

# SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

## A high-resolution gridded dataset to assess electrification in sub-Saharan Africa

Giacomo Falchetta <sup>1,2</sup>, Shonali Pachauri <sup>1</sup>, Simon Parkinson<sup>1,3</sup> & Edward Byers <sup>1</sup>

Received: 19 February 2019

Accepted: 31 May 2019

Published online: 03 July 2019

Spatially explicit data on electricity access and use are essential for effective policy-making and infrastructure planning in low-income, data-scarce regions. We present and validate a 1-km resolution electricity access dataset covering sub-Saharan Africa built on gridded nighttime light, population, and land cover data. Using light radiance probability distributions, we define electricity consumption tiers for urban and rural areas and estimate the by-tier split of consumers living in electrified areas. The approach provides new insight into the spatial distribution and temporal evolution of electricity access, and a measure of its quality beyond binary access. We find our estimates to be broadly consistent with recently published province- and national-level statistics. Moreover, we demonstrate consistency between the estimated electricity access quality indicators and survey-based consumption levels defined in accordance with the World Bank Multi-Tier Framework. The dataset is readily reproduced and updated using an open-access scientific computing framework. The data and approach can be applied for improving the assessment of least-cost electrification options, and examining links between electricity access and other sustainable development objectives.

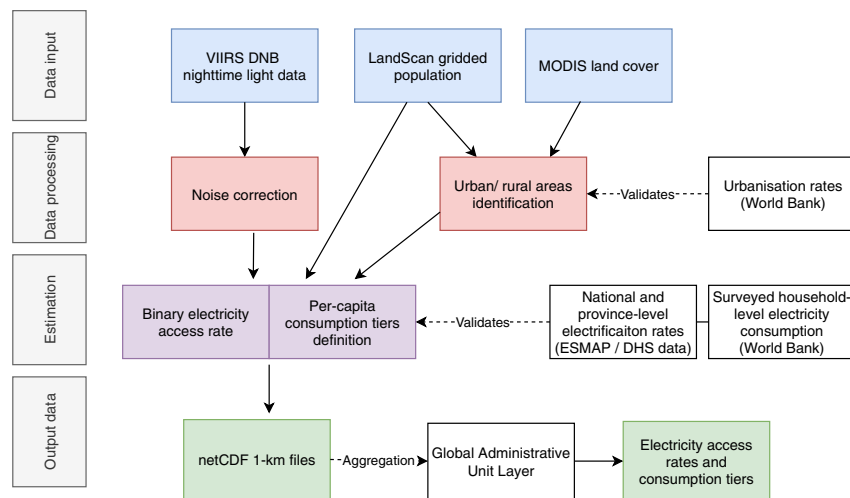
### Background & Summary

The United Nations' Sustainable Development Goal 7 (SDG 7) aims at ensuring access to affordable, reliable, sustainable and modern energy for all by 2030<sup>1</sup>. Access to electricity services (SDG targets 7.1.1 and 7.A and 7.B) is a key priority under this goal - also due to the strong interconnections it exhibits with other development objectives<sup>2-5</sup>, but achieving it globally presents significant challenges<sup>6</sup>. Spatially-explicit, up-to-date and harmonized information on electrification progress is needed to identify where local, national and international efforts can be focused to achieve the most benefits.

Tracking initiatives such as the *Progress Towards Sustainable Energy* portal<sup>7</sup>, the *Status of Electricity Access Report*<sup>8</sup>, and the publication of field-collected data<sup>9,10</sup> have quantified electrification progress over the last 20 years. It is estimated that in 2016 the global population without access dipped below 1 billion<sup>11</sup>, largely as a result of international investment flows<sup>12</sup>, declining costs of distributed access solutions (e.g., solar PV)<sup>13</sup>, and supporting policy frameworks<sup>14</sup>. Of those still without access, around 600 million live in sub-Saharan Africa (SSA), where 15 countries have electrification levels below 25%<sup>11</sup>.

Monitoring at national scales is important to track efforts regularly. Yet, assessing sub-national heterogeneity is essential to render a clearer picture of the electrification status of a country and improve the efficacy, design, and implementation of local-scale electrification plans<sup>15,16</sup>. This is because electrification efforts could be concentrated in specific regions, such as those nearer to existing grid-connections, whereas progress may remain slow or even stagnate in more remote areas that lack access to energy resources<sup>17,18</sup>. Moreover, data on differences in consumption levels are needed to assess the quality and inequality of access and subsequent implications of these access inequalities<sup>19</sup>. However, actual electricity consumption data are rarely available at a high level of spatial resolution. Using field-surveys to collect such data are costly and time-intensive. In this context, an ongoing debate is centered on how to best deliver good quality data to track energy access objectives and to lower its collection and processing costs, as well as to make sound decisions under data scarcity<sup>20</sup>.

<sup>1</sup>Energy Program, International Institute for Applied Systems Analysis (IIASA), Schossplatz 1, 2361, Laxenburg, Austria. <sup>2</sup>Future Energy Program, Fondazione Eni Enrico Mattei (FEEM), Corso Magenta 63, 20123, Milan, Italy. <sup>3</sup>Institute for Integrated Energy Systems, University of Victoria, PO BOX 3055 STN CSC, Victoria, Canada. Correspondence and requests for materials should be addressed to G.F. (email: [giacomo.falchetta@feem.it](mailto:giacomo.falchetta@feem.it))



Blue: gridded input data; red: data processing stages; purple: estimation stages; green: output data; white: supporting datasets.

**Fig. 1** Schematic of the open-source scientific computing framework. The flowchart represents the different stages, data inputs, and outputs in the generation and validation of the data. Blue: remotely-sensed variables; red: processing steps; purple: output metrics; green: output data; white: validation and supporting datasets; grey: stages.

Here, we present and validate an updatable, fully-reproducible 1-*km* resolution dataset of electricity access and use tiers for SSA built on regularly updated gridded nighttime light (NTL), population, and land cover data. The dataset is in the form of netCDF files with counts of the number of people without access in each 1-*km*<sup>2</sup> cell for each year between 2014 and 2018 and the estimated tier of consumption in each pixel for 2018. The estimates for each metric and country are validated with the most recent available figures. The data can be browsed interactively in a web-based interface. The entire code - from data generation, processing, and export - is available for results reproduction. New NTL composites are published each month, allowing continuous updates. Our estimates are broadly consistent with both field survey-based province-level electrification levels reported by DHS surveys, and aggregated national-level figures published yearly by the the ESMAP/World Bank *Tracking SDG 7* report. The dataset can be combined with other regional-level data and implemented in an array of studies for developing scenarios and modelling pathways of energy access and demand in Integrated Assessment Models<sup>21,22</sup>, assessing least-cost electrification options<sup>15</sup>, tracking development objectives at a local scale, designing infrastructure, and analysing links between electricity access and other socio-economic and well-being metrics.

Figure 1 illustrates a schematic of the open-source scientific computing framework, which is discussed in detail in the Methods section. Recent high-resolution data assembling efforts for nighttime light (NTL)<sup>23</sup>, population<sup>24,25</sup>, and land cover<sup>26</sup>, and cloud computing facilities for spatial data processing<sup>27</sup> enable estimating the evolution in electricity access and consumption tiers at the 1-*km*<sup>2</sup>-pixel-level. We employ VIIRS-DNB (*Visible Infrared Imaging Radiometer Suite, Day-Night Band*) stray-light corrected product for NTL radiance (450 m resolution at the equator) and LandScan (used as the main source) and WorldPop (for comparison) gridded population distribution datasets, both with a 1-*km* resolution. WorldPop was only used as a secondary source due to possible circularity issues since the underlying model also uses NTL as an input for estimating population. Refer to <https://www.popgrid.org/compare-data> for a comparative overview of inputs and methodology of different gridded population datasets. The use of Landscan and WorldPop supports accounting for recent population growth (on average 4.1% in urban areas vs. 2.7% for SSA as a whole in 2017, according to World Bank Data, <https://data.worldbank.org/>) and migration dynamics.

Nighttime lights have already been used as a proxy for electricity access<sup>28–31</sup>, residential consumption<sup>32–37</sup>, outages detection<sup>38–40</sup>, population fluctuations and migration<sup>41,42</sup>, regional GDP<sup>43,44</sup> and income inequality<sup>45</sup>. This is, however, the first time a fully reproducible and updatable dataset at this high level of resolution has been constructed and validated for SSA.

## Methods

**Data sources and processing tools.** The data sources - described in Table 1 - include VIIRS-DNB stray-light corrected NTL monthly composites for the years 2014–2018<sup>23</sup> (which measure radiance and filter cloud cover), LandScan 2014 and 2017 downscaled gridded population<sup>24</sup>, and WorldPop 2015 and 2020<sup>25</sup> for sensitivity analysis. National electricity access rates are drawn from the ESMAP/World Bank tracking SDG 7 portal<sup>7</sup>, while province-level figures are drawn from an array of field surveys through the USAID/DHS Program STATcompiler for sub-national benchmarking<sup>10</sup>. Household survey data to validate consumption tiers have been retrieved from the WorldBank Microdata Library (<http://microdata.worldbank.org>). For defining countries, we adopted the global administrative boundaries (GADM) dataset v3.6<sup>46</sup>.

Dataset	Source	Time resolution	Spatial resolution	Year(s)
VIIRS-DNB nighttime lights	<sup>23</sup>	1 month	0.008°	2014–2018
LandScan gridded population	<sup>24</sup>	1 year	1 km	2014–2017
WorldPop gridded population	<sup>25</sup>	1 year	1 km	2015 and 2020
MODIS Land Cover Type, University of Maryland classification	<sup>26</sup>	1 year	450 m	2017
GADM global administrative layers	<sup>46</sup>	—	—	2018
USAID/DHS StatCompiler surveys	<sup>10</sup>	1–2 years	Province-level	2014–2018, depending on country
ESMAP/World Bank electrification database	<sup>7</sup>	1 year	Country-level	2014–2016
World Bank MicroData Library household surveys	—	1–2 years	Household-level	2014–2018, depending on country

**Table 1.** Table of data inputs for electrification data generation and validation.

The Google Earth Engine platform<sup>27</sup> is used to process satellite imagery and extract data, which is subsequently used to produce the datasets and generate plots in the R scientific computing environment. The Code Availability and Data Records section links to the repositories that host the Earth Engine Javascript code and R script allowing for results reproduction, alteration of parameters for further sensitivity analysis, or other improvements as implemented by new users. Results are also dynamically visualized in a web interface, under <https://glessa.ene.iiasa.ac.at>.

**Urban and rural areas identification.** The definition of urbanisation is highly variable across countries and regions of both Africa<sup>47,48</sup> and the entire world<sup>28</sup>. The United Nations' definition of urban and rural population is tightly linked to political and administrative factors that make it challenging to generalize the human settlement structure across different countries<sup>49</sup>. Recently, new methods have been promoted to overcome such issues, especially in high-income countries<sup>50</sup>. Those rely on satellite imagery and land cover, but also on functional definitions, e.g. accounting for the entire commuting zones surrounding cities.

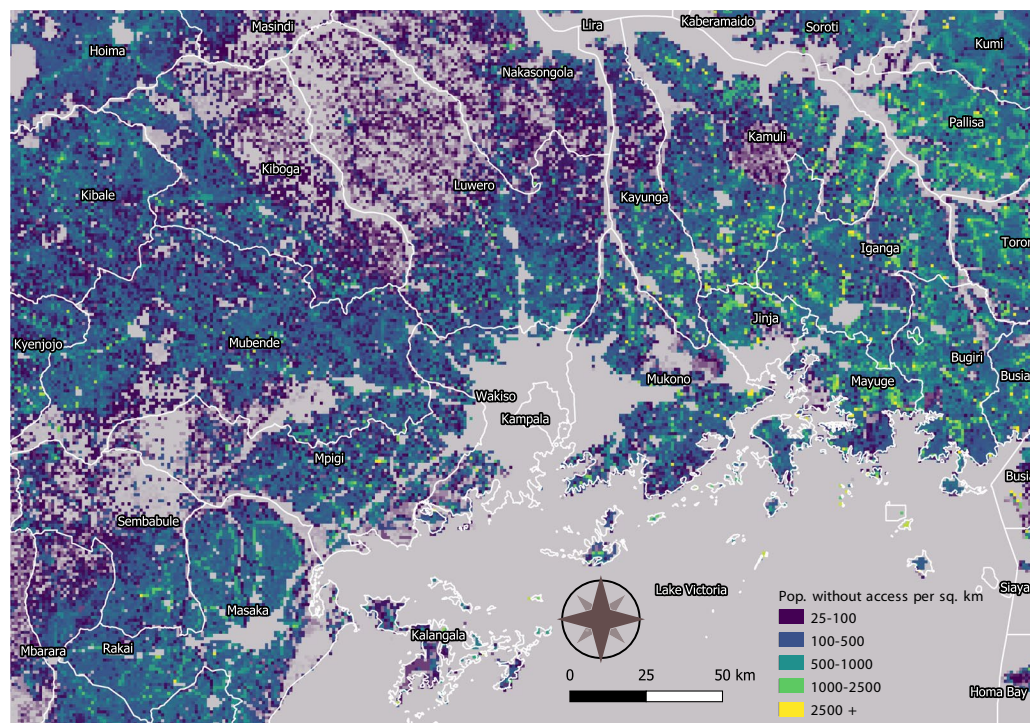
For identifying urban and rural settlements at the grid-cell level, we adopt the 2017 MCD12Q1 V6 MODIS Land Cover Type<sup>26</sup> from the Annual University of Maryland classification (having a resolution of 500 m) in combination with the 2017 LandScan population (1-km resolution). Both datasets are updated on an annual basis. For urban areas, land cover type 13 (urban and built-up lands) is used as a necessary condition, in combination with a population density higher than specific  $inhab \cdot km^{-2}$  thresholds.

Supplementary Table 1 lists the clusters of countries and the corresponding thresholds which have been defined and applied so as to minimize the discrepancy with World Bank reported urban-rural population split for 2017 (<https://data.worldbank.org/>). Threshold bands partially overlap with geographic regions, such as Southern Africa, where a threshold of more than 650  $inhab \cdot km^{-2}$  is found to effectively predict urban areas. Nonetheless, for some countries their threshold does not match their region and they were inserted in density clusters which allowed a better estimation of their reported urbanization rate. For rural settlements, land area with a population density lower than that of each cluster's threshold was considered. Non-populated land was discarded.

The approach results in a very consistent estimate ( $R^2 = 0.84$  and 0.067 residual standard error in a simple linear regression) vis-à-vis the World Bank/UN population division figures built on countries' official national statistics. Supplementary Fig. 1 shows the prediction obtained with the approach.

**Electricity access estimation.** A large body of literature has shown that locations where satellite sensors cannot detect visible light are likely to be areas where people live without access to electricity<sup>28–30,51</sup>. We calculated the median value of radiance within each pixel of the NPP-VIIRS monthly composites for both 2014 and 2018 within the Google Earth Engine Platform; a lower-bound noise floor was set at 0.25  $\mu W \cdot cm^{-2} \cdot sr^{-1}$  for 2014 and to 0.35  $\mu W \cdot cm^{-2} \cdot sr^{-1}$  for 2018 to remove calibration noise and ephemeral lights as discussed in the relevant literature<sup>52–55</sup>. The discrepancy in the noise caps is justified on the grounds of the increased noisiness of the data witnessed from 2017. According to expert opinion, this is owing to calibration adjustments in the lunar correction and to the launch of a new VIIRS-equipped satellite (JPSS J1) in 2017 (as discussed in<sup>56</sup>). The spread in the noise lower-bound cap set before and after 2017 was determined from the examination of by-definition zero radiance pixels - such as within large water bodies - which nonetheless present a systematically positive radiance for data since 2017, and a systematically zero radiance for previous data. Refer to the Supplementary Information for histograms comparing the distribution of background noise before and after 2017 for the same non-populated area. Population counts within each GADM level 1 provinces have been produced for both people living in lit (above 0.25/0.35  $\mu W \cdot cm^{-2} \cdot sr^{-1}$ ) and in dark areas ( $\leq 0.25/0.35 \mu W \cdot cm^{-2} \cdot sr^{-1}$ ). The gridded 1-km classification was used to aggregate gridded results into spatial polygon features representing provincial and national administrative boundaries. Aggregation of the gridded data is achieved assuming grid cells belong to the polygon that contain its centroid. To provide an example of the output data, Fig. 2 shows a snapshot of the estimated density of people without electricity access over Uganda in 2018.

**Classification of consumption tiers.** Previous studies have shown that residential electricity consumption can be accurately predicted with satellite derived NTL data at both the country<sup>57</sup> and the province-level<sup>58</sup>. Here, we considered a measure of radiance for those estimated to live in areas with electricity access. Based on the distribution of the quartile values of non-zero light radiance across SSA countries (Supplementary Fig. 2), four



**Fig. 2** Distribution of people without access over Uganda in 2018. The figure provides a sample representation of the output dataset. Colours represent the density of people without electricity access in each 1-km pixel. Administrative boundaries correspond to the first level of the GADM definition<sup>46</sup>.

Tier	Urban		Rural	
	Lower-bound	Upper-bound	Lower-bound	Upper-bound
1	>0	<0.40	>0	<0.38
2	≥0.40	<0.48	>0.38	<0.45
3	≥0.48	<0.88	>0.45	<0.68
4	≥0.88	—	>0.68	—

**Table 2.** Definition of radiance tiers ( $\mu W \cdot cm^{-2} \cdot sr^{-1}$ ) for electrified areas used to estimate consumption levels.

tiers of residential electricity consumption were defined, with thresholds set at the median value of each quartile distribution. In order to account for the strong urban-rural discontinuity in lighting, this was done separately for urban and rural settlements. Settlement classification - and thus indirectly also tiers - is based on population density data and accounts for the heterogeneity in densities across countries, as described previously. Corresponding tier thresholds are reported in Table 2.

The four tiers are validated against survey data consistent with the World Bank Multi-Tier framework<sup>59</sup>, as discussed in the Technical Validation Section. The tier data can be mapped onto a gridded population dataset to derive the share of population with access to electricity in each consumption tier in both urban and rural areas.

### Data Records

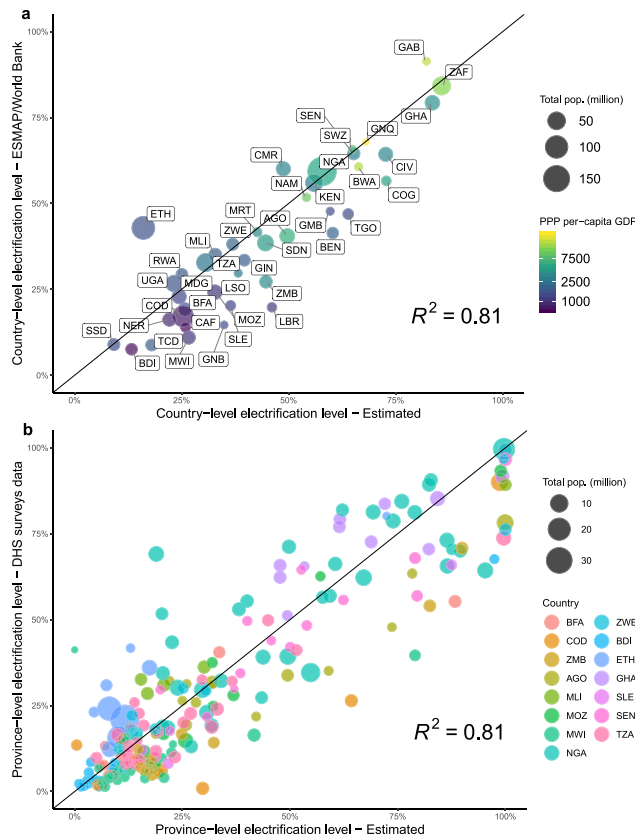
The data<sup>60</sup> are hosted on Mendeley Data at the permanent <https://doi.org/10.17632/kn4636mtvg>. The repository includes two netCDF v4 files containing the data output of the analysis, which are reported in Table 3. The first file reports the number of people without access in each grid cell between 2014 and 2018. The second reports the consumption tier in each populated, electrified cell for the year 2018. Both files have a 1-km spatial resolution and cover sub-Saharan Africa entirely. The data may be shared, used and adapted even for commercial use with the requirement of attributing to this original work as per a Creative Commons BY v4.0 (CC BY 4.0) License.

### Technical Validation

To assess the consistency of the estimated electrification levels, we produced a cross-sectional fit comparing the NTL estimate against the most recent official data source available, the ESMAP/World Bank *Tracking SDG 7* electrification data for the year 2016, as shown in Fig. 3a. The plot also includes a colour-legend for PPP per-capita GDP, which suggests that discrepancies tend to be larger in poorer countries. Overall, this reveals a rather consistent estimation of the heterogeneity in the data ( $R^2 = 0.81$ ). The remote sensing approach performs well in terms

File	Time resolution	Spatial resolution	Format
Density of people without electricity access	1 year (2014–2018)	1 km	netCDF v4
Tier of electricity consumption	1 year (2018)	1 km	netCDF v4

**Table 3.** Output data files, resolution, and format.



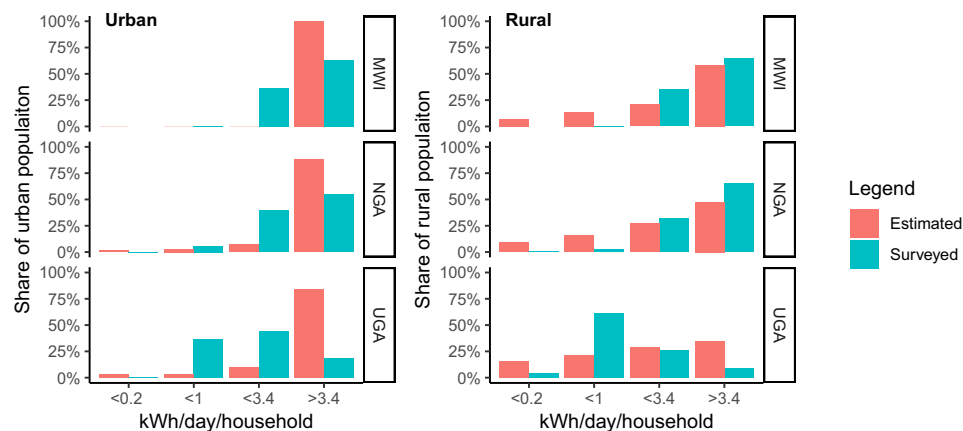
**Fig. 3** Scatterplot representing: **(a)** the national electricity access rates estimated at the pixel-level relative to values reported by ESMAP/World Bank for year 2016 (the point size is scaled to the national population and colours describe the PPP per-capita GDP of each country); **(b)** the province-level electricity access rates estimated at the pixel-level relative to USAID/DHS StatCompiler (with point size scaled to the province population and colours identifying the country of belonging). Various years are included between 2014 and 2017, depending on the survey data available for specific countries.

of reproducing the electrification levels - among others - in Botswana, Namibia, Tanzania, Uganda, Mali, Guinea, Madagascar, and Namibia.

The largest underestimations are found for Ethiopia (−26.8%), Cameroon (−11.3%), Gabon (−9.2%), while electrification levels are significantly overestimated vis- a-vis official figures in Liberia (+26%), Guinea Bissau (+20.2%), and Benin (+18.9%). Downward bias can be mainly attributed to - amongst other reasons - the under-detection of standalone access solutions such as solar-home-systems, or to the low level of lighting among highly distributed communities. For instance, the large discrepancy in Ethiopia could be owing to the very rapid diffusion of solar-home-systems among rural households over the last years (4 million off-grid solar devices sold between 2015 and 2017, with a penetration of 20% in 2017<sup>61</sup>), which might be scarcely detected by NTL data due to the inherent mismatch between solar power production and night-time light measurements in the absence of diurnal storage systems. Moreover, limited interior lighting might not be visible from outside and thus detectable using the satellite imagery.

Upward discrepancies (overestimations) are prone to be attributable to a number of factors. These encompass illegal connections (refer to<sup>62,63</sup> for a recent contribution to estimation of electricity thefts), other sources of light (e.g. from the effect of gas flaring, leading to very high and diffuse radiance detected by satellites and poorly filtered by data processing algorithms), or from the presence of a large number of people without access even in areas where there is electricity (such as in urban slums and peri-urban settlements).

Finally, it was assumed that a 1-km pixel being lit implies that everyone in that location has access. This assumption implies that these people are in proximity to connection possibilities, i.e. remoteness is not a



**Fig. 4** Validation of estimated consumption tiers compared to World Bank Multi-Tier framework thresholds in urban and rural areas, drawn from household surveys for selected countries with data availability.

significant problem. At the same time, a pixel being identified as dark still allows for the possibility that some form of (standalone) electricity generation is in place. Therefore, our estimate might be interpreted best as a measure of grid-based electricity access. Crucially, the accuracy of official electrification levels and the differing survey criteria by which a household is considered electrified across countries can affect the validation. In any case, both caveats are mitigated here by the high resolution of the data, which limits potential errors of misclassifying populations as with or without access.

Overall, the validation shows that our bottom-up approach to estimate electrification levels - derived from 1-km grid cell level estimates - represents a generally effective method for predicting national electrification levels. In an attempt to perform a more precise validation, we collected survey-derived province-level shares of households with access to electricity from the USAID/DHS Program STATcompiler (demographic and health surveys, and disease-related surveys)<sup>10</sup>. This was possible for fifteen SSA countries (Angola, Burundi, Burkina Faso, DR Congo, Ethiopia, Ghana, Mali, Malawi, Mozambique, Nigeria, Senegal, Sierra Leone, Tanzania, Zambia, and Zimbabwe). It is emphasized that the surveying period of data reported by the USAID/DHS STATcompiler often spans between two calendar years. Therefore, to produce a consistent validation, we referred to different years (between 2014 and 2017) dependent on the survey year in each country, and produce NTL-based estimates for each year. Also, USAID/DHS data are subject to smoothing processes due to the use of sampling weights (equivalent to the frequency that the sampling unit represents in the target population.), which might skew access rate estimates<sup>10</sup>.

Our methodology is found to be consistent with the specific electrification levels ( $R^2 = 0.81$ ), as shown in Fig. 3b, although a few inconsistencies exist. In particular, greater error is observed between our estimate and the USAID/DHS data for provinces with smaller populations. This suggests that the approach is generally more effective in estimating electrification in more populated provinces.

Last, to validate the estimated electricity consumption tiers in rural and urban areas, respectively, we retrieve household survey data from the World Bank Microdata Library (<http://microdata.worldbank.org>). Recent data on expenditure or consumption of electricity between 2014 and 2018 are available only for three countries: Malawi (2016–17), Nigeria (2015–16), and Uganda (2013–14). Thus, it must be highlighted that the estimate is built on data from year 2018, but it could only be validated on earlier field data. In all cases, the data collected refer to consumption in the month preceding the survey. Expenditure had to be divided by reference prices to obtain power demand. For Malawi, a price of MWK 26 in rural and MWK 35 in urban areas was used<sup>64</sup>, for Nigeria a price of NGN 7 in rural and NGN 12 in urban areas was considered<sup>65</sup>, while for Uganda a price of UGX 400 in rural and UGX 550 in urban areas was adopted<sup>66</sup>. Price discrepancies account for cross-subsidies and monthly fees exempted from rural households with low consumption levels<sup>67</sup>. Continuous consumption values have then been clustered into tiers of daily consumption following the thresholds as defined by the World Bank Multi-Tier Framework (WB-MTF)<sup>59</sup>, with WB-MTF Tiers 4 and 5 (the two highest) aggregated to construct a consistent set of four estimated tiers. Urban and rural areas are distinguished in the surveys for each household. Figure 4 reports the results of the validation, which show a rather consistent estimation.

The main discrepancies are observed for low consumption tiers in rural areas, where our estimate predicts a relevant fraction of the electrified households to live with less than 1 kWh/day, a result which does not find robust evidence in surveys. In all likelihood, the main reason behind this discrepancy is that survey questions about electricity expenditure and consumption only target grid-connected households, and thus results underrepresent communities relying on decentralised electricity access solutions, which often also have low consumption levels. Another significant discrepancy is observed for households consuming more than 3.4 kWh/day in urban Uganda, overestimated by our data. Here, a plausible source for the discrepancy concerns the high population density characterising Kampala and the provinces surrounding Lake Victoria, resulting in greater urban-rural classification uncertainty.

**Sensitivity analysis.** The analysis includes two main sources of uncertainty: the choice of the NTL noise floor, and the accuracy of the gridded population data across different years.

Supplementary Fig. 3a shows the change in the estimate in response to a  $\pm 25\%$  change in the NTL data noise floor (set at  $0.25 \mu W \cdot cm^{-2} \cdot sr^{-1}$  until 2016 and  $0.35 \mu W \cdot cm^{-2} \cdot sr^{-1}$  for later years, as discussed in the Methods section) and can be interpreted as a confidence interval for the estimate. Although the accuracy of the estimates are heterogeneous across countries for the three noise floors, a simple linear regression yields an adjusted  $R^2$  of 0.77 for the baseline, 0.76 for the  $+25\%$ , and 0.78 for the  $-25\%$ . Therefore, the overall fit does not vary substantially, and it is a country-specific question which noise floor minimizes the discrepancy with the World Bank reported electrification levels. In our analysis, for consistency purposes we keep the noise floor constant at the baseline. In advanced regional or country-level analysis, it is recommended to carry out ad-hoc sensitivity analysis on the parameter, by modifying the open-access scientific computing framework.

Supplementary Fig. 3b compares the validation obtained with LandScan and WorldPop data. In this case some countries' reported electrification levels are better estimated when the WorldPop population data are used despite a generally lower fit overall. However, the LandScan results in a *ceteris paribus* significantly more accurate prediction of national electrification level than the WorldPop. A simple linear regression produces a  $R^2$  of 0.77 for the LandScan and of 0.63 for the WorldPop data. Therefore, results underpin the choice of using LandScan as the default gridded population.

## Usage Notes

The dataset is in netCDF (Network Common Data Form) format, which is the standard default for gridded time-series data in the earth system modeling community. For the population without access dataset, the netCDF file features five layers, which describe years between 2014 and 2018. Each pixel contains the local count of the people estimated to live without access to electricity. For the tiers of consumption, the netCDF only contains data in pixels where people have access in year 2018. Pixel values range over integers between 1 and 4, where each refers to the tier of consumption estimated locally.

The data have a native resolution of 1-km, which allows for assessments at different scales. It can be analyzed without further edits, but also aggregated at the province level. Potential applications relate to the design of scenarios and the projection of future pathways of energy access and consumption in the sub-Saharan African continent. This is of particular relevance if the dataset is combined with gridded information on income and population to produce future scenarios of energy access and demand. As discussed in<sup>68</sup>, there is a need for improved data and representation of future energy demand in developing countries. The elaboration of such information in Integrated Assessment Models could substantially improve the understanding of the drivers, impacts, synergies and trade-offs between different development goals.

The data also have potential to be used in inequality and vulnerability assessments and scenarios. Including local-level information about electricity supply and quality offers new possibilities for assessing and correlating adaptation options at a finer spatial scale across different spheres, such as health, productivity, agricultural outcomes, and migration. Further uses include the improvement of the input data for long-term infrastructure optimization models at different spatial scales, or the assessment of national/village-level policies and the effects over the 2014–2018 period. At the same time, new NTL data are published on a monthly basis. This allows updating the dataset presented here and keeping track of progress with little delay as new data become released.

Finally, the methods and output datasets also exhibit significant potential for being used and improved through applications of machine learning and neural-network based methods for identifying patterns, clusters and dynamics. For instance, to date the Google Earth Engine platform<sup>27</sup> allows for the implementation of up to ten different machine learning algorithms on geospatial data. Using those methods to detect electricity-related infrastructure<sup>69</sup>, such as power lines (as recently shown in<sup>70</sup>), power plants, and substations, may allow further improving electricity access and consumption estimates. Furthermore, the forthcoming VIIRS Black Marble VNP46 new NTL data product<sup>71</sup> is likely to allow even a greater precision in the estimation of electrification levels and proxying of consumption levels thanks to less noisiness and more precise calibration of the data.

## Code Availability

The latest version of the code and guide for the dataset reproduction are hosted at [https://github.com/giacfalk/Electrification\\_SSA\\_data](https://github.com/giacfalk/Electrification_SSA_data).

The code is available in two files: a JavaScript file to be imported into Google Earth Engine (step 1), and an R script to be run after having run the JavaScript file in Earth Engine (step 2). In order to run the Earth Engine script, it is necessary to create a Google account and register for Earth Engine via <https://signup.earthengine.google.com>.

All the input data required for the analysis are openly available, with the exception of the LandScan gridded population dataset. To obtain access to population data, it is possible to either apply for an account to use the LandScan data at <https://landscan.ornl.gov/user/apply>, or use the WorldPop data, which are directly accessible through running the script in the repository.

Analysis was run in R 3.4.4 using packages raster (2.8–19), ncd4 (1.16), RnetCDF (1.9–1), googledrive (0.1.3), data.table (1.11.8), dplyr (0.7.8), plyr (1.8.4), ggplot2 (3.0.0), wbstats (0.2), ggrepel (0.8.0), sf (0.7–1), cowplot (0.9.3), reshape2 (1.4.3), rworldmap (1.3–6), rgdal (1.3–6), tidyr (0.8.3) and RColorBrewer (1.1–2). Furthermore, the checkpoint package (0.5.0) is used to provide a snapshot of and retrieve the required version of the packages and their dependencies.

The average processing time for the Google Earth Engine Javascript code (to run and export assets to Google Drive) is about 10 minutes, while the R script takes around 10 minutes to install the required packages and dependencies, and an additional 10 minutes to generate the netCDF files and the figures and save them (although this depends on the specifications of the hardware used). Note that the running time of the R script is also sensitive to the internet connection speed, since the script retrieves the assets from Google Drive.

The processing of satellite data and geospatial operations are performed at a 1000 m scale under a EPSG 4326 coordinate reference system.

## References

1. UN DESA. The Sustainable Development Goals Report 2018 *United Nations*. (2018).
2. McCollum, D. L. *et al.* Connecting the sustainable development goals by their energy inter-linkages. *Environmental Research Letters* **13**, 033006 (2018).
3. Riva, F., Ahlborg, H., Hartvigsson, E., Pachauri, S. & Colombo, E. Electricity access and rural development: review of complex socio-economic dynamics and causal diagrams for more appropriate energy modelling. *Energy for Sustainable Development* **43**, 203–223 (2018).
4. Nerini, F. F. *et al.* Mapping synergies and trade-offs between energy and the sustainable development goals. *Nature Energy* **3**, 10 (2018).
5. Gertler, P. J., Lee, K. & Mobarak, A. M. Electricity reliability and economic development in cities: A microeconomic perspective. *EEG State-of-Knowledge Paper Series, Oxford Policy Management Center for Effective Global Action*. (2017).
6. Bonan, J., Pareglio, S. & Tavoni, M. Access to modern energy: a review of barriers, drivers and impacts. *Environment and Development Economics* **22**, 491–516 (2017).
7. Angelou, N. *et al.* Global tracking framework. (The World Bank, 2013).
8. World Bank. State of Electricity Access Report 2017. (World Bank, Washington, DC, 2017).
9. Aklin, M., Harish, S. P. & Urpelainen, J. A global analysis of progress in household electrification. *Energy Policy* **122**, 421–428 (2018).
10. ICF. The DHS Program STATcompiler. Funded by USAID. (2012).
11. IEA. World Energy Outlook 2018. IEA, Paris (2018).
12. Eberhard, A., Gratwick, K., Morella, E. & Antmann, P. Independent power projects in sub-Saharan Africa: Investment trends and policy lessons. *Energy Policy* **108**, 390–424 (2017).
13. Ilas, A., Ralon, P., Rodriguez, A. & Taylor, M. Renewable power generation costs in 2017. *International Renewable Energy Agency (IRENA): Abu Dhabi, UAE* (2018).
14. Trotter, P. A., McManus, M. C. & Maconachie, R. Electricity planning and implementation in sub-Saharan Africa: A systematic review. *Renewable and Sustainable Energy Reviews* **74**, 1189–1209 (2017).
15. Mentis, D. *et al.* Lighting the world: the first application of an open source, spatial electrification tool (OnSSET) on sub-Saharan Africa. *Environmental Research Letters* **12**, 085003 (2017).
16. Korkovelos, A., Khavari, B., Sahlberg, A., Howells, M. & Arderne, C. The role of open access data in geospatial electrification planning and the achievement of SDG7. an OnSSET-based case study for Malawi. *Energies* **12**, 1395 (2019).
17. Scott, A. & Seth, P. The political economy of electricity distribution in developing countries. (Overseas Development Institute, 2013).
18. Sovacool, B. K. The political economy of energy poverty: A review of key challenges. *Energy for Sustainable Development* **16**, 272–282 (2012).
19. Trotter, P. A. Rural electrification, electrification inequality and democratic institutions in sub-Saharan Africa. *Energy for Sustainable Development* **34**, 111–129 (2016).
20. Cader, C., Pelz, S., Radu, A. & Blechinger, P. Overcoming data scarcity for energy access planning with open data - The example of Tanzania. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4/W8*, 23–26 (2018).
21. Dagnachew, A. G. *et al.* The role of decentralized systems in providing universal electricity access in Sub-Saharan Africa – A model-based approach. *Energy* **139**, 184–195 (2017).
22. Dagnachew, A. G., Lucas, P. L., Hof, A. F. & van Vuuren, D. P. Trade-offs and synergies between universal electricity access and climate change mitigation in sub-Saharan Africa. *Energy Policy* **114**, 355–366 (2018).
23. Mills, S., Weiss, & Liang, C. VIIRS day/night band (DNB) stray light characterization and correction. Proc. SPIE 8866, Earth Observing Systems XVIII, 88661P (23 September 2013).
24. Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C. & Worley, B. A. Landsat: a global population database for estimating populations at risk. *Photogrammetric engineering and remote sensing* **66**, 849–857 (2000).
25. Tatem, A. J. WorldPop, open data for spatial demography. *Scientific Data* **4**, 170004 (2017).
26. Friedl, M. & Sulla-Menashe, D. MCD12Q1 MODIS/Terra + Aqua Land Cover Type Yearly L3 Global 500 m SIN Grid V006 [Data set]. *NASA EOSDIS Land Processes DAAC* **10**, <https://doi.org/10.5067/MODIS/MCD12Q1.006> (2015).
27. Gorelick, N. *et al.* Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **202**, 18–27 (2017).
28. Doll, C. N. & Pachauri, S. Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery. *Energy Policy* **38**, 5661–5670 (2010).
29. Elvidge, C. D. *et al.* Who's in the dark – satellite based estimates of electrification rates. In Yang, X. (ed.) *Urban Remote Sensing*, 211–224 (2011).
30. Min, B., Gaba, K. M., Sarr, O. F. & Agalassou, A. Detection of rural electrification in Africa using DMSP-OLS night lights imagery. *International Journal of Remote Sensing* **34**, 8118–8141 (2013).
31. Dugoua, E., Kennedy, R. & Urpelainen, J. Satellite data for the social sciences: measuring rural electrification with night-time lights. *International Journal of Remote Sensing* **39**, 2690–2701 (2018).
32. Coscieme, L. *et al.* A thermodynamic geography: night-time satellite imagery as a proxy measure of energy. *Ambio* **43**, 969–979 (2014).
33. He, C., Ma, Q., Li, T., Yang, Y. & Liu, Z. Spatiotemporal dynamics of electric power consumption in Chinese Mainland from 1995 to 2008 modeled using DMSP/OLS stable nighttime lights data. *Journal of Geographical Sciences* **22**, 125–136 (2012).
34. Townsend, A. C. & Bruce, D. A. The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution. *International Journal of Remote Sensing* **31**, 4459–4480 (2010).
35. Fehrer, D. & Krarti, M. Spatial distribution of building energy use in the United States through satellite imagery of the earth at night. *Building and Environment* **142**, 252–264 (2018).
36. Shi, K., Yu, B., Huang, C., Wu, J. & Sun, X. Exploring spatiotemporal patterns of electric power consumption in countries along the belt and road. *Energy* **150**, 847–859 (2018).
37. Baldwin, H. B., Klug, M., Tapracharoen, K. & Visudchindaporn, C. Utilizing suomi NPP's day-night band to assess energy consumption in rural and urban areas as an input for poverty analysis. In *AGU Fall Meeting Abstracts* (2017).
38. Burke, A. *et al.* Verification and enhancement of VIIRS day-night band power outage detection product. In *AGU Fall Meeting Abstracts* (2017).
39. Wang, Z. *et al.* Monitoring disaster-related power outages using nasa Black Marble nighttime light product. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1853–1856 (2018).
40. Mann, M. L., Melaas, E. K. & Malik, A. Using VIIRS day/night band to measure electricity supply reliability: preliminary results from maharashtra, india. *Remote Sensing* **8**, 711 (2016).
41. Bharti, N., Djibo, A., Tatem, A. J., Grenfell, B. T. & Ferrari, M. J. Measuring populations to improve vaccination coverage. *Scientific Reports* **6**, 34541 (2016).
42. Bharti, N. & Tatem, A. J. Fluctuations in anthropogenic nighttime lights from satellite imagery for five cities in Niger and Nigeria. *Scientific Data* **5**, 180256 (2018).
43. Bickenbach, F., Bode, E., Nunnenkamp, P. & Söder, M. Night lights and regional GDP. *Review of World Economics* **152**, 425–447 (2016).
44. Wang, X. *et al.* Estimation and mapping of sub-national GDP in Uganda Using NPP-VIIRS imagery. *Remote Sensing* **11**, 163 (2019).
45. Mveyange, A. Night lights and regional income inequality in Africa. *Tech. Rep.*, WIDER Working Paper (2015).
46. Hijmans, R., Garcia, N. & Weiszorek, J. GADM: database of global administrative areas. *Version 3.6* (2018).



47. Potts, D. Urban data and definitions in sub-Saharan Africa: Mismatches between the pace of urbanisation and employment and livelihood change. *Urban Studies* **55**, 965–986 (2018).
48. Borel-Saladin, J. Where to draw the line: Data problems and other difficulties estimating urbanisation in Africa. *Development Southern Africa* **34**, 137–150 (2017).
49. Arellano, B. & Roca, J. Defining urban and rural areas: a new approach. In *Remote Sensing Technologies and Applications in Urban Environments II*, vol. 10431, 104310E (International Society for Optics and Photonics, 2017).
50. Dijkstra, L. & Poelman, H. A harmonised definition of cities and rural areas: the new degree of urbanisation. *Regional Working Paper 2014, European Commission* **1**, 2014 (2014).
51. Burlig, F. & Preonas, L. Out of the darkness and into the light? development effects of rural electrification. *Energy Institute at Haas WP* **268** (2016).
52. Román, M. O. & Stokes, E. C. Holidays in lights: Tracking cultural patterns in demand for energy services. *Earth's Future* **3**, 182–205 (2015).
53. Levin, N. & Zhang, Q. A global analysis of factors controlling VIIRS nighttime light levels from densely populated areas. *Remote Sensing of Environment* **190**, 366–382 (2017).
54. Shi, K. *et al.* Modeling and mapping total freight traffic in china using NPP-VIIRS nighttime light composite data. *GIScience & Remote Sensing* **52**, 274–289 (2015).
55. Wang, R., Wan, B., Guo, Q., Hu, M. & Zhou, S. Mapping regional urban extent using NPP-VIIRS DNB and MODIS NDVI data. *Remote Sensing* **9**, 862 (2017).
56. Miller, S. D. *et al.* The dark side of hurricane matthew: Unique perspectives from the VIIRS day/night band. *Bulletin of the American Meteorological Society* (2018).
57. Falchetta, G. & Noussan, M. Interannual variation in night-time light radiance predicts changes in national electricity consumption conditional on income-level and region. *Energies* **12**, 456 (2019).
58. Xiao, H. *et al.* Spatio-temporal simulation of energy consumption in china's provinces based on satellite night-time light data. *Applied Energy* **231**, 1070–1078 (2018).
59. Bhatia, M. & Angelou, N. *Beyond connections: energy access redefined.* (World Bank, 2015).
60. Falchetta, G., Pachauri, S., Parkinson, S. & Byers, E. A high-resolution gridded dataset to assess electrification in sub-Saharan Africa. *Mendeley Data*. <https://doi.org/10.17632/kn4636mtvg> (2019).
61. Dalberg Advisors & Lighting Global Off-grid solar market trends report 2018. *International Finance Corporation*. (2018).
62. De Bercegol, R. & Monstadt, J. The Kenya slum electrification program. local politics of electricity networks in Kibera. *Energy Research & Social Science* **41**, 249–258 (2018).
63. Yakubu, O., Babu, N. & Adjei, O. Electricity theft: Analysis of the underlying contributory factors in Ghana. *Energy Policy* **123**, 611–618 (2018).
64. ESCOM. Electricity charges with effect from 4th April, 2014, <http://www.escom.mw/4th-april-2015-tariffs.php> (2014).
65. Premium Times Nigeria Nigerian government hikes electricity tariffs; abolishes fixed charges, <https://www.premiumtimesng.com/business/195537-nigerian-government-hikes-electricity-tariffs-abolishes-fixed-charges.html>.
66. UMEME UMEME Power Tariff Structure, <https://www.umeme.co.ug/articles/TariffAdNV.pdf>.
67. Kojima, M., Bacon, R., & Trimble, C. Political economy of power sector subsidies: a review with reference to sub-Saharan Africa *World Bank Group*.
68. Riva, F., Tognollo, A., Gardumi, F. & Colombo, E. Long-term energy planning and demand forecast in remote areas of developing countries: Classification of case studies and insights from a modelling perspective. *Energy Strategy Reviews* **20**, 71–89 (2018).
69. Oshri, B. *et al.* Infrastructure quality assessment in Africa using satellite imagery and deep learning. Preprint at, <https://arxiv.org/abs/1806.00894> (2018).
70. Gershenson, D., Roher, B. & Lerner, A. Predictive model for accurate electrical grid mapping, <https://code.fb.com/connectivity/electrical-grid-mapping/> (2019).
71. Román, M. O. *et al.* Nasa's Black Marble nighttime lights product suite. *Remote Sensing of Environment* **210**, 113–143 (2018).

## Acknowledgements

Financial support from the MIUR (Italian Ministry of University and Research) through the Catholic University of Milan, from Fondazione Eni Enrico Mattei, and from the International Institute for Applied Systems Analysis is gratefully acknowledged. The authors would like to thank Olha Danylo for her precious help in developing the dynamic visualisation interface, and the entire Energy Program of the International Institute for Applied Systems Analysis.

## Author Contributions

G.F. conceptualised the framework, processed the data, and generated the figures; Sh.P., Si.P., and E.B. provided valuable input to the concept, analysis, and presentation of the results; G.F., Sh.P., Si.P., and E.B. wrote the paper.

## Additional Information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41597-019-0122-6>.

**Competing Interests:** The authors declare no competing interests.

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



**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2019