Nighttime light intensity as a proxy for economic and social development at the local level*

(Preliminary draft: Please do not cite or circulate without the authors' permission)

Anna Bruederle† and Roland Hodler‡

October 28, 2016

Abstract

We study whether, and up to which resolution, nighttime light reflects the key dimensions of human development. We construct indicators of health, education and wealth from geo-coded DHS data for 29 African countries and aggregate them into spatial grids with cells of roughly 10×10 km to 50×50 km. We find that nighttime light is associated with human development at the local level when comparing cells within countries, and also when comparing cells across countries. We conclude that nighttime light can be used to detect variation in socio-economic outcomes across small geographic units. Further, nighttime light proves to be a useful measure of development when we are concerned not only with a region's level of economic activity, but also with the level of social wellbeing of its people.

Keywords: Africa, local development, nighttime light, poverty measurement.

JEL classification: I14, I25, I32, O18, O55.

1 Introduction

Economic and social data for subnational geographical units, such as provinces or municipalities, ethnographic regions or ecological zones, are unavailable for most developing countries, and of poor quality if they exist. The lack of spatially disaggregated data has

*This research project has benefited from funding support by the Basic Research Fund (GFF) of the University of St. Gallen.

†SIAW-HSG, University of St.Gallen. Email: mirjamanna.bruederle@unisg.ch.

‡Department of Economics and SIAW-HSG, University of St.Gallen; OxCarre, University of Oxford; and CESifo, Munich. Email: roland.hodler@unisg.ch.

inhibited empirical research on many interesting and relevant questions in development economics, economic geography, growth theory, public economics and political economics. For example, subnational-level data are indispensable to understand what drives differences in economic growth between regions within the same country; whether certain political regimes tend to encourage regional favoritism; how natural resource extraction affects the local economy; how economic development is linked to ethnic divisions; and so forth. In addition, being able to observe economic and social outcomes at subnational level allows to exploit within-country variation in policy regimes or allocation of resources to draw causal conclusions. Within-country comparisons can be better suited than cross-country comparisons to evaluate policies or resource allocation effects because countries vary in many dimensions that are difficult to control for.

Because of the lack of reliable subnational statistics, social scientists have recently resorted to alternative measures that do not depend on data collection on the ground. Nighttime light intensity, calculated from weather satellite recordings, has been proposed by Henderson et al. (2012) as a good proxy for economic activity. They show that changes in nighttime light intensity are positively correlated with changes in GDP at the country level. A key benefit of nighttime light data is that they are measured with consistent quality across countries with very different institutional capacities, and are not susceptible to politically motivated manipulation. In addition, given their availability for all inhabited areas of the world in pixels corresponding to less than one square kilometer, they can be aggregated at the level of any subnational geographical unit as needed by the researcher.

For these reasons, nighttime light data have been used in many empirical studies to construct measures of economic activity in subnational regions for which disaggregated statistical data are not available. Some of these studies have been published in the most renowned economics journals. As within the community of social scientists both the recognition of nighttime light data as a measure for economic activity, and the competence in the use of specialized software to process them, are becoming more wide-spread, it is likely that we will see many more applications in the future.

On closer examination, our understanding of what nighttime lights do and what they do not capture is rather limited to date. The seminal paper by Henderson et al. (2012) has shown that changes in nighttime light intensity correlate with growth in GDP at the level of countries. Hodler and Raschky (2014a) have demonstrated that this correlation holds also for subnational provinces based on the province-level GDP data by Gennaioli et al. (2013). Chen and Nordhaus (2011) document a positive relationship between luminosity and GDP at the level of grid-cells of 1° latitude $\times$ 1° longitude (with 1° corresponding to approximately 110km at the equator). There are however at least two important gaps in the evidence that underpins the validity of nighttime light as a tool in the social sciences. First, we do not know whether nighttime light intensity is an accurate proxy for economic activity and economic development for small geographical units such as municipalities. Second, very few studies have looked at the relation between nighttime light and other dimensions of human development such as education and health. Much uncertainty remains around the question whether nighttime light captures welfare in a broader sense than economic output.

This study provides evidence to fill these gaps. We explore the relationship between nighttime light intensity and human development at the local level in Africa. We use geo-referenced data from the Demographic and Health Surveys (DHS) to construct local indicators of health, education and wealth, and relate them in time and space to nighttime light intensity. As a spatial structure for our analysis we use grids of different resolutions spanned over the African continent. We fill cells which intersect with DHS sample clusters with values for our local development indicators, and use cells where we have information from two or more survey waves to analyze change over time. Specifically, this approach allows us to address the following research questions: *(i)* What is the association between nighttime light intensity and indicators of human development at the local level within African countries?, and *(ii)* What is the association between nighttime light intensity and indicators of human development at the local level across Africa?

Our largest grid cells are aligned to those of the PRIO-GRID dataset (Tollefsen et al., 2012), which have a cell length of 0.5 decimal degrees (corresponding to roughly 55 km at the equator). The PRIO-GRID dataset provides time series of a rich set of socio-economic, resource, landuse and climatic variables in a standardized spatial grid structure. We believe that it will become a relevant data source for spatial studies in the social science, and hence constitutes a meaningful framework for our analysis. We replicate our analysis on two finer grids, with cells of 0.25×0.25 decimal degrees, and 0.1×0.1 decimal degrees, respectively. We focus our analysis on Africa because it is the region where the lack of spatially disaggregated data on economic and social development outcomes is most eminent.

We find that more intense nighttime light emissions are associated with better human development outcomes in terms of health, education and wealth at the local level. The associations hold when we compare grid cells within a country, and when we compare grid cells all over the African continent. We also find that the association is rather stable when we apply different grid resolutions, but smaller grid cells tend to yield smaller effect sizes. Importantly, after controlling for local electrification and urbanization rates, we find that nighttime light contains additional information about variation in local development, both within countries and across the continent. We conclude that nighttime light is a good proxy not only for economic development as measured in GDP, but also for other dimensions of human development related to health and education. This is evidence in support of future use of nighttime light data for studies that are concerned not only or not primarily with differences in economic outcomes across regions, but also with regional differences in health and education outcomes. It is also evidence in support of the use of nighttime light as a proxy for development in very fine-grained analyses with small spatial units of observation.

1.1 Related literature

Our study contributes to previous evidence on how nighttime light correlates with measures of economic and social development. Early studies that have proposed nighttime light intensity as a proxy for economic outcomes include Sutton and Costanza (2002); Doll et al. (2006); Sutton et al. (2007); Elvidge et al. (2009) and Ghosh et al. (2009). Among economists, nighttime light gained broad recognition as a measure of economic activity after Henderson et al. (2012) established a strong correlation between nighttime light and GDP at the level of countries. Hodler and Raschky (2014a) and Chen and Nordhaus (2010) confirmed this correlation for subnational provinces and grid cells of 1×1 decimal degrees. The most fine-grained analysis of how nighttime light is associated within economic outcomes is provided by Mellander et al. (2015) in a study on Sweden. They use geo-coded micro-data on population, enterprises, employment and wages at a resolution of 250×250 meter grid cells in urban, and $1,000 \times 1,000$ meter in rural areas. Data at this level of spatial accuracy are however not available for developing countries.

The correlation between nighttime light and a survey-based wealth index has first been studied by Noor et al. (2008), who use aggregates of nighttime light and wealth at the level of the largest subnational units. More recently, Weidmann and Schutte (2016) have analyzed the correlation between the DHS wealth index and nighttime lights within circular buffer zones of 2 and 5 km radius around urban and rural DHS clusters. In their sample of 39 developing countries, they find that light is a good predictor for wealth at the local level. In a recent publication, Jean et al. (2016) show that accurate estimates of consumption and wealth at the level of survey clusters can be produced through a novel machine learning technique which processes daytime satellite images and nighttime light. In order to increase the precision of estimations at lower expenditure levels where nightlights display little variation, the technique uses features visible in daytime satellite imagery to capture variation among poorer clusters. In an application to five African countries, their model can explain 55 to 75 % of the variation in average household wealth. Nighttime light has also been shown to be correlated with electrification rates at the level of countries and subnational units (Elvidge et al., 2011), and at the very local level by Min et al. (2013), using data from a household surveys in Senegal and Mali. One of the first attempts to use nighttime light as a predictor for a social development outcome beyond economic well-being at the subnational level has been made by Chen (2015). He finds that nighttime light is correlated with infant mortality and poverty rates at the level of provinces.

Comparisons of economic outcomes between various subnational units of interest, enabled by the use of nighttime lights, has recently helped answer research questions related to drivers of conflict, the economic importance of political institutions, regional inequality, and regional favoritism. Nighttime light aggregated at the level of ethnic homelands is used by Alesina et al. (2016) to measure inequality between ethnic groups; by Esteban et al. (2015) to study drivers of conflict; by Michalopoulos and Papaioannou (2013a) to examine the effect of contemporaneous national institutions; and by Michalopoulos and Papaioannou (2013b) to study the role of pre-colonial ethnic institutions on contemporary development. Rohner et al. (2013) assess the effect of civil conflict on trust and ethnic identity in Uganda by aggregating nighttime light at the level of counties to match their county-level data on social capital. Besley and Reynal-Querol (2014) use luminosity within 1×1 km grid cells to study the effects of historical conflict on contemporary conflict and development. In order to assess the casual impact of economic shocks on conflict in Africa, Hodler and Raschky (2014a) use lagged rainfall within administrative units of the second subnational level as an instrument for economic shocks, where their first stage is estimated with nighttime light as the dependent variable. By comparing how nighttime lights develop in birth regions of political leaders and in other regions, Hodler and Raschky (2014b) find evidence for regional favoritism in a sample of in 126 countries. Brown et al (2015) study how proximity to a microfinance bank affects financial inclusion and use nighttime light within a 5 km zone around a bank to control for local economic activity.

The recent proliferation of studies which build on units of observation other than national states and focus on regions where relevant data from statistical offices are sparse demonstrates the wealth of possible applications of nighttime light. Many of the research questions that could be approached empirically by using nighttime light, however, are not only, or not primarily, concerned with economic development as an outcome. Differences in other key dimensions of development, such as health and education, can often provide a more in-depth understanding of inequalities across groups and subnational regions, and how these inequalities are linked to, for example, institutions or conflict. To our knowledge, no in-depth and spatially comprehensive evidence has been produced on how nighttime light is associated with different indicators of human development at a local level. We aim to address this gap by assessing whether and how nighttime light can help predict local outcomes in health, education and wealth in African countries.

2 Data and methods

Our units of observation are quadratic cells of a spatial grid spanned over the African continent. We work with three different grid resolutions: our cell sizes are 0.5×0.5 decimal degrees, 0.25×0.25 decimal degrees, and 0.1×0.1 decimal degrees. We fill our cells with values for health, education and wealth calculated from geo-coded Demographic and Health Surveys (DHS), and with average nighttime light intensity calculated from the DMSP-OLS Nighttime Lights Time Series issued by the National Oceanic and Atmospheric Administration (NOAA). Our samples cover 29 African countries and a period of 22 years from 1992 to 2013. Within this period, we have one to five geo-coded DHS survey waves per country. For the first part of our analysis, with work with a pooled cross-section of approximately 7,500 cells in our coarsest grid, more than 17,000 cells for our intermediary grid, and more than 47,000 cells for our finest grid.

For the second part of our analysis which is concerned with change over time, we construct a panel where cells are the panel dimension and years are the time dimension. The panel is highly unbalanced because the number of survey waves and intervals between waves vary across countries, and because a new sample of clusters is selected for each DHS wave. Hence, even in countries for which we have two or more survey waves, only a subset of cells is observed more than once for each grid resolution.

We align our coarsest grid to the PRIO-GRID dataset (Tollefsen et al., 2012). The PRIO-GRID dataset provides time series of a rich set of socio-economic, resource, landuse and climatic variables, including nighttime light, in a standardized spatial grid structure. It has a resolution of 0.5×0.5 decimal degrees, corresponding to cells of roughly 55 x 55 km at the equator, and covers all terrestrial cells on the globe. Given the broad applicability of the PRIO-GRID dataset, and the fact that it can easily be merged with other geo-referenced data, it will probably be used by many social scientists for spatial analyses in the future. We therefore choose it as a reference grid structure for our study.

2.1 Nighttime light as the main explanatory variable

Our nighttime light variable is based on satellite images collected by the Operational Linescan System (OLS) sensors installed on satellites of the Defense Meteorological Satellite Program (DMSP). These weather satellites circle the earth several times per day and collect a digital stream of images relevant for weather observation and forecasting. OLS sensors are designed to help identify cloud coverage at night through detecting moonlight reflections, but on cloud-free nights they record light emissions from the earth's surface. These images are processed by the National Geophysical Data Center of the National Oceanic and Atmospheric Administration (NOAA-NGDC) into global annual composites of cloud-free nighttime light. Of the different nighttime light series produced by NOAA, we choose the Stable Lights series for our analysis because this is the product that is used most frequently in the economic literature. The Stable Lights series is designed as a global map showing the relative nighttime light intensities on the earth's surface, where transient lights that are deemed ephemeral have been filtered out and non-lit areas are set to zero. Several steps of screening and filtering of the raw images, both mannual and by automated algorithms, are involved in obtaining the Stable Lights product, which are described in detail in Baugh et al. (2013) and briefly summarized here. First, from all the images collected throughout a year, only nighttime observations not affected by lunar illuminance are selected to feed into the annual composite. Further, images affected by cloud coverage are removed because the presence of clouds can either obscure lights completely, or diffuse the signals so that they appear larger but dimmer. As a next step, transient light is separated from persistent light in order to filter out lights from fires or fishing boats. This is done by an algorithm which analyzes the composite histograms and removes bright outliers. The selected segments are then averaged into a global annual cloud-free outlier-removed image. This annual average still contains background noise, i.e., non-zero light intensity values in areas where lights are not present. Therefore, as a final step, areas that are below a locally computed threshold for background noise are deemed as areas where lights are not present and set to zero.

The NOAA currently provides annual data for the time period from 1992 to 2013, in gridded format with a cells of 30×30 arc seconds. This cell site corresponds to less than one square kilometer at the equator. For each of the cells, annual average light intensity is reported in digital number (DN) range from 0 to 63, with higher values implying more intense nighttime light.

The Stable Lights series has two caveats that are a concern for many applications in the social science. First, the digital numbers are often saturated and top-coded at DN 63 in the centers of metropolitan areas due to calibration of OLS sensors which allows to detect very low levels of illuminance. The Stable Lights maps hence do not allow to distinguish between bright urban centers and their periphery (Letu et al. (2012) provide a detailed discussion of the saturation problem). However, for the African continent the fraction of cells in the original data with DN 63 is less than 0.06 %, which is why we do not consider it a concern for our study.

The second caveat is that permanent light sources of low intensity might be inappropriately set to zero by the processing steps involved in the preparation of the Stable Lights series. This concern has been pointed out by Chen and Nordhaus (2010), and by Henderson et al. (2012), who note that pixels with DN 1 and 2 seem to be underrepresented in the data. For our samples, the question whether to interpret a DN of zero as complete absence of light emission, or as possibly very low light emission that has been filtered away, is very relevant. If we clip the 2013 Stable Lights to the areas from which our survey data originate, i.e., to the buffer zones of 5 km around rural and 2 km around urban DHS clusters, we obtain a share of 86 % of pixels with DN zero.[1] By contrast, these areas contain no pixels of DN values 1, 2 or 3. Given that we know that these areas are inhabited, we interpret a DN of zero as none to very low light emission, rather than no light emission with certainty.

To construct a nighttime light variable at the level of our grid cells (*light*), we calculate average light values for each cell for each year in which a DHS survey was carried out.[2] We use GIS software for these calculations. The share of cells with an average light value of zero is 42 % for our coarsest grid resolution; 58 % for the intermediary resolution; and 72 % for the finest resolution.

[1]To calculate this figure, we first merge the 5 km (for rural clusters) and 2 km (for urban clusters) circular zones around DHS clusters in our sample, where we use only the latest DHS wave for each country. We then clip the Stable Lights 2013 raster to this mask and analyze the distribution of pixel values within the mask.

[2]If DHS data collection extended over two calendar years, we use the earlier year, because most of the survey questions we use to construct our indicators refer to the recent past.

2.2 Construction of dependent variables from DHS data

Our dependent variables are constructed from the Demographic and Health Surveys (DHS), which are large periodic household surveys that have been carried out in low-income countries in Africa and elsewhere since the 1980s (ICF International, 1992-2013).[3] These surveys primarily collect information from women at childbearing age on a wide range of topics related to health, nutrition, fertility and education, as well as a set of household characteristics such as access to infrastructure and ownership of household assets. In each country, households are selected to produce nationally representative samples. Usually, sampling is done through a stratified cluster design, based on the country's most recent population census, in two stages. At the first stage, clusters are drawn from official listings of census enumeration areas, which in most countries correspond to small villages or blocks within larger villages or cities. At the second stage, a sample of households is drawn randomly from a list of households in each cluster. Mean outcomes for a cluster hence provide a picture of the developmental state at local level. DHS follow a largely consistent methodology and structure across countries and years, which is why they qualify for the analysis of local development across countries and over time. Nevertheless, amendments are made to the questionnaires from time to time, and the coding of some items is country-specific.

For those DHS that are geo-referenced, the data contain geo-coordinates of the cluster center, usually recorded with Global Positioning System (GPS) receivers. To ensure confidentiality of the respondents, some noise is added to the coordinates by displacing each locality in a random direction and by a random distance of 0 to 2 km for urban clusters, and 0 to 5 km for rural clusters, with 1 percent of rural clusters displaced by up to 10 km.

Our samples cover 71 survey waves across 29 countries in Africa. This includes all DHS with GPS-measured geo-coordinates that have been carried out in African countries over the time period for which nighttime light data are currently available, i.e., 1992 to 2013.[4]

In order to capture the key dimensions of human development in our sample localities, we construct from the DHS data a set of 6 indicators related to health, education, and wealth. Our choice of indicators is restricted by the information available across all DHS in our sample. Specifically, we construct the following indicators:

1. Primary school attendance (*primary school attendance*).
   We calculate the net attendance ratio in primary education as the ratio of the number of children of official school age (as defined by the national education system) who have attended primary school in the year preceding the survey to the total population of children of official school age.

[3] For further details on DHS methodology and data, see http://dhsprogram.com.

[4] We exclude from our analysis DHS which identify survey locations through gazetteers. This is because gazetteers identify locations by the center of the respective village, town or city. This does not allow to distinguish between the periphery of larger metrolpolitan areas and the presumably wealthier and more developed center. Clusters located by gazetteers make up less than 1 percent of all geo-coded DHS clusters in Africa in our sample period.

2. Number of years of schooling of household members aged 18 or older (*years of education*).
3. Infant mortality rate (*infant mortality*).
   Infant mortality rate is typically defined as the number of infants that die before reaching the age of one year per 1,000 live births. In order to make full use of records for children born within a relatively short period preceding the survey, we follow the DHS methodology and derive infant mortality rates through a synthetic cohort life table approach. This approach calculates mortality probabilities for small age segments (0 months, 1-2 months, 3-5 months, 6-12 months) and combines them into the one-year age segment. Because we aim to generate a relatively near-term picture of infant mortality, but need to feed a reasonable number of births into our computation, we use birth records from the three-years period preceding the survey.
4. Proportion of births attended by skilled health personnel (*profess. birth assist*).
   We calculate the percentage of deliveries attended by a doctor, a nurse or a midwife among all births in the 3-years period preceding the survey, for the same considerations spelled out under infant mortality rates.
5. Wealth index (*wealth*).
   Because the DHS do not collect data on household income or consumption, we proxy wealth by an index of household asset indicators. We use the DHS wealth index when it is available. The DHS wealth index is constructed as a linear combination of indicators of whether the household owns selected assets; of the type of water, sanitation and energy facilities; and of the housing quality. Weights for each of the components are derived by principal component analysis (PCA).[5] For surveys where the DHS wealth index is not available, we compute an analogous index following the DHS methodology. We then include those wealth indicators that are available across all surveys: ownership of radio, television, refrigerator, bicycle, motorcycle, and car; floor materials; type of drinking water source; type of toilet facility; and access to electricity. Given that the absolute value of the PCA-based combination is not meaningful per se, we allocate all interviewed households into wealth quintiles within the survey wave, and use the quintile as the dependent variable. Note that this methodology does not allow for comparisons of wealth outcomes across survey waves.
6. Electricity-independent wealth index (*electr.indep. wealth*).
   Several of the components of the DHS wealth index depend on electricity access for their use, and electricity access is inherently correlated with nighttime light emission. We therefore construct a second wealth index, again following the DHS wealth index

[5] For a detailed description of how the DHS wealth index is constructed, see Rutstein and Johnson (2004). For a general discussion on the use of asset indices to capture household wealth, see for example Filmer and Scott (2012).

methodology, in which we include only indicators that do not depend on electricity: ownership of bicycle, motorcycle, and car; floor materials; type of drinking water source; and type of toilet facility. Again, we compute for each household its respective quintile within the survey wave and use the quintile as the dependent variable.

For each of these indicators, we derive cell-level aggregates from the geo-coded DHS. We do this by allocating households to cells, taking into account that the reported geo-codes represent only the approximate center of an area within which the surveyed households are actually located. For those DHS clusters whose center points are reported to lie close to a cell boundary, or a vertex, we allocate a respective part of households to the neighboring cell(s).

More specifically, we use the following four-step procedure to aggregate development outcomes at cell level. First, we calculate a mean value at cluster level for each indicator, and retain the number of observations on which this mean is based. Note that the number of observations per cluster varies across our indicators depending on the relevant base, i.e., births during the 3 years preceding the survey, children at school age, adults over age 18, or households. Second, we draw circular zones around the reported cluster center points, where we use a 2 km radius for urban clusters, and a 5 km radius for rural clusters. These zones account for the random displacement of cluster coordinates, while at the same time reflecting the spread of respondent households around cluster center points. We clip the cluster zones to country borders and coast lines, because we know that the surveyed households reside on terrestrial areas within the national boundaries of the respective DHS country. Third, we intersect the zones with our spatial grid, which leaves us with patches of DHS cluster zones that can each be uniquely attributed to a specific grid cell. We divide the number of observations in each cluster into these patches, in proportion to the area share of the patch in the full circle.[6] As a final step, we can then compute average indicator values for grid cells: We compute a weighted mean from the respective cluster-level indicators, using as weights the number of observations that fall within the cell. This procedure yields cell-level values for our development indicators for all grid cells that intersect with at least one DHS cluster zone. The number of observations underlying the cell mean, and the number of clusters from which these observations stem, vary across cells and across indicators. Steps 3 to 5 are replicated for each of our three grid resolutions. We use geographical information system (GIS) software for steps two and three above.

For 26 countries we have more than one geo-coded DHS in our sample period. We calculate cell-level development indicators from each survey wave for these countries. The cells that intersect with at least one DHS cluster zone from more than one survey wave constitute the basis for our analysis over time.

[6] We thereby assume that households are distributed uniformly across space within the cluster zone.

2.3 Independent and control variables

We use three control variables in some of our regressions. First, we control for population density in the year 2010 within the cell (*population*), which we compute from the CIESIN Gridded Population of the World (GPWv4) dataset (Center for International Earth Science Information Network, 2016). We use population density in 2010 as a proxy for population density in any year within our sample period and hence do not account for possible differences in population trends across cells. This is because the CIESIN GPWv4 grid draws on population data from the 2010 round of censuses, and extrapolation to earlier years is done at different administrative levels across countries, depending on data availability from previous census rounds. Since we cannot be sure about the resolution at which differences in population growth are captured, we prefer to work with a static variable.

Second, we control for the electrification rate within the cell (*electricity*), which we derive from the DHS data using the same four-step procedure applied to generate other cell-level development indicators. Third, we control for the urbanization rate within the cell (*share urban*), which we also derive from the DHS data. Along with the geo-coordinates, the DHS report for each cluster whether it is classified as urban or rural according to the national statistical administration.[7] We compute the share of clusters within the cell that are classified as urban.

3 Cross-sectional analysis

3.1 Empirical specification

For our cross-sectional analysis, we pool all geo-coded DHS in African countries within our sample period. For each of our grid resolutions, we regress each cell-level development indicator for a specific year on the logarithm of mean nighttime light for that cell in the same year (or in the first year of data collection if it stretched over two calendar years), and on population density within the cell. In our main model, we include country-year fixed effects. We thereby account for differences in survey methodology in the assessment of household wealth across countries and waves. Regression estimates in this model hence capture the association between development and light within countries and at a given point in time. We introduce as additional controls in this model the cell-level electrification and the urbanization rates. Our coefficient on the light variable then measures the additional information contained in local nighttime light intensity that is not yet reflected in local population density, electricity access, and country-specific features that distinguish urban from rural settlements.

[7] The rural-urban classification reported with the DHS geo-codes varies between countries and is therefore a highly imperfect indicator of differences in urban-rural living standards. This is of course not a concern in specifications which include country fixed effects.

In order to find out whether nighttime light is associated with development across countries, we run the same regressions without country fixed effects. In these regressions, we still include year fixed effects to absorb progress that is common to all units in our sample, and to account for innovations in the remote sensing technology over time. We exclude wealth indicators from this part of the analysis because they are not based on a coherent methodology across countries. We cluster standard errors at the country level in all our models.

We use the logarithm of nighttime light and of population density because the distribution of these variables are strongly right-skewed in our samples. The logarithmic transformation allows us to give more weight to the variation in light and population density in the lower segments where observations are concentrated. Given the high share of observations with a nighttime light value of zero in our samples, we add a constant of 0.01 before taking the logarithm in order not to drop these cells from the samples. This is in line with our assumption that values of zero in the Stable Lights data can still occur in places with very low levels of light emission, as discussed above. The coefficient of interest in a log-level model can be interpreted as the increase in the dependent variable, in our case DHS outcomes, by one unit from a one percent increase in the log-transformed regressor, in our case nighttime light intensity. However, once we add a small constant before taking logs, the coefficient only corresponds to an approximation to this effect. Nevertheless, we show that our results are robust to using a logarithmic transformation of mean nighttime light without any constant added.

3.2 Results

Table 1 presents our main results, which are OLS regressions of our DHS-based development indicators on the logarithm of nighttime light with a small constant added, country-year fixed effects, and various controls. Panel A presents estimates based on grid cells of 0.5×0.5 decimal degrees, aligned to the PRIO-GRID, and panel B presents estimates based on our smallest grid cells of 0.1×0.1 decimal degrees. We find that nighttime light is associated with all indicators of human development at the local level within survey waves for either of the grid resolutions. The effects are small in size, but statistically significant. For cells of 0.5×0.5 decimal degrees, when controlling only for population density (odd columns in panel A), a 10 % increase in mean nighttime light within the a cell is associated with an increase in the primary school attendance rate by 0.4 percentage points; with an increase in adults' years of education by 0.06; with a drop in infant mortality by 0.25 per 1000 births; with an increase in the share of births assisted by professional health personnel by 0.6 percentage points; with an increase in the average wealth quintile within the cell by 0.035 for the standard wealth index, and by 0.027 for the index of electricity-independent assets. For all of our indicators, the size of these effects corresponds to less than one tenth of a standard deviation. When adding the local electrification and urbanization rates as controls, we obtain statistically significant positive coefficients on these variables (except for the regression on infant mortality) which are also larger in size than the coefficients on light in the simplest model. Better electricity access and the higher degrees of urbanization are positively associated with human development outcomes at the local level. When introducing these additional controls, the coefficients on light drop by more than half. We also lose significance on the light coefficient for the regression on infant mortality. This means that part of the variation in nighttime light is absorbed by variation in electrification and urbanization, but nighttime light still contains information about the local level of development that goes beyond these two explanatory variables.

When we compare results for the 0.5×0.5 decimal degrees grid to those for the 0.1×0.1 decimal degrees grid (panel B of Table 1), we can see that smaller grid cells yield smaller effect sizes across all models, which are still statistically significant. Overall we find very similar patterns of results for both grid resolutions.

We conclude that local variation in nighttime light within a country at a given point in time captures the state of human development at the local level. A level of light emission that is relatively high compared to the country mean is indicative of a relatively healthy, well educated and wealthy local population. Nighttime light is still informative of local development after controlling for the degree of urbanization and electrification.

Next, we run the same regressions, but instead of country-year fixed effects we include year fixed effects only. Given that the methodology used to derive our wealth variables is specific to country and survey wave, we have to exclude wealth outcomes from this part of the analysis. Results are presented in Table 2. We find that across countries, local nighttime light intensity is also positively correlated with health and education of the local population. Effect sizes are slightly larger than in the models with country-year fixed effects. Analogous to our results within countries, a substantial part of the association between light and development is explained by local electrification rates and levels of urbanization. Contrary to the within-country model, the effect of nighttime light on infant mortality is still statistically significant after introducing additional controls. When comparing different grid resolutions, effects for the 0.1×0.1 decimal degree cells again turn out smaller in magnitude than for the larger cells, but are still statistically significant for most of the models. We conclude that variation in nighttime light across the continent is a proxy for the local developmental status in terms of health and education. Even if information on local electrification and urbanization is available, nighttime light contains additional information about variation in local development.

We run two robustness tests to check whether the relationship between nighttime light and development holds also for variations of our model specifications. First, we use as the explanatory variable in our within-country model the logarithmic transformation of nighttime light without adding a constant. This implies that we drop from our samples those cells that have a mean nighttime light value of zero. In spite of the substantially reduced sample, we find very similar results to those in our main specification. For the fine grid resolution of 0.1×0.1 decimal degrees, coefficients on light are no longer significant for primary school attendance and professional birth assistance after controlling for levels of electrification and urbanization.

Second, we replicate our main specification while accounting for the substantial variation in the numbers of observations per cell across cells. Cells in densely populated areas tend to intersect with several clusters, while cells in remote rural areas often carry information from only a small number of households in a single cluster. We therefore run weighted regressions where we use as weights the number of observations for each development indicator within a cell. Results are presented in Table 4. Our main results are robust to weighting, but we lose significance on the light coefficient for some of the models when we control for electrification and urbanization.

4 Conclusions

Social scientists are increasingly using nighttime light intensity, calculated from weather satellite recordings, as a proxy for economic activity in regions for which disaggregated data from statistical offices are not available. Previous literature has shown that nighttime light intensity correlates with GDP at the level of countries, at the level of subnational provinces, and at the level of grid cells of 1×1 decimal degrees (approx. 110×110 km at the equator). Our paper complements the evidence that underpins the use of nighttime light in the social sciences by two important aspects: First, we apply a higher spatial resolution to study whether nighttime light intensity is an accurate proxy for development within small geographical units. Specifically, we analyze grid cells of roughly 10 to 50 square kilometers. Second, we examine whether nighttime light also captures other dimensions of human development, notably education and health, at the local level. We construct indicators of health, education and wealth from geo-coded DHS data for 29 African countries and 71 survey waves and aggregate them into a spatial grid structure. We use grids of a resolution of 0.5×0.5 decimal degrees; 0.25×0.25 decimal degrees; and 0.1.1 decimal degrees. Further, we calculate the mean nighttime light and the mean population density, as well as measures of electricity access and urbanization, within these grid cells. We then regress our development indicators on local nighttime light in the year when the survey was carried out, and population density. We find that primary school attendance, adults years of education, infant mortality, and the rate of births assisted by a health-care professional are all correlated with nighttime light. This correlation is present both within countries and across countries. Wealth measured as an index of household assets is also correlated with nighttime light at the local level within countries. For regressions on most of our indicators, nighttime light retains a statistically significant coefficient after controlling for local electrification rates and urbanization: Part of the variation in local development is explained by electricity access and urbanization, but nighttime light contains additional information about the local development status not reflected in these two variables. Further, we find that the relationships are overall stable across different grid resolutions, but as grid cells get smaller, effects also become smaller in magnitude. We conclude from this evidence that nighttime light can serve as a proxy for development even at very high spatial resolutions. For example, it might be used to detect variation in socio-economic outcomes across small geographic units such as municipalities, zones directly affected by natural resource extraction, or areas surrounding physical infrastructure. Second, we conclude that nighttime light captures not only economic development as measured in GDP, but also human development as measured in health, education and household wealth outcomes. This implies that future studies concerned with subnational differences in human development may well use nighttime light to measure these differences in absence of other reliable disaggregated data. It also implies that some of the existing evidence on subnational inequalities derived with the use of nighttime light might be revised to broaden the interpretation of results. Differences in nighttime light across regions are indicative not only of differences in economic activity, but also of differences in peoples social wellbeing.

References

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

Baugh, K., Elvidge, C., Ghosh, T., & Ziskin, D. (2013). Development of a 2009 stable lights product using dmsp-ols data. *Proceedings of the Asia-Pacific Advanced Network*, *30*, 114–130.

Besley, T., & Reynal-Querol, M. (2014). The legacy of historical conflict: Evidence from africa. *American Political Science Review*, *108*(02), 319–336.

Center for International Earth Science Information Network. (2016). Gridded population of the world, version 4 (gpwv4). Available from `http://www.ciesin.columbia.edu/data/gpw-v4`

Chen, X. (2015). Explaining subnational infant mortality and poverty rates: What can we learn from night-time lights? *Spatial Demography*, *3*(1), 27–53.

Chen, X., & Nordhaus, W. (2010). *The value of luminosity data as a proxy for economic statistics*. Cambridge, MA: National Bureau of Economic Research.

Doll, C., Muller, J.-P., & Morley, J. (2006). Mapping regional economic activity from nighttime light satellite imagery. *Ecological Economics*, *57*(1), 75–92.

Elvidge, C., Baugh, K., Sutton, P., Bhaduri, B., Tuttle, B., Ghosh, T., et al. (2011). Who's in the dark—satellite based estimates of electrification rates. In *Urban remote sensing* (pp. 211–224). John Wiley & Sons, Ltd.

Elvidge, C., Sutton, P., Ghosh, T., Tuttle, B., Baugh, K., Bhaduri, B., et al. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, *35*(8), 1652–1660.

Esteban, J., Morelli, M., & Rohner, D. (2015). Strategic mass killings. *Journal of Political Economy*, *123*(5), 1087–1132.

Filmer, D., & Scott, K. (2012). Assessing asset indices. *Demography*, *49*(1), 359–392.

Ghosh, T., Anderson, S., Powell, R., Sutton, P., & Elvidge, C. (2009). Estimation of mexico's informal economy and remittances using nighttime imagery. *Remote Sensing*, *1*(3), 418.

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

Hodler, R., & Raschky, P. (2014a). Economic shocks and civil conflict at the regional l evel. *Economics Letters*, *124*(3), 530–533.

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

ICF International (Ed.). (1992-2013). *Demographic and health surveys (various) [datasets]*. Calverton, Maryland: ICF International.

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

Letu, H., Hara, M., Tana, G., & Nishio, F. (2012). A saturated light correction method for dmsp/ols nighttime satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(2), 389–396.

Michalopoulos, S., & Papaioannou, E. (2013a). National institutions and subnational development in africa. *The Quarterly Journal of Economics*, *129*(1), 151–213.

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

Min, B., Gaba, K. M., Sarr, O. F., & Agalassou, A. (2013). Detection of rural electrification in africa using dmsp-ols night lights iimagery. *International Journal of Remote Sensing*, *34*(22), 8118–8141.

Rohner, D., Thoenig, M., & Zilibotti, F. (2013). Seeds of distrust: Conflict in uganda. *Journal of Economic Growth*, *18*(3), 217–252.

Rutstein, S., & Johnson, K. (2004). *The dhs wealth index: Dhs comparative reports no. 6.* Calverton, Maryland: ORC Macro.

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

Sutton, P., Elvidge, C., & Ghosh, T. (2007). Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. *International Journal of Ecological Economics & Statistics*, *8*(S07), 5–21.

Table 1: OLS egression of development indicators on nighttime light, pooled cross-section with country-year fixed effects.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	primary school attendance		years of education		infant mortality		profess. birth assist.		wealth		electr.indep. wealth	
Panel A: Grid resolution 0.5 × 0.5 decimal degrees												
ln(light+0.01)	0.039***	0.008**	0.612***	0.166***	-2.479***	-0.728	0.064***	0.014**	0.346***	0.085***	0.268***	0.117***
	(0.005)	(0.003)	(0.053)	(0.039)	(0.563)	(0.760)	(0.006)	(0.005)	(0.026)	(0.014)	(0.042)	(0.023)
ln(population)		0.022*		0.086		1.669		0.008		0.031		-0.009
		(0.011)		(0.070)		(1.152)		(0.007)		(0.046)		(0.046)
electricity		0.250***		3.558***		-14.047**		0.377***		2.039***		1.213***
		(0.060)		(0.279)		(5.593)		(0.036)		(0.110)		(0.394)
share urban		0.087***		1.228***		-4.762		0.153***		0.726***		0.390*
		(0.019)		(0.284)		(3.373)		(0.025)		(0.108)		(0.193)
Observations	7,436	7,424	7,442	7,429	7,498	7,498	7,123	7,123	7,144	7,144	7,275	7,275
Panel B: Grid resolution 0.1 × 0.1 decimal degrees												
ln(light+0.01)	0.024***	0.004*	0.412***	0.100***	-1.070***	-0.093	0.044***	0.010***	0.237***	0.057***	0.157***	0.059***
	(0.004)	(0.002)	(0.030)	(0.022)	(0.367)	(0.339)	(0.003)	(0.003)	(0.013)	(0.008)	(0.033)	(0.014)
ln(population)	0.026***	0.027**	0.126	0.148**	-0.278	-0.350	0.019**	0.021***	0.054	0.065*	0.022	0.029
	(0.009)	(0.010)	(0.080)	(0.061)	(0.657)	(0.710)	(0.008)	(0.006)	(0.062)	(0.038)	(0.054)	(0.040)
electricity		0.221***		3.329***		-10.571**		0.347***		1.915***		1.051***
		(0.047)		(0.212)		(4.134)		(0.029)		(0.108)		(0.351)
share urban		0.062***		1.138***		-3.423		0.145***		0.640***		0.334**
		(0.020)		(0.245)		(2.788)		(0.024)		(0.092)		(0.159)
Observations	46,786	46,647	46,847	46,699	47,205	47,205	44,914	44,914	44,576	44,576	45,061	45,061
Country-Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses; standard errors clustered at country level. *** p<0.01, ** p<0.05, * p<0.1

Table 2: OLS regression of development indicators on nighttime light, pooled cross-section without country fixed effects.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Grid resolution 0.5 × 0.5 decimal degrees								
	primary school attendance		years of education		infant mortality		profess. birth assist.	
ln(light+0.01)	0.057***	0.025**	0.742***	0.318***	-5.208***	-2.994***	0.077***	0.027**
	(0.010)	(0.011)	(0.114)	(0.111)	(0.807)	(0.730)	(0.012)	(0.012)
ln(population)		-0.004		-0.237		4.944***		-0.036
		(0.012)		(0.158)		(0.943)		(0.022)
electricity		0.224***		2.487***		-20.631***		0.283***
		(0.063)		(0.727)		(5.041)		(0.060)
share urban		0.052		1.373**		4.960		0.183***
		(0.039)		(0.584)		(4.254)		(0.037)
Observations	7,422	7,410	7,428	7,415	7,484	7,484	7,109	7,109
Panel B: Grid resolution 0.1 × 0.1 decimal degrees								
ln(light+0.01)	0.033***	0.008	0.508***	0.154**	-3.072***	-1.464***	0.050***	0.013
	(0.008)	(0.006)	(0.080)	(0.072)	(0.553)	(0.362)	(0.009)	(0.008)
ln(population)	0.013	0.016	-0.118	-0.074	2.635**	2.397**	-0.018	-0.014
	(0.015)	(0.014)	(0.191)	(0.166)	(1.125)	(1.090)	(0.022)	(0.019)
electricity		0.236***		2.904***		-18.953***		0.284***
		(0.048)		(0.557)		(3.699)		(0.048)
share urban		0.038		1.291**		3.926		0.180***
		(0.024)		(0.487)		(2.546)		(0.034)
Observations	46,622	46,483	46,683	46,535	47,041	47,041	44,750	44,750
Year FE	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses; standard errors clustered at country level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3: OLS regression of development indicators on nighttime light, pooled cross-section, no constant added to light DN before log transformation.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	primary school attendance		years of education		infant mortality		profess. birth assist.		wealth		electr.indep. wealth	
Panel A: Grid resolution 0.5 × 0.5 decimal degrees												
ln(light)	0.037***	0.008**	0.560***	0.130**	-1.541**	-0.008	0.058***	0.012**	0.311***	0.070***	0.228***	0.094***
	(0.006)	(0.003)	(0.060)	(0.049)	(0.601)	(1.078)	(0.006)	(0.005)	(0.021)	(0.011)	(0.039)	(0.022)
ln(population)		0.012		0.111		0.597		0.007		0.011		-0.045
		(0.008)		(0.070)		(1.170)		(0.007)		(0.033)		(0.035)
electricity		0.281***		3.683***		-12.922**		0.354***		2.006***		1.254***
		(0.052)		(0.279)		(5.679)		(0.038)		(0.121)		(0.364)
share urban		0.047**		1.176***		-4.385		0.160***		0.702***		0.252
		(0.018)		(0.342)		(4.410)		(0.028)		(0.133)		(0.264)
Observations	4,308	4,304	4,309	4,305	4,340	4,340	4,139	4,139	4,100	4,100	4,154	4,154
Panel B: Grid resolution 0.1 × 0.1 decimal degrees												
ln(light)	0.026***	0.003	0.566***	0.132**	-1.776*	-0.616	0.051***	0.007	0.301***	0.068***	0.175***	0.067***
	(0.006)	(0.003)	(0.065)	(0.058)	(0.970)	(0.747)	(0.005)	(0.005)	(0.013)	(0.015)	(0.047)	(0.015)
ln(population)	0.008	0.011*	0.052	0.110**	0.111	-0.047	0.008	0.014*	0.006	0.034**	-0.028	-0.016
	(0.007)	(0.006)	(0.085)	(0.050)	(0.669)	(0.640)	(0.011)	(0.007)	(0.040)	(0.015)	(0.042)	(0.032)
electricity		0.210***		3.340***		-11.581**		0.318***		1.789***		1.115***
		(0.043)		(0.292)		(5.464)		(0.029)		(0.096)		(0.342)
share urban		0.033*		1.175***		-1.014		0.133***		0.624***		0.044
		(0.016)		(0.371)		(3.833)		(0.031)		(0.144)		(0.217)
Observations	12,915	12,881	12,928	12,893	13,004	13,004	12,696	12,696	11,862	11,862	11,874	11,874
Country-Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses; standard errors clustered at country level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Weighted LS regression of development indicators on nighttime light, pooled cross-section, observations per cell used as weights.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	primary school attendance		years of education		infant mortality		profess. birth assist.		wealth		electr.indep. wealth	
	Panel A: Grid resolution 0.5 × 0.5 decimal degrees											
ln(light+0.01)	0.047***	0.009*	0.796***	0.068	-2.193***	0.304	0.076***	0.011	0.420***	0.065	0.394***	0.219***
	(0.007)	(0.005)	(0.058)	(0.073)	(0.678)	(0.945)	(0.008)	(0.008)	(0.025)	(0.038)	(0.058)	(0.075)
ln(population)		0.005		0.143***		-0.602		0.008		0.027		-0.141*
		(0.006)		(0.034)		(1.060)		(0.006)		(0.029)		(0.076)
electricity		0.176***		2.807***		-6.138		0.248***		1.323***		1.779***
		(0.048)		(0.463)		(7.167)		(0.064)		(0.250)		(0.481)
share urban		0.128***		2.855***		-14.308***		0.303***		1.374***		-0.461
		(0.041)		(0.680)		(3.458)		(0.066)		(0.392)		(0.721)
Observations	7,422	7,410	7,428	7,415	7,484	7,484	7,109	7,109	7,133	7,133	7,264	7,264
	Panel B: Grid resolution 0.1 × 0.1 decimal degrees											
ln(light+0.01)	0.025***	0.004	0.412***	0.014	-1.024***	0.572	0.044***	0.010	0.229***	0.028	0.237***	0.159**
	(0.004)	(0.004)	(0.039)	(0.048)	(0.330)	(0.447)	(0.004)	(0.006)	(0.023)	(0.027)	(0.040)	(0.061)
ln(population)	0.017***	0.010**	0.342***	0.192***	-1.298**	-0.680	0.029***	0.015**	0.118**	0.056***	-0.041	-0.048
	(0.005)	(0.005)	(0.071)	(0.022)	(0.505)	(0.472)	(0.006)	(0.006)	(0.047)	(0.016)	(0.079)	(0.052)
electricity		0.183***		2.751***		-8.931*		0.219***		1.353***		1.361***
		(0.043)		(0.291)		(4.643)		(0.047)		(0.170)		(0.341)
share urban		0.086***		2.262***		-10.959***		0.219***		1.151***		-0.296
		(0.031)		(0.593)		(2.209)		(0.054)		(0.307)		(0.577)
Observations	46,622	46,483	46,683	46,535	47,041	47,041	44,750	44,750	44,451	44,451	44,936	44,936
Country-Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses; standard errors clustered at country level. *** p<0.01, ** p<0.05, * p<0.1