best split), max_depth (the maximum tree length), and whether bootstrap samples are used when building trees or the full dataset is used for each tree.

of our method's out-of-sample performance (we train and calibrate the model in 2012 and evaluate it using data from 2013) and compares with Li *et al.* (2020) and Nechaev *et al.* (2021).

First, the left panels plot the scatters of the predicted values (in the vertical axis) against the actual DMSP values (in the horizontal axis) for our method (panel a), the Li *et al.* (2020) sigmoid function method (panel c), and the Nechaev *et al.* (2021) convoluted neural network approach (panel e). Our method's root mean square error (RMSE) is 0.641, considerably lower than the 3.11 of the approach of Li *et al.* (2020) and lower than the 0.72 of Nechaev *et al.* (2021). Our method performs better at the middle and mainly the lower end of the luminosity distribution, particularly useful when studying regions in a low-income or (lower) middle-income setting. Despite performing, on average, significantly better, our method does slightly worse compared to Nechaev *et al.* (2021) in very high luminosity areas, which are rare in Africa.

Second, because low-light regions are chief in Africa and many works apply binary transformation, we also examine performance at the extensive margin. Panels (b), (d), and (f) report "confusion matrices" of lit and unlit pixels with the three methods. The rows correspond to actual adjusted DMSP values, and the columns to out-of-sample predicted DMSP values for 2013. The top-left counts pixels classified correctly as unlit, and the bottom-right pixels correctly classified as lit. For our method, the share of lit pixels correctly classified (*recall*) is 0.90 [697890/(697890 + 79415)]; the share of all predicted as lit pixels correctly classified (*precision*) is 0.72 [697890/(697890 + 266227)]. These statistics depend on the distribution of lit and unlit pixels, which is highly skewed. The actual share of lit pixels is just 2%. Simply classifying all pixels as lit would get a recall score of 100%, but a precision of just 2%. The figure thus also reports the $F1$ score, a widely used metric to evaluate the success of binary classifiers when one class is rare (Lipton *et al.*, 2014). The $F1$ score takes the harmonic mean of the recall and precision scores; $F1 = 2 * recall * precision/(recall + precision)$. Higher $F1$, bounded between 0 and 1, indicates a better accuracy. Our method yields an $F1$ of 0.80, much higher than the 0.18 of the sigmoid approach and slightly better than the 0.74 of the CNN approach.

The Appendix shows that our merging is plausible, as there are no discontinuities in 2012-2014. Appendix Figure A1 uncovers a strong co-evolution between the harmonized luminosity series and the share of the population with electricity in Mozambique, Kenya, the Democratic Republic of Congo, Ghana, Tanzania, and Nigeria without a jump when moving from DMSP to VIIRS (see also Henderson *et al.* (2012)). Appendix Figure A2 also reveals no major swings in the spatial distribution of the harmonized luminosity series in $2013 - 2014$ in Ghana.

3 Cross-Country Patterns

While our focus is on the use of lights in local economic activity, we commence the analysis examining the cross-country association between nighttime luminosity and GDP, using data from World Bank's World Development Indicators Database.[9]

Cross-sectional Association Figure 2 Panels (a) and (b) illustrate the strong cross-sectional association between GDP and the sum of the harmonized nighttime lights in 2005 (adjusted DMSP) and 2015 (downgraded VIIRS to adjusted DMSP). Appendix Table B1 - Panel *A* explores in more detail the cross-sectional association. The coefficient is highly significant, showing that luminosity is a good proxy of output across the 48 African countries. The fit is strong in both periods with an adjusted R^2 exceeding 0.9. The elasticity is stable across periods (around 0.53), and robust to dropping outliers, excluding island nations, and augmenting the specifications with broad regional constants.

Within-Country over Time Association We run panel and long-run differences specifications to examine the dynamic association between nighttime lights and GDP. Figure 2 - Panel (c) illustrates the panel association at the yearly frequency, plotting (residuals of) GDP and luminosity netting out country fixed-effects and year constants. As the noise in GDP and nighttime lights gets magnified at the yearly frequency, panel (d) gives the panel elasticity using five-year averages of GDP and luminosity. Panels (e) and (f) plot the association between changes in GDP and the adjusted and harmonized series over $1992-2019$ and over $1992-2013$. Taking long-run differences, the error-in-variables fall, and the estimates strengthen. The highly significant elasticity hovers around 0.34 and 0.36 (median regression estimates are between 0.31 and 0.33). These estimates are similar to ones $[0.30-0.33]$ reported by Henderson *et al.* (2012) across 188 countries for $1992/3-2005/6$.

4 Local Development

As applied research on growth, long-run development, economic history, and political economy has moved over the past years from cross-country approaches to designs that exploit spatial variation within countries (Michalopoulos and Papaioannou, 2018), we explore the potential of luminosity to capture regional well-being. We use georeferenced Demographic and Health Surveys (DHS) for our validation, following a considerable body of research (e.g., Young (2012), Lu and Vogl (2022)) Across all tests, we compare the newly compiled adjusted and merged VIIRS-DSMP lights series with unadjusted ones, as doing so illustrates a part of our contribution.

[9] We use all available country-years but Equatorial Guinea, which is an outlier (see also Henderson *et al.* (2012)). We require that every country is observed in the DMSP and the VIIRS era; imposing this restriction drops Eritrea, Djibouti, and Somalia. Appendix Tables B2-B3 provide further evidence.

4.1 Data and Specification

We obtain all geo-referenced Demographic and Health Surveys from Africa. Appendix Table B5 reports the country-survey years, while Appendix Table B6 gives summary statistics. The 34 countries are from all parts of the continent, relatively richer and poorer. Most surveys were conducted in the 2000s and 2010s, but we also have over a dozen surveys in the 1990s. We extract the following outcomes: public goods, like access to electricity and availability of a flush or a pit toilet, and a composite household wealth index based on a principal component aggregation of household characteristics (like the quality of the household roof and the ownership of assets). We also look at education using the mean years of schooling of respondents aged 15-39.[10] The geo-referenced DHS gives information across survey clusters (enumeration areas), typically cities, towns, and villages. We match these points to three spatial units as one of our objectives is to examine the usefulness of the luminosity series at various resolutions. First, we aggregate the DHS to gridcells of 0.25×0.25 degrees, roughly $27km$ by $27km$ at the equator; see Appendix Figure B3. We match the responses' cluster to the gridcell where their coordinate falls and then aggregate across all gridcell-year. Second, we match DHS data to admin-1 (one level below national), for example, Nigerian states or South African provinces. Third, we aggregate across admin-2 units based on administrative boundary data from the Global Administrative Areas (GADM, `https://gadm.org/`).

We associate the development proxies with luminosity running the following specification:

$$Y_{g,c,t} = \beta NL_{g,c,t} + \gamma ln(area)_g + \mu_{c(g),t}[+\delta_g] + \epsilon_{g,c,t} \tag{1}$$

$Y_{g,c,t}$ denotes the average of the socio-economic outcome (schooling, composite wealth index, access to flush/pit toilet, and electricity) in gridcell (or administrative unit) g in country c, in a survey conducted in year t. We standardize all outcomes to have a mean zero and a standard deviation of one to enable comparisons of the coefficients. $NL_{g,c,t}$ is either the log sum of nightlights plus half of the minimum positive value or an indicator that equals one if the gridcell is lit. The specifications include country-year fixed effects $\mu_{c(g),t}$. The cross-sectional specifications also control for the logarithm of the gridcell's/unit' area, $ln(area)_g$, which is absorbed by the unit (pix, admin-area) fixed-effects $[\delta_g]$ in the panel estimation.

4.2 Cross-sectional Estimates

Figure 3 panels (a)-(b) plot the cross-sectional estimates with the harmonized and adjusted DMSP-VIIR series (red diamonds) and the unadjusted ones (blue squares). Two results emerge. First,

[10] We choose this age range because, in the panel specifications, we want to capture the 'flow' of education. We get similar results using an upper age of 65.

nightlights are a suitable proxy for local development, as we obtain significant correlations across all specifications with both the harmonized and adjusted series and the unadjusted ones. The estimates in panel (b) hover around $0.5-0.6$, suggesting that lit areas have a half standard deviation higher schooling and access to electricity and pit/flush toilets; the coefficient on the wealth index suggests differences of one standard deviation. The correlation, however, is far from perfect as the binary luminosity index cannot fully capture the considerable spatial variation in development (Appendix Figure B4). Second, with all outcomes, we obtain *more robust and less noisy* correlations with the newly compiled, adjusted, and merged VIIRS-DMSP series compared to the unadjusted ones. Likewise, when we look at the extensive margin of luminosity, all coefficients increase with the harmonized series, consistent with a measurement error interpretation, as using less noisy explanatory variables yields less attenuated estimates and a higher R^2 (Wooldridge, 2010).

4.3 Panel Estimates

To examine the dynamic correlation between development and luminosity, we augmented the cross-sectional specification with admin-unit or gridcell-level constants (δ_g). All panel specifications reported in panels (c) and (d) of Figure 3 yield positive coefficients. [Appendix Table B8 reports the estimates for non-standardized measures.] The coefficients with the adjusted series are *always* larger than the analogous ones with the unadjusted ones, telling of the reduction in measurement error that our adjustments to the DMSP series and merging to the downgraded VIIRS achieve. For example, the estimate on log lights in the specification with years of schooling is around 0.02 when we use the newly compiled series, about double the coefficient with the unadjusted series near 0.01. Similar results hold when the outcome is the composite wealth index or the share of households in the gridcell with electricity access. The comparison of the specifications examining the association between development and the extensive margin of lights with the new and the unadjusted series in panel (d) yields starker patterns. All specifications with the unadjusted series yield indistinguishable from zero estimates, while almost all permutations with the adjusted series yield significant correlations. The harmonized series suggests that mean years of schooling increase, on average, by 0.05 standard deviations in gridcells turning lit, compared to unlit; this translates into 0.125 schooling years (Appendix Table B8). When gridcells turn lit with the newly-compiled series, the Wealth Index and Electricity Access increase by 0.06 standard deviations.

4.4 Further Evidence[11]

4.4.1 Spatial Aggregation

Applied research uses luminosity data across spatial units of various sizes, some coarse (Alesina *et al.*, 2016 at admin-1 and admin-2 units and linguistic areas), some granular (Henderson *et al.*, 2018 at small gridcell level; Storeygard, 2016 at city-level). Figure 4 panels (a) and (b) provide graphical illustrations of the luminosity-wealth elasticity across spatial units of various sizes to explore the implications of aggregation. The furthest to the left point gives the coefficient from a specification across small units, 2×2 gridcells (0.5×0.5 decimal degrees). As one moves along the x-axis, the data is aggregated into larger units, with the largest being 12×12 gridcells.

The cross-sectional estimates (panel a) are highly significant, around 0.18, across all aggregation levels. All panel estimates (panel b) are more than two standard deviations larger than zero, showing that luminosity approximates well variation in household assets and public good access. Besides, the coefficients are fairly stable across aggregation levels, around 0.07. The estimates are almost always larger with the newly-compiled harmonized luminosity series that fuses the VIIRS into the adjusted for top-coding, sensor calibration, and blooming DSMP series (red markers) compared to the merging of VIIRS to the unadjusted DSMP. The "correction" and the associated reduction in measurement error are more important at the lowest levels of aggregation. As research moves into more granular analyses, it must use the harmonized series and carefully consider measurement error. In contrast, aggregating light data to coarser spatial units reduces the impact of measurement error in the unadjusted lights series, as evidenced by cross-country analyses showing minimal differences in GDP-luminosity elasticity using adjusted and unadjusted lights series.

4.4.2 Localized Variation

Researchers commonly use identification designs, such as local fixed effects models Wantchekon *et al.* (2015)) or spatial regression discontinuity designs Michalopoulos and Papaioannou (2014)), to advance on causation by comparing proximate areas. We thus assess how well luminosity captures local development, focusing on estimates within increasingly proximate areas, adding fixed effects of increasing spatial resolution (in the panel estimates interacted with year constants) to partial-out localized hard-to-account-for features related to location, ecology, and culture.

Figure 4 panels (c) and (d) plot the luminosity - wealth correlation with fixed effects of various sizes. Moving along the x-axis, we plot estimates with larger (coarser) fixed effects. The furthest to the left specification includes fixed-effects of blocks of 2×2 gridcells (0.5×0.5 decimal degrees); the furthest to the right specification includes 12×12 gridcell block fixed-effects. The cross-sectional estimates are quite stable; the coefficients with the harmonized and adjusted series are

[11] The results, reported below, with the DHS composite wealth index, are similar to the ones with mean years of schooling and access to clean piped water and electricity. We thus report them in the Appendix Section B.2

$0.17 - 0.19$, somewhat larger than with the unadjusted series (about 0.15). The within-gridcell estimates highlight the improvement the harmonized DMSP-VIIRS series achieves vis a vis more agnostic approaches. The coefficient is significantly positive and stable. Even when comparing nearby areas, changes in luminosity approximate changes in household wealth. In contrast, the coefficients of the unadjusted luminosity series are smaller and, in many permutations, statistically indistinguishable from zero. As shown in the Appendix Section B.2, the patterns are similar with schooling and public goods.

4.4.3 Urban and Rural

Another issue with light data regards their accuracy in explaining well-being in urban and rural areas. Appendix Figure B8 plots the luminosity coefficients for the four development outcomes separately for urban and rural households, using DHS classification. All cross-sectional estimates are highly significant, suggesting that nighttime lights proxy well schooling, household wealth, and public goods access across both urban (higher luminosity) and rural (low luminosity) areas. Besides, the estimates are similar in the rural and urban samples. The panel specifications yield somewhat different patterns. First, the coefficients are statistically significant only when using the newly compiled harmonized and adjusted for top-coding, blooming, and sensor calibration lights series. Second, the urban sample estimates are consistently larger than the ones in the rural sample, showing that the development-luminosity nexus is stronger in urban areas, an asymmetry that echoes the recent findings across India of Asher *et al.* (2021).

5 Mozambique (Census-based Estimates)

Many studies use luminosity from specific countries to explore various inquiries, such as the geographic impact of demonetization in India (Chodorow-Reich *et al.*, 2020), landmine clearance in Mozambique (Chiovelli *et al.*, 2019), and the flattening of the government hierarchy in China (Li *et al.*, 2016). Recent work in Namibia shows that luminosity approximates better local development, as reflected in census data, than surveys (I. *et al.*, 2022). We thus examine the association between the newly compiled luminosity series and local development using all Mozambican Censuses that allow us to zoom in at high spatial resolution with many observations.

Specification and Sample We estimate linear specifications across Mozambican administrative units linking development to luminosity. We have retrieved, processed, and digitized the full censuses of 1997, 2007, and 2017. We estimate the following equation:

$$Y_{i,t} = \beta L_{i,t} + \gamma_a ln(area)_i + \mu_{j(i),t} + [\delta_i +]\epsilon_{i,t} \tag{2}$$

$Y_{i,t}$ denotes the average years of schooling of Mozambicans 15-19 years and non-agriculture employment of 15-24-year-olds in administrative unit i in period t.[12] $L_{i,t}$ is either the log sum of nightlights plus half of the minimum positive value or a lit indicator. $\mu_{j(i),t}$ denote broader administrative units j-year fixed effects. The cross-sectional specification also controls for log geographic area $ln(area)_i$. δ_i are administrative unit fixed effects that account for geography, location, and other time-invariant factors in the panel estimation. Appendix Table B9 gives summary statistics across $1,126$ admin-4 units (*localides*).

Cross-Sectional Estimates Figure 5 panels (a) and (b) report the cross-sectional estimates with the two transformations of luminosity. The specifications reveal a strong luminosity-development correlation, further illustrating the usefulness of luminosity to approximate localized differences in education and employment in the "modern" sector. Besides, the coefficients with the harmonized and adjusted series are stronger than the analogous ones with the "unadjusted" light data, showing the reduction in measurement error from top-coding and blooming. The estimates imply that years of schooling and employment in the modern sectors are (at least) half a standard deviation higher in lit as compared to unlit localities, about 0.5 years and 10 percentage points, respectively. The luminosity-development correlation is also present when we add admin-3 unit fixed effects (*postos*) to exploit localized variations across proximate localities. Appendix Tables B10-B11 reports cross-sectional estimates also across 142 admin-2 areas (*distritos*), including admin-1 (provinces) fixed effects and across 403 admin-3 units (*postos*), including admin-2 fixed effects. Luminosity is a significant proxy of education and non-agriculture employment across all administrative splits. As with the DHS analysis, the improvement in estimates from the harmonized and adjusted for top-coding, blooming, and sensor calibration series is mostly noticeable at finer spatial resolution.

Dynamic Correlations Figure 5 panels (c) and (d) give panel estimates (with admin-4 unit fixed-effects) that explore the dynamic association between luminosity and development. Within-locality changes in luminosity correlate significantly with swings in schooling and out-of-agriculture employment. As shown in Appendix Tables B12-B13, the correlation is strong across all levels of spatial aggregation (across admin-2 and admin-3 units). Luminosity co-moves with schooling and modern-sector employment even when we augment the specifications with interactions between census-year constants and admin-3 unit fixed-effects that allow us to zoom within geographically proximate areas and account for quite localized unobserved trends. Besides, the panel estimates for non-agricultural employment are considerably larger with the newly compiled fused VIIRS to the adjusted DMSP series luminosity series. Appendix Figure B10 illustrates the within-locality patterns plotting the increase in schooling years for four groups of localities; initially (in 1997)

[12] Employment status is not available for the 2017 census. We work with young employment and schooling to better approximate changes in economic conditions in the panel estimation.

unlit admin-4 units that either stay unlit (by 2007 or 2017) or turn lit and initially lit localities that either stay lit or turn unlit. The difference in schooling is about half a year when comparing localities turning lit (from unlit) or staying lit (rather than becoming unlit), even when we compare nearby localities with the inclusion of admin-3 fixed-effects.

6 Discussion

While satellite nighttime lights have gained widespread popularity in applied research, it is still unclear when and where the luminosity data is a dependable proxy of economic development. The debate is especially pertinent in situations involving high-resolution analyses in low development regions, where a significant portion of pixels are dark. But it is in low-income areas that satellite imagery are *a priori* needed, as in such environments there are limited high quality data.

This paper compiles a novel annual series of nighttime light data encompassing all African countries across roughly one square kilometer pixels. Employing ensemble methods, we standardize luminosity data from various satellites with sensors of differing resolutions and accuracy. The new series accounts for various intricacies related to variations in sensor quality, top-coding, blooming phenomena, and the transition from the DMSP-OLS to the VIIRS satellite systems in 2013.

The second part of our study explores the relationship between luminosity and local development proxies such as education, household wealth, sectoral employment, and access to public services. We draw upon geo-referenced survey data from 34 African countries and full-census data from Mozambique. By harnessing both cross-sectional and temporal variation, we show that the newly harmonized luminosity series effectively encapsulates local development (dynamics). This holds true even at highly granular levels, where the challenge of excess-zero observations is pronounced. The adjusted and harmonized luminosity series correlate much stronger with all development proxies, even more so in changes and at granular levels, telling of the reduction in measurement error that the newly compiled series achieves. We view our work as offering insights into the ongoing discourse concerning the reliability and utility of satellite nighttime light data in economic development in regions grappling with data limitations. Besides, as more satellite data become available to researchers, future research can blend the newly compiled nighttime lights series with other granular data, such as daytime economic activity (traffic) and imagery of structures to provide high-quality mappings of well-being in low-income settings.

References

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

Asher, S., Lunt, T., Matsuura, R. and Novosad, P. (2021). Development Research at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India using the SHRUG Open Data Platform. *World Bank Economic Review*, **35** (4).

Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, **11** (1), 685–725.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A. and Zhang, Q. (2017). Roads, railroads, and decentralization of chinese cities. *The Review of Economics and Statistics*, **99** (3), 435–448.

Bilicka, K. and Seidel, A. (2022). Measuring Firm Activity from Outer Space, nBER WP 29945.

Bleakley, H. and Lin, J. (2012). Portage and path dependence. *Quarterly Journal of Economics*, **127**, 587–644.

Bluhm, R. and Krause, M. (2020). Top lights: Bright cities and their contribution to economic development, soDa Laboratories Working Paper Series No. 2020-08.

— and McCord, G. C. (2022). What can we learn from nighttime lights for small geographies? measurement errors and heterogeneous elasticities. *Remote Sensing*, **14** (5).

Breiman, L. (2001). Random forests. *Machine Learning*, **45**, 5–32.

Cao, X., Hu, Y., Zhu, X., Shi, F., Zhuo, L. and Chen, J. (2019). A simple self-adjusting model for correcting the blooming effects in dmsp-ols nighttime light images. *Remote Sensing of Environment*, **224**, 401—411.

Ch, R., Martin, D. A. and Vargas, J. F. (2021). Measuring the size and growth of cities using nighttime light. *Journal of Urban Economics*, **125**.

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

Chiovelli, G., Michalopoulos, S. and Papaioannou, E. (2019). Landmines and Spatial Development, nBER WP 24758.

Chodorow-Reich, G., Gopinath, G., Mishra, P. and Narayanan, A. (2020). Cash and the economy: Evidence from india's demonetization. *The Quarterly Journal of Economics*, **135** (1).

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

Dreher, A., Fuchs, A., Hodler, R., Parks, B. C., Raschky, P. A. and Tierney, M. J. (2019). African leaders and the geography of china's foreign assistance. *Journal of Development Economics*, **140**, 44–71.

Elvidge, C. D., Zhizhin, M., T., G., FC, H. and J, T. (2021). Annual time series of global viirs nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing of Environment*, **68** (1), 77–88.

Geurts, P., Ernst, D. and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, **63**, 3—-42.

Gibson, J. (2021). Better night lights data, for longer*. *Oxford Bulletin of Economics and Statistics*, **83** (3), 770–791.

—, Olivia, S., Boe-Gibson, G. and Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, **149**, 102–602.

Harari, M. (2020). Cities in bad shape: Urban geometry in india. *American Economic Review*, **110**, 2377—-2421.

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

—, Storeygard, A. and Deichmann, U. (2017). Has climate change driven urbanization in africa? *Journal of Development Economics*, **124**, 60–82.

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

Hjort, J. and Poulsen, J. (2019). The arrival of fast internet and employment in africa. *American Economic Review*, **109** (3), 1032—1079.

Hodler, R. and Raschky, P. (2014). Regional favouritism. *Quarterly Journal of Economics*, **129** (2), 995–1023.

I., M., Ferreira, T. and Lessmann, C. (2022). Nighttime lights and wealth in very small areas. *Review of Regional Research*, **42**, 161–190.

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

Khachiyan, A., Thomas, A., Zhou, H., Hanson, G., Cloninger, A. and Rosing, T. (2022). Using neural networks to predict micro-spatial economic growth. *American Economic Review: Insights*, forthcoming.

Levin, N., Kyba, C. C., Zhang, Q., Sánchez de Miguel, A., Román, M. O., Li, X., Portnov, B. A., Molthan, A. L., Jechow, A., Miller, S. D., Wang, Z., Shrestha, R. M. and Elvidge, C. D. (2020). Remote sensing of night lights: A review and an outlook for the future. *Remote Sensing of Environment*, **237**.

Li, P., Lu, Y. and Wang, J. (2016). Does flattening government improve economic performance? evidence from china. *Journal of Development Economics*, **123**, 18–37.

Li, X. and Zhou, Y. (2017). A stepwise calibration of global dmsp/ols stable nighttime light data (1992–2013). *Remote Sensing*, **9** (6).

—, —, Zhao, M. and Zhao, X. (2020). A harmonized global nighttime light dataset 1992–2018. *Nature: Scientific Data*, **7** (168).

Lipton, Z. C., Elkan, C. and Naryanaswamy, B. (2014). Optimal thresholding of classifiers to maximize f1 measure. In T. Calders, F. Esposito, E. Hüllermeier and R. Meo (eds.), *Machine Learning and Knowledge Discovery in Databases*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 225–239.

Lu, F. and Vogl, T. (2022). Intergenerational Persistence in Child Mortality. *American Economic Review: Insights*, forthcoming.

Martinez, L. (2022). How much should we trust the dictator's gdp growth estimates? *Journal of Political Economy*, **130** (10), 2731–2769.

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

— and — (2014). National institutions and sub-national development in africa. *Quarterly Journal of Economics*, **129** (1), 151–213.

— and — (2018). Spatial patterns of development: A meso approach. *Annual Review of Economics*, **10**, 383–410.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, **31** (2), 87–106.

Nechaev, D., Zhizhin, M., Poyda, A., Ghosh, T., Hsu, F.-C. and Elvidge, C. (2021). Cross-sensor nighttime lights image calibration for dmsp/ols and snpp/viirs with residual u-net. *Remote Sensing*, **13** (24).

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

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

Sagi, O. and Rokach, L. (2018). Ensemble learning: A survey. *WIREs Data Mining Knowl Discov.*, **8**.

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

Wantchekon, L., Klašnja, M. and Novta, N. (2015). Education and human capital externalities: Evidence from colonial benin. *The Quarterly Journal of Economics*, **130** (2), 703–758.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

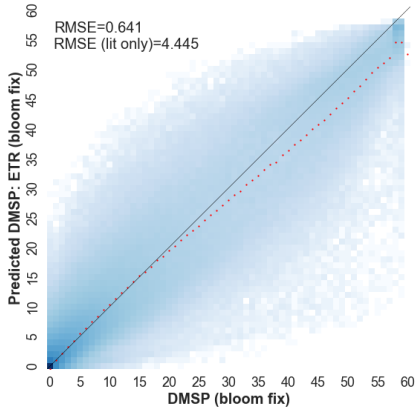
Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S. and Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature Communications*, **11** (2583).

Young, A. (2012). The african growth miracle. *Journal of Political Economy*, **120** (4), 696—739.

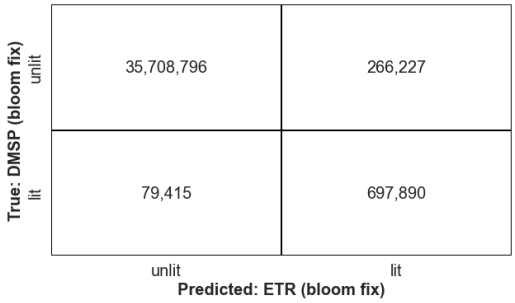
Figure 1: Mapping Nighttime Lights. VIIRS and DMSP in 2013

Extremely Randomized Trees (Ensemble Method)

(a) Scatter Ensemble Method



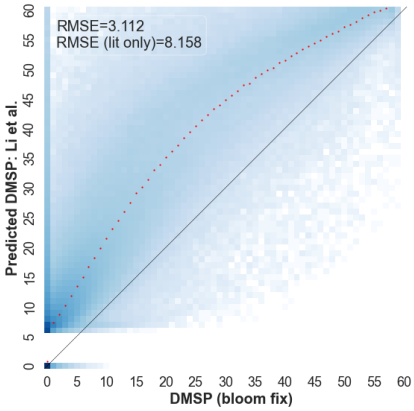
(b) Confusion Matrix Ensemble Method



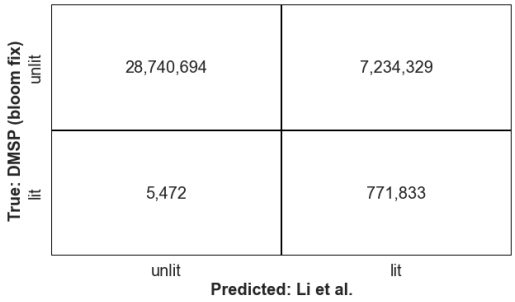
Recall=0.90, Precision=0.72, F1 score=0.80

Sigmoid Function Approach Li *et al.* (2020)

(c) Scatter Sigmoid 'DMSP-like'



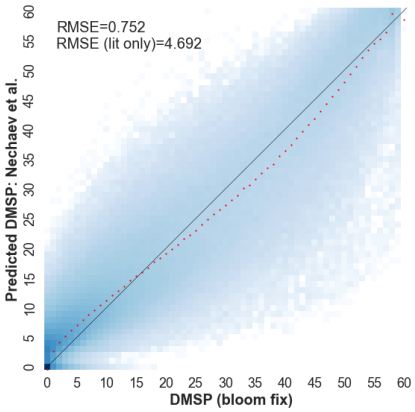
(d) Confusion Matrix Sigmoid 'DMSP-like'



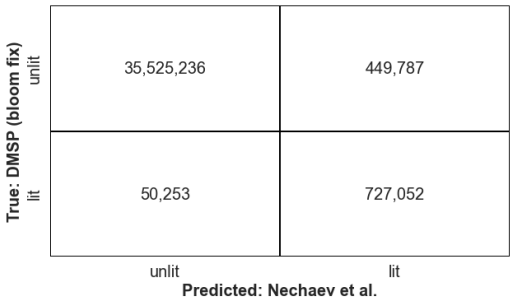
Recall=0.99, Precision=0.10, F1 score=0.18

Convolutional Neural Network Approach Nechaev *et al.* (2021)

(e) Scatter Neural Network 'DVNL'



(f) Confusion Matrix Neural Network 'DVNL'

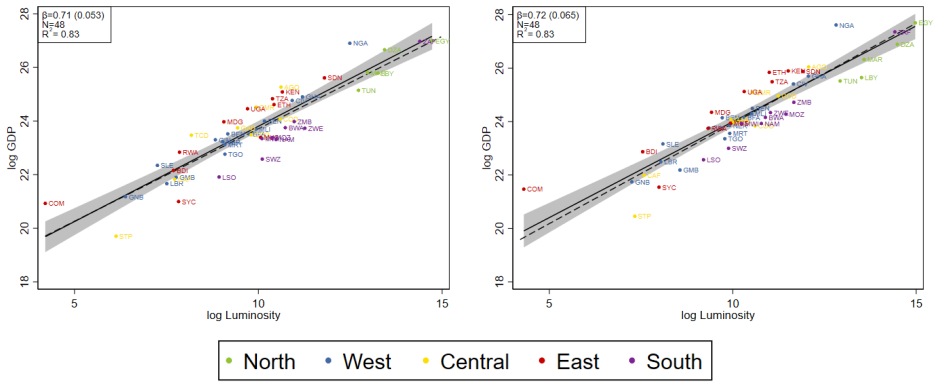


Recall=0.94, Precision=0.62, F1 score=0.74

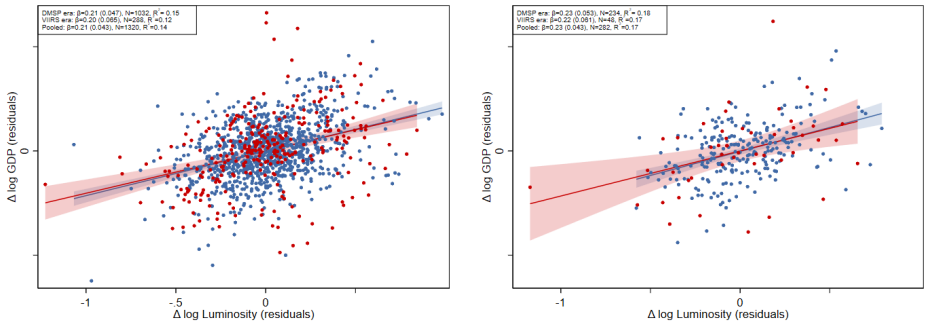
This figure gives illustrations of the pixel-level mapping of luminosity in 2013, when both the VIIRS and the DSMP series are available. Panels (a)-(b) report out-of-sample estimates with our extremely randomized trees, ensemble, method. Panels (c)-(b) give estimates with the sigmoid function of Li *et al.* (2020) that yields 'DMSP-like' downgraded VIIRS series. Panels (e) and (f) give the tabulation of the convolutional neural network method of Nechaev *et al.* (2021), 'DVNL'. Panels (b), (d), and (f) give confusion matrices looking at the extensive margin of luminosity in 2013.

Figure 2: GDP-Luminosity Elasticity. Country-level Estimates

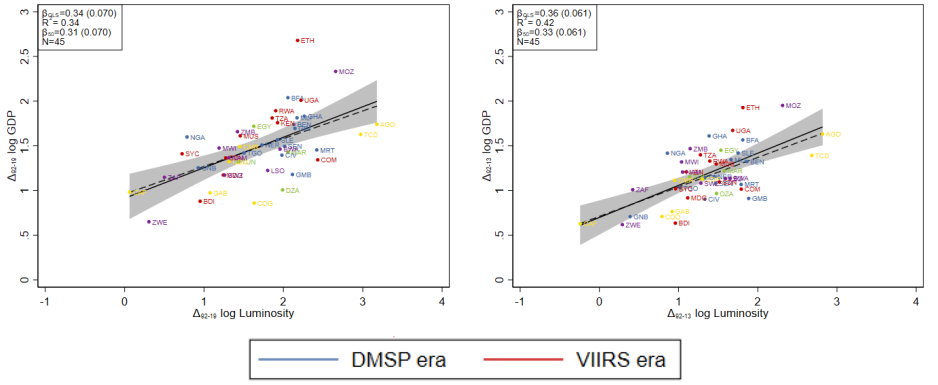
(a) Cross-sectional - log GDP 2005 (b) Cross-sectional - log GDP 2015



(c) Panel Estimates - Annual (d) Panel Estimates - 5-year



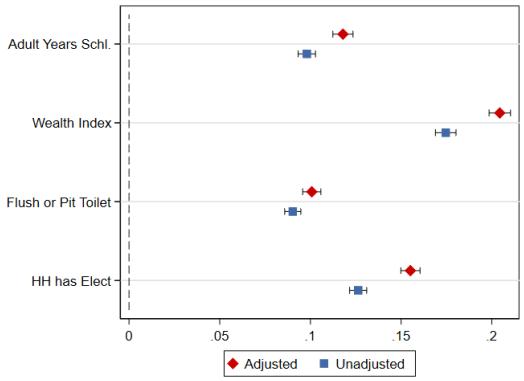
(e) Long-Differences - Δ 1992-2019 (f) Long-Differences - Δ 1992-2013



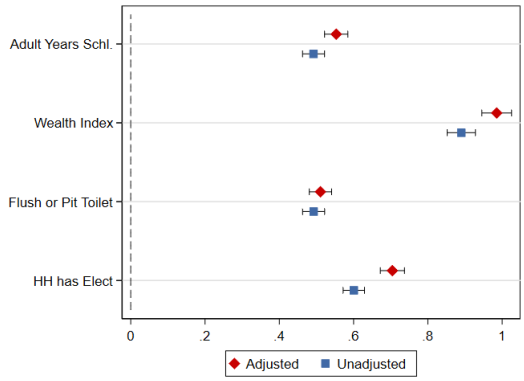
Panels (a) and (b) plot the cross-country association between the log of GDP and the log of nighttime lights (luminosity) across African countries in 2005 (DMSP period) and in 2015 (VIIRS period), alongside the LS regression line (solid line) and the median regression line (dashed). Panels (c) and (d) plot the panel association between log GDP and the log nighttime lights (luminosity) across African countries and years during the period 1992-2019. Panel (c) uses all observations; panel (d) uses five-year average values. The panels plot the residuals of log GDP and log luminosity on country fixed-effects and year fixed-effects. Blue dots indicate country-year residuals during 1992-2012 (DMSP period) and the red dots indicate residuals during 2013-2019 (VIIRS period). The figure also gives the LS and median regression (dashed) lines and associated standard errors during the DMSP period (blue) and the VIIRS period (red). Panels (e) and (f) plot the long-difference association between log GDP and log luminosity over 2019-1992 and 2013-1992, respectively. All specifications use the (logarithm of the) newly compiled luminosity series harmonized VIIRS-DMSP after adjusting the DMSP data for top coding, sensor calibration, and blooming. Countries in panels (a), (b), (e), and (f) are colored according to their broad African region (East, West, North, Central, and Southern).

Figure 3: Local Development-Luminosity Correlation. DHS Analysis

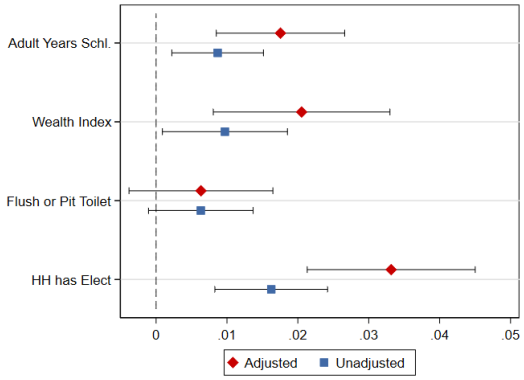
(a) Cross-sectional Estimates - Log Nightlights



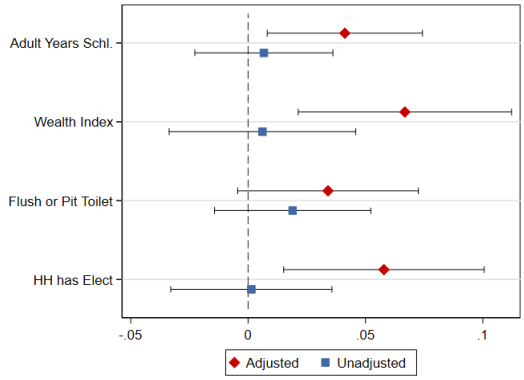
(b) Cross-sectional Estimates - Lit Indicator



(c) Panel Estimates - Log Nightlights



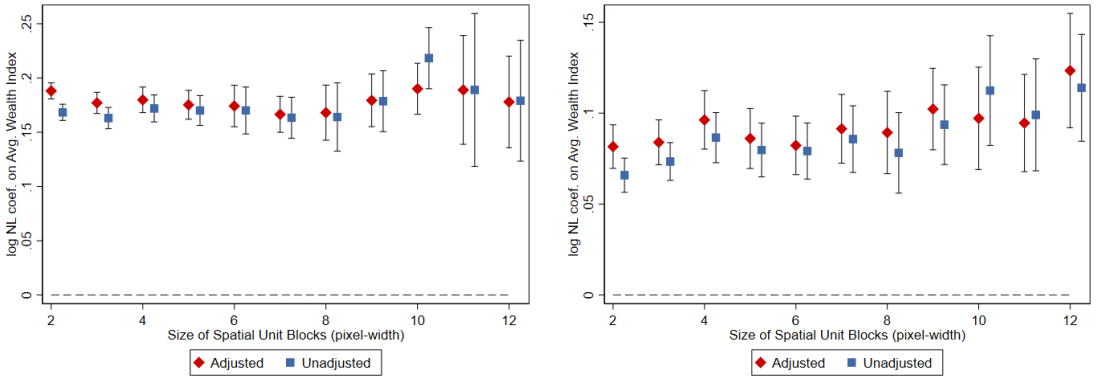
(d) Panel Estimates - Lit Indicator



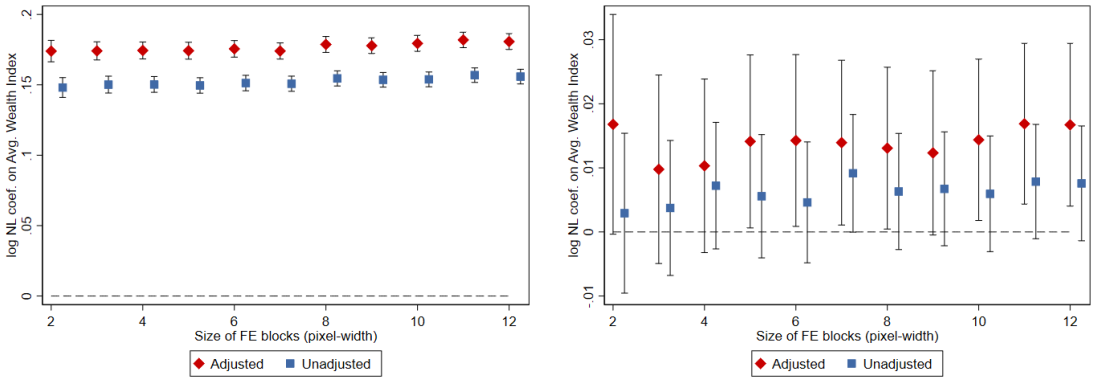
The figure plots coefficients from regressions associating proxies of development from the Demographic and Health Surveys (DHS) on night-time lights (luminosity). Panels (a) and (c) use log luminosity; panels (b) and (d) use an indicator that equals one when the gridcell is lit and zero otherwise. Panels (a) and (b) control for country-survey-year fixed effects and log cell area. Panels (c) and (d) also include grid-cell fixed effects. All outcome variables, mean years of schooling of the population aged 15-39, a composite wealth index, and access to piped water and electricity are standardized to have a mean of zero and a standard deviation of one. The bars give 95% confidence intervals based on standard errors clustered at the grid-cell level.

Figure 4: Household Wealth-Luminosity Correlation. Further Evidence

(a) Cross-sectional, Varying Spatial Unit Size (b) Panel Estimates, Varying Spatial Unit size

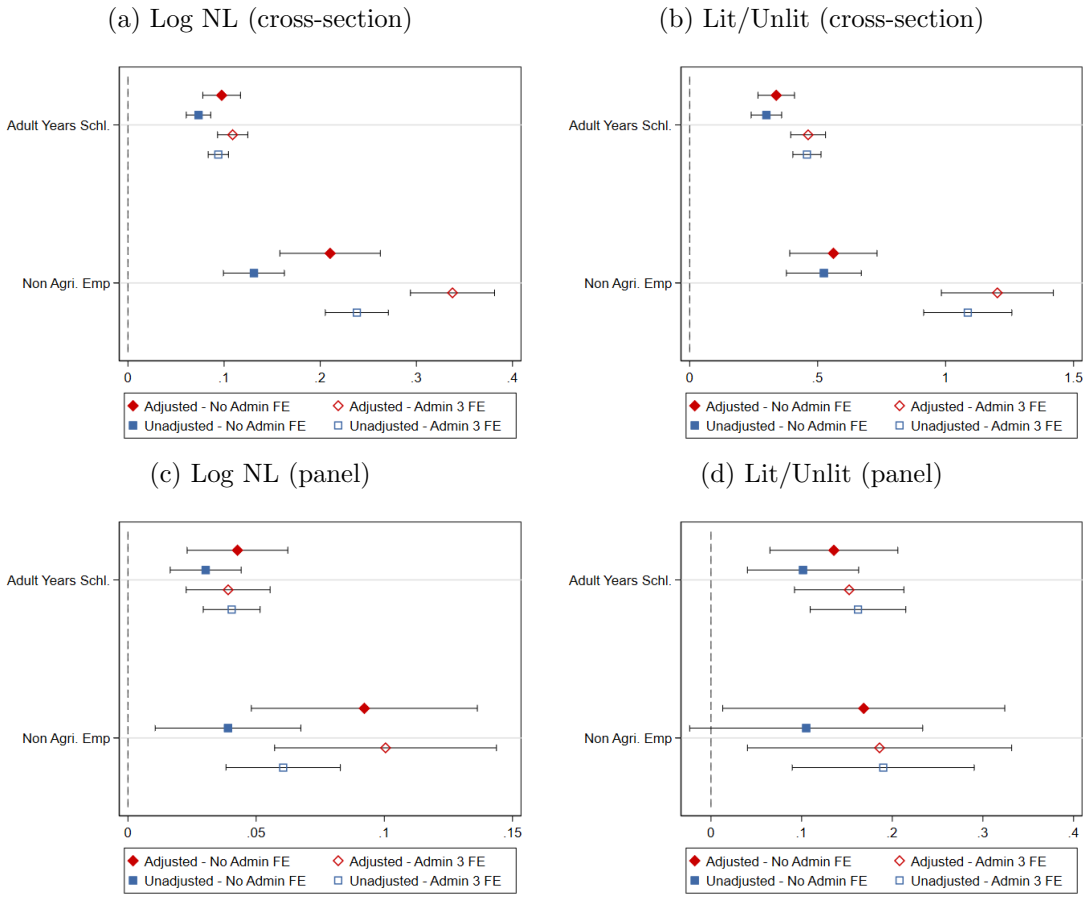


(c) Cross-sectional, Varying Fixed-Effects Size (d) Panel Estimates, Varying Fixed-Effects Size



The figure plots coefficients from regressions associating the DHS Composite Wealth Index on Log Luminosity. Panels (a) and (c) give cross-sectional estimates with country-survey year constants, controlling also for the gridcell's log land area. Panels (b) and (d) give panel estimates that, besides the country-year constants, also include gridcell fixed effects. Panels (a)-(b) plot coefficients of log luminosity varying the (gridcell) size unit of the empirical analysis. Panel (c) plots cross-sectional coefficients of log luminosity holding gridcell size fixed and augmenting the specification with block fixed effects of increasing size. Panel (d) plots panel coefficients of log luminosity augmenting the specification with interactions between country-survey-year constants with block fixed effects of increasing sizes. Red markers denote estimates using the harmonized VIIRS-DMSP luminosity series, adjusting the DMSP for top coding, sensor calibration, and blooming. Blue markers denote estimates using the unadjusted merged VIIRS-DMSP luminosity series. The bars denote 95% confidence intervals, based on standard errors clustered at the grid-cell level for panels (c) and (d) and at the spatial unit level for panels (a) and (b).

Figure 5: Luminosity and Local Development. Mozambique Census Analysis



The figure plots coefficients from regressions associating mean years of schooling of individuals aged 15-39 and employment outside agriculture (in services, manufacturing, and mining) with night-time lights luminosity across Mozambican localities, level-4 administrative units. Panels (a) and (c) use the natural logarithm of nightlights, adding a small number. Panels (c) and (d) employ a luminosity indicator variable that equals one if the administrative unit is lit and zero otherwise. Schooling years are computed for all three Mozambican censuses (1997, 2007, and 2017). The share of non-agricultural employment is calculated using the 1997 and 2007 censuses. Panels (a) and (b) give cross-sectional estimates. Solid red diamonds and solid blue squares condition on the log admin area and year constants. The hollow circle/square also conditions on interactions between year constants and admin-3 fixed effects. Panels (c) and (d) give panel estimates with locality (admin-4) fixed-effects and year constants. The hollow circle/square specifications also condition on interactions between year constants and admin-3 fixed effects. The two outcome variables are standardized to have a mean of zero and a standard deviation of one. The bars represent 95% confidence intervals, and standard errors are clustered at the admin-3 level.

Supplementary Data Appendix
Illuminating Africa?

Giorgio Chiovelli Stelios Michalopoulos
Elias Papaioannou Tanner Regan

November 23, 2023

Table of Contents

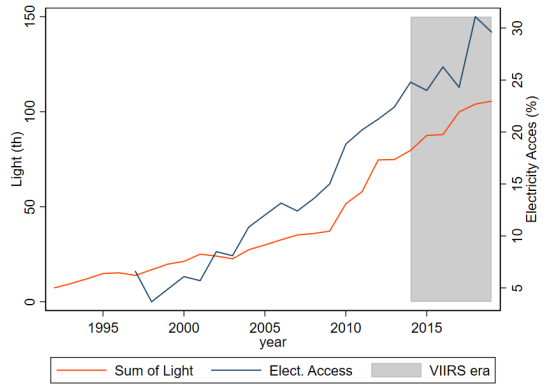
A Data and Methodology

Appendix Figure A1 demonstrates the significant within-country over-time correlation between the newly-compiled harmonized luminosity series, based on the downgrading and merging of the VIIRS series to the adjusted for senor quality, top-coding, and blooming DMSP series. The figure plots the harmonized nighttime light data alongside the share of the population with electricity in Mozambique, Kenya, the Democratic Republic of Congo, Ghana, Tanzania, and Nigeria using data from the World Bank's Development Indicators Database. Two results emerge. First, luminosity correlates with electricity access in all countries. Second, there was no major change in luminosity from 2012 to 2014, when we switched from the DMSP to the VIIRS satellite system.

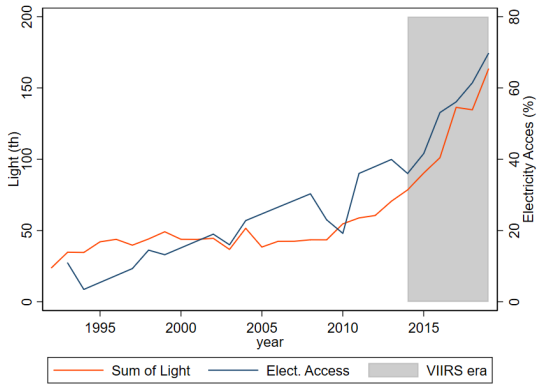
Appendix Figure A2 plots the spatial distribution of the newly compiled luminosity series harmonized across VIIRS and DMSP in Ghana in $2012 - 2015$. There are no major swings in the harmonized and adjusted VIIRS-DSMP luminosity series in $2012 - 2014$, when we switch from the DMSP to the VIIRS satellite system.

Figure A1: Electricity Access and Harmonized Luminosity Series across African Countries

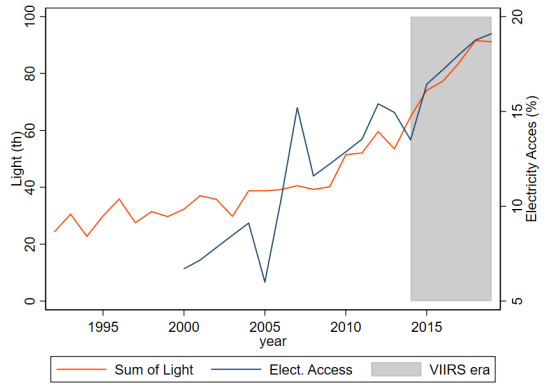
(a) Mozambique



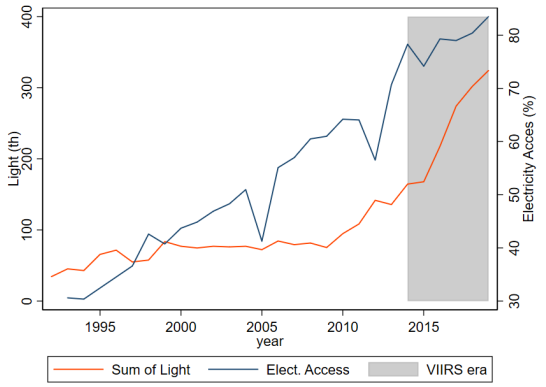
(b) Kenya



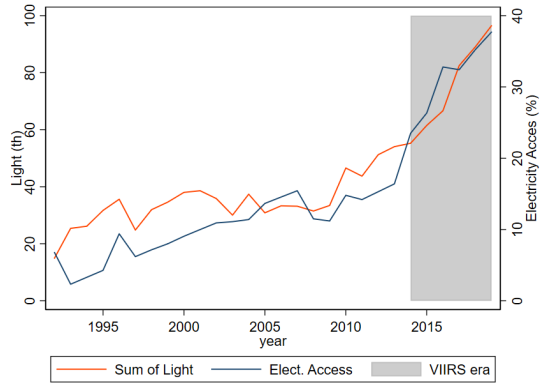
(c) Democratic Republic of Congo



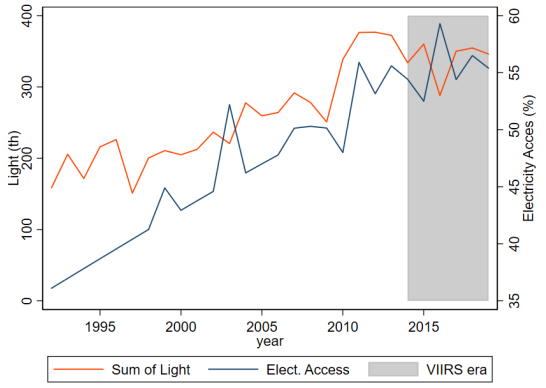
(d) Ghana



(e) Tanzania



(f) Nigeria

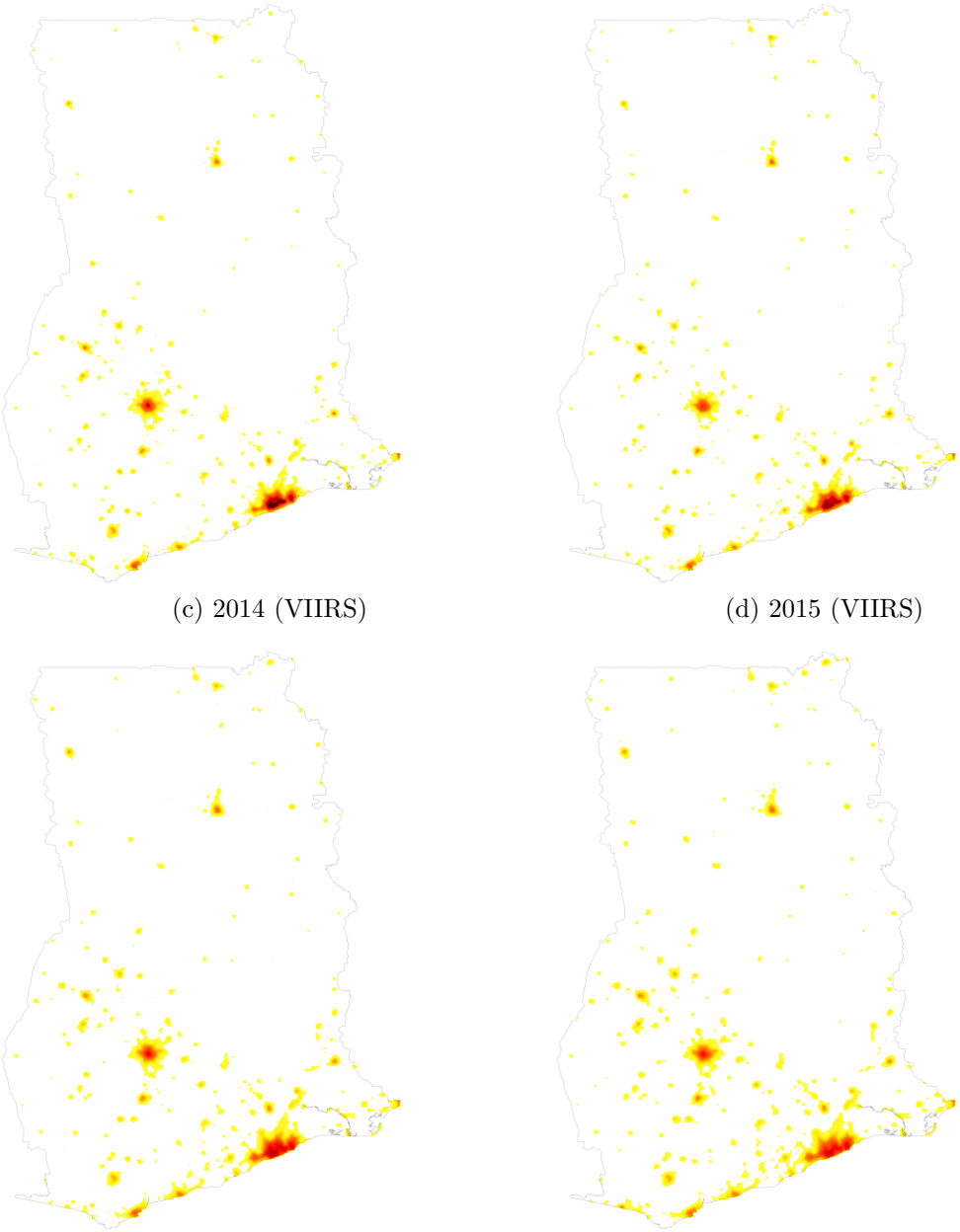


The figure plots trends in Electricity Access (as a share of the total population) and the sum of the harmonized and adjusted VIIRS-DSMP luminosity series (in thousands of DNs) across six African countries. Data on electricity access come from World Bank's World Development Indicators Database.

Figure A2: Spatial Distribution of Economic Activity, Ghana 2012-2015

Harmonized Nighttime Lights Series

(a) 2012 (DMSP) (b) 2013 (DMSP)



(c) 2014 (VIIRS) (d) 2015 (VIIRS)

The figure plots the distribution of the harmonized across satellite systems (VIIRS and DMSP) and adjusted for blooming, sensor quality, and top-coding nighttime lights series in Ghana over 2012 – 2015.

B Supplementary Evidence

B.1 Luminosity - GDP Elasticity

Appendix Tables B1-B3 report cross-country regression estimates exploring the correlation between GDP (Gross Domestic Product) and the newly compiled luminosity series that fuse the VIIRS data (2013−2020) to the DMSP data (1992−2013). The estimates, therefore, complement the (graphical) analysis of the cross-country GDP-luminosity elasticity in Section 3 of the main paper. Columns (1)-(3) give the estimates when we use the harmonized and adjusted for top-coding, sensor calibration, and blooming/bleeding VIIRS-DSMP luminosity series. Columns (4)-(6) give, for comparability, estimates using the luminosity series that fuses VIIRS to the DMSP series without adjusting for blooming and top-coding.

Cross-sectional Patterns Appendix Table B1 - Panel *A* reports pooled across years cross-sectional regression estimates of the following form:

$$\ln GDP_{c,t} = \beta \ln NL_{c,t} + \gamma_a \ln(area)_c + \gamma_p \ln pop_{c,t} + \mu_t + \epsilon_{c,t} \tag{3}$$

$ln(GDP)_{c,t}$ denotes the logarithm of current GDP (in PPP terms) of country c in year t; $ln(NL)_{c,t}$ is the log sum of the merged DMSP-VIIRS nightlights; $ln(pop)_{c,t}$ is the log of population while $ln(area)_c$ denotes log land area. All specifications include year constants, μ_t, that capture the increase in development and luminosity over time. Column (1) gives the lights-GDP elasticity across the entire period, $1992 - 2019$; columns (2) and (3) look at the periods where only DMSP and only VIIRS are available, respectively. The luminosity GDP elasticity is about 0.53; the coefficient is highly significant, showing that luminosity is a good proxy of output across the 48 African countries. The fit is strong with an adjusted R^2 around 0.93. As shown in Appendix Table B2, the GDP-luminosity elasticity is stable when we exclude island nations (Comoros, Equatorial Guinea, Mauritius, Mayotte, Reunion, Sao Tome and Principe, and Seychelles) and augment the specifications with broad regional constants (southern, central, western, eastern, and northern Africa). As shown in columns (4)-(6) of Tables B1-B2, the lights-GDP elasticity of the merged series adjusting for the deficiencies of DMSP is somewhat smaller as compared to the unadjusted DMSP data (about 0.62), though the fit is similarly strong.

Panel Estimates Appendix Table B1 - Panel *B* examines the dynamic association between nighttime lights and GDP. To do so, we augment the cross-sectional specification with country-fixed effects (δ_c). The regression equation reads:

$$\ln GDP_{c,t} = \beta \ln NL_{c,t} + \mu_t + \delta_c + \epsilon_{c,t} \tag{4}$$

The regression analysis, therefore, complements the graphical illustrations in Figure 2 panels (c)-(d) in the main paper. GDP-luminosity elasticity with the newly compiled harmonized luminosity series that fuses the higher quality and more granular VIIRS series to the DMSP series adjusted for top-coding, blooming, and sensor quality in column (1)-(3) hovers around $0.21 - 0.26$. The elasticity is quite stable when we drop island nations or interact the year constants with broad African region fixed effects to account for regional trends (Appendix Table B3). When we use the merged VIIRS to the unadjusted DMSP (in (4)-(6)), we obtain again a highly significant elasticity, which is somewhat higher (around $0.26 - 0.31$).

Long-Run Differences Appendix Table B1 - Panel *C* plots the correlation between long-run changes in GDP and night-time lights in long differences. Taking the long-run differences (over 2019-1992 and 2013-1992) reduces noise in both GDP and nighttime lights, which is likely considerable in the annual frequency. The specification reads.

$$\Delta \ln GDP_c = \beta \Delta \ln NL_c + \delta \Delta \ln Pop_c + [\mu_r] + \epsilon_c \tag{5}$$

where $\Delta ln(GDP)$ is the change in log GDP, $\Delta ln(NL)_c$ the change in log nightlights, $\Delta ln(Pop)_c$ the change in log population, and $\mu_{r(c)}$ are broad region r constants. Consistent with the graphical illustrations in Figure 2, panels (e)-(f), the elasticity hovers around 0.3 (cols 1-3), close to the ones reported by Henderson *et al.* (2012) across more than 170 countries for $1992/3 - 2005/6$ $[0.30 - 0.33.]$ The median regression estimates, reported in Appendix Table B4, are quite similar (around 0.3).

Table B1: GDP - Luminosity Elasticity. Cross-Country Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) ln(GDP)	(2) ln(GDP)	(3) ln(GDP)	(4) ln(GDP)	(5) ln(GDP)	(6) ln(GDP)
Panel A: Cross-sectional estimates						
ln(NL)	0.530***	0.530***	0.536***	0.616***	0.618***	0.617***
	(0.0388)	(0.0406)	(0.0382)	(0.0311)	(0.0316)	(0.0389)
ln(area)	-0.0147	0.00359	-0.0834*	-0.00648	0.0119	-0.0744
	(0.0449)	(0.0459)	(0.0491)	(0.0420)	(0.0418)	(0.0494)
ln(Pop.)	0.402***	0.377***	0.485***	0.359***	0.337***	0.436***
	(0.0699)	(0.0722)	(0.0716)	(0.0653)	(0.0671)	(0.0695)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1320	1032	288	1320	1032	288
Countries	48	48	48	48	48	48
R^2	0.926	0.922	0.933	0.936	0.934	0.935
Panel B: Panel Estimates						
ln(NL)	0.216***	0.219***	0.167***	0.270***	0.270***	0.202***
	(0.0442)	(0.0488)	(0.0556)	(0.0546)	(0.0555)	(0.0677)
ln(Pop.)	0.590***	0.688***	0.841	0.553**	0.654**	0.800
	(0.215)	(0.246)	(0.524)	(0.213)	(0.244)	(0.540)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1320	1032	288	1320	1032	288
Countries	48	48	48	48	48	48
R^2	0.993	0.995	0.999	0.993	0.995	0.999
within R^2	0.283	0.280	0.159	0.300	0.286	0.164
Panel C: Long-difference Estimates						
Δ ln(NL)	0.297***	0.323***	0.232***	0.348***	0.364***	0.254***
	(0.0812)	(0.0606)	(0.0818)	(0.101)	(0.0911)	(0.0868)
Δ ln(Pop.)	0.392	0.461*	0.252	0.355	0.428	0.281
	(0.272)	(0.266)	(0.448)	(0.318)	(0.359)	(0.461)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Countries	45	45	45	45	45	45
R^2	0.460	0.486	0.403	0.445	0.410	0.412

The table reports OLS regressions associating the logarithm of national GDP (Gross Domestic Product) in current PPP USD on the logarithm of the sum of nighttime lights (luminosity) across African countries. Panel *A* gives cross-sectional estimates, including year constants and log country land area. Panel *B* gives panel estimates with country fixed effects and year fixed effects. Panel *C* reports cross-sectional long differences specifications. Columns (1)-(3) use the newly compiled nighttime lights series that fuses the downgraded VIIRS data into the DMSP data after adjusting the latter for cross-sensor calibration, top-coding, and blooming with the extremely randomized forest method detailed in Section 2. Columns (4)-(6) use the merged VIIRS-DMSP series without adjusting the latter for top-coding and blooming-bleeding. Columns (1) and (4) give estimates across the entire period, 1992-2019; columns (2) and (5) give estimates over the DMSP satellite system period, 1992-2013, and columns (3) and (6) give estimates in the VIIRS period, 2014-2019. Heteroskedasticity-adjusted standard errors clustered at the country level are given in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B2: GDP - Luminosity Elasticity. Sensitivity. Cross-Country Cross-Sectional Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) ln(GDP)	(2) ln(GDP)	(3) ln(GDP)	(4) ln(GDP)	(5) ln(GDP)	(6) ln(GDP)
Panel A: All African countries, adding Region FEs						
ln(NL)	0.604***	0.605***	0.619***	0.732***	0.737***	0.729***
	(0.0708)	(0.0732)	(0.0665)	(0.0508)	(0.0490)	(0.0637)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1320	1032	288	1320	1032	288
Countries	48	48	48	48	48	48
R^2	0.945	0.943	0.951	0.958	0.958	0.956
Panel B: All African countries, adding Region FEs and Island Dummy						
ln(NL)	0.587***	0.590***	0.597***	0.734***	0.748***	0.707***
	(0.0759)	(0.0792)	(0.0679)	(0.0634)	(0.0614)	(0.0676)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1320	1032	288	1320	1032	288
Countries	48	48	48	48	48	48
R^2	0.947	0.945	0.954	0.958	0.958	0.957
Panel C: Dropping island nations						
ln(NL)	0.542***	0.546***	0.534***	0.601***	0.602***	0.600***
	(0.0283)	(0.0312)	(0.0266)	(0.0327)	(0.0355)	(0.0309)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1217	953	264	1217	953	264
Countries	44	44	44	44	44	44
R^2	0.928	0.922	0.939	0.931	0.926	0.939

The table reports cross-sectional OLS regressions associating the logarithm of national GDP (Gross Domestic Product) in current PPP USD on the logarithm of the sum of nighttime lights (luminosity) across African countries. Panel *A* gives cross-sectional estimates, including year constants, broad African region (North, South, Central, East, and West) constants, and log country land area. Panel *B* also controls for an indicator variable for island nations (Comoros, Mauritius, Mayotte, Reunion, Sao Tome and Principe, and Seychelles), while panel *C* excludes island nations. Columns (1)-(3) use the newly compiled nighttime lights series that fuses the downgraded VIIRS data into the DMSP data after adjusting the latter for cross-sensor calibration, top-coding, and blooming with the extremely randomized forest method detailed in Section 2. Columns (4)-(6) use the merged VIIRS-DMSP series without adjusting the latter for top-coding and blooming-bleeding. Columns (1) and (4) give estimates across the entire period, 1992-2019; columns (2) and (5) give estimates over the DMSP satellite system period, 1992-2013, and columns (3) and (6) give estimates in the VIIRS period, 2014-2019. Heteroskedasticity-adjusted standard errors clustered at the country level are given in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B3: GDP - Luminosity Elasticity. Sensitivity. Panel Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) ln(GDP)	(2) ln(GDP)	(3) ln(GDP)	(4) ln(GDP)	(5) ln(GDP)	(6) ln(GDP)
Panel A: All African countries						
ln(NL)	0.216***	0.219***	0.167***	0.270***	0.270***	0.202***
	(0.0442)	(0.0488)	(0.0556)	(0.0546)	(0.0555)	(0.0677)
ln(Pop.)	0.590***	0.688***	0.841	0.553**	0.654**	0.800
	(0.215)	(0.246)	(0.524)	(0.213)	(0.244)	(0.540)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1320	1032	288	1320	1032	288
Countries	48	48	48	48	48	48
R^2	0.993	0.995	0.999	0.993	0.995	0.999
within R^2	0.283	0.280	0.159	0.300	0.286	0.164
Panel B: exclude island nations						
ln(NL)	0.239***	0.243***	0.215***	0.322***	0.298***	0.275***
	(0.0466)	(0.0500)	(0.0683)	(0.0502)	(0.0562)	(0.0816)
ln(Pop.)	0.721***	0.763***	1.188**	0.668***	0.753***	1.179**
	(0.218)	(0.262)	(0.563)	(0.197)	(0.253)	(0.555)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1217	953	264	1217	953	264
Countries	44	44	44	44	44	44
R^2	0.992	0.994	0.998	0.992	0.994	0.998
within R^2	0.320	0.309	0.211	0.354	0.318	0.228
Panel C: region-year FEs and exclude islands						
ln(NL)	0.261***	0.226***	0.171**	0.315***	0.265***	0.252***
	(0.0449)	(0.0490)	(0.0682)	(0.0522)	(0.0568)	(0.0882)
ln(Pop.)	0.995***	1.478***	0.714	1.065***	1.532***	0.876
	(0.266)	(0.360)	(0.864)	(0.249)	(0.350)	(0.773)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1217	953	264	1217	953	264
Countries	44	44	44	44	44	44
R^2	0.993	0.995	0.999	0.994	0.995	0.999
within R^2	0.373	0.393	0.094	0.386	0.396	0.123

The table reports OLS panel regressions associating the logarithm of national GDP (Gross Domestic Product) in current PPP USD on the logarithm of the sum of nighttime lights (luminosity) across African countries. Panel A gives cross-sectional estimates, including year constants and log country land area. All specifications include country fixed effects and and year fixed effects. The specification in Panel C also include broad African region specific year fixed effects. Panel B drops from the estimation island nations (Comoros, Equatorial Guinea, Mauritius, Mayotte, Reunion, Sao Tome and Principe, and Seychelles). Columns (1)-(3) use the newly compiled nighttime lights series that fuses the downgraded VIIRS data into the DMSP data after adjusting the latter for cross-sensor calibration, top-coding, and blooming with the extremely randomized forest method detailed in Section 2. Columns (4)-(6) use the merged VIIRS-DMSP series without adjusting the latter for top-coding and blooming-bleeding. Columns (1) and (4) give estimates across the entire period, 1992-2019; columns (2) and (5) give estimates over the DMSP satellite system period, 1992-2013, and columns (3) and (6) give estimates in the VIIRS period, 2014-2019. Heteroskedasticity-adjusted standard errors clustered at the country level are given in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B4: National GDP Long Differences Estimates

	1992/93-2018/19	1992-2013		
	(1) DMSP$^+$-VIIRS	(2) DMSP-DMSP	(3) DMSP$^+$-DMSP$^+$	(4) DMSP$^+$-VIIRS
Panel A: All African countries. OLS				
Δ ln(NL)	0.297***	0.364***	0.323***	0.286***
	(0.0812)	(0.0911)	(0.0606)	(0.0620)
Δ ln(Pop.)	0.392	0.428	0.461*	0.419
	(0.272)	(0.359)	(0.266)	(0.299)
Countries	45	45	45	45
R^2	0.375	0.317	0.405	0.371
Panel D: All African countries (median regression)				
Δ ln(NL)	0.209***	0.310**	0.279***	0.327***
	(0.0707)	(0.116)	(0.0774)	(0.0605)
Δ ln(Pop.)	0.369	0.615	0.800**	0.261
	(0.288)	(0.367)	(0.300)	(0.313)
Countries	45	45	45	45
pseudo-R^2	0.277	0.208	0.270	0.247

The table presents regressions associating changes in the logarithm of national GDP (Gross Domestic Product) in current PPP USD on changes in the logarithm of the sum of nighttime lights (luminosity) across African countries. Panel *A* reports OLS estimates, and panel *B* median regressions estimates. Column (1) gives estimates from the long difference over the entire period (1992-2019) specification, while columns (2)-(4) take the difference over the DSMP period (1992-2013). Column (2) uses the DMSP nighttime light data without adjusting for top-coding and blooming. Column (3) uses the adjusted for top-coding, blooming, and sensor calibration DSMP series (denoted DMSP$^+$), and Column (4) uses DMSP$^+$ for the base period (1992) and downgraded VIIRS for the end period (2013). All specifications control for changes in countries' log populations and for broad African region constants (Northern, Central, Western, Eastern, and Southern). Standard errors robust to heteroskedacticity are reported in parentheses. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B.2 DHS Analysis

This Appendix Section complements the regional analysis linking development outcomes from the Demographic and Health Surveys with the newly compiled harmonized VIIRS-DSMP luminosity series in section 4 of the main paper.

Data Aggregation Appendix Figure B3 illustrates the aggregation of the DHS data to gridcells of 0.25×0.25 degrees, roughly $27km$ by $27km$ at the equator, zooming into central Mozambique. Circles report DHS enumeration areas or clusters, typically (large) villages, towns, and cities. The map also displays gridcells (black lines) to which we aggregate the underlying DHS data.

The choice of gridcell size (0.25×0.25 degrees) is a product of DHS displacement. The DHS cluster GPS points are displaced by up to 10km in order to maintain anonymity. So by choosing gridcells that are $27km$ wide and tall we ensure that any DHS cluster located at the center of a gridcell will be assigned to that gridcell even after displacement. This is especially important when building the panel data, since gridcells that are too small will not have repeat observations due to random DHS displacement. Of course, even with our choice of gridcell size, it is always possible that we 'miss' DHS clusters that repeat over time but are displaced to adjacent gridcells.

Sample Appendix Table B5 gives the survey years for all counties in the DHS analysis. The 34 countries come from all parts of Africa; some are landlocked (e.g., Burkina Faso, Central African Republic, Burundi), some coastal (e.g., Mozambique, Sierra Leone), and a few are island nations (Madagascar, Comoros). The sample spans relatively richer and poorer countries and includes former British, French, Portuguese, and Belgian colonies.

Summary Statistics Appendix Table B6 gives summary statistics for the four outcomes from the DHS (years of schooling, composite wealth index, access to flush or pit toilet, and access to electricity); the luminosity series (with and without the adjustments in the DMSP for top-coding and blooming), and area.

Preliminary Cross-Sectional Patterns Appendix Figure B4 illustrates the significant, although far from perfect, cross-sectional association between luminosity and the four proxies of local development. The four panels plot the histogram of mean years of schooling, the composite wealth index, access to electricity, and availability of flush or pit toilets for lit and unlit small gridcells, netting out country-year fixed effects and conditioning on log gridcell area. The histogram for lit gridcells (red vertical bars) is evidently to the right of the analogous one for unlit gridcells (blue bars). The two distributions are different for all development outcomes, especially with education and household wealth. Nonetheless, there is also overlap, as the binary transformation of

luminosity cannot fully capture the wide spatial variation in well-being within African countries.

Baseline Regression Analysis Appendix Tables B7 and B8 report cross-sectional and panel estimates linking the four DHS outcome measures (education, composite wealth index, access to electricity, and pit/flush toilet) to log luminosity and the lit indicator. In these tables, we do not standardize the dependent variables (as we do in the graphical illustrations in the main paper). In both tables, panels A and C report estimates with the newly compiled, adjusted for top-coding, blooming, and sensor quality DMSP series merged to the downgraded VIIRS. For comparability, panels B and D give the results with the merged VIIRS-DMSP series without adjusting the pre-2013 DMSP series for top-coding and blooming. The two tables, therefore, report the regression analogs to the coefficients reported in Figure 3. The cross-sectional specifications in Panel C of Appendix Table B7 suggest a significant increase of 1 point (one standard deviation) in the DHS standardized composite wealth index and about 1.7 more schooling years between lit and unlit areas. The panel estimates, while yielding statistically significant correlations, imply smaller effects. The results in panel C of Appendix Table B8 imply an increase of 0.124 years of schooling and 0.06 in the composite wealth index for administrative units turning from unlit to lit.

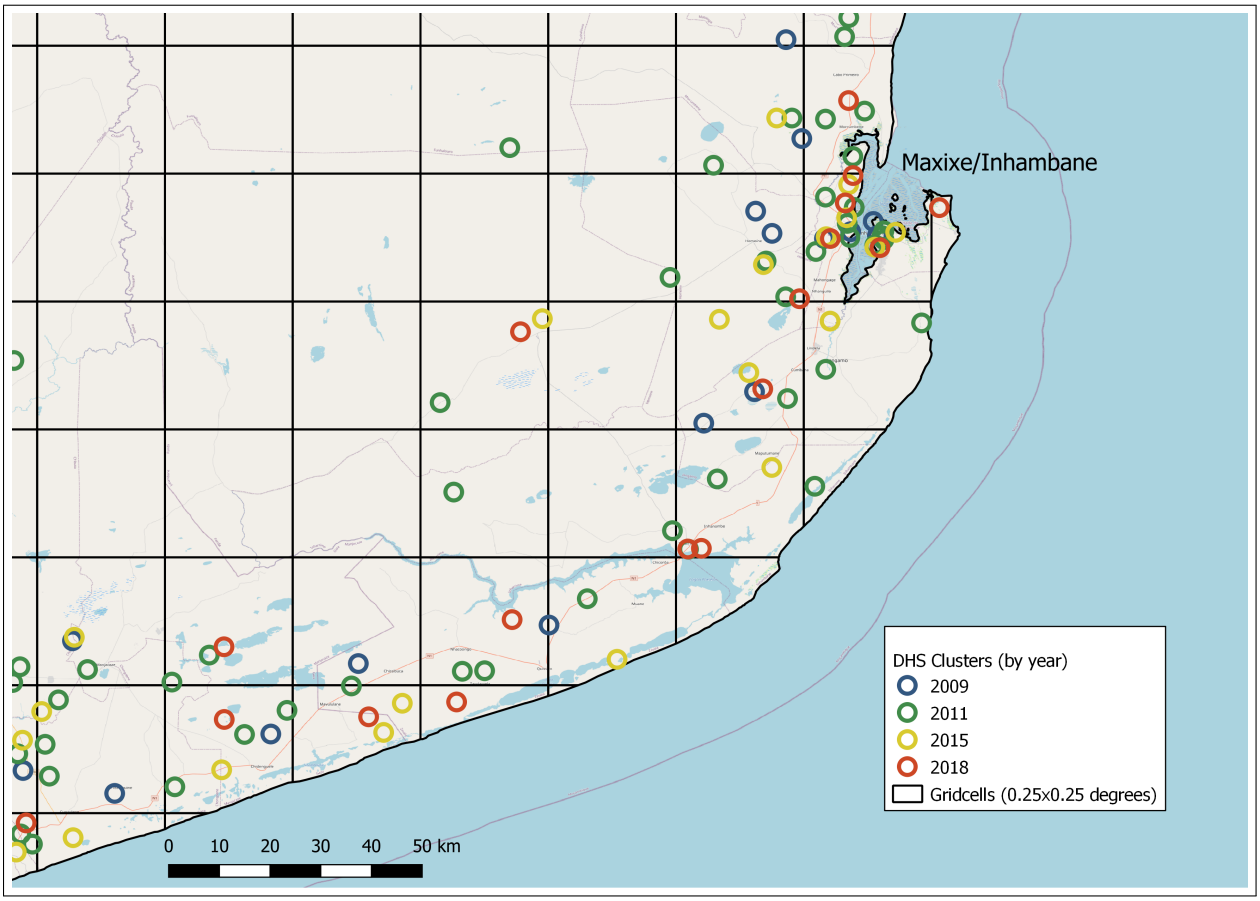
Further Evidence A. Varying Spatial Unit Appendix Figures B5, B6, and B7 panels (a) and (b) explore the association between luminosity and education, access to electrification, and to flush/pit toilet varying the size of the spatial unit. The results, therefore, complement the analysis in Figure 4 - Panels (a)-(b), in the main paper with the composite wealth index. The patterns with the three development proxies are similar to the ones with the DHS wealth index, based on household assets. First, cross-sectionally, differences in education, access to electricity and flush and pit toilet correlate with log luminosity across both small, medium, and larger areas. Besides, the coefficients are similar. Second, changes in log luminosity within gridcells over time correlate with changes in the three development proxies, although the dynamic correlations are weaker and more noisy. Third, the use of the harmonized and adjusted merged VIIRS-DSMP data yield stronger and with smaller standard errors correlations, especially in the panel specifications.

Further Evidence B. Varying Localized Variation Panels (c)-(d) of Appendix Figures B5, B6, and B7 report coefficients on log luminosity with education, access to electrification, and to flush/pit toilet, varying the localized variation with fixed-effects of varying coarseness. These specifications are, thus, similar to the ones in panels (c) and (d) of Figure 4 the composite wealth index as the outcome variable. The qualitative takeaways are mostly similar as with the composite wealth index. Notably, the luminosity years of schooling correlation turns significant only when adding mid-size fixed effects, of blocks 8×8 cells or larger. Besides, the luminosity access to electrification correlation is highly significant with the newly harmonized series, even when exploiting very gran-

ular variability with fine fixed-effects, illustrating again the reduction in measurement error from our ensemble method that fuses a downgraded vintage of VIIRS into the DMSP after adjusting them for top-coding, blooming, and sensor calibration (see Section 2).

Further Evidence C. Rural-Urban Appendix Figure B8 plots the coefficients on luminosity distinguishing across DHS respondents in rural and urban households (using the DHS classification). Panels (a) and (b) give cross-sectional estimates with log luminosity and the lit indicator (conditioning for log land area and country-survey-year constants). Panels (c) and (d) give panel estimates (with unit fixed effects and country-survey-year fixed effects). Red markers [diamonds] give the estimates with the harmonized and adjusted VIIRS-DMSP series, while blue markers [squares] report analogous estimates with the unadjusted for top-coding and blooming series. The cross-sectional analysis suggests that within-country across space) differences in luminosity correlate significantly with schooling, access to public goods, and household assets (as captured in the composite wealth index). The estimates appear similar in urban and rural locations. Besides, the adjustment for top-coding and blooming slightly improves the coefficient's magnitude. The panel specifications yield somewhat different patterns. First, the coefficients are statistically indistinguishable from zero when one uses the 'raw' luminosity series. In contrast, the coefficients are higher and pass standard statistical significance thresholds with the newly compiled harmonized and adjusted for top-coding, blooming, and sensor calibration lights series. Second, the coefficients in the urban sample of survey respondents are always larger than the ones in the rural sample, telling that the local development-luminosity nexus is stronger in urban areas.

Figure B3: Aggregation Example. DHS Clusters and Gridcells



The figure maps the southern coast of Mozambique near the city of Maxixe/Inhambane. The circles represent DHS clusters (enumeration areas), colored by the survey year. The grid gives the cells at which we aggregate and analyze the DHS data. The background imagery is from OpenStreetMap.

Table B5: DHS Sample

	Country	N years	N cell-years	Sample years
1	Angola	3	486	2006, 2011, 2015
2	Benin	4	366	1996, 2001, 2012, 2017
3	Burkina Faso	6	980	1993, 1999, 2003, 2010, 2014, 2017
4	Burundi	3	129	2010, 2012, 2016
5	Cameroon	3	633	2004, 2011, 2018
6	CAR	1	64	1994
7	Chad	1	330	2014
8	Comoros	1	10	2012
9	Cote d'Ivoire	3	382	1994, 1998, 2012
10	Democratic Republic of the Congo	2	586	2007, 2013
11	Egypt	7	839	1992, 1995, 2000, 2003, 2005, 2008, 2014
12	Gabon	1	131	2012
13	Ghana	7	1069	1993, 1998, 2003, 2008, 2014, 2016, 2019
14	Guinea	4	621	1999, 2005, 2012, 2018
15	Kenya	4	826	2003, 2008, 2014, 2015
16	Lesotho	3	157	2004, 2009, 2014
17	Liberia	6	529	2007, 2009, 2011, 2013, 2016, 2019
18	Madagascar	5	1077	1997, 2008, 2011, 2013, 2016
19	Malawi	7	809	2000, 2004, 2010, 2012, 2014, 2015, 2017
20	Mali	7	1341	1996, 2001, 2006, 2010, 2012, 2015, 2018
21	Morocco	1	223	2003
22	Mozambique	4	840	2009, 2011, 2015, 2018
23	Namibia	3	572	2000, 2006, 2013
24	Niger	2	228	1992, 1998
25	Nigeria	6	2297	2003, 2008, 2010, 2013, 2015, 2018
26	Rwanda	4	166	2005, 2008, 2010, 2014
27	Senegal	11	1379	1993, 1997, 2005, 2008, 2010, 2014, 2015, 2016, 2017, 2018, 2019
28	Sierra Leone	4	392	2008, 2013, 2016, 2019
29	Swaziland	1	32	2006
30	Tanzania	7	1962	1999, 2003, 2007, 2010, 2012, 2015, 2017
31	Togo	3	220	1998, 2013, 2017
32	Uganda	8	1232	2000, 2006, 2009, 2010, 2011, 2014, 2016, 2018
33	Zambia	3	867	2007, 2013, 2018
34	Zimbabwe	4	814	1999, 2005, 2010, 2015

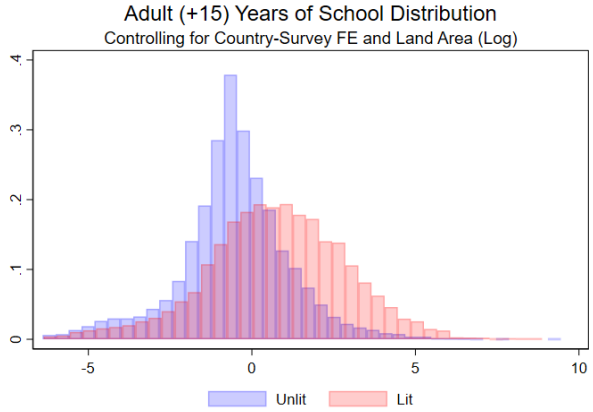
Table B6: Descriptive Statistics - Nightlights and DHS Outcomes

	Min	p10	p50	p90	Max	Mean	SD	N
Sensor, blooming, & topcode fixes	0.00	0.00	0.00	698.00	304,748.00	840.72	5,572.68	22589
Sensor calibration only	0.00	0.00	15.00	1,973.00	55,947.00	1,196.10	4,468.00	22589
Log of sensor, blooming, & topcode fixes	-0.69	-0.69	-0.69	6.55	12.63	1.61	3.15	22589
Log of sensor calibration only	-0.69	-0.69	2.74	7.59	10.93	2.75	3.55	22589
Sensor, blooming, & topcode fixes (dummy)	0.00	0.00	0.00	1.00	1.00	0.42	0.49	22589
Sensor calibration only (dummy)	0.00	0.00	1.00	1.00	1.00	0.53	0.50	22589
Gridcell area in km^2	4.00	689.59	753.95	768.53	769.31	727.21	102.70	22589
Log of gridcell area in km^2	1.39	6.54	6.63	6.64	6.65	6.56	0.31	22589
Years of Schooling(15-39)	0.00	0.76	5.07	8.97	13.83	4.99	3.02	18996
DHS Composite Wealth Index	1.00	1.35	2.45	3.99	5.00	2.57	0.97	19215
Piped Water Access	0.00	0.00	0.05	0.88	1.00	0.25	0.33	22503
Household Electricity Access	0.00	0.00	0.04	0.86	1.00	0.23	0.33	22507

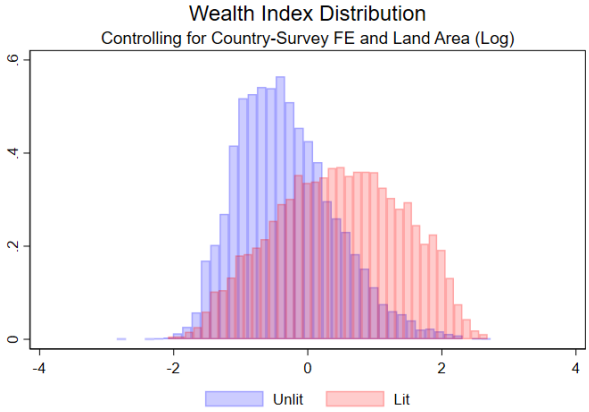
The table reports summary statistics for the nightlights and DHS data employed. The observations that come from 34 African countries are at the grid level. For the log of nightlights, we take ((half of the minimum value of positive NL) + NL) before taking the log.

Figure B4: DHS Development Outcomes across Lit and Unlit Grid-cells

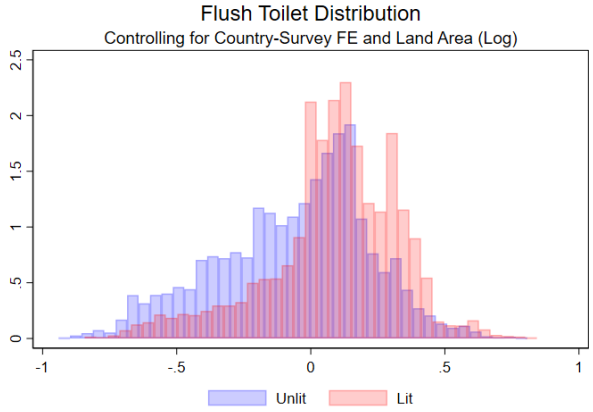
(a) Mean Adult (15-39) years schl.



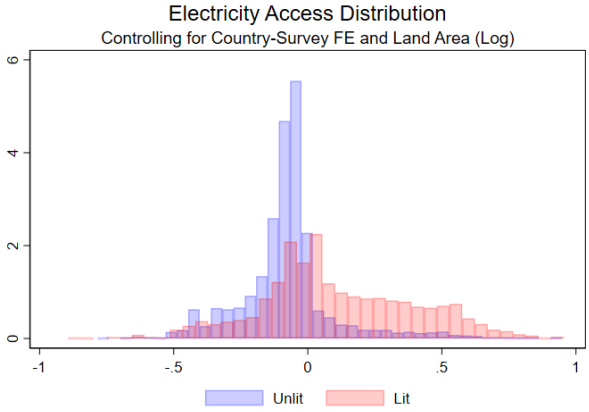
(b) Avg. Wealth Index



(c) Flush or Pit Toilet



(d) HH has Elect.



The figure plots the histograms of four proxies of local development for lit and unlit 0.25×0.25 gridcells after netting country-survey-year fixed-effects and log land area. Panel (a) gives average years of schooling for individuals between 15 and 39 years old. Panel (b) gives the DHS composite wealth index, based on household assets. Panel (c) gives the tabulations on access to flush or pit toilets. Panel (d) gives the histograms for household access to electricity.

Table B7: DHS Cross-sectional Estimates

	(1) Mean Adult (15-39) years schl.	(2) Avg. Wealth Index	(3) Flush or Pit Toilet	(4) HH has Elect.
Panel A: Log sum of nightlights - sensor, blooming, & topcode fixes				
ln(minNL/2+NL)	0.356***	0.198***	0.0355***	0.0508***
	(0.0160)	(0.00550)	(0.00205)	(0.00178)
Obs	18996	19215	22506	22507
Obs(NL=0)	11055	11286	13148	13149
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gc1km-yr	gc1km-yr	gc1km-yr	gc1km-yr
R^2	0.578	0.392	0.437	0.594
Panel B: Log sum of nightlights - sensor calibration only				
ln(minNL/2+NL)	0.296***	0.169***	0.0318***	0.0414***
	(0.0151)	(0.00594)	(0.00196)	(0.00180)
Obs	18996	19215	22506	22507
Obs(NL=0)	8917	8847	10459	10459
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gc1km-yr	gc1km-yr	gc1km-yr	gc1km-yr
R^2	0.569	0.377	0.441	0.573
Panel C: lit dummy - sensor, blooming, & topcode fixes				
1(NL>0)	1.670***	0.953***	0.180***	0.231***
	(0.0880)	(0.0361)	(0.0120)	(0.0108)
Obs	18996	19215	22506	22507
Obs(NL=0)	11055	11286	13148	13149
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gc1km-yr	gc1km-yr	gc1km-yr	gc1km-yr
R^2	0.545	0.301	0.422	0.532
Panel D: lit dummy - sensor calibration only				
1(NL>0)	1.485***	0.861***	0.173***	0.197***
	(0.0955)	(0.0390)	(0.0122)	(0.0106)
Obs	18996	19215	22506	22507
Obs(NL=0)	8917	8847	10459	10459
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gc1km-yr	gc1km-yr	gc1km-yr	gc1km-yr
R^2	0.533	0.265	0.418	0.506

Note: This table presents regressions of economic indicators from the DHS on nightlights. Observations are 0.25 x 0.25 degree gridcell-years. Each panel is done with a different definition for nightlights: Panel A uses the log sum of nightlights that have been adjusted for cross-sensor calibration and to fix blooming and topcoding, panel B uses the log of sum nightlights that have only been adjusted for cross-sensor calibration, Panel C uses a dummy if the cell is lit based on nightlights that have been adjusted for cross-sensor calibration and to fix blooming and topcoding, and Panel D uses a dummy if the cell is lit based on nightlights that have been adjusted for cross-sensor calibration. All specifications include the log area of the cell, and fixed effects for country-year. Standard errors in parentheses are clustered at the block level, where each block is 10x10 gridcells. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

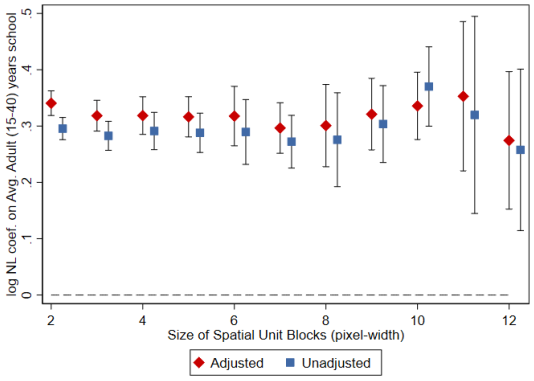
Table B8: DHS Panel Estimates

	(1) Mean Adult (15-39) years schl.	(2) Avg. Wealth Index	(3) Flush or Pit Toilet	(4) HH has Elect.
Panel A: Log sum of nightlights - sensor, blooming, & topcode fixes				
ln(minNL/2+NL)	0.0530***	0.0199***	0.00223	0.0109***
	(0.0125)	(0.00574)	(0.00186)	(0.00195)
Obs	14714	15068	18367	18367
Obs(NL=0)	7620	8037	9803	9803
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gc1km-yr	gc1km-yr	gc1km-yr	gc1km-yr
R^2	0.893	0.787	0.812	0.807
Panel B: Log sum of nightlights - sensor calibration only				
ln(minNL/2+NL)	0.0262***	0.00940**	0.00222	0.00532***
	(0.00893)	(0.00396)	(0.00136)	(0.00136)
Obs	14714	15068	18367	18367
Obs(NL=0)	5884	6002	7509	7509
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gc1km-yr	gc1km-yr	gc1km-yr	gc1km-yr
R^2	0.892	0.787	0.812	0.807
Panel C: lit dummy - sensor, blooming, & topcode fixes				
1(NL>0)	0.124***	0.0646***	0.0120*	0.0189***
	(0.0418)	(0.0204)	(0.00676)	(0.00720)
Obs	14714	15068	18367	18367
Obs(NL=0)	7620	8037	9803	9803
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gc1km-yr	gc1km-yr	gc1km-yr	gc1km-yr
R^2	0.892	0.787	0.812	0.807
Panel D: lit dummy - sensor calibration only				
1(NL>0)	0.0203	0.00590	0.00668	0.000453
	(0.0428)	(0.0186)	(0.00582)	(0.00597)
Obs	14714	15068	18367	18367
Obs(NL=0)	5884	6002	7509	7509
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gc1km-yr	gc1km-yr	gc1km-yr	gc1km-yr
R^2	0.892	0.786	0.812	0.807

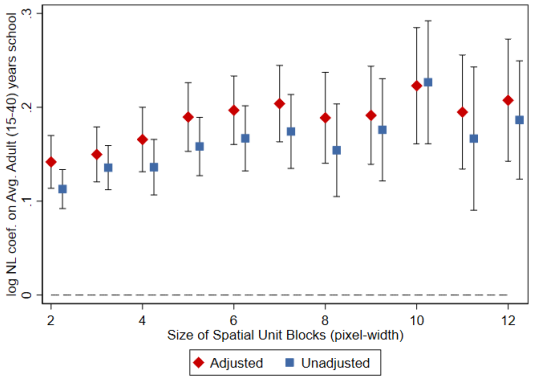
Note: This table presents regressions of economic indicators from the DHS on nightlights. Observations are 0.25 x 0.25 degree gridcell-years. Each panel is done with a different definition for nightlights: Panel A uses the log sum of nightlights that have been adjusted for cross-sensor calibration and to fix blooming and topcoding, panel B uses the log of sum nightlights that have only been adjusted for cross-sensor calibration, Panel C uses a dummy if the cell is lit based on nightlights that have been adjusted for cross-sensor calibration and to fix blooming and topcoding, and Panel D uses a dummy if the cell is lit based on nightlights that have been adjusted for cross-sensor calibration. All specifications include fixed effects for country-year, and gridcell. Standard errors in parentheses are clustered at the block level, where each block is 10x10 gridcells. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure B5: Schooling-Luminosity Correlation. Further Evidence

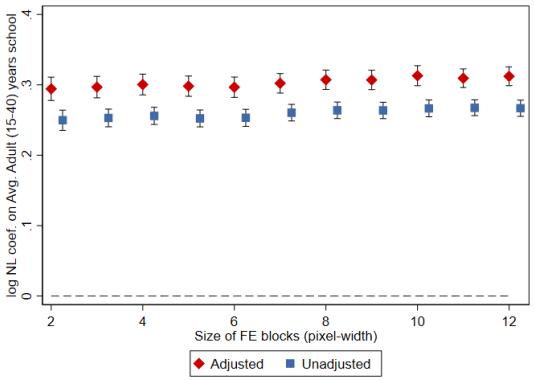
(a) Cross-sectional, varying Spatial Unit size



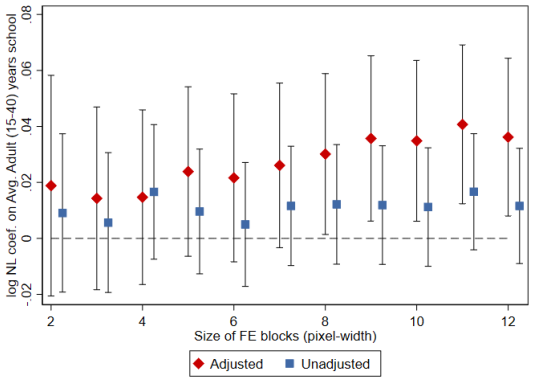
(b) Panel, varying Spatial Unit size



(c) Cross-sectional, varying FE size



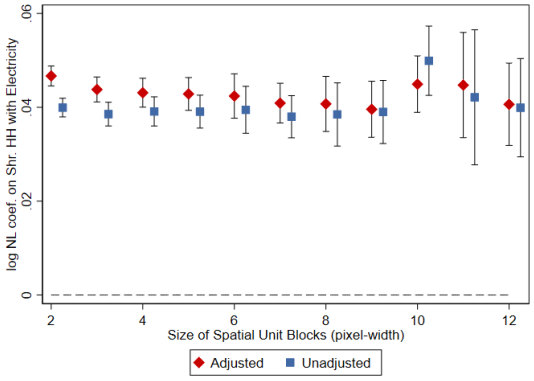
(d) Panel, varying FE size



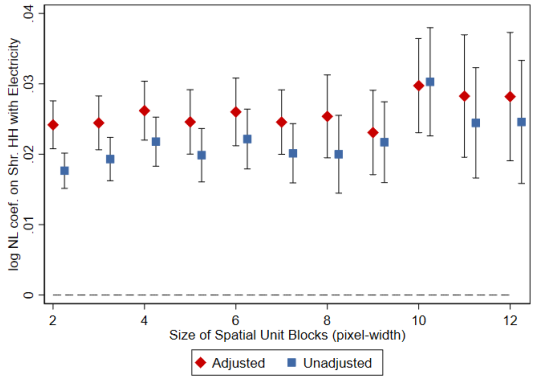
The figure plots coefficients from regressions associating mean years of schooling of individuals aged 15-39 on Log Luminosity. Panels (a) and (c) give cross-sectional estimates with country-survey year constants, controlling also for the gridcell's log land area. Panels (b) and (d) give panel estimates that, besides the country-year constants, also include gridcell fixed effects. Panels (a)-(b) plot coefficients of log luminosity varying the (gridcell) size unit of the empirical analysis. Panel (c) plots cross-sectional coefficients of log luminosity augmenting the specification with block fixed effects of various sizes. Panel (d) plots panel coefficients of log luminosity augmenting the specification with interactions between country-survey-year constants with block fixed effects of various sizes. Red markers denote estimates using the harmonized VIIRS-DMSP luminosity series, adjusting the DMSP for top coding, sensor calibration, and blooming. Blue markers denote estimates using the unadjusted merged VIIRS-DMSP luminosity series. The bars denote 95% confidence intervals, based on standard errors clustered at the grid-cell level for panels (c) and (d) and at the spatial unit level for panels (a) and (b).

Figure B6: Household Access to Electricity-Luminosity Correlation. Further Evidence

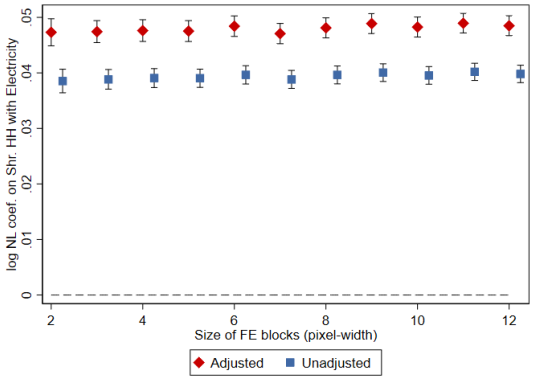
(a) Cross-sectional, varying Spatial Unit size



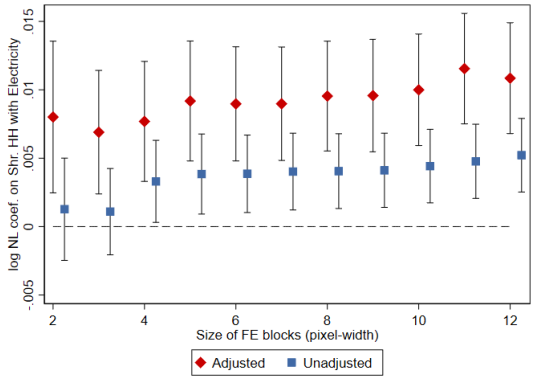
(b) Panel, varying Spatial Unit size



(c) Cross-sectional, varying FE size



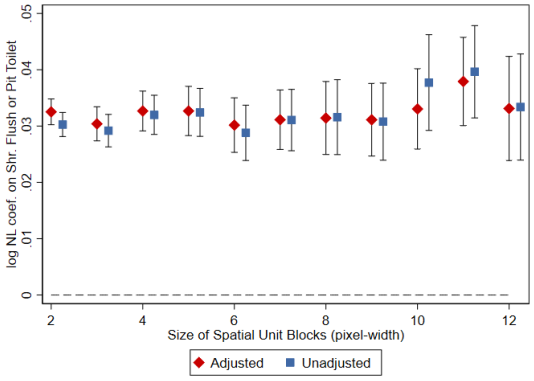
(d) Panel, varying FE size



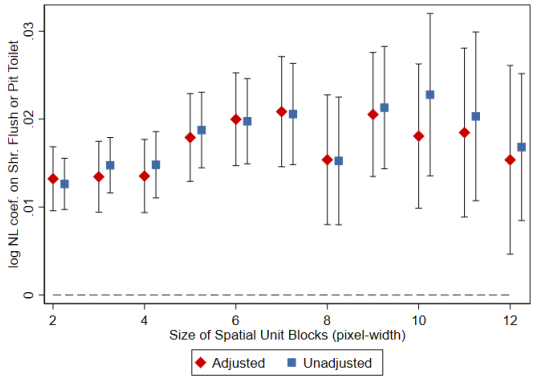
The figure plots coefficients from regressions associating household access to electricity on Log Luminosity. Panels (a) and (c) give cross-sectional estimates with country-survey year constants, controlling also for the gridcell's log land area. Panels (b) and (d) give panel estimates that, besides the country-year constants, also include gridcell fixed effects. Panels (a)-(b) plot coefficients of log luminosity varying the (gridcell) size unit of the empirical analysis. Panel (c) plots cross-sectional coefficients of log luminosity augmenting the specification with block fixed effects of various sizes. Panel (d) plots panel coefficients of log luminosity augmenting the specification with interactions between country-survey-year constants with block fixed effects of various sizes. Red markers denote estimates using the harmonized VIIRS-DMSP luminosity series, adjusting the DMSP for top coding, sensor calibration, and blooming. Blue markers denote estimates using the unadjusted merged VIIRS-DMSP luminosity series. The bars denote 95% confidence intervals, based on standard errors clustered at the grid-cell level for panels (c) and (d) and at the spatial unit level for panels (a) and (b).

Figure B7: Household Access to Flush/Pit Toilet-Luminosity Correlation. Further Evidence

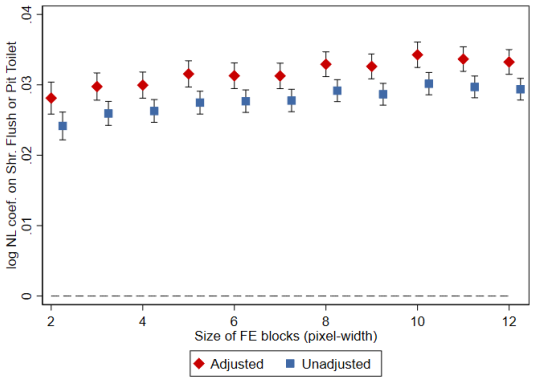
(a) Cross-sectional, varying Spatial Unit size



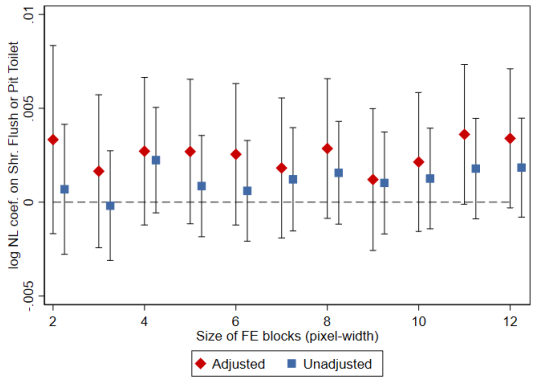
(b) Panel, varying Spatial Unit size



(c) Cross-sectional, varying FE size



(d) Panel, varying FE size

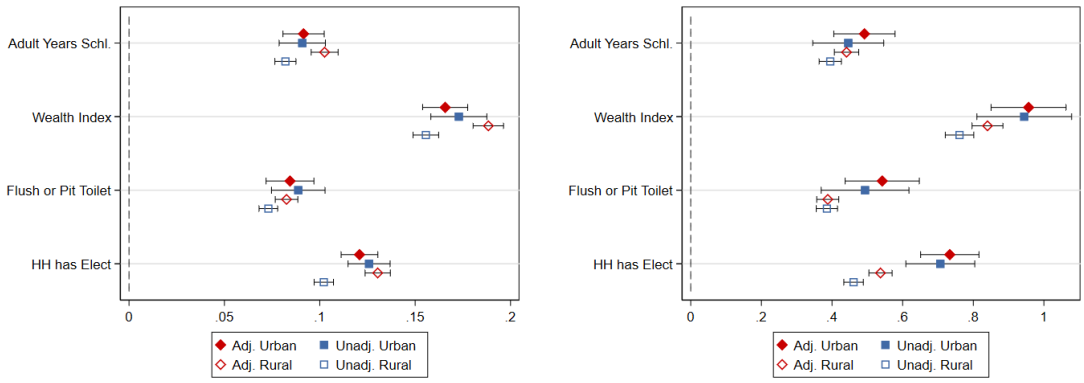


The figure plots coefficients from regressions associating household access to flush or pit toilet on Log Luminosity. Panels (a) and (c) give cross-sectional estimates with country-survey year constants, controlling also for the gridcell's log land area. Panels (b) and (d) give panel estimates that, besides the country-year constants, also include gridcell fixed effects. Panels (a)-(b) plot coefficients of log luminosity varying the (gridcell) size unit of the empirical analysis. Panel (c) plots cross-sectional coefficients of log luminosity augmenting the specification with block fixed effects of various sizes. Panel (d) plots panel coefficients of log luminosity augmenting the specification with interactions between country-survey-year constants with block fixed effects of various sizes. Red markers denote estimates using the harmonized VIIRS-DMSP luminosity series, adjusting the DMSP for top coding, sensor calibration, and blooming. Blue markers denote estimates using the unadjusted merged VIIRS-DMSP luminosity series. The bars denote 95% confidence intervals, based on standard errors clustered at the grid-cell level for panels (c) and (d) and at the spatial unit level for panels (a) and (b).

Figure B8: Local Development - Luminosity Association. Urban and Rural Areas

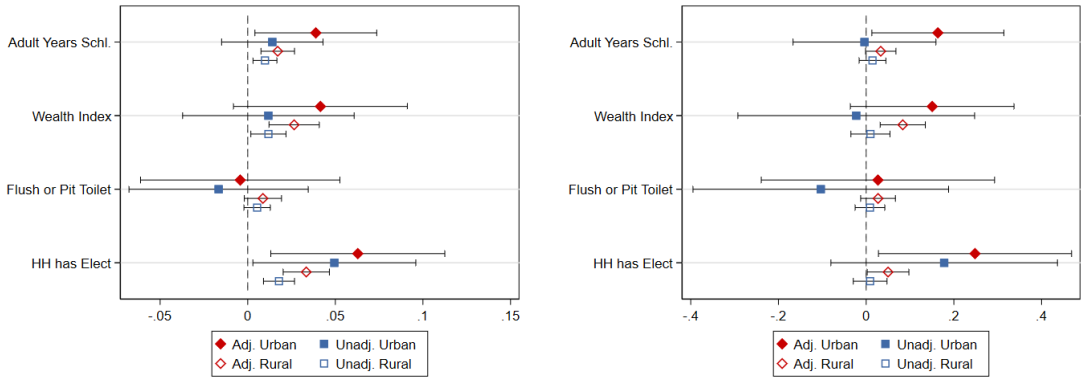
(a) Cross-section - Log Nightlights

(b) Cross-section - Lit dummy



(c) Panel - Log Nightlights

(d) Panel - Lit dummy



Notes: This figure plots coefficients from regressions of standardized DHS measures on nightlights. For the luminosity variables, panels (a) and (c) use log nightlights and panels (b) and (d) use an indicator equal to one for positive lights and zero otherwise. The top panels (a) and (b) control for country by year fixed effects. Panels (a) and (b) control for log grid cell area, while panels (b) and (c) control for grid cell fixed effects. In each panel, estimates from the subset of urban grid cells (red diamonds) are compared with estimates from the subset of rural grid cells (blue squares). The solid markers denote estimates using our corrected nightlight series, and hollow markers denote estimates using the unadjusted series. For these figures, all DHS outcomes are standardized to have a mean of zero and a standard deviation of one. The bars represent 95% confidence intervals, and standard errors are clustered at the gridcell level.

B.3 Mozambique Census

Summary Statistics Appendix Table B9 gives summary statistics of local development measures and nighttime luminosity across Mozambican localities (admin-4 units). We proxy local development with the mean years of schooling of the population aged 15-39 years, as recorded in the Censuses of 1997, 2007, and 2017; and with the share of employment outside agriculture for 15-24 years old, using information from the 1997 and 2007 Censuses (as the data is missing for the 2017 Censuses). We take the mean values among the young, as in the panel estimates, since we want to capture the "flow" of these variables more accurately. We have information for the full census for 1997, 2007, and 2017. The table gives summary statistics of various luminosity transformations using the newly-compiled harmonized and adjusted VIIRS-DMSP series and with the merged VIIRS-DMSP series without adjusting the later for top-coding and blooming. Panel A gives the statistics pooling across all census years. Panels B, C, and D give the statistics for the 1997, the 2007, and the 2017 Census, respectively.

Patterns Appendix Figure B9 illustrates the spatial distribution of luminosity across $1,184$ Mozambican localities in 1997, 2007, and 2017, the Census years. The maps use the newly-compiled lights series that fuses the VIIRS data (post-2013) to the adjusted for blooming, top-coding, and sensor calibration DMSP series (1992-2013) with the extremely randomized forest (ensemble) method, detailed in Section 2 of the paper.

Cross-Sectional Estimates Appendix Tables B10 and B11 report cross-sectional estimates associating mean years of schooling of the population aged 15-39 years and the share of youth employment outside agriculture with log luminosity and a lit indicator, respectively across the census years. Panel A gives estimates across admin-2 units (*distritos*). Panel B gives estimates across admin-3 units (*postos*). Panel C gives estimates across admin-4 units (*localidades*). All specifications condition on log land area and census-year constants. Panel A conditions also on census-year specific admin-1 (province) fixed-effects. Panel B conditions also on census-year specific admin-2 (district) fixed-effects. Panel C conditions also on census-year specific admin-3 (posto) fixed-effects. The estimates suggest that employment outside of agriculture is about 10 percentage points and mean schooling about 0.6 years higher in lit admin-2, admin-3, or admin-4 units than unlit ones.

Panel Estimates Appendix Tables B12 and B13 report panel estimates that explore the dynamic correlation between local development and luminosity across Mozambican administrative units. All specifications include administrative unit fixed effects and census year fixed effects. Panel A gives estimates across admin-2 units (districts). Panel B gives estimates across admin-3 units (postos). Panel C gives estimates across admin-4 units (localities). Panel A conditions also on census-year

specific admin-1 (province) fixed-effects. Panel B conditions also on census-year specific admin-2 (district) fixed-effects. Panel C conditions also on census-year specific admin-3 (posto) fixed-effects. The estimates suggest an increase in mean years of schooling of about $0.25-0.37$ and an increase of non-agriculture employment of about three percentage points for localities that turn lit than those that stay unlit.

Visual Illustration. Dynamic Correlation Appendix Figure B10 illustrates the within-locality co-movement of luminosity and mean years of schooling of 15-39-year-old Mozambicans over $1997-2007$ (panels (a)-(b)) and $1997-2017$ (panels (c)-(d)). The green bars plot the increase in schooling across $1,028$ unlit in 1997 localities. Dark green bars in panel (a) reveal an increase in average schooling of 2.3 years in the 89 localities that turned lit by 2007, much higher than in the 939 localities that remained unlit by 2007 (1.76). The difference in schooling between initially unlit locations that either stay unlit or turn lit over the twenty years ($2017-1997$) in panel (c) is 0.5 years (4.26 vs 3.85). Blue bars plot the increase in mean schooling for the 98 localities lit in 1997. Schooling increased by 2.44 years for the 85 that remained lit in 1997, while schooling increased by 2 years in the 13 localities that turned unlit by 2007. Panels (b) and (d) plot changes in schooling years for the four categories of localities [unlit - unlit (light green), unlit - lit (dark green), lit - unlit (light blue) and lit - lit (dark blue), conditional on admin-3 fixed effects, to control for the considerable differences in local development across Mozambique and compare nearby localities. Differences in schooling correlate with differences in nighttime lights.

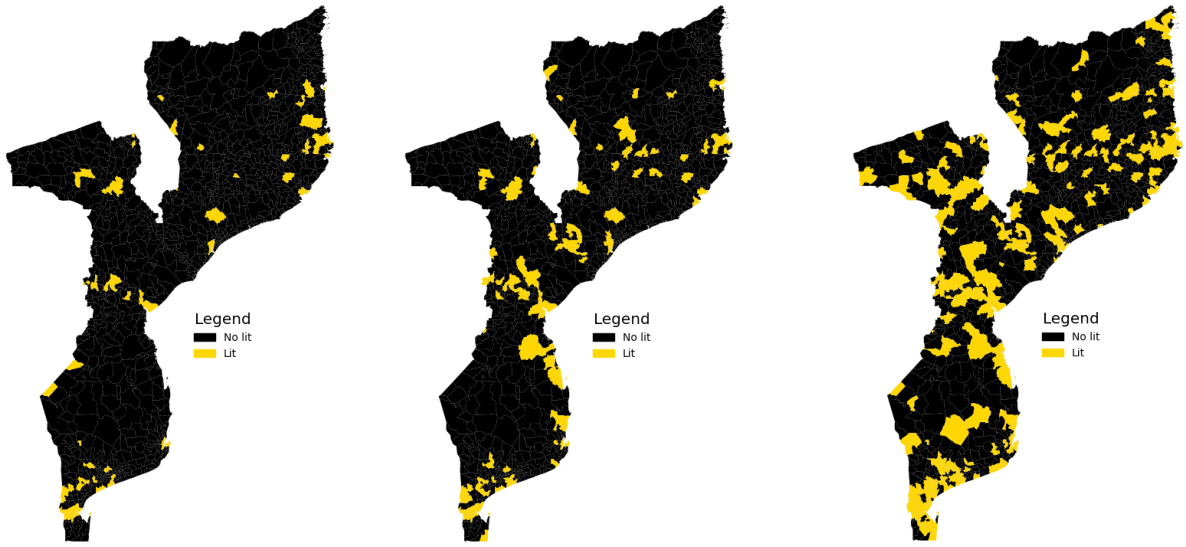
Table B9: Summary Statistics - Mozambique Census and Nighttime Lights

	Min	p10	p50	p90	Max	Mean	SD	N
Panel A: 1997, 2007, and 2017								
Sensor, blooming, & topcode fixes	0.00	0.00	0.00	27.00	20,053.00	52.66	577.71	3378
Sensor calibration only	0.00	0.00	0.00	183.00	22,014.00	113.01	744.21	3378
Log of sensor, blooming, & topcode fixes	-0.69	-0.69	-0.69	3.31	9.91	0.11	1.83	3378
Log of sensor calibration only	-0.69	-0.69	-0.69	5.21	10.00	0.79	2.51	3378
Sensor, blooming, & topcode fixes (dummy)	0.00	0.00	0.00	1.00	1.00	0.19	0.39	3378
Sensor calibration only (dummy)	0.00	0.00	0.00	1.00	1.00	0.29	0.45	3378
Gridcell area in km^2	0.96	86.17	416.78	1,458.47	8,349.15	668.26	854.70	3378
Log of gridcell area in km^2	-0.04	4.46	6.03	7.29	9.03	5.91	1.19	3378
Years of adult (15-40) schooling	0.04	0.47	2.45	5.53	8.70	2.82	1.95	3377
Share of youth (15-24) emp. out agriculture	0.00	0.02	0.09	0.34	0.99	0.15	0.16	2252
Panel B: 1997								
Sensor, blooming, & topcode fixes	0.00	0.00	0.00	0.00	5,158.00	15.04	201.37	1126
Sensor calibration only	0.00	0.00	0.00	24.00	7,618.00	36.12	335.53	1126
Log of sensor, blooming, & topcode fixes	-0.69	-0.69	-0.69	-0.69	8.55	-0.37	1.17	1126
Log of sensor calibration only	-0.69	-0.69	-0.69	3.20	8.94	-0.04	1.76	1126
Sensor, blooming, & topcode fixes (dummy)	0.00	0.00	0.00	0.00	1.00	0.09	0.28	1126
Sensor calibration only (dummy)	0.00	0.00	0.00	1.00	1.00	0.13	0.34	1126
Gridcell area in km^2	0.96	86.17	416.78	1,458.47	8,349.15	668.26	854.96	1126
Log of gridcell area in km^2	-0.04	4.46	6.03	7.29	9.03	5.91	1.19	1126
Years of adult (15-40) schooling	0.04	0.27	0.68	1.78	4.66	0.89	0.68	1126
Share of youth (15-24) emp. out agriculture	0.00	0.02	0.07	0.29	0.99	0.12	0.14	1126
Panel C: 2007								
Sensor, blooming, & topcode fixes	0.00	0.00	0.00	15.00	9,523.00	36.64	417.87	1126
Sensor calibration only	0.00	0.00	0.00	121.00	13,269.00	82.65	617.15	1126
Log of sensor, blooming, & topcode fixes	-0.69	-0.69	-0.69	2.74	9.16	-0.06	1.64	1126
Log of sensor calibration only	-0.69	-0.69	-0.69	4.80	9.49	0.47	2.30	1126
Sensor, blooming, & topcode fixes (dummy)	0.00	0.00	0.00	1.00	1.00	0.15	0.36	1126
Sensor calibration only (dummy)	0.00	0.00	0.00	1.00	1.00	0.22	0.41	1126
Gridcell area in km^2	0.96	86.17	416.78	1,458.47	8,349.15	668.26	854.96	1126
Log of gridcell area in km^2	-0.04	4.46	6.03	7.29	9.03	5.91	1.19	1126
Years of adult (15-40) schooling	0.32	1.53	2.46	4.48	7.25	2.75	1.19	1126
Share of youth (15-24) emp. out agriculture	0.01	0.03	0.11	0.38	0.97	0.17	0.17	1126
Panel D: 2017								
Sensor, blooming, & topcode fixes	0.00	0.00	0.00	86.00	20,053.00	106.30	884.39	1126
Sensor calibration only	0.00	0.00	1.00	390.00	22,014.00	220.26	1,072.72	1126
Log of sensor, blooming, & topcode fixes	-0.69	-0.69	-0.69	4.46	9.91	0.74	2.31	1126
Log of sensor calibration only	-0.69	-0.69	0.41	5.97	10.00	1.93	2.91	1126
Sensor, blooming, & topcode fixes (dummy)	0.00	0.00	0.00	1.00	1.00	0.33	0.47	1126
Sensor calibration only (dummy)	0.00	0.00	1.00	1.00	1.00	0.52	0.50	1126
Gridcell area in km^2	0.96	86.17	416.78	1,458.47	8,349.15	668.26	854.96	1126
Log of gridcell area in km^2	-0.04	4.46	6.03	7.29	9.03	5.91	1.19	1126
Years of adult (15-40) schooling	0.66	3.14	4.81	6.48	8.70	4.83	1.34	1125

The table reports summary statistics for the variables employed in Section 5 of the main paper that associates regional development, as recorded in the three post-independence Mozambican censuses (1997, 2007, and 2017) and nightlights lights (luminosity) across 1126 admin-4 units (localities).

Figure B9: Mozambique Lit/Unlit Localities in 1997, 2007, 2017

(a) 1997 (b) 2007 (c) 2017



The figure plots the spatial distribution of the lit indicator using the newly complied harmonized and adjusted VIIRS-DSMP lights series across Mozambican localities in 1997, 2007, and 2017, the years of the three Mozambican censuses.

Table B10: Mozambique Cross-Sectional Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) Mean Adult (15-39) years schl. (97&07)	(2) Share Youth (15-24) Emp. out Ag. (97&07)	(3) Mean Adult (15-39) years schl. (17 only)	(4) Mean Adult (15-39) years schl. (97&07)	(5) Share Youth (15-24) Emp. out Ag. (97&07)	(6) Mean Adult (15-39) years schl. (17 only)
Panel A: Admin Level 2						
ln(minNL/2+NL)	0.194***	0.0253***	0.0441	0.188***	0.0241***	0.0483
	(0.0250)	(0.00431)	(0.0620)	(0.0240)	(0.00433)	(0.0892)
Obs	282	282	141	282	282	141
Obs(NL=0)	159	159	14	135	135	2
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
R^2	0.844	0.722	0.182	0.837	0.714	0.180
Panel B: Admin Level 3						
ln(minNL/2+NL)	0.177***	0.0316***	0.211***	0.133***	0.0212***	0.159***
	(0.0183)	(0.00337)	(0.0237)	(0.0136)	(0.00233)	(0.0192)
Obs	774	774	387	774	774	387
Obs(NL=0)	627	627	190	564	564	114
FEs	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
units	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
R^2	0.915	0.770	0.588	0.916	0.759	0.565
Panel C: Admin Level 4						
ln(minNL/2+NL)	0.213***	0.0338***	0.172***	0.138***	0.0210***	0.145***
	(0.0267)	(0.00456)	(0.0318)	(0.0166)	(0.00291)	(0.0239)
Obs	2124	2124	1061	2124	2124	1061
Obs(NL=0)	1903	1903	724	1784	1784	520
FEs	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R^2	0.894	0.779	0.616	0.894	0.775	0.624

Note: This table presents regressions of economic indicators from the Mozambique census on nightlights. Each panel is done at a different administrative level: panel A used admin 2 units, panel B admin 3, and panel C admin 4. Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as the log sum of light in a district, the log area of the district, and fixed effects for year interacted with the admin unit one level above (e.g. in panel C, units are admin 4 and so we include admin 3 by year fixed effects). Standard errors in parentheses are clustered at the admin 2 level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B11: Mozambique Cross-Sectional Estimates - Lit Indicator

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) Mean Adult (15-39) years schl. (97&07)	(2) Share Youth (15-24) Emp. out Ag. (97&07)	(3) Mean Adult (15-39) years schl. (17 only)	(4) Mean Adult (15-39) years schl. (97&07)	(5) Share Youth (15-24) Emp. out Ag. (97&07)	(6) Mean Adult (15-39) years schl. (17 only)
Panel A: Admin Level 2						
1(NL>0)	0.710***	0.0864***	0.406	0.675***	0.0878***	0.253
	(0.116)	(0.0207)	(0.287)	(0.113)	(0.0210)	(0.472)
Obs	282	282	141	282	282	141
Obs(NL=0)	159	159	14	135	135	2
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
R^2	0.794	0.668	0.184	0.789	0.670	0.177
Panel B: Admin Level 3						
1(NL>0)	0.629***	0.107***	0.849***	0.604***	0.0888***	0.630***
	(0.0772)	(0.0143)	(0.113)	(0.0764)	(0.0121)	(0.145)
Obs	774	774	387	774	774	387
Obs(NL=0)	627	627	190	564	564	114
FEs	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
units	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
R^2	0.904	0.739	0.549	0.905	0.731	0.507
Panel C: Admin Level 4						
1(NL>0)	0.607***	0.0902***	0.687***	0.555***	0.0842***	0.608***
	(0.0895)	(0.0143)	(0.117)	(0.0777)	(0.0144)	(0.113)
Obs	2124	2124	1061	2124	2124	1061
Obs(NL=0)	1903	1903	724	1784	1784	520
FEs	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R^2	0.888	0.764	0.613	0.889	0.766	0.609

Note: This table presents regressions of economic indicators from the Mozambique census on nightlights. Each panel is done at a different administrative level: panel A used admin 2 units, panel B admin 3, and panel C admin 4. Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as an indicator for positive values of luminosity, the log area of the district, and fixed effects for year interacted with the admin unit one level above (e.g. in panel C, units are admin 4 and so we include admin 3 by year fixed effects). Standard errors in parentheses are clustered at the admin 2 level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B12: Mozambique Panel Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) Mean Adult (15-39) years schl. (97&07)	(2) Share Youth (15-24) Emp. out Ag. (97&07)	(3) Mean Adult (15-39) years schl. (97,07&17)	(4) Mean Adult (15-39) years schl. (97&07)	(5) Share Youth (15-24) Emp. out Ag. (97&07)	(6) Mean Adult (15-39) years schl. (97,07&17)
Panel A: Admin Level 2						
ln(minNL/2+NL)	0.0357**	0.00441**	0.125***	0.0264*	0.00315*	0.136***
	(0.0166)	(0.00213)	(0.0397)	(0.0141)	(0.00188)	(0.0485)
Obs	282	282	423	282	282	423
Obs(NL=0)	159	159	173	135	135	137
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
R^2	0.983	0.982	0.869	0.983	0.982	0.870
Panel B: Admin Level 3						
ln(minNL/2+NL)	0.0547***	0.00989***	0.0946***	0.0292**	0.00401**	0.0467***
	(0.0174)	(0.00285)	(0.0193)	(0.0137)	(0.00190)	(0.0162)
Obs	774	774	1161	774	774	1161
Obs(NL=0)	627	627	817	564	564	678
FEs	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
units	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
R^2	0.987	0.971	0.962	0.987	0.970	0.961
Panel C: Admin Level 4						
ln(minNL/2+NL)	0.113***	0.0148***	0.0835***	0.0737***	0.00626**	0.0592***
	(0.0273)	(0.00444)	(0.0240)	(0.0146)	(0.00283)	(0.0174)
Obs	2124	2124	3185	2124	2124	3185
Obs(NL=0)	1903	1903	2627	1784	1784	2304
FEs	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R^2	0.978	0.965	0.963	0.978	0.965	0.963
Panel D: Admin Level 4 - No FEs						
ln(minNL/2+NL)	0.168***	0.0161***	0.0763***	0.132***	0.00972***	0.0790***
	(0.0246)	(0.00388)	(0.0169)	(0.0166)	(0.00180)	(0.0115)
Obs	2250	2250	3374	2250	2250	3374
Obs(NL=0)	1978	1978	2734	1854	1854	2397
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R^2	0.943	0.937	0.907	0.944	0.936	0.908

Note: This table presents regressions of economic indicators from the Mozambique census on nightlights. Each panel is done at a different administrative level: panel A used admin 2 units, panel B admin 3, panel C admin 4, panel D also uses admin 4 units but does not add fixed effects for admin 3 by year. Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as the log sum of light in a district and fixed effects for year interacted with the admin unit one level above (e.g. in panel C, units are admin 4 and so we include admin 3 by year fixed effects). Also included are fixed effects for the admin level denoted in the panel title, and therefore these coefficients reflect changes. Standard errors in parentheses are clustered at the admin 2 level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B13: Mozambique Panel Estimates - Lit Indicator

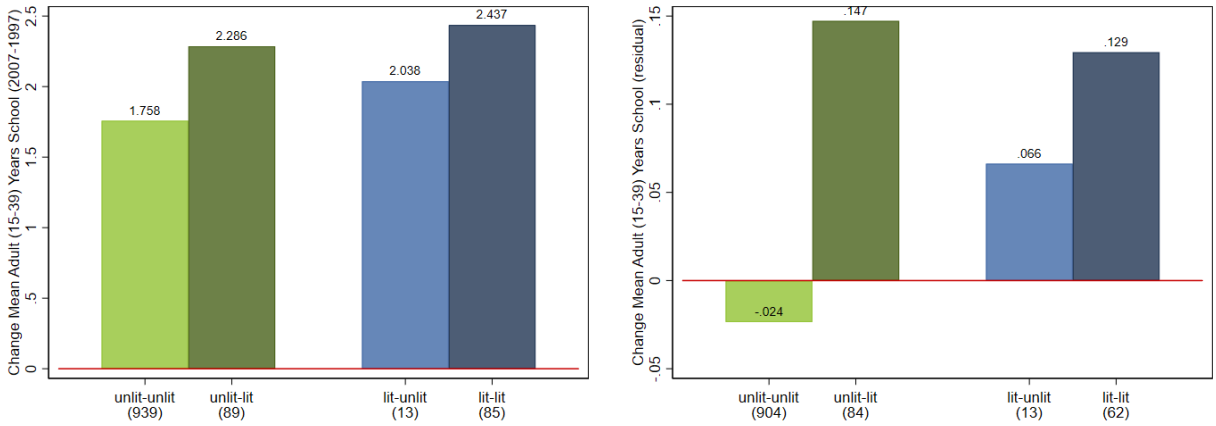
	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) Mean Adult (15-39) years schl. (97&07)	(2) Share Youth (15-24) Emp. out Ag. (97&07)	(3) Mean Adult (15-39) years schl. (97,07&17)	(4) Mean Adult (15-39) years schl. (97&07)	(5) Share Youth (15-24) Emp. out Ag. (97&07)	(6) Mean Adult (15-39) years schl. (97,07&17)
Panel A: Admin Level 2						
1(NL>0)	0.0879	0.0131*	0.432**	0.0679	0.00908	0.417**
	(0.0672)	(0.00777)	(0.172)	(0.0657)	(0.00813)	(0.192)
Obs	282	282	423	282	282	423
Obs(NL=0)	159	159	173	135	135	137
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
R^2	0.983	0.982	0.868	0.982	0.982	0.868
Panel B: Admin Level 3						
1(NL>0)	0.0962	0.0187	0.230**	0.0786	0.00629	0.0497
	(0.0824)	(0.0115)	(0.0922)	(0.0727)	(0.0108)	(0.0941)
Obs	774	774	1161	774	774	1161
Obs(NL=0)	627	627	817	564	564	678
FEs	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
units	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
R^2	0.987	0.970	0.961	0.987	0.970	0.961
Panel C: Admin Level 4						
1(NL>0)	0.253***	0.0270*	0.265***	0.243***	0.0169	0.198**
	(0.0881)	(0.0156)	(0.0766)	(0.0660)	(0.0128)	(0.0761)
Obs	2124	2124	3185	2124	2124	3185
Obs(NL=0)	1903	1903	2627	1784	1784	2304
FEs	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R^2	0.978	0.964	0.963	0.978	0.964	0.963
Panel D: Admin Level 4 - No FEs						
1(NL>0)	0.372***	0.0298**	0.298***	0.489***	0.0305***	0.317***
	(0.0885)	(0.0132)	(0.0589)	(0.0686)	(0.00846)	(0.0543)
Obs	2250	2250	3374	2250	2250	3374
Obs(NL=0)	1978	1978	2734	1854	1854	2397
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R^2	0.940	0.935	0.907	0.942	0.935	0.908

Note: This table presents regressions of economic indicators from the Mozambique census on nightlights. Each panel is done at a different administrative level: panel A used admin 2 units, panel B admin 3, panel C admin 4, panel D also uses admin 4 units but does not add fixed effects for admin 3 by year. Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as an indicator for positive values of luminosity and fixed effects for year interacted with the admin unit one level above (e.g. in panel C, units are admin 4 and so we include admin 3 by year fixed effects). Also included are fixed effects for the admin level denoted in the panel title, and therefore these coefficients reflect changes. Standard errors in parentheses are clustered at the admin 2 level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure B10: Mozambique Δ Mean Years Schooling by Changes in Lit/Unlit

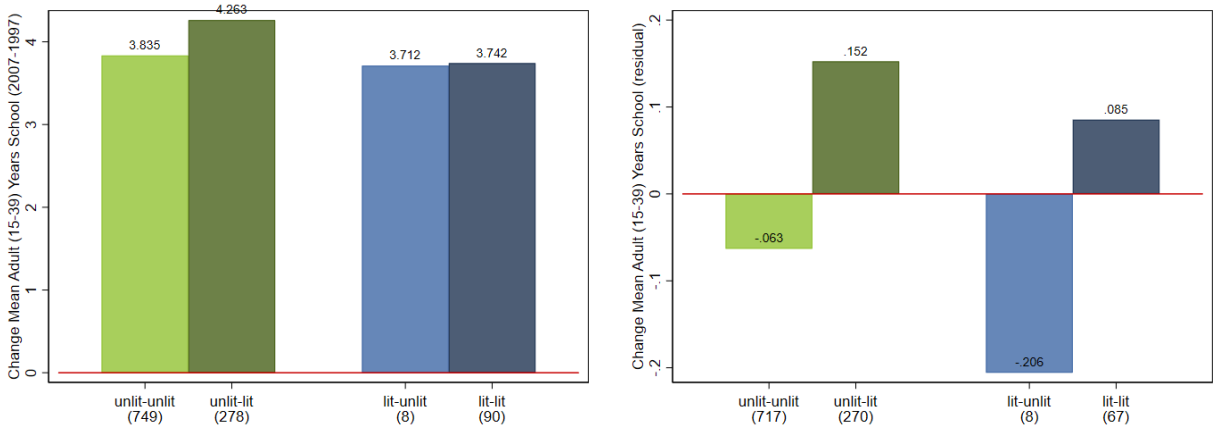
(a) 1997-2007

(b) 1997-2007 (residuals)



(c) 1997-2017

(d) 1997-2017 (residuals)



The figure plots the change in average years of schooling for 15-39-year-olds by changes in the extensive margin of luminosity across Mozambican localities (admin-4 units). Green bars plot the mean years of schooling change for initially unlit localities. Dark green bars plot the change in schooling for localities that are turning lit, while light green bars plot the change in schooling for localities remaining unlit. Blue bars plot the mean years of schooling change for initially lit localities. Dark blue bars plot the change in schooling for localities that remain lit, while light blue bars plot the change for localities that turn from lit to unlit. Panels (a) and (c) plot unconditional changes in mean schooling years over 1997 and 2007 and over 1997 and 2017, respectively; panels (b) and (d) plot changes in mean years of schooling over 1997-2007 and 1997-2017, conditional on admin-3 fixed effects.