

Määttä, Ilari; Ferreira, Thomas; Leßmann, Christian

**Article — Published Version**

## Nighttime lights and wealth in very small areas

Review of Regional Research

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Määttä, Ilari; Ferreira, Thomas; Leßmann, Christian (2021) : Nighttime lights and wealth in very small areas, Review of Regional Research, ISSN 1613-9836, Springer, Berlin, Heidelberg, Vol. 42, Iss. 2, pp. 161-190,  
<https://doi.org/10.1007/s10037-021-00159-6>

This Version is available at:

<https://hdl.handle.net/10419/286866>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



# Nighttime lights and wealth in very small areas: Namibian complete census versus DHS data

Ilari Määttä<sup>1</sup> · Thomas Ferreira<sup>2</sup> · Christian Leßmann<sup>3</sup>

Accepted: 10 September 2021 / Published online: 14 October 2021  
© The Author(s) 2021

**Abstract** Nighttime lights observed from satellites are a widely accepted proxy measure for economic development. This is mainly based on cross-country evidence that finds strong correlations between lights and Gross Domestic Product. Yet, the evidence on the correlations at local levels is scarce, and it often relies on

---

**Availability of data and material** Stable light data is publicly available at <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>. Demographic and health surveys data is publicly available for registered users at <https://www.dhsprogram.com/data/available-datasets.cfm>. Gridded population of the world data is publicly available at <https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11>. The complete census data is not public. We thank Namibian National Planning Commission with whom Thomas Ferreira worked previously for supplying us with the data. We only share final aggregated data sets to protect the respondents' anonymity in the census. For the publicly available data we provide R code to process the data. In addition, we provide R code for replicating the analysis using the aggregated data sets.

---

**Authors' contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ilari Määttä and Thomas Ferreira. The first draft of the manuscript was written by Ilari Määttä and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

---

**Consent for publication** All authors consent publication.

✉ Ilari Määttä  
[i.maeettae@tu-braunschweig.de](mailto:i.maeettae@tu-braunschweig.de)

Thomas Ferreira  
[tferreira@povertyactionlab.org](mailto:tferreira@povertyactionlab.org)

Christian Leßmann  
[christian.lessmann@tu-dresden.de](mailto:christian.lessmann@tu-dresden.de)

<sup>1</sup> Institute of Economics, TU Braunschweig, Braunschweig, Germany

<sup>2</sup> J-PAL Africa, University of Cape Town, Cape Town, South Africa

<sup>3</sup> ifo Institute & CESifo, TU Dresden, Dresden, Germany

randomly sampled survey data. We contribute by enhancing the understanding of the relationship between light and development at local levels. First, we use complete (non-publicly available) census data from Namibia to evaluate the findings based on the randomly sampled Demographic and Health Surveys data. We find that the census data provides a stronger association between light and wealth at local levels. Second, we criticize the practice of aggregating light from buffers around survey cluster locations. Instead, we recommend aggregating data in grid cells, and studying the relationship in different grid sizes. In our study correlations based on grid cells remain significant from a 0.5 degree grid to the smallest 0.0083 degree grid ( $\sim 1 \text{ km}^2$ ) allowed by the nighttime light data. Third, we supplement the commonly used relative wealth index by using individual asset variables as proxies for the total stock of wealth. The stock variables reveal a significant association between changes in light and wealth which cannot be found using our relative wealth index. Altogether, our results show that nighttime lights provide an even stronger signal of economic development at local levels than the current survey-based results in the literature suggest.

**Keywords** Nighttime Lights · Wealth · Spatial Data · Demographic and Health Surveys

**JEL** R1 · O1

### Nachtlichter und Wohlstand in Kleinräumigen Daten:

Zensus Daten vs DHS Daten in Namibia

**Zusammenfassung** Satellitenbilder der Erde bei Nacht sind ein inzwischen weit verbreiteter Indikator für wirtschaftliche Entwicklung. Insbesondere Länderdaten weisen eine starke Korrelation zwischen Bruttoinlandsprodukt und Lichtemissionen auf. Weniger eindeutig ist der Zusammenhang auf lokaler Ebene, wo in weniger entwickelten Ländern auf Haushaltsbefragungen zur Generierung von Wohlstandsmaßen zurückgegriffen werden muss. Breit verfügbare, geo-referenzierte Haushaltsdaten bieten die Demographic and Health Surveys, welche jedoch in dünn besiedelten Gebieten nicht-repräsentativ sind und zugleich aus Gründen der Anonymisierung keine exakten Koordinaten bereitstellen. In unserer Untersuchung können wir auf die nicht-öffentlichen vollständigen Zensus-Daten für Namibia mit mehr als 2 Mio. Haushalten zurückgreifen, um bestehende Ergebnisse auf Basis der DHS Daten zu evaluieren. Dabei zeigt sich, dass in den feinteiligen Zensus-Daten der Zusammenhang zwischen Lichtemissionen und Wohlstandsindikatoren stärker als in den DHS Daten ist. Darüber hinaus ist die Aggregation von Lichtdaten um die Koordinaten der Befragungsdaten keine triviale Entscheidung. Unsere Ergebnisse favorisieren eine Lösung, bei der Lichtdaten zunächst in einem Raster aggregiert werden, anstelle des üblichen Verfahrens von kreisförmigen Umgebungszonen. Mit diesem Verfahren finden wir einen signifikanten Zusammenhang zwischen Lichtemissionen und Wohlstandsindikatoren selbst in sehr kleinräumigen Zellen von nur 0.0083 Grad ( $1/120$  Grad, ca.  $1 \text{ km}^2$ ). Zuletzt vergleichen wir unterschiedliche Wohlstandsmaße. Häufig verwendet wird der aggregierte Wohlstandsindex aus den DHS Daten. Unse-

re Untersuchung einzelner Vermögensgegenstände auf Basis der Zensus-Daten kann im Gegensatz zu den DHS-Daten jedoch sogar auf kleinräumiger Ebene einen signifikanten Zusammenhang zwischen Wohlstand und Lichtemissionen nachweisen. Insgesamt zeigen unsere Untersuchungen der hoch-aufgelösten Zensus-Daten von Namibia, dass Licht ein noch besserer Indikator für wirtschaftliche Entwicklung auf lokaler Ebene ist als vorhandene Studien auf Basis von DHS Daten vermuten lassen.

## 1 Introduction

Data quality is a long-standing obstacle for research applications in development economics. To address this, proxy measures of economic development such as nighttime lights observed via satellites have gained popularity in the past decade. It is a valuable addition to traditional data sources because it has global coverage and spatial granularity. The link between lights and economic output has been well established at a national level (Henderson et al. 2012). However, the local level evidence is incomplete due to the lack of high quality geographically disaggregated data on economic development that we could use to validate the light data.<sup>1</sup>

Based on Demographic and Health Surveys (DHS) data, Bruederle and Hodler (2018) and Weidmann and Schutte (2017) analyse the relationship between light and various development measures at the local level. Their results show strong correlations across many developing countries. However, the DHS data lacks detail because of random sampling and random offsets of survey cluster locations to protect respondent anonymity. Therefore, we extend this literature by using two waves of the complete non-publicly available Namibian census, which provides data on the full population and exact borders of the enumeration areas.<sup>2</sup>

The census allows us to answer more detailed research questions about the light-wealth association at a local level. First, we can evaluate the accuracy of DHS based results in Namibia. Second, the enumeration area borders allow us to study how the results differ across different spatial aggregations. We use buffers around point locations, administrative area borders and different sized grids cells as units of observation. Third, the census contains information on the whole population, which allows us to study total light in relation to the total stock of wealth in addition to widely used relative wealth measures—mainly wealth indices generated using principal components analysis (Filmer and Pritchett 2001).

The second question relates to the discussion on the modifiable areal unit problem (MAUP), which states that the results may vary depending on the spatial aggregation of the data (Doll et al. 2006). In the context of nighttime lights and economic development, the aggregation has been done, for example, by using national borders, buffers around cluster points or varying grid sizes (Henderson et al. 2012; Bruederle and Hodler 2018). However, we are not aware of any study that systematically tracks the difference in results across different spatial aggregations. Furthermore, it is not

---

<sup>1</sup> Note that with local levels we imply very small spatial areas such as enumeration areas.

<sup>2</sup> We want to thank the Namibian National Planning Commission with whom Thomas Ferreira worked previously for supplying us with the data. The census data was still deidentified to protect anonymity.

clear what is the smallest aggregation, where light still provides a meaningful signal of economic development. We contribute to the literature by studying the same underlying data in different local areal units including enumeration areas, buffers around survey clusters, and varying grid sizes.

Our third research question focuses on the proxy measurement of economic development using nighttime lights. The choice is data driven because survey data on asset ownership is often the only available indicator at local levels in developing countries. DHS, for example, provides rich data on household assets but not on income. In DHS, wealth indices based on principal component analysis allow researchers to then rank households, and thus clusters, in terms of wealth but this cannot provide indications of total wealth at local levels due to sampling designs. We suggest more straightforward measures to proxy the total stock of wealth, like the amount of people with access to electricity. Intuitively, this measure should be best correlated to the sum of light in an area. We can construct the stock of wealth based on DHS, but then we need a third data source for population density as DHS surveys are not representative at cluster levels. In contrast, we get the total population and information on their access to electricity directly from the full census data. Using this as a proxy for the total stock of wealth allows us to search for a signal of economic development in the light data, which might be missed due to the constraints in DHS data.

Our results show a significant association between the total sum of light and the total stock of wealth across all surveys, aggregation methods and grid cell sizes. The same is true for measures of relative levels of wealth. In general, the association is stronger in most of the specifications for the census data. Therefore, we conclude that nighttime lights are an even better proxy for wealth on micro level, than the findings based on DHS data suggest. Furthermore, the significant association remains even on the smallest grid size that the light data allows, which is roughly 1 km<sup>2</sup> at the equator. In other words, looking at a single pixel in the light data, carries a meaningful signal of wealth in that area.

For the spatial units, we argue that the common practice of aggregating light from buffers around DHS survey cluster locations is prone to errors. The buffers fail to aggregate light that is relevant to survey respondents. The extent of the failure depends on population density, and therefore it leads to systematic bias between rural and urban areas. Additionally, the buffers often overlap, which leads to non-independent observations. Instead of the buffers, we recommend a grid-based approach.

Literature on changes in wealth over time on a local level is also scarce. It is difficult to derive meaningful indicators of wealth changes from DHS data. We do not find a significant association between the change in relative wealth and light in DHS or census data. However, when we use the change in the total stock of wealth as a dependent variable, we do find a significant association in all census specifications, but not in the DHS data. The difference is most likely due to DHS requiring population density data from another source, which introduces measurement error in the results. Therefore, we provide new evidence that changes in light values provide a significant signal of changes in wealth even at the smallest possible

grid level provided in nighttime lights data in Namibia. The signal, however, tends to get weaker in smaller spatial units.

In Sect. 2 we introduce the relevant previous literature on nighttime lights. Sect. 3 discusses the data that we use. Sect. 4 presents our methodology and Sect. 5 discusses the results. Sect. 6 concludes.

## 2 Literature

The economic literature has established that nighttime lights observed in satellite images can be used to measure economic activity. The connection was first observed by Croft (1978) who studied the Defense Meteorological Satellite Program (DMSP) data. Later, the National Geophysical Data Center (NGDC) produced a user-friendly cleaned stable light data set (Elvidge et al. 1997, 2009; Baugh et al. 2010). The improvements paved the way for researchers to study the link between light and economic activity in more detail (Doll et al. 2006; Ghosh et al. 2009; Sutton and Costanza 2002). This new data source was widely accepted in economic literature after the influential studies of Chen and Nordhaus (2011) and Henderson et al. (2012), who established the connection between light and gross output measures at the national level.

The nighttime lights data collected by DMSP is available in digital format for years 1992–2013. The technology has since been replaced by Visible Infrared Imaging Radiometer Suite (VIIRS), which improves on many aspects of the previous series (Elvidge et al. 2017). However, the older DMSP remains highly relevant for purposes of economic research. The questions related to economic development often depend on data with a long time series. Therefore, DMSP and VIIRS can complement each other by allowing researchers to study long term economic development. For example, Li et al. (2020) generate a consistent time series of DMSP and VIIRS. This and other products might be helpful in future research which is not based on a single data source.

The DMSP nighttime lights data overcomes some of the limitations of traditional data sources. It has global coverage, a yearly frequency and it is not subject to national or regional borders. This allows researchers to freely define spatial units of interest. Furthermore, the process of data collection is less dependent on national statistical capacity (see Määttä and Lessmann (2019) for discussion). These advantages in the nighttime lights data have opened a wide range of research opportunities in economic research that utilizes spatial data. Just a few examples of the light data applications include research on institutions (Michalopoulos and Papaioannou 2013), corruption (Hodler and Raschky 2014), ethnic inequality (Alesina et al. 2016), geography and trade (Storeygard 2016; Henderson et al. 2017), natural disasters (Fabian et al. 2019) and conflict (Lessmann and Steinkraus 2019).

Even though the connection between light and economic development is well established at the macro level (Henderson et al. 2012), the evidence at local levels is scarce. Bruederle and Hodler (2018) and Weidmann and Schutte (2017) are two studies that provide evidence at local levels. The resolution of the stable light data is approximately 1 km<sup>2</sup> at the equator. The lack of local level evidence is therefore

understandable because we do not have global ground truth data on a such high spatial resolution. One option for studying the relation between nighttime lights and economic development is to use geocoded survey data.

A further complication is how we define meaningful economic indicators. Output generating activities may have wide variation in terms of light emissions. Aggregating output and light at a national level abstracts away from these details but at a local level a clear definition is essential. For example, we have no information on how the economic output of a farmer manifests in the amount of light that a satellite records in that area in the evening. In a similar way, a highly paid programmer probably does not emit more light in his work than the farmer. However, they tend to live in differently lit areas, which shows up as a significant correlation between economic output and light intensity on the macro scale.

One can also argue that economic output is not the best indicator of economic development in a local setting. Our indicator choices are bound by the information available in the survey data sources. In many instances, survey data only provides information on respondent's wealth, instead of economic output or income. Capturing information on income is challenging for numerous reasons including non-response, measurement error and questionnaire length. A widely used alternative to income and consumption are wealth indices where questions relating to the ownership of household assets and other indicators of well-being—such as access to electricity—are combined into a single index through principal components analysis as described in Filmer and Pritchett (2001). We use the assets and wealth indices as proxies of economic development, and expect them to contribute to the light intensity in an area. However, we recognize that other aspects, such as public infrastructure, surely have an effect as well. Based on the available data, we define economic development as the level and growth of respondent's assets in a given area, and we seek to answer how well the nighttime lights are associated with it. For different aspects of economic development in local level, such as education or health, we refer to Bruederle and Hodler (2018).

One challenge in working with wealth indices is that they are only relevant within the context and time of the survey—wealth indices from different countries are not comparable and similarly wealth indices within the same country are not comparable over time (Rutstein and Staveteig 2014). A simple solution to this problem is to pool data and generate wealth indices based on assets that are comparable over time and space, however, methodological concerns remain such as the relative importance of certain assets over time (Harttgen et al. 2013). We mitigate this concern by using single asset variables to check for robustness. Furthermore, our data sample has a shorter time span (13 years) than DHS data in general (30 years), which reduces the chance of large shifts in asset importance.

The most notable studies on the correlation between light and wealth at a local level are based on DHS data. DHS program supports a series of nationally representative randomly sampled surveys. They are well suited for evaluating the light data. Its standardized survey questions allow comparison over long time horizon and many developing countries. The program has been running since 1985 and it has collected data from over 90 countries. The surveys include questions on household assets, which are combined into a wealth index. Furthermore, some of the DHS sur-

veys provide coordinates of cluster locations. These points are randomly offset by 2km in urban regions and mostly 5km but sometimes even 10km in rural regions. This leads researchers to aggregate light in a radius of the random offset distance around those cluster points and analyze it in relation to the wealth index. Bruederle and Hodler (2018) and Weidmann and Schutte (2017) carry out such studies across developing countries and conclude that nighttime lights are a good proxy of wealth at the micro level.

Another relevant strand of literature for our research is the modifiable areal unit problem (MAUP) which points out that spatial statistics are sensitive to how data is aggregated. This was first noted by Gehlke and Biehl (1934) and later described in Openshaw (1983). Chen and Nordhaus (2019), for example, show how correlations between VIIRS nighttime lights and economic data is sensitive to the size of the unit of observation. They show that correlations are weaker at the state level than at metropolitan level. They find that the correlation coefficient of light and annual GDP is 0.85 using states and close to 1 for metropolitans using United States GDP statistics. We are not aware of similar research for DMSP nighttime lights at local level in developing countries. The lights data has been applied to different spatial aggregations, but only few studies systematically gauge a sensitivity of the results to different spatial aggregations. Bruederle and Hodler (2018) provide an excellent comparison of differences between circular buffers and 0.5 degree grid cells while Weidmann and Schutte (2017) try to incrementally increase the size of the circular buffer. However, they can only evaluate the results in the context of the DHS data that they are using. We have the advantage of seeing how the changes affect the results using DHS and census data. This allows us to evaluate if the shortcomings in DHS data lead to systematic bias. Furthermore, we are not constrained by the random displacement of survey clusters, which allows us to study the light-wealth relationship on the finest resolution of the light data.

### 3 Data

We study the association between light and wealth using multiple sources of survey- and spatial data for Namibia. DMSP Nighttime light data is our main explanatory variable of interest. We also use Gridded Population of the World (GPW) population density data to support the data-analysis. DHS and census survey data provides us with the dependent asset variables and geocoded locations for small areas. We discuss each of these sources in more detail in this section starting with the survey data.

The availability of geocoded data for small areas from different household surveys presents a unique opportunity to study light-wealth correlations as household surveys, apart from the DHS, do not generally have geocoded information at small areas. Household surveys that track poverty are generally only statistically representative at aggregated geographic levels such as the first level of administration and maybe an urban/rural subdivision. This is mainly due to random sampling, and the need to protect the anonymity of respondents. DHS surveys have small-area indi-



cators with a random offset (to protect anonymity) which explains why they have been used to study light-wealth correlations.

The Demographic and Health Survey program facilitates a series of standardized household surveys across developing countries. It is funded by the United States Agency for International Development (USAID). The program has been running since 1984, and by now it has provided support to more than 400 surveys in over 90 countries. The program aims to keep the survey questions unchanged, which makes them comparable across time and countries. The surveys use nationally representative random samples that provide information on health, fertility, education, wealth, and many other socio-economic questions. Some of the DHS surveys include GPS coordinates of cluster locations, which we will discuss in more detail in the method section.

For Namibia, the standard DHS survey has been conducted in 1992, 2000, 2009 and 2013. We use the survey rounds of 2000 and 2013 because those years are nearest to the census data years (2001 and 2011). From the survey questions, we use multiple variables that are related to household wealth. These include access to electricity, radio, television, water piped into the residence and flush toilet for year 2000 (Wealth index B). The year 2013 survey includes additional variables on the ownership of a car, computer, motorbike, bicycle, landline telephone, mobile telephone, refrigerator, stove, internet, and a microwave oven (Wealth index A). The selection of these variables is based on the census data, where we can find matching survey questions. We use some of the variables separately as indicators for the total stock of wealth, and we combine all of them together to construct a relative wealth index. The construction of the wealth indexes is illustrated in appendix A1. Altogether, the DHS data for the year 2013 provides information on 41,646 individuals in 550 clusters, and the data for the year 2000 includes 31,675 individuals in 260 clusters.

Specialized survey data can get us even closer to the ground truth than DHS data. We use non-publicly available complete Namibian census data from 2001 (CBS 2001) and 2011 (NSA 2013). Unlike the randomly sampled DHS data, the census data has information on all people in Namibia. This makes it an optimal ground truth comparison because it is representative at the level of an enumeration area.

Namibia serves as an interesting case to explore measurement issues when relating wealth to light. It is one of the least densely populated countries in the world and one of the most unequal with a GINI coefficient in 2015 of 59.1 (World Bank 2020). Wealth is also geographically unequal. Under colonial rule, native populations were generally marginalized and geographically segregated from colonists (Odenaal 2011). These facts help highlight our concerns with using DHS buffers. The high levels of geographic inequality increase the likelihood of assigning the wrong light to households with offset buffers. Similarly, with low population density the offsets are more likely not to capture light relevant to households. We illustrate these ideas in the next section.

The wealth related questions in the census data are matched with the DHS wealth variables listed earlier. The census data has a further advantage of providing the exact borders of the enumeration areas, which also provides a more precise association between light and wealth. For both years we study conventional households. This

leaves 2,059,530 individuals living in 463,874 households across 5490 enumeration areas in 2011 and 1,773,235 individuals living in 346,455 households across 4084 enumeration areas for 2001.<sup>3</sup>

We use the Stable Lights product, which is the most used nighttime light product in the economic literature (Baugh et al. 2010). It is a spatial raster data set that provides an annual average of light intensity value between 0 and 63 in 30 arc seconds resolution, which corresponds to roughly 1 km<sup>2</sup> sized area at the equator. We use the Stable Light data from satellite F18 for years 2011 and 2013, and satellite F15 for years 2000 and 2001. The Light gain settings may change over time and between satellites, but this is not a concern for us because we use changes. Therefore, we assume that change in light intensity due to the satellite would affect all areas in Namibia equally.

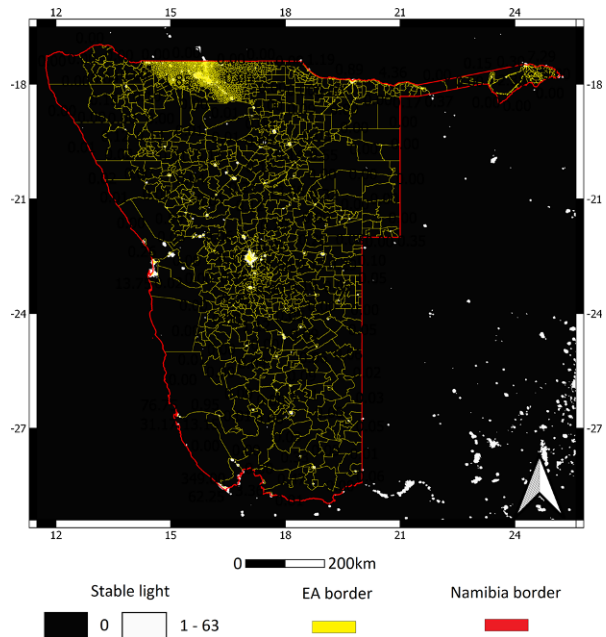
The light products also have shortcomings that are important to keep in mind. First, Bluhm and Krause (2018) discuss the problem of “top coding”, which means that some city centers reach the maximum light intensity value of 63. The capability of the DMSP-Operational Linescan System (OLS) is saturated in those pixels, and therefore the data cannot distinguish between light levels in New York and Windhoek for example. This is not a large problem in Namibia because there are not many saturated pixels. Second, Abrahams and Oram Lozano-Gracia (2018) give a detailed demonstration of “blurring”. Shortly, the on-board processing of the light images in the satellite blurs the light sources so, that some light spreads into the neighbouring pixels of the light source. This problem is rather relevant in our application and we will discuss it in the method section. Third, Määttä and Lessmann (2019) explain how the process of removing background and transitional light in the Stable Lights product is very strict. Light gets filtered out especially in rural areas in Africa, which can be seen in Fig. 1. This is also a relevant problem for our study, and we note that a less strict filtering method might find a stronger signal of wealth in the nighttime lights data. Even with all the shortcomings, the nighttime lights data provides a very useful proxy for spatial economic development, and it is important to understand how well it is correlated with wealth at a local level.

We have the exact population count in the enumeration areas for the census data, but for the DHS we need an additional data source for population density. We use the Gridded Population of the World version 4 (CIESIN 2018) population density raster data for two reasons. First, it does not rely on the nighttime lights data in the process of spatially distributing the population counts unlike some other alternatives. Second, GPWv4 is also used in the paper from Bruederle and Hodler (2018) that we are building on. The data is available in 5-year intervals, and we use the dataset for year 2000. For 2001 and 2013 we infer population counts using linear interpolation between the available years.

We re-project and re-sample all geospatial data into the World Geodetic System 1984 (WGS84) coordinate reference system. It is used in the nighttime lights data, and in many of our reference literature (see Bruederle and Hodler (2018) and Henderson et al. (2012)). In the WGS84 projection, data is reported on a latitude-longitude grid. This means that map units are based on degrees. One grid pixel in

<sup>3</sup> The total population in the census was 1,830,330 and 2,113,077 for the years 2001 and 2011 respectively.

**Fig. 1** Stable lights and enumeration areas in Namibia (2011)



the light data is  $30 \times 30$  arc seconds. Due to the curvature of the earth, the size of one pixel measured in square kilometers depends on the latitude. The pixel size varies between  $\sim 0.82 \text{ km}^2$  in northern Namibia and  $\sim 0.75 \text{ km}^2$  in the south. We follow the recommendations in the literature to weigh the pixel light values by the size of an area (Elvidge et al. 2020). We refer to pixels and raster data when we use the base layers of light and population data. In contrast, we refer to grid cells and varying degree grid data when we aggregate the base layer data into larger spatial units.

## 4 Methodology

Our main goal is to assess the light-wealth correlations from DHS by comparing them to census data. We carry out regression analysis using multiple indicators of wealth at different levels of spatial aggregation. Furthermore, the census data is not collected exactly in the same year as DHS data, and the number of observations differ. Therefore, we run detailed analysis to pinpoint where the differences in DHS and census-based results stem from. In this section, we will explain the process of disentangling these methodological differences.

The wealth index in DHS is the commonly chosen dependent variable in the literature (Bruederle and Hodler 2018; Weidmann and Schutte 2017). As alluded to earlier, wealth indices are generated by combining information from multiple survey questions about household assets and other measures of well-being. The most common used method to combine these different indicators into a single measure is through principal component analysis (PCA). PCA is a statistical technique that

extracts from a given set of variables, uncorrelated linear combinations of them, that captures the largest amount of variation in the variables (Filmer and Pritchett 2001). This helps to account for the fact that wealthier households are more likely to have a combination of assets that poorer households do not have. The PCA linear combination that describes the greatest amount of variance is then used as the asset index. The resulting index values are then divided into five quintiles that reflect households' relative wealth (1 = poor, 5 = rich).

We do not use the ready-made wealth index provided in the DHS because it would entail further problems when comparing the results to our census data. First, the DHS wealth index components differ from variables that are available in the census data. Secondly, the DHS wealth index values are relative to respondents in one survey, which makes comparison over time unfeasible. The problem of using the DHS wealth index across time or countries is well documented in Rutstein and Staveteig (2014). We alter the DHS wealth index by pooling the respondent and adjusting variables. When we analyze changes between the two surveys, we pool respondents from both surveys together prior to constructing the wealth index. In our data sets, we have more asset variables available for the later years. Therefore, we construct two wealth indices with different components and use them depending on if we analyse level or growth effects. For our levels analysis we can use more variables which introduce greater variation in the wealth indices because there are no concerns with assets being added or removed from surveys over time. After adjusting for the pool of respondents and variable composition, we carry out PCA to create the wealth indices. The data section and appendix A1 discuss wealth index variable compositions.

The use of the relative wealth index is well justified in the DHS based analysis, but the census data provides us with reliable information on the total stock of wealth as well. For example, we know the exact number of people with access to electricity in the census enumeration areas, instead of average of respondents with access to electricity in the DHS. Intuitively, the total stock of wealth should be better correlated with light, which is a product of the amount of people and their light generating activities. Therefore, using individual assets as dependent variables contributes to the literature by exploring the kind of wealth signals we are able to detect in the nighttime lights data. We choose the amount of people with access to electricity as our main proxy for the stock of wealth because it is the most likely variable to provide a signal in the light data. Furthermore, it is highly correlated with many of the other wealth index components such as televisions and radios that require electricity.

It is possible that access to electricity may provide stronger results than other wealth components which are not related to electricity. Therefore, we check the robustness of the results using the survey questions on access to flush toilet, piped water in the household and car ownership. The car ownership variable is only available for the later survey years.

The stable lights data shows annual average values of light intensity for each pixel. These values can be spatially aggregated in multiple ways. To get comparable results, we first follow the example from Bruederle and Hodler (2018) by taking the average light in a spatial unit. Furthermore, we use a logarithmic transformation to

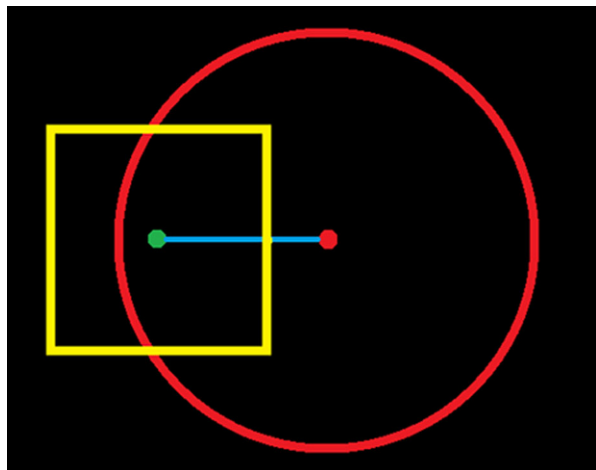
account for the skewness of light data. We also add 0.01 to the mean light values before taking the logarithm to keep non-lit areas in the analysis. However, we also report results for specifications, where areas with no light are excluded from the analysis.

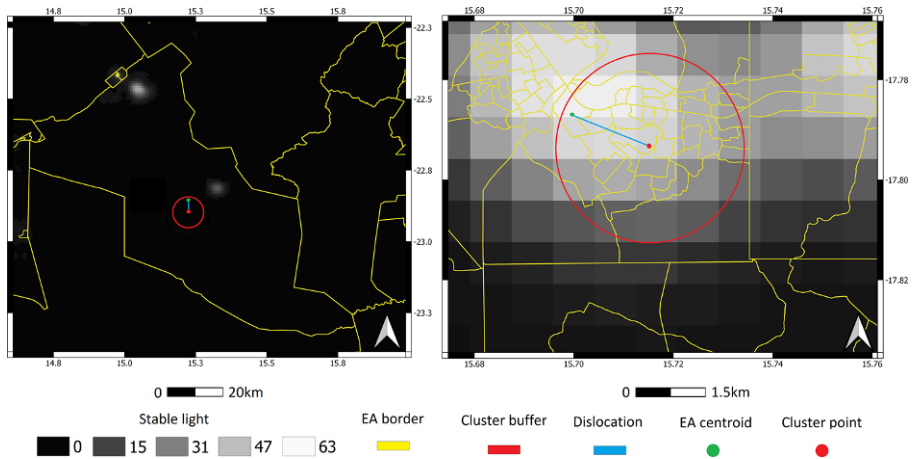
We also use the sum of light as an independent variable. It has a straightforward connection to the total stock of wealth, which is also a sum in a given area. This way, we could even make a prediction of wealth based on light without knowing the size and shape of the aggregation area. This will help us in generalizing the findings on light-wealth correlations across different spatial units. We can also add control variables for area size and population, but they might not always be available for researchers. Therefore, it is advantageous to assess what we can say about the stock of wealth in a region, without knowing anything else but the sum of light intensity. Like the mean of light, a sum of light leads to high outliers for city areas. Therefore, we use the logarithmic transformation as explained in the previous paragraph.

We need a few additional steps of variable processing to analyze the correlation between the change of wealth and the change of light. For relative wealth, we take the difference of the index values between the years. For all other variables, we take the difference of log transformed values between the years.

Our second large difference to the previous DHS-light studies is the way of spatial aggregation. The usual approach is to aggregate light from buffer areas around DHS cluster coordinates. Those coordinates are centroids of the enumeration areas that are dislocated by 2 (urban) or 5 (rural) kilometers to protect respondents' anonymity. A radius of a buffer around the cluster points is usually chosen to match the 2 or 5 km dislocation (Bruederle and Hodler 2018), or even 10 km (Jean et al. 2016), to ensure that all relevant light is aggregated. The buffering process is illustrated in Fig. 2. The stylized example highlights two problems in aggregating light with the buffers. First, the survey respondents may live anywhere within the yellow enumeration area borders, and therefore the buffer might not aggregate light from all relevant areas. Second, the buffer may aggregate light that is outside the area relevant to the survey respondents.

**Fig. 2** DHS cluster buffer example. The *yellow line* represents enumeration area borders. The *green dot* is the enumeration area centroid. The *blue line* is the dislocation of the enumeration area centroid. The *red dot* is the cluster point reported in DHS. The *red circle* is the buffer used to aggregate light





**Fig. 3** Light aggregation using rural and urban buffers

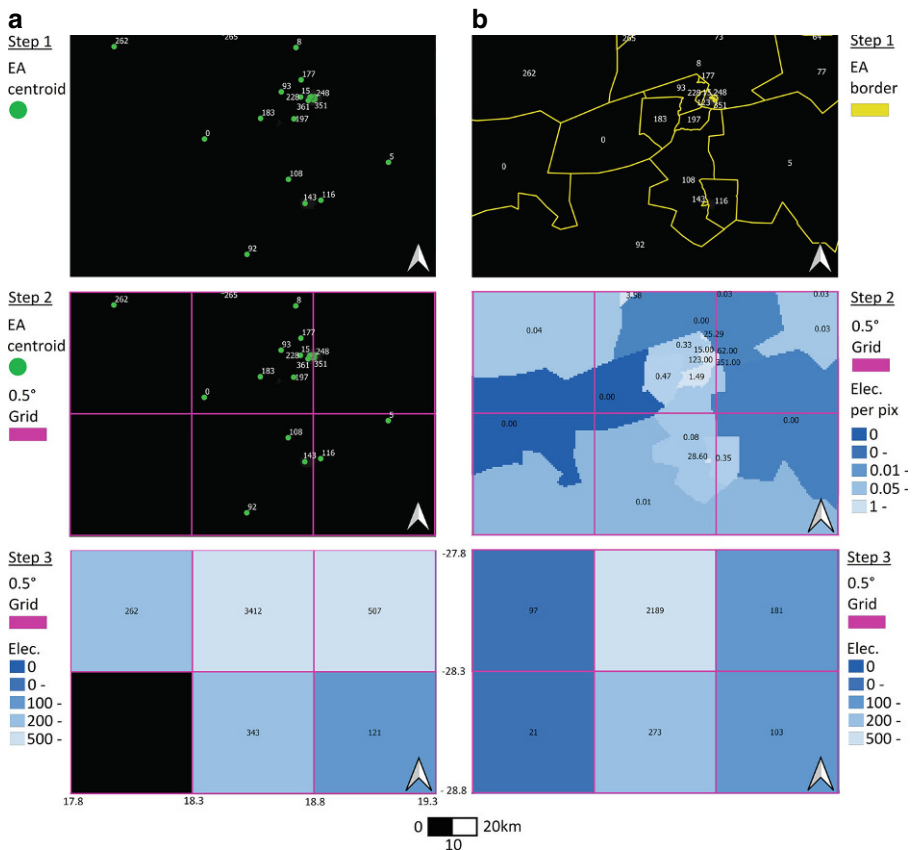
The two problems of aggregating irrelevant light and not aggregating all relevant light, play a different role depending on the size of an enumeration area. Note that DHS enumeration areas are based on previous census enumeration areas, which makes it easy for us to compare them. Fig. 3 illustrates two examples of rural and urban areas in Namibia. On the left, the rural enumeration area is much larger than the 5 km buffer. The survey respondents probably live in the two lighted areas, but they are assigned zero light because the buffer lands on a possibly uninhabited area. On the right, urban enumeration areas are much smaller than the 2 km buffers. The relevant light in the example enumeration area is high. However, because of including pixels with lower light values in the buffer, the respondents are assigned with a lower light average.

Light aggregation from the buffers and enumeration area borders cause a further problem of non-independent observations. In the urban case, if we would draw a buffer for each enumeration area in Fig. 3, many of them would overlap with each other. In the worst case, depending on the dislocation, they might overlap completely. A related problem manifests also when using the actual enumeration area borders. As can be seen in the urban image, some enumeration areas are so small that they both fit into the same light pixel. Therefore, the resolution of the light data is too coarse to take full advantage of the detailed census data. This leads to a violation of assuming independent observations in econometric regression models. The problem may be solved by weighing the overlapping observations or using grids as is done in Bruederle and Hodler (2018). An additional drawback of using the census enumeration area borders is that they are continuously adjusted based on the population count. Therefore, it is not possible to measure changes over time.

An alternative for aggregating light from enumeration area borders and buffers is to divide the underlying data into grid cells. The grid approach abstracts away from the highest resolution that the data allows. This solution removes the problem of non-independent observations and may help with the problem of aggregating light from

the wrong area to the survey respondents. However, the latter problem remains to a smaller extent, because an enumeration area might be located over two grid cells. A common choice of the grid cell size is 0.5 degrees (Bruederle and Hodler 2018). However, the studies using 0.5 degree grid focus on multiple countries. A smaller grid size could be more useful for single country studies. Therefore, we also study the relationship between light and wealth at 0.25, 0.1 and 0.00833 degree levels. The smallest level is the size of one pixel in the light data.

The main challenge in using grid cells is to decide how the survey data should be distributed in them. We demonstrate two different methods depending on whether we know the enumeration area borders or not—(a) a Point data method and (b) a polygon data method. (a) The *Point data method* simulates the case of DHS data, where we only know the cluster locations. (b) The *Polygon data method* takes advantage of the enumeration area borders, and therefore is likely to lead to more accurate results. Both methods process the data in three steps that are presented in Fig. 4. Step 1 presents the data as it is before any processing takes place. In this example, the numbers represent people with access to electricity in an enumeration area. In step 2, we add the 0.5 degree grid. For the polygon method (b), we also



**Fig. 4** Aggregating survey data into grid cells. **a** Point data method, **b** Polygon data method



rasterize the values on the light data resolution to get electricity access per pixel values. In step 3, the point method sums up the people with access to electricity in each enumeration area centroid that falls within a grid cell. For the polygon method, all electricity access per pixel values are summed up to get the final amount of people with electricity access per pixel. As the results show, even though we start with the same data, utilizing the borders leads to quite different grid values. We also note that the point method produces missing values for a grid cell, if there are no centroids in it.

## 5 Results

This section reports the results of a regression analysis based on all the data and method choices that were presented in the previous section. Variable choices, survey data sampling, survey years and the method of aggregating light and survey data, all have an impact on the results. In the interest of keeping focus, we will only present results that show the highest impact on the conclusions of our comparisons between the full census and DHS data light correlations. Furthermore, we show results with and without control variables.

### 5.1 Level of wealth and light

The replication of the DHS buffer specification in Namibia, and comparison to actual enumeration area border-based results, is provided in the appendix A2. The association between wealth and light is not as strong as is found in the literature across developing countries. In Namibia, the regression  $R^2$  is 42.1% (A2.2, column 1), when using the logarithmically transformed mean of stable light as the only explanatory variable for the wealth index. Using the same specification in other developing countries, Bruederle and Hodler (2018) find  $R^2$  of 52.7%. The lower than average explanatory power in Namibia is supported by the findings in Weidmann and Schutte (2017). They rank order DHS clusters by light and wealth for each country, and then compare the rank correlations. The correlation in the 2007 Namibian DHS was placed 46 (highest rank equals 1) out of 57 surveys.

When we switch from randomly sampled DHS data to the full census data, and still aggregate light using the simulated buffers from enumeration area centroids, the association is slightly stronger with a  $R^2$  of 43.9% (A2.2, column 3). This is to be expected because the census data is closer to the ground truth and eliminates measurement error introduced in the DHS sampling process. Furthermore, when we use the census data and aggregate light within the enumeration area borders instead of the buffers, the  $R^2$  increases to 48.3% (A2.2, column 5). This is also to be expected because the borders aggregate light that is most relevant for the survey respondents. Controlling for population density does not have large impact on the results. Altogether, the association between wealth and light is significant, and slightly stronger when using the full census than the DHS data.

Next, we apply the point and polygon aggregation methods for 0.5 degree grid cells. It allows us to compare results from the DHS to the census step by step to



**Table 1** Descriptive statistics for 0.5 degree grid cells

Variable name	Description	DHS 2013 Point	Census 2011 Point	Census 2011 Polygon
–	Grid cells (count)	152	275	338
Wealth_A	Wealth index with 15 components	2.63	2.74	2.67
Elec	Amount of people with access to electricity	7385	2805	798
Stb_mean	Average stable light	0.21	0.10	0.15
Stb_sum	Sum stable light	605	296	245
popD	Population per pixel	–	3.12	1.68
Pop	Population	–	7480	4017
GPW_popD	GPW population per pixel	5.76	3.54	2.99
GPW_pop	GPW population (GPW_popD * pixels)	16777	9499	7754
Pixels	Pixels in Grid cells	3199	3260	3047

see where the differences arise. Table 1 shows descriptive statistics for the different surveys and methods. The first notable difference is that the amount of grid cells considered is lower in the point method. This is caused by the fact that enumeration area centroids do not fall in each grid cell. In DHS, the amount is even lower due to random sampling. One full grid cell is 3600 pixels but note that cells at country borders are cut smaller. The higher mean and sum of light in DHS most likely reflects the exclusion of remote areas, development between the two years or differences in the satellite light gain settings. In column two and three, we can see that changing the point method to the polygon method lowers the population amount and density. This happens even though the numbers are derived from the same underlying data. The difference is down to how we assume the population is distributed in each enumeration area (all in a central point or equally divided in each pixel). We will not know for sure which one is closer to the actual distribution, but we can check if the choice influences our conclusions.

The wealth index does not vary much across the columns. In contrast, the electricity access variable illustrates how large the difference can be when using different data and methods. The large gap is partly explained by the reduction in grid cells and development over two years. However, the bias gets amplified when calculating the amount of people with access to electricity. It is the product of GPW\_pop and the share of access to electricity. As the table shows, the GPW\_pop variable is probably an over-estimate as well. As a result, the calculation compounds the biases in both variables, and widens the gap compared to census-based figures that use the actual count. Our analysis demonstrates how important the population data and aggregation method are. These are the key factors when drawing the conclusions on the association between wealth and light.

We start the 0.5 degree grid analysis by using similar regression specifications as in Bruederle and Hodler (2018) in Table 2. The light variable is significant at a 99% confidence level in all specifications. In column (1), the magnitude of the light variable (0.242) is slightly lower for Namibia than what Bruederle and Hodler (2018) find across developing countries (0.326). The interpretation of the coefficient

**Table 2** Regressions using 0.5 degree grid cells and relative wealth as dependent variable

OLS regression	(1)	(2)	(3)	(4)	(5)	(6)
Survey	DHS	DHS	Census	Census	Census	Census
Year	2013	2013	2011	2011	2011	2011
Method	Point	Point	Point	Point	Polygon	Polygon
Dep. var	Wealth_A	Wealth_A	Wealth_A	Wealth_A	Wealth_A	Wealth_A
Log_stb_mean	0.242 (0.046)***	0.578 (0.06)***	0.204 (0.030)***	0.476 (0.036)***	0.088 (0.028)***	0.308 (0.0260)***
Log_GPW_popD	–	–0.498 (0.067)***	–	–0.329 (0.031)***	–	–
Log_popD	–	–	–	–	–	–0.35 (0.022)***
Observations	152	152	275	275	338	338
R2	0.158	0.387	0.147	0.394	0.029	0.437

Note: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Units with zero light are included

is that a one percent increase in mean light is associated with 0.252 change in the wealth index in a grid cell. In column (1), the light alone explains 15.8% of the variation in the wealth index, which is lower than 35.7% found in Bruederle and Hodler (2018). This result also supports the finding in Weidmann and Schutte (2017) that the association of light and wealth is relatively weaker in Namibia. In column (2), we control for population density and get a higher light coefficient and  $R^2$  than Bruederle and Hodler (2018). In their study, the population density variable is negative (–0.042) but not significant. This leads us to believe that the role of population density might play a largely different role across countries. In Namibia, which is one of the least densely populated countries in the world, it is highly significant, and the magnitude is almost on the same level as light.

In Table 2, we also compare how the results differ between DHS and census. Columns (3) and (4) use the same point method but switch the data from DHS to census. This changes the year of the survey and the number of observations, but the results are similar.

Columns (5) and (6) consider the polygon method instead of the point method. It seems that aggregating a relative wealth variable by taking the average of the pixels is a worse practice than taking the average of enumeration area centroid points. However, once we control for the actual population count from the census in column (6), the association is stronger. Altogether, the conclusions that we can draw based on DHS data, are well in line with the census-based results when using a relative wealth index as is done in the literature.

Now that we have established a baseline comparison to the literature, we take a different approach by using the stock of wealth indicators. This utilizes the full scope of the census data where we know the actual amount of people with access to electricity, piped water, flush toilets and car ownership. In principle, we can construct the stock of wealth variables also for DHS by using a third source for population density. However, as we demonstrated in the Table 1, this leads to biased values.

**Table 3** Regressions using 0.5 degree grid cells and the stock of wealth as dependent variable

OLS regression	(1)	(2)	(3)	(4)	(5)	(6)
Survey	DHS	DHS	Census	Census	Census	Census
Year	2013	2013	2011	2011	2011	2011
Method	Point	Point	Point	Point	Polygon	Polygon
Dep. var	Log_elec	Log_elec	Log_elec	Log_elec	Log_elec	Log_elec
Log_stb_sum	0.470 (0.06)***	0.410 (0.076)***	0.457 (0.035)***	0.321 (0.042)***	0.387 (0.032)***	0.263 (0.029)***
Area_size	–	0.001 (0.000)**	–	0.001 (0.000)***	–	0.001 (0.000)***
Log_GPW_pop	–	0.371 (0.247)	–	0.668 (0.122)***	–	–
Log_pop	–	–	–	–	–	0.608 (0.066)***
Observations	152	152	275	275	338	338
R <sup>2</sup>	0.289	0.313	0.380	0.471	0.300	0.563

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Units with zero light are included

Therefore, it is understandable that this approach is often avoided when only using DHS data.

Table 3 uses the amount of people with access to electricity as a stock dependent variable of wealth. It is the most obvious variable that should show up as a signal in nighttime lights. It is also highly correlated with other household assets. Our primary goal is to determine if the nighttime lights provide a further signal of wealth, compared to what can be concluded from relative wealth-based DHS studies. Specifications (1), (3) and (5) provide an intuition of the relationship, in which we simply sum up the light values in a grid cell. In this case the point method (3) provides the strongest association, but the polygon method (5) is also highly significant. Considering the  $R^2$ , the association is weaker for DHS in column (1). In columns (2), (4) and (6) we include control variables for the grid cell size and the amount of population. They are all highly significant. Furthermore, the difference in the  $R^2$  between the DHS and census widens substantially, which shows the advantage of using the actual population data instead of another source.

In contrast to the relative wealth regressions, population density has a positive sign. The intuition is that average wealth is lower in densely populated areas, but the absolute amount of wealth can still be higher. In column (6), the  $R^2$  climbs as high as 56.3%. However, this high association might just be due to the choice of electricity access as an indicator of wealth. Therefore, we repeat the regressions specifications (5) and (6) with different individual stock variables of wealth in appendix A3. The variables piped water, flush toilet and car ownership are not directly linked to electricity, but the results show an equally strong, or an even stronger association with light. Therefore, we conclude that nighttime lights carry a stronger signal of the total stock of wealth in a grid cell compared to what can be concluded from the DHS data.

So far we have aggregated data within enumeration area borders, buffers and 0.5 degree grid cells. The latter is the favored cell size in the literature, but arbitrary, nevertheless. Any other cell size could also be justified, especially as we focus on Namibia only. Therefore, we repeat the regressions for 0.25, 0.1 and 0.00833 (one light pixel) grid cell sizes. In order to get an overview of the results, we pack all of the light coefficients, their significance and  $R^2$  into Table 4. The results differ from earlier regressions because we have excluded observations that have no light emissions. This practice is debatable because it may be that the non-lit areas do emit light, which is just not detected in the satellites, as argued by Bruederle or Hodler (2018). Alternatively, it may be that the non-lit areas are simply uninhabited, and therefore justifiably removed. The truth is probably somewhere in the middle, and therefore we provide analysis with and without the non-lit areas for robustness. In general, excluding non-lit areas leads to a stronger association between nighttime lights and wealth. Furthermore, in the census point method, we control for the population count reported in the census instead of GPW population. Note that in the previous analysis, we aimed to be comparable to the existing literature but at this stage we are seeking the specifications that provide best signal of wealth using the nighttime lights data.

**Table 4** Regressions with varying spatial units and the stock of wealth as dependent variable

	Survey	DHS	DHS	Census	Census	Census	Census
Year		2013	2013	2011	2011	2011	2011
Method		Point	Point	Point	Point	Polygon	Polygon
Dep. var		Log_elec	Log_elec	Log_elec	Log_elec	Log_elec	Log_elec
Enumeration area borders/buffers	Log_stb_sum	1.734***	1.218***	0.277***	0.262***	0.899***	0.850***
	Control vars	No	Yes	No	Yes	No	Yes
	R2	0.159	0.198	0.046	0.063	0.059	0.070
	N	370	370	3215	3215	3215	3215
0.5 degree grid	Log_stb_sum	1.201***	0.918***	0.894***	0.514***	0.857***	0.318***
	Control vars	No	Yes	No	Yes	No	Yes
	R2	0.359	0.397	0.653	0.768	0.361	0.808
	N	104	104	129	129	146	146
0.25 degree grid	Log_stb_sum	1.255***	1.270***	1.013***	0.553***	0.865***	0.318***
	Control vars	No	Yes	No	Yes	No	Yes
	R2	0.226	0.226	0.537	0.680	0.289	0.781
	N	119	119	175	175	214	214
0.1 degree grid	Log_stb_sum	1.485***	1.592***	0.973***	0.654***	0.799***	0.361***
	Control vars	No	Yes	No	Yes	No	Yes
	R2	0.219	0.221	0.375	0.508	0.207	0.762
	N	157	157	278	278	445	445
One pixel grid	Log_stb_sum	—	—	—	—	1.346***	0.785***
	Control vars	—	—	—	—	No	Yes
	R2	—	—	—	—	0.101	0.520
	N	—	—	—	—	9602	9602

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Control variables are: area\_size, log\_GPW\_pop (DHS) and log\_pop (census). Units with zero light are not included

Table 4 shows that light provides a highly significant association with the total stock of wealth across all spatial units, surveys and methods. The association tends to get weaker in smaller area units, but the signal remains clear even at the smallest possible pixel level in the nighttime lights data. This is an interesting result when considering the MAUP in the context of nighttime lights. Chen and Nordhaus (2019) find that the association becomes stronger at a smaller spatial scale. However, they move from US state level to metropolitan areas. Our largest spatial scale is the 0.5 degree grid, which is closer to the metropolitan area size. Therefore, this might be the area size with the strongest association, while moving to larger or smaller units weakens it. However, given the large differences between the countries and light data, this finding remains a point for future work.

In Table 4, the association between lights and wealth is stronger in DHS when the data is aggregated within enumeration area borders or buffers. However, the enumeration area level results are unreliable due to overlapping buffers and blooming, as discussed earlier. When we switch to the grids, then the census results are stronger, especially with the control variables. Using the census data, it is not clear whether the point or polygon method should be preferred. The point method is stronger without the controls, while the polygon method is stronger with the controls. We recommend the polygon method if population data and borders are available because it retains more observations and allows for pixel level analysis.

## 5.2 Change of wealth and light

Now we turn attention to measuring changes in nighttime lights and wealth. Relating growth in nighttime lights to growth in GDP, as done by Henderson et al. (2012), provided the benchmark for using nighttime lights in economic research. Again, we provide our most detailed results for 0.5 degree grid cells, but we also present an overview for other spatial aggregations. We report on changes in the total stock of wealth to changes in the sum of light as it is a more intuitive comparison.

For completeness, we provide analysis on relative wealth change as well (appendix A4). The wealth index B is using a smaller set of asset components as explained in the data section. The wealth change is constructed simply by subtracting the earlier year values from later year values. The other variables show the difference between log transformed earlier year values and log transformed later year values. The results are all insignificant, which means that at least for Namibia between 2000 and 2013 (DHS) or 2001 and 2011 (census), there does not seem to be any evidence of relative wealth change being associated with stable lights change. Grid cells that have no light in both years are excluded. The results did not improve when we included non-lit grid cells. This is an important result since economists often exploit the time series properties of (panel) data. Using DHS would suggest that it is difficult to capture the relevant variation.

Table 5 shows the results for the association between changes in the amount of people with access to electricity and changes in stable lights. The association is significant, but weak with the exception of column (6). It seems that the polygon method and actual population count are required for explaining more variation in the change of wealth. Furthermore, most of the explanatory power comes from changes

**Table 5** Regressions using 0.5 degree grid cells and the change in the stock of wealth as dependent variable

OLS regression	(1)	(2)	(3)	(4)	(5)	(6)
Survey	DHS	DHS	Census	Census	Census	Census
Year	2000–2013	2000–2013	2001–2011	2001–2011	2001–2011	2001–2011
Method	Point	Point	Point	Point	Polygon	Polygon
Dep. var	Elec_log_change	Elec_log_change	Elec_log_change	Elec_log_change	Elec_log_change	Elec_log_change
Stb_sum_log_change	0.467 (0.210)**	0.434 (0.208)**	0.114 (0.034)***	0.113 (0.034)***	0.069 (0.021)***	0.067 (0.018)***
Area_size	–	–0.001 (0.001)	–	0.000 (0.000)	–	0.000 (0.000)
GPW_Pop_log_change	–	–4.425 (3.087)	–	–1.816 (1.028)*	–	–
Pop_log_change	–	–	–	–	–	1.138 (0.138)***
Observations	74	74	146	146	172	172
R2	0.064	0.114	0.072	0.093	0.060	0.334

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Units with zero light in both years are not included

**Table 6** Regressions with varying spatial units and the change in the stock of wealth as dependent variable

	Survey	DHS		DHS		Census		Census		Census	
	Year	2000–2013	2000–2013	2000–2013	2000–2013	2001–2011	2001–2011	2001–2011	2001–2011	2001–2011	2001–2011
	Method	Point	Point	Point	Point	Point	Point	Polygon	Polygon	Polygon	Polygon
0.5 degree grid	Dep. var	Elec_log_change	Elec_log_change	Elec_log_change	Elec_log_change	Elec_log_change	Elec_log_change	Elec_log_change	Elec_log_change	Elec_log_change	Elec_log_change
	Stb_sum_log_change	0.467**	0.434**	0.434**	0.114***	0.113***	0.113***	0.069***	0.067***	0.067***	0.067***
	Control vars	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes
	R2	0.064	0.114	0.114	0.072	0.093	0.093	0.06	0.334	0.334	0.334
0.25 degree grid	N	74	74	74	146	146	146	172	172	172	172
	Stb_sum_log_change	0.223	0.157	0.157	0.204***	0.196***	0.196***	0.066***	0.075***	0.075***	0.075***
	Control vars	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes
	R2	0.015	0.085	0.085	0.136	0.187	0.187	0.043	0.281	0.281	0.281
0.1 degree grid	N	73	73	73	192	192	192	256	256	256	256
	Stb_sum_log_change	0.450*	0.337	0.337	0.226***	0.226***	0.226***	0.088***	0.085***	0.085***	0.085***
	Control vars	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes
	R2	0.039	0.106	0.106	0.079	0.097	0.097	0.038	0.238	0.238	0.238
One pixel grid	N	74	74	74	251	251	251	531	531	531	531
	Stb_sum_log_change	–	–	–	–	–	–	0.058***	0.063***	0.063***	0.063***
	Control vars	–	–	–	–	–	–	No	Yes	Yes	Yes
	R2	–	–	–	–	–	–	0.008	0.371	0.371	0.371
	N	–	–	–	–	–	–	11706	11706	11706	11706

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Control variables are area\_size, GPW\_pop\_log\_change (DHS) and pop\_log\_change (census). Units with zero light in both years are not included

in population rather than changes in light. Appendix A5, shows that there is no evidence for a significant association between changes in light and changes in other stock of wealth variables (piped water and flush toilets). Including non-lit areas did not improve the results either.

Finally, we compare changes in light and wealth across the different spatial units in Table 6. The results based on DHS are not robust across different grid sizes (columns 1 and 2). The results based on the census follow a similar pattern that was found in 0.5 degree cells. The light variable is significant on every level with or without controls. However, the explanatory power is low, unless control variables are included. In this case, it is mostly the population variable that contributes to explaining changes in wealth. Even though the wealth signal in light is low, it is remarkable that it remains significant even on a single pixel level. This is good news for researchers who want to apply light data on local levels or draw economic conclusions based on the effects of changes in nighttime lights. We can conclude that there is a significant association between changes in nighttime lights and changes in wealth on a local level. This association does not show in DHS based studies due to use of relative wealth variables and a third source population data. However, using the full census of Namibia, we find that light change is associated with wealth change even at highly disaggregated spatial levels.

## 6 Conclusions

Nighttime lights data has established its place as a proxy for economic development on a macro level. Consequently, a large influx of new studies applied the data in economic research during the past decade (e.g. Michalopoulos and Papaioannou 2013; Hodler and Raschky 2014; Alesina et al. 2016; Storeygard 2016; Henderson et al. 2017). However, the evidence is not complete at a local level. Due to data constraints, we do not know if nighttime lights provide a meaningful signal of economic development in the smallest spatial units. One approach is to relate lights to DHS wealth data (Bruederle and Hodler 2018). It retains a large sample of developing countries, but compromises on the data coverage due to random sampling. We evaluate their findings by taking an advantage of complete census data in Namibia.

We raise concerns with the practice of aggregating light data from a buffer around survey locations. The buffers fail to aggregate light that is relevant to survey respondents or, the buffers aggregate too much of it. The type and extent of the failure depends on the spatial size of an enumeration area and therefore it leads to a systematic bias between densely populated and remote areas. Furthermore, the buffers can overlap, which violates the assumption of independent observations in basic regression analysis. As a solution, we recommend aggregating survey and nighttime lights data using grids. The distribution of survey data into grid cells depends on whether it contains point or polygon data for the enumeration areas. Our results do not give a clear indication on whether it is better to use the point or polygon method. However, the latter provides a more realistic distribution of the underlying survey data and keeps every grid cell in the sample. Therefore, if the border data is available, we recommend using the polygon method.



Altogether, we have found that nighttime lights data is an even stronger proxy for economic development at a local level than previous literature suggests. The largest differences in the results can be attributed to using the total stock of wealth, a grid data aggregation approach and data coverage provided by the full census.

It is not clear how generalizable the results are outside Namibia. However, based on the previous literature, the association between light and wealth is stronger in most developing countries than in Namibia. Therefore, we are optimistic that similar results can be found in other countries as well. We encourage future research to study the correlation between nighttime lights and different indicators of economic development in different regions to Namibia. Based on our findings, we can also encourage the use of nighttime lights data in applications that use very small spatial units.

## Appendix

### A1. Wealth index components

	Asset variable	Availability in DHS	Availability in census
Wealth index B	Access to electricity*	2000	2001
	Radio	2000	2001
	Television	2000	2001
	Water piped into residence*	2000	2001
	Flush toilet*	2000	2001
Wealth index A	Car*	2000 & 2013	2001 & 2011
	Motorcycle	2000 & 2013	2001 & 2011
	Bicycle	2000 & 2013	2001 & 2011
	Computer	2000 & 2013	2001 & 2011
	Landline telephone	2000 & 2013	2001 & 2011
	Mobile telephone	2000 & 2013	2001 & 2011
	Refridgerator	2000 & 2013	2001 & 2011
	Stove	2000 & 2013	2001 & 2011
	Internet	2000 & 2013	2001 & 2011
	Microwave oven	2000 & 2013	2001 & 2011

*Note: Wealth index A and B are relative indicators constructed with principal component analysis. \*In addition, we use access to electricity, car ownership, flush toilet and water piped into residence as absolute wealth indicators.*

**Fig. 5** Wealth index components

## A2. Relative wealth and light in enumeration areas and buffers.

**Table 7** Enumeration area and buffer descriptive statistics

Variable name	Description	DHS 2013	Census 2011
–	Clusters/Enumeration Areas	550	5468
Wealth_A	Wealth index with 15 components	3.00	3.01
Stb_EA_mean	Average stable light in EA	–	16.27
Stb_BUF_mean	Average stable light in buffer	16.56	13.74
EA_popD	Population per pixel in EA	–	176
GPW_BUF_popD	GPW population density per pixel in buffer	832	708

**Table 8** Regressions with enumeration area borders/buffers and relative wealth as dependent variable

OLS regression	(1)	(2)	(3)	(4)	(5)	(6)
Survey	DHS	DHS	Census	Census	Census	Census
Year	2013	2013	2011	2011	2011	2011
Unit	Buffer	Buffer	Buffer	Buffer	EA border	EA border
Dep. var	Wealth_A	Wealth_A	Wealth_A	Wealth_A	Wealth_A	Wealth_A
Log_stb_mean	0.219 (0.011)***	0.297 (0.024)***	0.205 (0.003)***	0.274 (0.006)***	0.204 (0.003)***	0.232 (0.004)***
Log_GPW_BUF_popD	–	–0.108 (0.029)***	–	–0.104 (0.008)***	–	–
Log_EA_popD	–	–	–	–	–	–0.061 (0.006)***
Observations	550	550	5468	5468	5468	5468
R <sup>2</sup>	0.421	0.435	0.439	0.457	0.483	0.492

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Units with zero light are included

### A3. Regressions with 0.5 degree grid cells and the stock of wealth as dependent variable

**Table 9** Regressions with 0.5 degree grid cells and the stock of wealth as dependent variable

OLS regression	(1)	(2)	(3)	(4)	(5)	(6)
Survey	Census	Census	Census	Census	Census	Census
Year	2011	2011	2011	2011	2011	2011
Method	Polygon	Polygon	Polygon	Polygon	Polygon	Polygon
Dep. var	Log_piped	Log_piped	Log_flush	Log_flush	Log_car	Log_car
Log_stb_sum	0.339 (0.026)***	0.234 (0.022)***	0.268 (0.024)***	0.170 (0.019)***	0.235 (0.023)***	0.070 (0.010)***
Area_size	—	0.001 (0.000)***	—	0.001 (0.000)***	—	0.000 (0.000)
Log_pop	—	0.514 (0.051)***	—	0.479 (0.043)***	—	0.813 (0.023)***
Observations	342	342	342	342	342	342
R2	0.328	0.625	0.271	0.656	0.23	0.889

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Units with zero light are included

#### A4. Regressions with 0.5 degree grid cells and the change in relative wealth as dependent variable

**Table 10** Regressions with 0.5 degree grid cells and the change in relative wealth as dependent variable

OLS regression	(1)	(2)	(3)	(4)	(5)	(6)
Survey	DHS	DHS	Census	Census	Census	Census
Year	2000–2013	2000–2013	2001–2011	2001–2011	2001–2011	2001–2011
Method	Point	Point	Point	Point	Polygon	Polygon
Dep. var	Wealth_B change	Wealth_B change	Wealth_B change	Wealth_B change	Wealth_B change	Wealth_B change
Stb_mean_log_change	–0.042 (0.167)	–0.058 (0.158)	0.055 (0.049)	0.054 (0.049)	0.000 (0.034)	–0.005 (0.035)
GPW_Pop_D_log_change	–	–1.390 (0.448)***	–	–0.291 (0.255)	–	–
Pop_D_log_change	–	–	–	–	–	0.088 (0.067)
Observations	74	74	134	134	172	172
R2	0.001	0.120	0.010	0.019	0.000	0.010

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Units with zero light for both years are not included

### A5. 0.5 Regressions using 0.5 degree grid cells and the change in the stock of wealth as dependent variable

**Table 11** 0.5 Regressions using 0.5 degree grid cells and the change in the stock of wealth as dependent variable

OLS regression	(1)	(2)	(3)	(4)
Survey	Census	Census	Census	Census
Year	2001–2011	2001–2011	2001–2011	2001–2011
Method	Polygon	Polygon	Polygon	Polygon
Dep. var	Piped_log_change	Piped_log_change	Flush_log_change	Flush_log_change
Stb_sum_log_change	0.011 (0.017)	0.009 (0.015)	0.010 (0.012)	0.008 (0.008)
Area_size	–	0.000 (0.000)	–	0.000 (0.000)
Pop_log_change	–	0.908 (0.117)***	–	0.842 (0.067)***
Observations	172	172	172	172
R2	0.003	0.266	0.005	0.488

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Units with zero light in both years are not included

**Funding** The research was supported by authors' home institutions. The authors are grateful for funding provided by the TWAS-DFG Cooperation Visits Programme for Scientist from sub-Saharan Africa allowing Thomas Ferreira the opportunity to spend time at TU Braunschweig, Institute of Economics, Germany.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abrahams A, Oram Lozano-Gracia CN (2018) Deblurring DMSP nighttime lights: a new method using Gaussian filters and frequencies of illumination. *Remote Sens Environ* 210:242–258. <https://doi.org/10.1016/j.rse.2018.03.018>
- Alesina A, Michalopoulos S, Papaioannou E (2016) Ethnic Inequality. *Polit Econ* 124(2):428–488. <https://doi.org/10.1086/685300>
- Baugh K, Elvidge C, Ghosh T, Ziskin D (2010) Development of a 2009 stable lights product using DMPS-OLS data. *Proc Asia Pac Netw* 30(0):114–130
- Bluhm R, Krause M (2018) Top lights-bright cities and their contribution to economic development. CESifo Working Paper Series No 7411. Center for Economic Studies, Munich
- Bruederle A, Hodler R (2018) Nighttime lights as a proxy for human development at the local level. *PLoS ONE* 13(9):e0202231. <https://doi.org/10.1371/journal.pone.0202231>
- CBS (2001) Namibia 2001 population and housing census. Central Bureau for Statistics,
- Chen X, Nordhaus WD (2011) Using luminosity data as a proxy for economic statistics. *Proc Natl Acad Sci* 108(21):8589–8594. <https://doi.org/10.1073/pnas.1017031108>
- Chen X, Nordhaus WD (2019) VIIRS nighttime lights in the estimation of cross-sectional and time-series GDP. *Remote Sens* 11(9):1–11. <https://doi.org/10.3390/rs11091057>
- CIESIN (2018) Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 10. Center for International Earth Science Information Network—Columbia University. NASA Socioeconomic Data and Applications Center, Palisades <https://doi.org/10.7927/H49C6VHW>
- Croft TA (1978) Night-time images of the earth from space. *Sci Am* 239(1):86–101
- Doll CNH, Muller J, Morley JG (2006) Mapping regional economic activity from night-time light satellite imagery. *Ecol Econ* 57(1):75–92. <https://doi.org/10.1016/j.ecolecon.2005.03.007>
- Elvidge CD, Baugh KE, Kihn EA, Kroehl HW, Davis ER (1997) Mapping city lights with nighttime data from the DMSP operational linescan system. *Photogramm Eng Remote Sensing* 63(6):727–734
- Elvidge CD, Ziskin D, Baugh KE, Tuttle BT, Ghosh T, Pack DW, Erwin EH, Zhizhin M (2009) A fifteen year record of global natural gas flaring derived from satellite data. *Energies* 2(3):595–622. <https://doi.org/10.3390/en20300595>
- Elvidge CD, Baugh KE, Zhizhin M, Hsu FC, Ghosh T (2017) VIIRS night-time lights. *Int J Remote Sens* 38(21):5860–5879. <https://doi.org/10.1080/01431161.2017.1342050>
- Elvidge CD, Ghosh T, Hsu FC, Zhizhin M, Bazilian M (2020) The dimming of lights in China during the COVID-19 pandemic. *Remote Sens* 12(17):2851. <https://doi.org/10.3390/rs12172851>
- Fabian M, Lessmann C, Sofke T (2019) Natural disasters and regional development—the case of earthquakes. *Envir Dev Econ* 24(5):479–505. <https://doi.org/10.1017/S1355770X19000159>
- Filmer D, Pritchett LH (2001) Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography* 38(1):115–132
- Gehlke CE, Biehl K (1934) Certain effects of grouping upon the size of the correlation coefficient in census tract material. *J Am Stat Assoc* 29(185):169–170. <https://doi.org/10.2307/2277827>
- Ghosh T, Anderson S, Powell R, Sutton P, Elvidge CD (2009) Estimation of Mexico's informal economy and remittances using nighttime imagery. *Remote Sens* 1(3):418–444. <https://doi.org/10.3390/rs1030418>

- Harttgen K, Klasen S, Vollmer S (2013) An African growth miracle? Or: What do asset indices tell us about trends in economic performance? *Rev Income Wealth* 59:537–561. <https://doi.org/10.1111/roiw.12016>
- Henderson JV, Storeygard A, Weil DN (2012) Measuring economic growth from outer space. *Am Econ Rev* 102(2):994–1028. <https://doi.org/10.1257/aer.102.2.994>
- Henderson JV, Squires T, Storeygard A, Weil DN (2017) The global distribution of economic activity: nature, history, and the role of trade. *Q J Econ* 133(1):357–406. <https://doi.org/10.1093/qje/qjx030>
- Hodler R, Raschky P (2014) Regional favoritism. *Q J Econ* 129(2):994–1033. <https://doi.org/10.1093/qje/qju004>
- Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S (2016) Combining satellite imagery and machine learning to predict poverty. *Science* 353(6301):790–794. <https://doi.org/10.1126/science.aaf7894>
- Lessmann C, Steinkraus A (2019) The geography of natural resources, ethnic inequality and civil conflict. *Eur J Polit Econ* 59:33–51. <https://doi.org/10.1016/j.ejpoleco.2019.01.005>
- Li X, Zhou Y, Zhao M, Zhao X (2020) A harmonized global nighttime light dataset 1992–2018. *Sci Data* 7(1):1–9
- Määttä I, Lessmann C (2019) Human lights. *Remote Sens* 11(19):2194. <https://doi.org/10.3390/rs11192194>
- Michalopoulos S, Papaioannou E (2013) National institutions and subnational development in Africa. *Q J Econ* 129(1):151–213. <https://doi.org/10.1093/qje/qjt029>
- NSA (2013) Namibia 2011 population and housing census. Namibia Statistics Agency,
- Odendaal W (2011) Land Grabbing in Namibia: A case study from the Omusati region, northern Namibia. Land Deals Politics Initiative, University of Sussex, 6–8 April 2011. Paper presented at the international conference on global land grabbing
- Openshaw S (1983) The modifiable areal unit problem. Geobooks, Norwick Norfolk
- Rutstein S, Staveteig S (2014) Making the demographic and health surveys wealth index comparable. DHS methodological reports no. 9. ICF International, Rockville
- Storeygard A (2016) Farther on down the road: transport costs, trade and urban growth in sub-saharan africa. *Review of Economic Studies* 83(3):1263–1295. <https://doi.org/10.1093/restud/rdw020>
- Sutton PC, Costanza R (2002) Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecol Econ* 41(3):509–527. [https://doi.org/10.1016/S0921-8009\(02\)00097-6](https://doi.org/10.1016/S0921-8009(02)00097-6)
- Weidmann N, Schutte S (2017) Using night light emissions for the prediction of local wealth. *J Peace Res* 54(2):125–140. <https://doi.org/10.1177/0022343316630359>
- World Bank (2020) World Development Indicators. GINI index (World Bank estimate) (SI.POV.GINI)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.