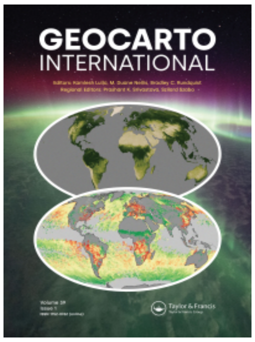

Geocarto International

ISSN: 1010-6049 (Print) 1752-0762 (Online) Journal homepage: www.tandfonline.com/journals/tgei20

Spatio-temporal estimation of electricity consumption in Bolivian municipalities using nighttime lights

Oana M. Garbasevschi , Andrea Sofia Garcia de León , Lykke E. Andersen , Guillermo Guzmán Prudencio , Michael Wurm & Hannes Taubenböck

To cite this article: Oana M. Garbasevschi , Andrea Sofia Garcia de León , Lykke E. Andersen , Guillermo Guzmán Prudencio , Michael Wurm & Hannes Taubenböck (2026) Spatio-temporal estimation of electricity consumption in Bolivian municipalities using nighttime lights, Geocarto International, 41:1, 2657705, DOI: 10.1080/10106049.2026.2657705

To link to this article: https://doi.org/10.1080/10106049.2026.2657705

GEOCARTO INTERNATIONAL
2026, VOL. 41, NO. 1, 2657705
https://doi.org/10.1080/10106049.2026.2657705

RESEARCH ARTICLE

OPEN ACCESS

Spatio-temporal estimation of electricity consumption in Bolivian municipalities using nighttime lights

Oana M. Garbasevschi[a,b], Andrea Sofia Garcia de León[c], Lykke E. Andersen[d], Guillermo Guzmán Prudencio[e,f], Michael Wurm[b] and Hannes Taubenböck[b,c]

[a]ifo Institute – Leibniz Institute for Economic Research at the University of Munich, Germany; [b]German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Germany; [c]Institute of Geography and Geology, Julius-Maximilians University Wurzburg (JMU), Germany; [d]Sustainable Development Solutions Network (SDSN Bolivia), Universidad Privada Boliviana (UPB), Obrajes, La Paz, Bolivia; [e]Postgraduate Center in Development Sciences, Major University of San Andres (CIDES-UMSA), La Paz, Bolivia; [f]Universidad Privada Boliviana (UPB), Obrajes, La Paz, Bolivia

ABSTRACT

Few research studies have focused on the nature of the relationship between nighttime lights and electricity consumption at subnational levels in South America, a region with heterogeneous geography and urbanization levels and complex socioeconomic dynamics. This study shows that it is possible to estimate, for Bolivia, a wide range of indicators of electricity consumption at the municipality level and two temporal scales using features derived from nighttime lights and other spatial data sources, in combination with readily available municipality characteristics. The prediction errors for annual electricity consumption range between 13% MAPE for average residential consumption and 59% MAPE for average commercial consumption. Similar accuracies are obtained when predicting monthly values. For both annual and monthly electricity consumption, we highlight the variation in estimation accuracy for various municipality subsets and show that prediction can be significantly improved when selecting municipalities based on population size, energy poverty, or levels of sustainable development.

ARTICLE HISTORY
Received 23 November 2025
Accepted 6 April 2026

KEYWORDS
Electricity forecast; machine learning; remote sensing; sustainable development; time series analysis

1. Introduction

Electricity plays an essential role as an enhancer of social and economic development. In Latin America and the Caribbean, fourteen million people still lacked access to electricity in 2018, and most of them were living in rural and isolated areas, where electricity was only one of the missing basic services. It is widely thought that reducing energy poverty and ensuring energy security lie at the heart of overcoming major challenges such as poverty, climate change, food insecurity, or inequalities in income and health (Andersen and Branisa, 2023; Santillán et al. 2020; Siksnelyte-Butkiene et al. 2021).

In Bolivia, 93% of the population had access to electricity in 2018. The population with no access to electricity or other basic services lives in small and isolated rural communities, with low incomes derived from agriculture and livestock (Eras-Almeida et al. 2019; Fernandez Fuentes et al. 2021). In line with the sustainable development goals, the aim of the Bolivian government is to achieve 100% electrification by 2025, the target year of the current energy policy plan (Fernandez Fuentes et al. 2021). Covering electrification needs is only one side of the issue, with growing pressure on existing infrastructure being another. Energy demand is predicted to double by 2030, following the pattern of population growth and increased use of electricity in economic sectors such as industry, transport or heating (Cheng et al. 2022). Furthermore, the energy sector is the second largest contributor to the total GHG emissions in Bolivia, which leads to the development of initiatives to increase sustainability in production and consumption, with national-level plans that aim at increasing the proportion of renewable electricity production to 75% and electrifying 10% of the public transport sector by 2030 (Fernandez Vazquez et al. 2024). Finally, energy supply systems in developing countries function under uncertain future configurations (Moksnes and Rozenberg, 2019). Lower

CONTACT Michael Wurm michael.wurm@dlr.de

rainfall could result in a diminished capacity to generate hydroelectricity (Moksnes and Rozenberg, 2019) while natural resource depletion could lead to decreased capacity for natural gas extraction (Fernandez Vazquez et al. 2024), an important resource for energy generation for export in Bolivia.

Accurate and up-to-date information on the existing electricity network and consumption could support multiple goals: planning for universal electricity access, managing demand, sustainable development of energy sources and designing region-appropriate electricity networks. These goals are especially relevant for countries like Bolivia with large variations in population density and where many rural areas are difficult to reach. Spatial and temporal electricity consumption data also harbour great potential to act as proxy indicators of socioeconomic conditions and inform the creation and monitoring of public welfare policies. This type of data, however, is seldom freely available at the most critical spatial scales – local or regional – or if available, it is corrupted by issues of time lag and spatial coarseness (Ivan et al. 2020; Lin and Raza 2021; Wang and Lu 2021).

New data sources and methods have been increasingly explored over the past two decades for the estimation of missing socioeconomic data. One of the most important data sources is satellite-borne Earth Observation (EO) products and geospatial data. A prime example of such data is nighttime lights (NTL) images, which are among the most widely used EO products for approximating the spatial distribution of the portion of electricity consumption (EC) that produces visible light (Elvidge et al. 2021). Furthermore, NTL images are empirically validated as robust resources for generating geocoded socioeconomic indicators in areas where data are missing or outdated (Kuffer et al. 2018; Wurm et al. 2019; Andreano and Simona, 2021; Georg et al. 2023).

Different sensors record NTL data, with various temporal frequencies and spatial resolutions. DMSP NTL data have the advantage of extended temporal coverage from 1992 to 2013, which enables long-term time series analysis, while VIIRS DNB data were first collected in 2012, at a higher spatial resolution and frequency than the DMSP sensor. The Black Marble Product Suite (Román et al. 2018) is a recent global NTL product, calibrated and scientifically validated, with similar spatial and temporal granularity as the VIIRS source. Newer NTL products include Luojia 1-01 data (Lu et al. 2025) and SDGSAT-1, which provides multi-band data with high spatiotemporal resolution (Chen et al. 2025). Since the relationship between NTL and EC can be affected by external parameters such as the economic activities performed in the area and environmental conditions (Li et al. 2019), accounting for the variation expressed by local land cover and land use conditions helps address spatial heterogeneity and improve estimation accuracy. For this reason, NTL data is used together with land use/land cover classification data and derived indicators such as the Enhanced Vegetation Index (EVI), total urban area, patch density of rural settlement (Xie and Weng 2016; Wang and Lu 2021), or urban functional zoning (Lu et al. 2025).

As a proxy for electricity consumption, the use of nighttime lights supports the evaluation of policies for rural electrification (Chindarkar and Goyal 2023), the identification and estimation of energy poverty (Min et al. 2024; Wang et al. 2021), assessing electrification needs (Falchetta and Pachauri, 2019; Alabi et al. 2022; Chakraborty et al. 2025), or estimating CO2 emissions (Qin et al. 2025). While numerous studies have analyzed the relationship between NTL and EC, most studies have focused on country-wide analysis, with only a few focusing on the subnational scale (Fehrer and Krarti 2018; Jasiński 2019; Wang and Lu 2021). Even fewer recent studies have addressed the fine-scale disaggregation of electricity consumption and other inherently related information, such as carbon emissions, using nighttime lights data (Cao et al. 2025; Qin et al. 2025).

To the best of our knowledge, the relation between nighttime lights and electricity consumption has not been analyzed at the subnational level for Latin American countries. The region is an especially interesting study area, as it consists of low- and middle-income countries with distinctive land cover and urbanization patterns that can influence both NTL accuracy and EC patterns (Min et al. 2013; Wang and Lu 2021). Another focus of research that only a few studies have adopted is the investigation of different types of electricity consumption. In addition to their intrinsic value as input for diverse energy planning scenarios, these electricity consumption indicators act as proxies for socioeconomic life conditions, an important dimension of societal heterogeneity in Bolivia. Residential energy consumption can be linked to household income and consumption power, with wealthier households owning and using a larger set of electric appliances (Solis et al. 2022). Commercial energy consumption is a direct output of income generation activities and can indicate local potential for the development of business and commercial activities

(Sanchez Solis et al. 2023). The number of streetlights is an indicator of the quality of public service, of the level of infrastructure development, and more generally, of the quality of life.

With this study, we address the literature gap and contribute to existing research in several ways. The key goal of the analysis is to investigate the reliability of NTL data to estimate multiple indicators of electricity consumption at the municipal level in Bolivia. We first accomplish this by building a comprehensive set of variables that describe the spatial variability of NTL radiance by combining NTL data with land cover, building, and human settlement locations. This allows us to quantify the diverse geography and density of population and economic activity specific to the country. Second, we estimate a series of models that predict monthly and annual electricity consumption in the period between 2012 and 2016 using time-matched NTL-derived variables. The combination of spatially and temporally disaggregated data is infrequently encountered, owing to data availability issues, in studies focusing on electricity consumption in Latin American countries. Furthermore, our results provide evidence of the relevance of NTL data to act as a proxy for various types of electricity consumption – electricity used for residential purposes, for commerce, and for public lighting – both in their aggregated form and as average consumption per user. Finally, the study examines the variation in the accuracy of electricity consumption estimates when considering various subsets of municipalities. Our data-driven machine learning approach leads to the identification of socioeconomic municipality characteristics that help define optimal subsets for both electricity nowcast and forecast tasks.

2. Materials and methods

2.1. Study area

The Plurinational State of Bolivia is a country in the western-central region of South America. The geography of the country is defined by diverse climate and ecological regions: the Andean region in the West, the sub-Andean region, characterized by valleys and temperate climes, and the lowlands, covering areas ranging from the Amazon rainforest in the North to the dry Chaco Forest in the South-East (INE 2020). The country consists of nine departments–the primary administrative division–subdivided into 339 municipalities. In 2020, the estimated population reached 11.6 million inhabitants, with 32.7% living in rural areas (CEPAL 2022). The urban population is growing rapidly, but is concentrated in three metropolitan areas: Santa Cruz, La Paz-El Alto and Cochabamba. According to CEPAL (2022), 31.1% of the Bolivian population were living in poverty in 2019, and 12.2% in extreme poverty.

The primary source of energy in Bolivia is fossil fuels, with natural gas, gasoline, and diesel covering three-quarters of the energy needs. Another quarter of energy consumption is covered from biomass and electricity (Fernandez Vazquez et al. 2024). Out of the total installed electricity generation capacity, in 2021, thermo-electric power plants contributed the highest share (71%), followed by 20.4% from hydropower plants, and 8.6% from other renewable sources such as biomass, solar and wind (Jimenez Zabalaga et al. 2025). Following the goal of achieving universal access to electricity by 2025, Bolivia invested heavily in fossil fuel energy power plants until 2017 (Navia et al. 2022). Nevertheless, various efforts have been initiated that aim at reducing carbon emissions and, more generally, the use of conventional energy sources (Fernandez Vazquez et al. 2024). The addition of renewable energy sources to the electricity distribution network has been slower than in other countries, such as Uruguay, Brazil, Argentina, or Chile (Navia et al. 2022). Current expansion strategies include the future addition of mostly hydroelectric plants and, to a lesser measure, of wind, solar or geothermal plants, even though solar energy in particular could represent a suitable primary energy source for a large part of the country, since, owing to its altitude and dry climate, the country contains regions with the highest level of solar ratio worldwide (Navia et al. 2022). Long-term development plans are necessary to fulfil the high potential of a cost-effective mix of energy sources and mitigate the negative effects that the dependency on fossil fuels entails (Fernandez Vazquez et al. 2024). Updated energy demand data and forecasting are essential in developing such strategies (Jimenez Zabalaga et al. 2025).

2.2. Nighttime lights data

This study used moonlight-adjusted NTL composites and associated quality flags from the Black Marble Product Suite (Román et al. 2018). Black Marble data have been successfully used to detect power outages

(Wang et al. 2018) and for electricity restoration (Román et al. 2019), showing good potential to proxy electricity consumption. Figure 1 displays the annual NTL radiance in Bolivia in 2016; the highlighted areas are the cities of Santa Cruz de la Sierra, La Paz, and Tarija.

Black Marble is an NTL product from NASA derived from the VIIRS Day-Night band and accessible as open-source data in the LAADS DAAC platform from the VNP46A collection (NASA 2021). It consists of Top-of-Atmosphere (TOA) DNB radiance (nWcm-2sr-1) with a spatial resolution of 15 arc-seconds, available from January 2012 to the present day. For this analysis, we used the VNP46A3 and VNP46A4 products, which provide monthly and annual all-angles nighttime radiance composites for 2012–2016. The all-angle composites are particularly suitable for this study because nighttime radiance is influenced by both sensor view angle and land-use characteristics. These products were generated from daily lunar and bidirectional reflectance distribution function (BRDF)-corrected radiance observations across three view angles and two snow conditions, enabling a comprehensive overview of multiple illumination sources, such as indoor lights and streetlights (Wang et al. 2022).

Before data distribution, the Black Marble preprocessing algorithm reduces the influence of external contamination by detecting clouds, correcting for atmospheric effects, and removing stray light, lunar illumination, background noise, and ephemeral light sources such as fires and aurora (Román et al. 2018; Wang et al. 2022; Wang et al. 2021). To further reduce surface reflectance effects, we selected snow-free composites to minimize the albedo's impact in the Andean regions. We addressed cloud contamination by applying the quality masks defined by the NASA-provided quality assurance flags. Only pixels flagged as high-quality and cloud-free in the VNP46A2 products were used for analysis (Wang et al. 2022). No

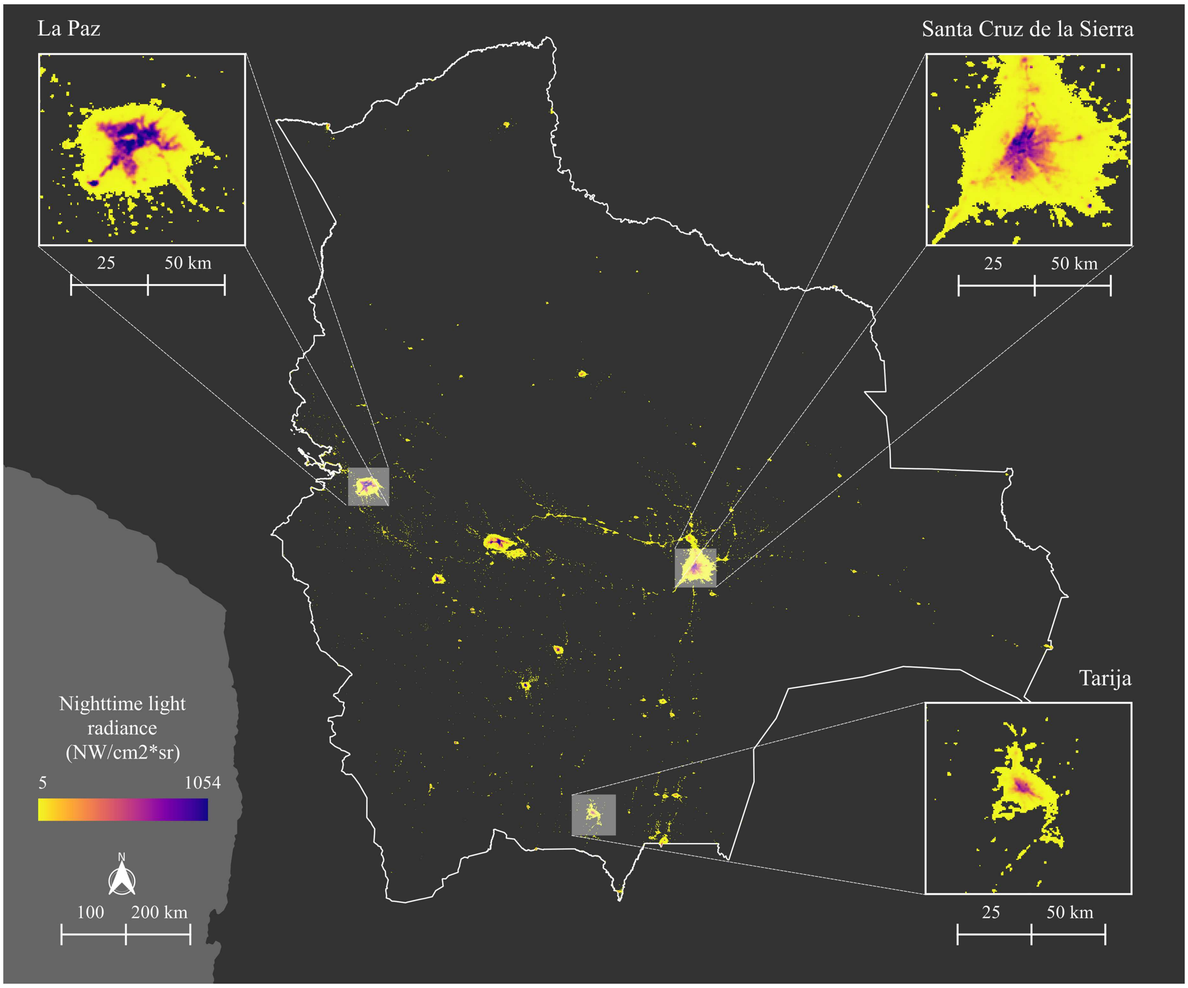

Figure 1. Distribution of NTL radiance in Bolivia, in 2016, with three departmental capitals highlighted.

additional cloud or atmospheric corrections were applied beyond those implemented in the Black Marble processing chain.

2.3. Spatial data

We explored NTL data in conjunction with other relevant open-source remote sensing data, which served several purposes. First, we described the topography of municipalities using summaries of spatial attributes. Second, we delineated areas of human settlements and economic activity. Finally, we investigated the differences in NTL radiance across different types of settlements and land uses and assessed their impact on EC estimation. These auxiliary data include elevation (Takaku et al. 2014), land cover classification (Buchhorn et al. 2020), delineation of built-up areas (Esch et al. 2012; Deutsches Zentrum für Luft- und Raumfahrt (DLR) 2016; JRC 2019), and building footprints and points of interest (OSM 2023). Table 1 summarizes the spatial data, including their resolution, year of collection, and sources.

Bolivia is a country with diverse geography, encompassing mountain ranges, high plateaus, and lowland regions, including flatlands, hills and valleys. We summarized the elevation features of each municipality using the Digital Surface Model (DSM) generated from sensors onboard the Advanced Land Observing Satellite (ALOS) (Takaku et al. 2014). We further explored the topography of the country using land cover/land use data from the Copernicus Global Land Service (CGLS) (Buchhorn et al. 2020). In addition to the built-up areas identified in CGLS-LC100, three other data sources were used to identify settlement areas: 1) the Global Urban Footprint (GUF) - a world map of built areas at 12m resolution (Esch et al. 2012; Deutsches Zentrum für Luft- und Raumfahrt (DLR) 2016); 2) the built-up area grid (GHS-BUILT), published by the European Joint Research Center (JRC 2019); 3) buildings and points of interest from the open global crowd-sourced dataset, Open Street Map (OSM).

For estimating the most likely maximum extent of human settlements and economic activity between 2012 and 2016 (with the denomination *built* versus *non-built* areas), we have combined the various information on built-up areas with population maps (WorldPop 2020) into a single spatial layer, as illustrated in Figure 2. Before merging, small areas such as individual buildings or points of interest have been expanded by a buffer with a radius equal to the NTL data resolution of 500 m. Due to irregular spatial and temporal resolutions, the layer is time-agnostic and has been used as a fixed mask for all temporal analysis. This layer can be used to filter NTL from sources such as fires, gas flares, or other irregular light activities and can also be used to identify eventual industrial or mining sites.

2.4. Electricity database and socioeconomic data

The electricity consumption database (ELBOL) contains observations of monthly consumption (MWh) for most Bolivian municipalities from January 2012 to December 2016 (Andersen et al. 2023). The electricity database allows the inspection of various uses of electricity: for residences, for commerce/services, for street lights, and for mining activities. Residential and commerce/services consumption account for approximately 20% of all EC, with the largest consumption registered by the transport and industry sectors (Fernandez Vazquez et al. 2024). The descriptive statistics in Table 2 show that for all usage types, the number of users and average EC per user show an increasing trend throughout the study period, a trend confirmed by official statistics for recent years (Cheng et al. 2022). The highest values of

Table 1. Spatial data description and provenance.

Layer type	Year	Resolution	Format	Source name
Elevation	2011	30 m	Raster	ALOS Global Digital Surface Model (DSM) Takaku et al. (2014)
Land cover	2015	100 m	Raster	Copernicus Global Land Service (CGLS) Buchhorn et al. (2020)
Built-up area	2012	30 m	Raster	Global Urban Footprint (GUF) [46, 47]
	2015	250 m	Raster	Global Human Settlement Layer (GHSL) JRC (2019)
	2022*	–	Vector	Building footprints from OSM "OpenStreetMap" (2022)
Points of interest	2022*	–	Vector	POI from OSM "OpenStreetMap" (2022)
Population count	2012–2016	100 m	Raster	WorldPop WorldPop (2020)
Settlements	2012–2022*	–	Vector	Own calculation
Nighttime lights	2012–2016	500 m	Raster	Black Marble Román et al. (2018)

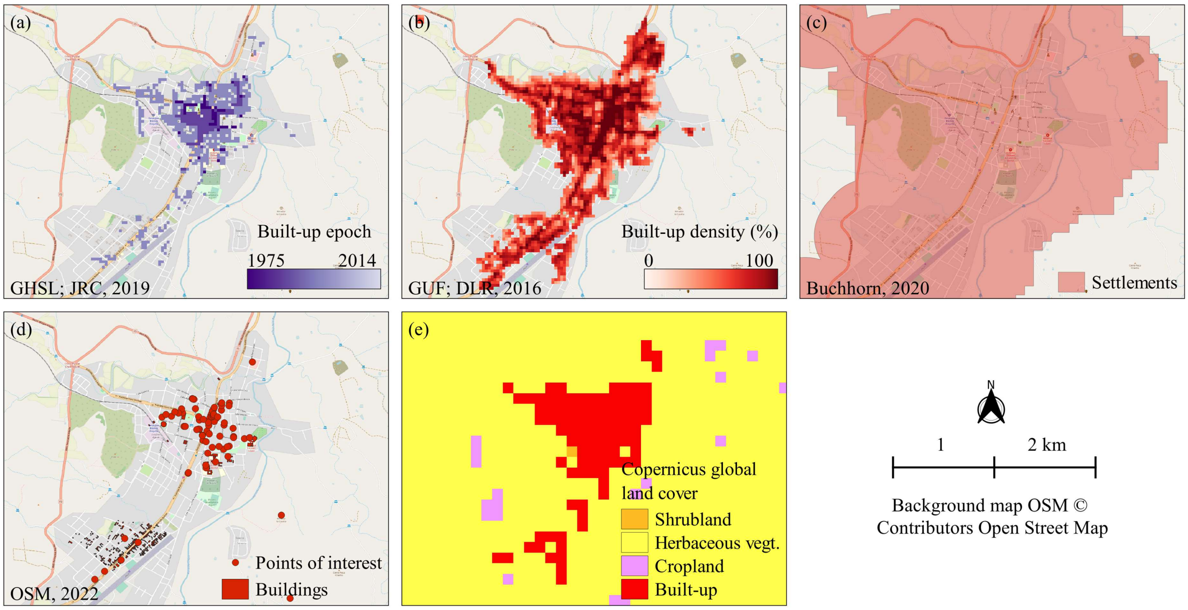

Figure 2. Illustration of the spatial data used for the creation of a human settlement layer (also labelled as *built* area).

Table 2. Summary of selected indicators in the electricity consumption database (ELBOL) (Andersen et al. 2023).

EC indicators	2012	2013	2014	2015	2016
Municipalities with EC	247	293	314	323	329
Residential EC users (millions)	1.89	2.07	2.20	2.41	2.55
Commercial EC users (millions)	0.26	0.28	0.30	0.32	0.35
Residential EC (GWh)	2.43	2.65	2.82	3.03	3.20
Commercial EC (GWh)	3.10	3.25	3.39	3.54	3.67
Municipalities with street lights EC	176	200	235	221	237
Street lights EC (GWh)	0.30	0.32	0.36	0.40	0.42

consumption are concentrated in the largest and most populated cities: Santa Cruz de la Sierra, La Paz, and Cochabamba. Figure 3 illustrates the spatial co-occurrence of low EC and high poverty rates.

Another important source of information on Bolivian municipalities is the Municipal Atlas of Sustainable Development Goals in Bolivia (Andersen et al. 2020). The Atlas compiles a wide range of indices informing on various socioeconomic living conditions at the municipality level. We have extracted for the purpose of analysis in conjunction with electricity consumption a set of municipality indicators, such as population size, degree of urbanization, levels of energy poverty, and an overall index expressing how well the different goals of sustainable development are covered in each municipality.

2.5. *Multivariate analysis*

The main goal of our analysis was to investigate the relationship between electricity consumption and explanatory variables in two temporal dimensions, annually and monthly, using non-linear regression models. Supervised learning models such as random forests (Breiman 2001) have been shown to exhibit high accuracies for diverse types of regression outcomes and to minimize nowcast and forecast errors, in comparison with linear models. Moreover, it has been shown that for small temporal data series, Random Forests perform better than more complex, deep machine learning models. Variable prediction importance was judged based on Shapley values computed with the SHAP approach (Lundberg et al. 2020). To measure model goodness-of-fit, we used the coefficient of determination R^2 and, to evaluate prediction accuracy, the Mean Absolute Percentage Error (MAPE). The model metrics we consistently report throughout the paper are test model errors, where all suitable data for training and prediction are considered.

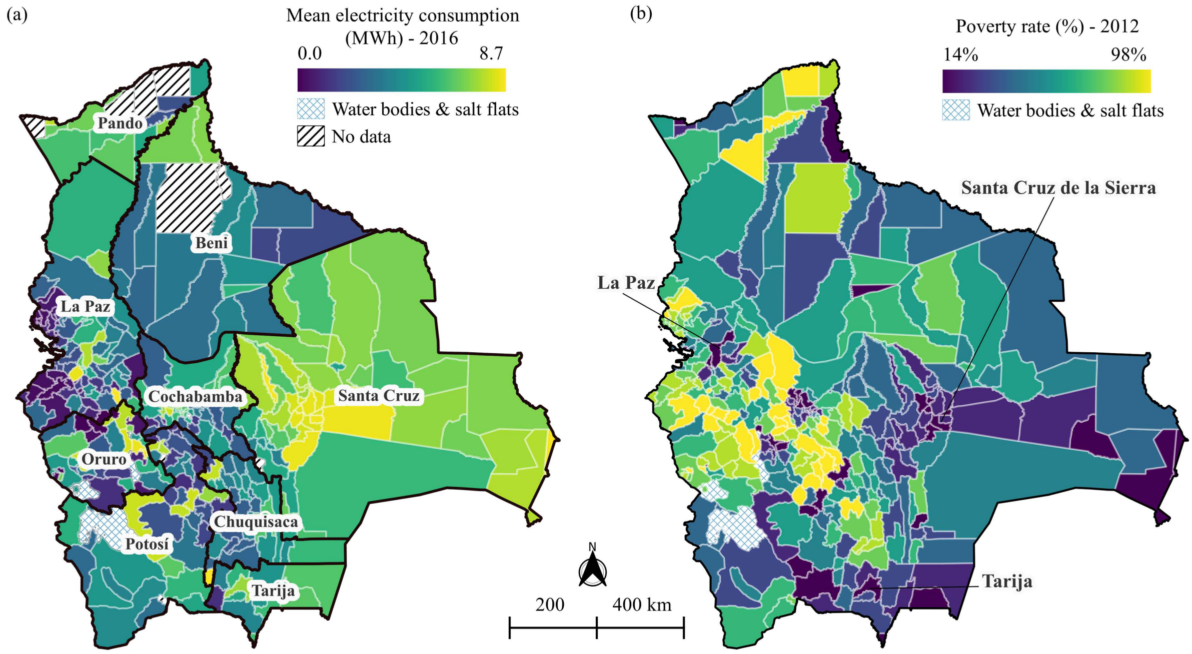

Figure 3. Average electricity consumption per electricity user (a) and poverty levels expressed as the level of unsatisfied basic need (Andersen et al. 2020) (b) in Bolivian municipalities. The nine departments are delineated on the left map, with three of the nine departmental capitals highlighted on the right map.

With a dataset spanning annual and monthly values over 6 years, it is difficult to implement cross-validation scenarios specific to time series data, such as blocked or extended cross-validation. To estimate model performance under different scenarios of data availability for training, we report on an adapted repeated cross-validation technique. We sample independently and without repetition *k-1* folds in the data from the training year and *1* fold from the test year. This implies that it is possible, but not necessary, that the same municipality appears both in the training and test sets. The procedure is repeated 1,000 times for various numbers of folds (2, 5, and 10). We also report prediction intervals using the Python module *forest-confidence-interva*l (Polimis et al. 2017), an implementation of the algorithm developed by Wager et al. (2014). The algorithm estimates the sampling variance of bagged predictors such as random forest trees using a bias-corrected method of the jackknife-after-bootstrap method that relies on the bootstrap replicates, which were used to form the bagged predictions (Wager et al. 2014). Figure 4 depicts a high-level overview of the data layers and main processing steps in the analysis.

We distinguished between different groups of explanatory variables to estimate the ability of readily available indicators to proxy electricity consumption, with or without the use of historical electricity data. The baseline model for comparison is the one where the explanatory variable is the EC of the previous year, or the EC of the same month in the previous year. Equations (1) and (2) summarize the different groups of variables in our predictive models. Predictions based on the model for year *y* are obtained using as training data the model from year *y-1*, or earlier years.

$$EC_y = M + NTL_y + EC_{y-1} \tag{1}$$

$$EC_{m,y} = M + \sum_{k=0}^{3} NTL_{m-k,y} + NTL_{y-1} + EC_{y-1} + EC_{m,y-1} \tag{2}$$

where *EC* are the different types of electricity consumption, *NTL* is the set of variables extracted from NTL data, and *M* is a set of fixed municipality attributes; *y* denotes the year and *m* the month.

The outcomes of interest are all the different types of electricity consumption made available in the electricity database: electricity used in residences (*EC-R*), used for commercial purposes – including companies, neighbourhood stores, city hall services and other (*EC-C*), used for street lightning (*EC-SL*),

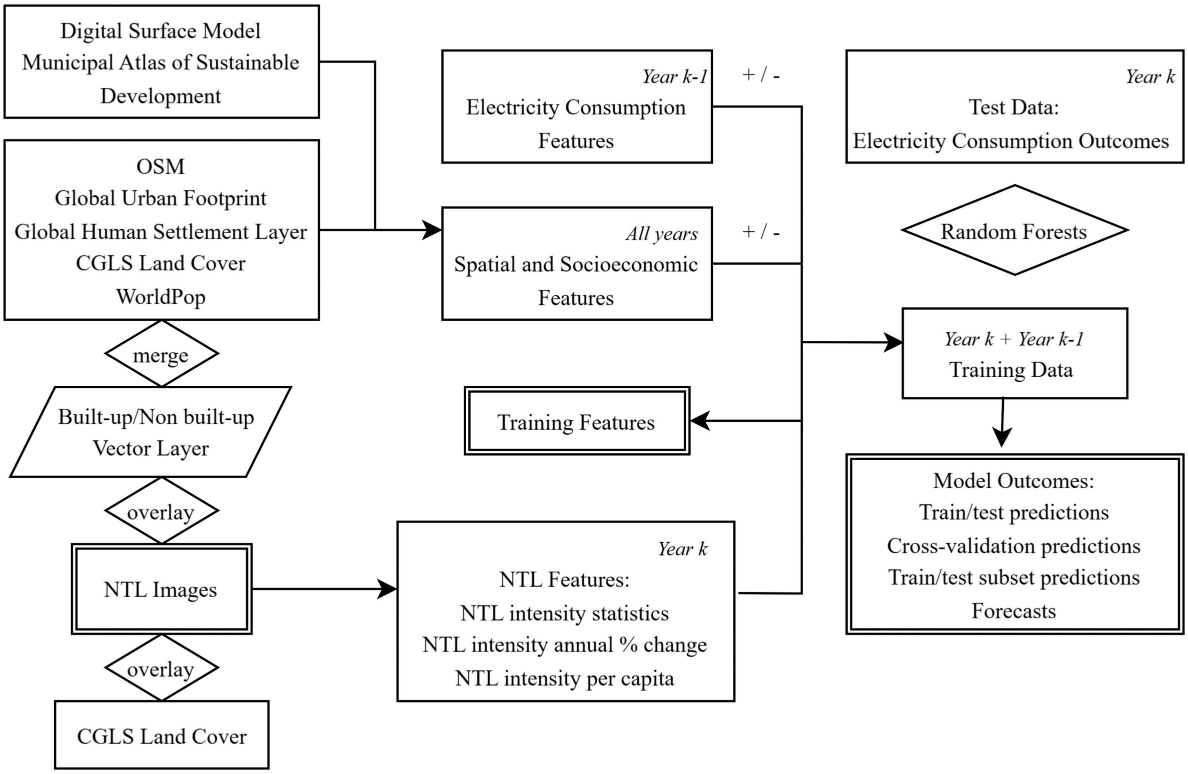

Figure 4. Overview of the most important processing steps in the analysis of annual data. The methodology is representative of the analysis of monthly data as well.

and the total consumption for these three uses (*EC Total*); we also predicted the average residential consumption per user (*Avg(EC-R)*) and the average commercial consumption per user (*Avg(EC-C)*). We excluded EC from mining activities, as it is concentrated in only a few municipalities, with extreme values that distort the municipality's overall EC consumption. All regression outcome variables were log-transformed prior to analysis, as well as the municipality population size.

The most important variables for estimating EC were derived from NTL data. We aggregated NTL luminosity values within specific spatial boundaries: at the level of the municipality, settlement areas (built), non-settlement areas (non-built), as well as within areas defined by CGLS land cover classes, such as built-up, forests, cropland, shrubland, and combinations thereof. The first forms of aggregation are the total number of pixels with non-null luminosity values and their total sum. Then, we computed various descriptive metrics to summarize the distribution of luminosity values within each type of spatial boundary: the mean, standard deviation, coefficient of variation, and skewness. Distribution metrics are computed both when including zeros in the set of values (e.g. *mean*) and when excluding them (e.g. *zmean*). The feature set initially included distribution statistics obtained by trimming the set of pixel values to exclude the top and bottom 5% and 10% outlier values. Model performance tests have proven, however, that these additional features do not improve the prediction capacity.

Other metrics derived from NTL data include the percentage change in luminosity between consecutive years or months, and luminosity per capita, within built and non-built areas. Municipality time-fixed characteristics include both spatial (elevation statistics, extent of land cover, and settlement areas) and socioeconomic (population size and density, degree of urbanization or degree of poverty) attributes. In this manner, we obtained a large set of 280 features, out of which over 230 were derived from the combination of NTL and various spatial data layers. As a feature selection technique, since Random Forests handles large-dimensional data well, solely highly correlated features (Pearson $p = 0.85$) were filtered out. Finally, to eliminate errors caused by outliers, both in the training and prediction steps, we eliminate observations where the outcome variable is either null or within the lowest 5% of non-null values.

Bolivian municipalities are characterized by extreme values in both population size and development levels. The same high variability characterizes electricity consumption. For this reason, to highlight the best scenarios for EC prediction, we tested various subsets of municipalities. We classified municipalities based on one of the following dimensions: population size, degree of urbanization, degree of sustainable development and energy poverty. Global scenario models usually estimate energy demand based on different levels associated with assumptions regarding GDP, population, urbanization rates and market penetration of electric vehicles (Moksnes and Rozenberg, 2019). The dimensions selected are in line with these general drivers of electricity demand. Except for population size, all indicators are published in the Municipal Atlas of Sustainable Development Goals (Andersen et al. 2020). The municipal sustainable development index is a multi-dimensional index that indicates the degree to which a municipality is developed from the point of view of multiple goals, among which are access to public and infrastructure services, to education, safety, or resilience against climate change. Energy poverty is expressed as the number of households that consume less than 25% of a nationally proposed electricity consumption target (*Tarifa Dignidad*), which is used as a threshold for applying discounted electricity tariffs (Andersen et al. 2019). For every dimension independently, municipalities were ordered according to increasing values of the chosen metric. We then extracted subsets starting at different values in the distribution of the ordering variable, from 0 with increments of 5 percentiles. We tested subsets of different sizes, ranging from 30% to 75% of the total number of municipalities.[1] The various subsets cover the entire distribution of municipality characteristics on the chosen dimension, i.e. from the most rural municipalities to the most populated, or the least energy poor. Figure 5 illustrates the distribution of values in 2016 for all types of electricity consumption considered in our analysis, as well as the distributions of the four indicators chosen as selection criteria for municipality subsets.

The availability of long time series of NTL luminosity values enables important forecast opportunities. Unfortunately, locally disaggregated information on electricity consumption has not been openly available for more recent years. For this reason, we have designed an exercise of forecasting annual residential energy consumption between 2017 and 2021 and comparing model predictions with the reported percentage change in national residential electricity consumption in those years with respect to 2016 values: an increase of 1.3% in 2017, 4.3% in 2018, 3.5% in 2019, 7.7% in 2020 and 9.4% in 2021 (International Energy Agency 2025).

2.6. Temporal analysis

In addition to the supervised learning approach, the relationship between monthly EC and monthly NTL radiance was explored using an unsupervised learning approach to detect general trends in EC and compare them with seasonal trends of NTL.

To observe the seasonal variation in total EC, we explored the monthly EC data using *k-means* clustering (MacQueen 1967) with the Euclidean distance as a measure of similarity between observations. Studies show that k-means with Euclidean distance is an efficient and robust method of time series clustering (Paparrizos and Gravano 2015), even when compared with more complex techniques of unsupervised learning. We evaluated the choice of the best number of clusters using the Davies–Bouldin index (Davies and Bouldin 1979), a metric based on the principle of minimizing within-cluster differences and maximizing between-cluster separation. Each cluster is represented by its centroid, which was obtained with Dynamic Time Warp (DTW) Barycentric Averaging (DBA) (Petitjean et al. 2011). DTW distance (Sakoe and Chiba 1978) aims at identifying an alignment between two sequences by flexibly matching sequence indices. DBA averaging is an iterative procedure that computes the average of a set of time series sequences by minimizing the DTW distance between the average and the sequences. Compared with the simple arithmetic mean, barycentric averaging better captured the monthly variation in the cluster centroid. The time series analysis was performed with the *tslearn* Python package (Tavenard et al. 2020).

For capturing both between- and within-year trends in electricity consumption, we analyzed the monthly values between January 2015 and December 2016. First, we excluded municipalities with at least one null monthly EC value in this interval. This was done to prevent drawing interferences from noisy data, where a sudden drop to zero in EC was considered a possible indicator of an aberration in the monthly time series, and the data for the municipality in question were flagged for further data processing.

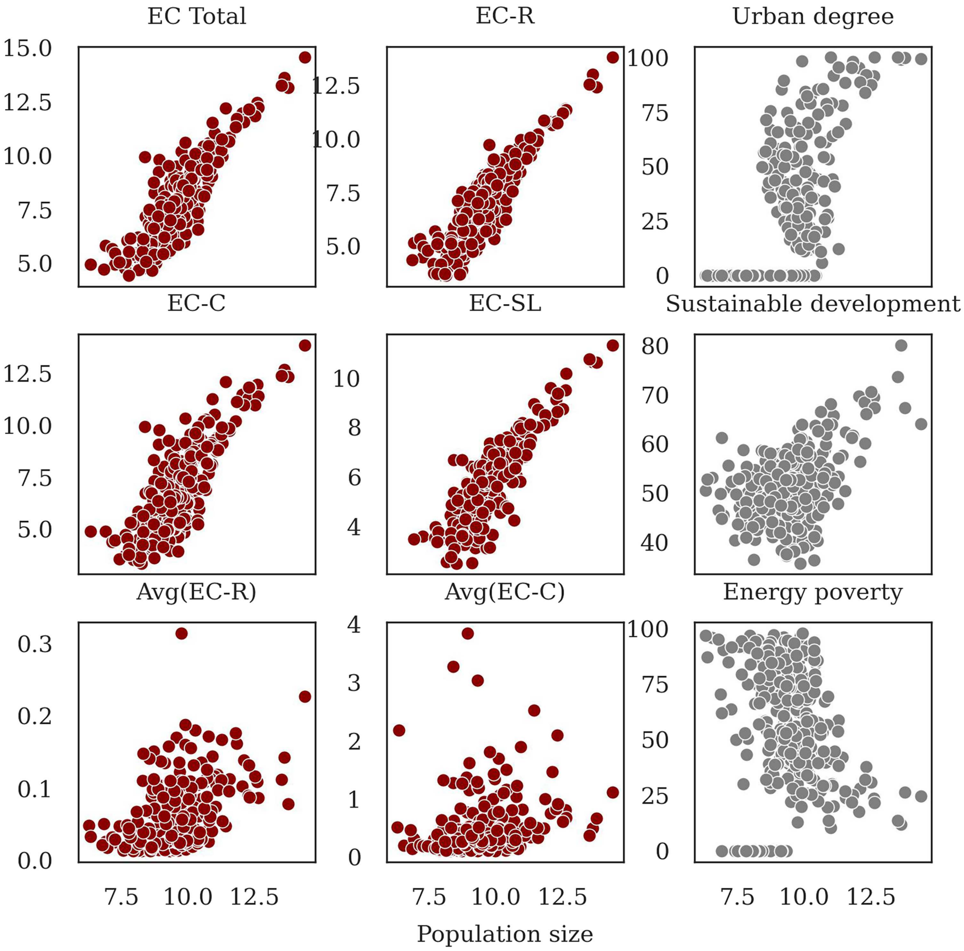

Figure 5. Distribution of 2016 annual values of different types of electricity consumption: electricity used in residences (*EC-R*), used for commercial purposes – including companies, neighborhood stores, city hall services and other (*EC-C*), used for street lightning (*EC-SL*), the total consumption for these three uses (*EC Total*), average residential consumption per user (*Avg(EC-R)*), and average commercial consumption per user (*Avg(EC-C)*). The distributions of significant municipality attributes are also displayed. The electricity consumption variables and the municipality population size have been log-transformed.

Our sample for the monthly analysis consisted of 295 municipalities (87% of all municipalities, home to some 95% of the population). Second, pre-processing steps included applying a rolling average in a window of two months and then normalizing each municipal time series through min–max scaling. These techniques were intended to smooth noise in the data and bring all time series to the same scale, between zero and one. The clustering features are the 24 monthly values. Each cluster is represented by a single time series, the centroid or profile. Then, starting from the clusters of the EC time series, we extracted NTL profiles from the NTL monthly values within the same municipality cluster. Finally, the representative EC profiles and corresponding NTL profiles were compared based on seasonality patterns over the two years under analysis.

3. Results

3.1. Spatio-temporal distribution of nighttime lights

Bolivia's landscape heterogeneity and predominance of rural areas, small villages, and peripheral mountainous regions naturally produce none to very low NTL radiance for large parts of the country. Between 2% (2016) and 5% (2012) of the municipalities show no NTL radiance. These municipalities are located in departments with the lowest GDP (INE 2021). A significant number of municipalities also register no electricity consumption: between 34% (2012) and 6% (2016) of all municipalities. For 2016, most of these municipalities are rural communities with low population counts and high levels of poverty.

NTL radiance is generally concentrated within settlement areas. Nevertheless, in 11% of all municipalities, more than 25% of the radiance is recorded in out-of-settlement areas, which are mostly located in areas of herbaceous vegetation and in the immediate vicinity of a settlement. Since these are small rural communities, we can deduce that the perimeter of economic activity is larger than what the available built-up datasets might indicate. Other areas with out-of-settlements NTL radiance are located around petrol and gas extraction fields, such as Margarita or San Alberto in the Tarija department; gas and condensate extraction fields, such as Incahuasi in the Santa Cruz department; and gold mines, such as Lake Suches in the La Paz department. The identification and monitoring of areas with mining activity with remote sensing data enables the evaluation of their environmental impact and sustainability, an important link with the socioeconomic development of the regions concerned (Rodríguez López et al. 2020).

Between 2012 and 2016, the median NTL radiance within settlement areas increased by 270%. The greatest increase was recorded in very rural and poor communities, which were predominantly located in La Paz. Among the few urban municipalities that recorded a high increase in radiance are the municipalities of Rurrenabaque and Reyes in Beni. The latter experienced a boom in tourism during that period due to important natural attractions (the Madidi National Park and the Pampas of Beni). There is no correlation pattern between the change in NTL radiance in settlement and non-settlement areas. In non-settlement areas, the NTL median luminosity increased by 290% between 2012 and 2016. Most of the municipalities that experienced the highest increase are highly urban municipalities, indicating that the high increase in NTL is very likely to be associated with urban developments located outside the core city areas – driven by population growth and socioeconomic factors such as immigration and house prices (Parés-Ramos et al. 2013).

We generally found the relationship between the percentage change in NTL radiance and the percentage change in EC is weak, as illustrated in Figure 6. A moderate correlation can be observed between the percentage change between 2012 and 2016 in total EC and NTL radiance outside settlement areas. The percentage changes between consecutive years are more closely related than the percentage changes across longer periods, as shown in Table 3. The most important correlations found concern the changes taking place between 2013 and 2015. For this period, we observe that the change in average commercial electricity consumption increases with the change in NTL intensity. Finally, we find that changes in NTL intensity, either as total or average, are more spatially concentrated than changes in any of the electricity consumption outcomes.

3.2. Annual models

Predicting current annual EC consumption using past annual values as input is moderately successful, as illustrated in Table 1 in the Appendix. For the 2016 predictions, the errors range from 20% for predicting total electricity consumption to 35% for predicting average commercial consumption per user. Using data from the previous year yields higher accuracy than other historical annual data. Model goodness-of-fit tests showed that the NTL-derived variables explained between 83% (2012) and 87% (2016) of the variance in total annual EC, as illustrated in Table 4. The cross-validation results reported in Table 5 show the impact that the choice of training and test data has on the prediction accuracy. Larger training sets lead to lower overall errors, while smaller test sets lead to higher variability in errors because of higher probabilities of choosing extreme values for prediction. The highest variability in model performance is observed for the prediction of commercial electricity consumption.

As expected, population size, degree of urbanization and poverty levels are important predictors for all types of electricity consumption. In terms of NTL-derived predictors, highlighted are the average aggregates of luminosity values over both built and non-built areas, as well as aggregates that express the variation in the distribution of luminosity values. These variables and their Shapley importance values are listed in Figure 7.

Prediction errors are considerably high when the predictor variables are solely extracted from NTL data: from 41% for predicting average residential EC to 88% for predicting average commercial EC. Nevertheless, the combination of NTL-derived variables and fixed municipality characteristics lowers

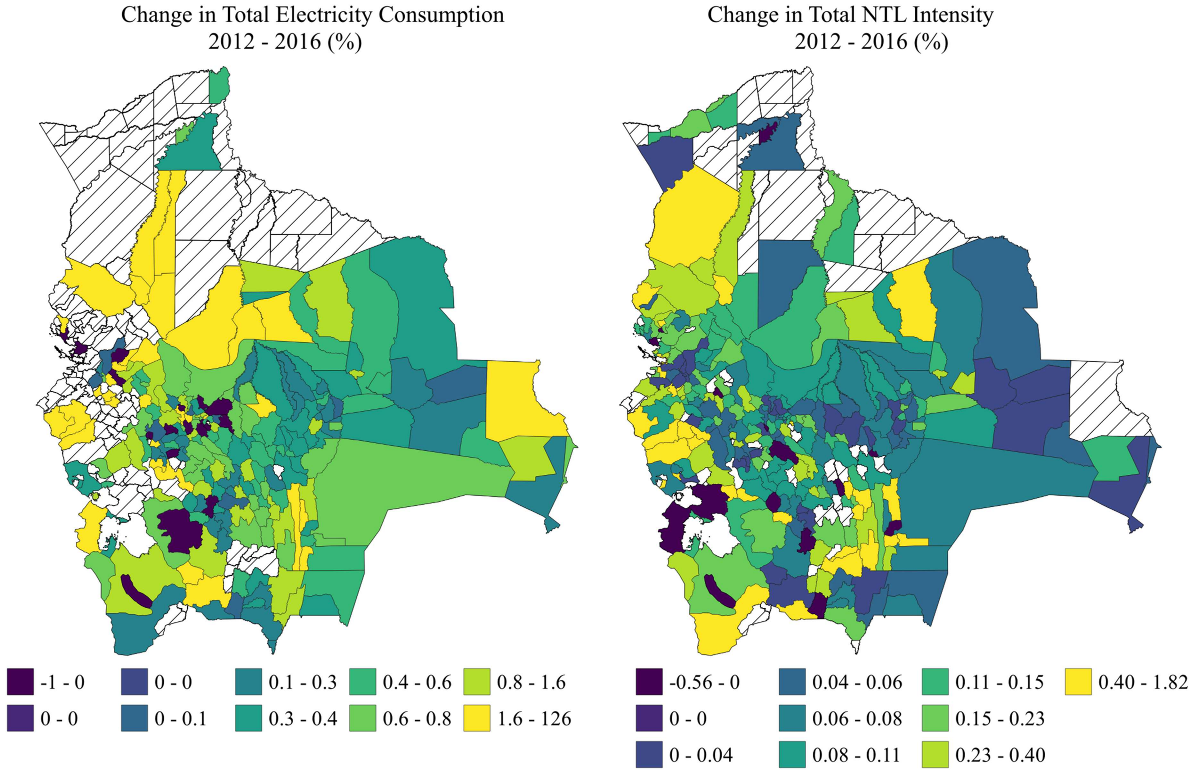

Figure 6. Percentage change in annual total EC (a) and NTL radiance (b) between 2012 and 2016, at the municipality level.

Table 3. The most important correlations (Pearson statistic $p > 0.2$) between changes in EC and changes in NTL intensity, considering the most important variables of each type, and all temporal ranges. The degree of spatial autocorrelation of each variable has been computed using Moran's I statistic.

EC variable	NTL variable	Year Start	Year End	Pearson (EC, NTL)	Moran I (EC)	Moran I (NTL)
Avg(EC-C)	Sum[NTL(Municipality)]	2013	2015	0.43	0.04	0.03
EC-SL	Sum[NTL(Not_Built)]	2014	2015	0.41	0.11	0.19
Avg(EC-C)	Sum[NTL(Built)]	2013	2015	0.36	0.05	0.06
EC-C	Zmean[NTL(Built)]	2013	2015	0.36	0.06	0.13
EC-SL	Sum[NTL(Municipality)]	2014	2015	0.31	0.11	0.01
EC-C	Zmean[NTL(Not_Built)]	2013	2015	0.31	0.06	0.19
EC-SL	Zmean[NTL(Built)]	2014	2016	0.28	0.14	0.04
EC-SL	Sum[NTL(Not_Built)]	2014	2016	0.28	0.14	0.07
Avg(EC-C)	Sum[NTL(Municipality)]	2013	2016	0.26	0.03	0.10
EC-C	Zmean[NTL(Municipality)]	2013	2015	0.26	0.06	0.15
EC-SL	Zmean[NTL(Municipality)]	2014	2015	0.26	0.11	0.18
EC-SL	Zmean[NTL(Municipality)]	2014	2016	0.26	0.14	0.01
EC-R	Zmean[NTL(Not_Built)]	2012	2015	0.26	0.05	0.21
EC Total	Sum[NTL(Not_Built)]	2014	2015	0.25	0.08	0.14
EC-R	Sum[NTL(Not_Built)]	2012	2015	0.25	0.05	0.17
EC-C	Zmean[NTL(Not_Built)]	2013	2016	0.25	0.04	0.14
Avg(EC-C)	Sum[NTL(Built)]	2013	2016	0.24	0.03	0.07
EC-SL	Sum[NTL(Municipality)]	2014	2016	0.24	0.14	0.02
EC-SL	Zmean[NTL(Not_Built)]	2013	2016	0.23	0.04	0.21
EC Total	Zmean[NTL(Not_Built)]	2012	2015	0.22	0.06	0.17
EC-SL	Zmean[NTL(Not_Built)]	2013	2015	0.22	0.11	0.17
EC-C	Zmean[NTL(Built)]	2013	2016	0.21	0.05	0.14
EC-R	Sum[NTL(Not_Built)]	2012	2016	0.21	0.05	0.18
EC-R	Zmean[NTL(Built)]	2012	2015	0.20	0.05	0.12
EC-SL	Zmean[NTL(Not_Built)]	2014	2015	0.20	0.11	0.28

prediction errors by as much as 30%. The highest errors are recorded for rural municipalities, located predominantly in the departments of Oruro and La Paz, which experienced a sharp increase in EC between 2014 and 2016. The most accurate estimates are obtained for residential EC per user, while the least accurate estimates are obtained for commercial EC per user. Figure 8 shows the geographic distribution of

Table 4. Model goodness of fit (R^2) and mean absolute percentage error (MAPE) for the prediction of annual values of different types of electricity consumption. Regression models include NTL-derived variables, extended by municipality attributes (model 1) and past values of electricity consumption (model 2). Including/excluding in the model each additional group of variables is marked by 1/0.

				Outcome											
Model				EC Total		EC-R		EC-C		EC-SL		Avg (EC-R)		Avg(EC-C)	
(1)	(2)	Train	Predict	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE
0	0	2012	2013	0.83	107	0.79	121	0.82	139	0.69	400	0.86	52	0.37	191
			2014	0.80	85	0.81	68	0.78	97	0.74	114	0.95	52	0.10	223
			2015	0.78	91	0.79	66	0.75	107	0.72	104	0.94	55	0	231
			2016	0.78	78	0.80	68	0.75	93	0.71	102	0.93	56	0	229
		2013	2014	0.79	75	0.82	58	0.74	93	0.77	97	0.87	55	0	136
			2015	0.76	83	0.81	56	0.72	95	0.76	87	0.87	51	0.20	138
			2016	0.76	69	0.79	58	0.72	77	0.75	85	0.87	50	0.09	133
		2014	2015	0.86	67	0.88	53	0.82	88	0.82	72	0.95	39	0.31	106
			2016	0.85	58	0.87	58	0.81	81	0.80	72	0.95	41	0.11	112
		2015	2016	0.87	55	0.88	56	0.83	78	0.81	77	0.95	41	0.28	88
	1	2012	2013	0.87	92	0.87	82	0.85	121	0.83	96	0.89	29	0.38	234
			2014	0.87	52	0.87	43	0.85	62	0.81	81	0.98	30	0	282
			2015	0.83	58	0.85	42	0.81	77	0.79	76	0.97	28	0	302
			2016	0.82	52	0.83	44	0.81	63	0.75	74	0.97	28	0	272
		2013	2014	0.91	38	0.90	34	0.89	44	0.84	70	0.89	31	0.35	122
			2015	0.86	43	0.86	37	0.85	51	0.82	63	0.90	24	0.26	136
			2016	0.84	40	0.85	40	0.84	44	0.80	60	0.89	25	0.27	112
		2014	2015	0.94	33	0.96	28	0.91	50	0.90	46	0.98	16	0.44	81
			2016	0.93	29	0.96	27	0.89	45	0.87	49	0.98	17	0.29	83
		2015	2016	0.94	30	0.96	25	0.91	44	0.89	48	0.99	14	0.44	65
1	1	2012	2013	0.87	92	0.87	82	0.85	121	0.83	96	0.89	29	0.38	234
			2014	0.87	52	0.87	43	0.85	62	0.81	81	0.98	30	0	282
			2015	0.83	58	0.85	42	0.81	77	0.79	76	0.97	28	0	302
			2016	0.82	52	0.83	44	0.81	63	0.75	74	0.97	28	0	272
		2013	2014	0.91	58	0.88	19	0.93	37	0.88	54	0.92	20	0.48	56
			2015	0.87	57	0.82	28	0.91	43	0.84	63	0.93	16	0.35	112
			2016	0.84	60	0.82	30	0.90	48	0.84	51	0.92	18	0.58	64
		2014	2015	0.97	21	0.97	18	0.97	28	0.95	28	0.99	9	0.67	44
			2016	0.95	24	0.97	19	0.93	39	0.91	31	0.99	11	0.46	59
		2015	2016	0.96	21	0.97	20	0.95	32	0.97	20	0.99	9	0.72	33

Table 5. Model goodness of fit (R^2) and mean absolute percentage error (MAPE) for the prediction of annual values of different types of electricity consumption. Model metrics are listed as average and standard deviation values over repeated (1000 repeats) cross-validation with a varying number of validation folds. The training data contains *k-1* folds of all available data in the training year and the test data *1* fold of all available data in the prediction year, sampled independently and without repetition. Regression models include NTL-derived variables, extended by municipality attributes and past values of electricity consumption.

			Outcome											
			EC Total		EC-R		EC-C		EC-SL		Avg(EC-R)		Avg(EC-C)	
Train	Predict	K-folds	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE
2013	2014	2	0.82 ± 0.05	63 ± 29	0.79 ± 0.06	30 ± 3	0.84 ± 0.04	45 ± 9	0.81 ± 0.04	63 ± 14	0.97 ± 0.03	22 ± 8	0.45 ± 0.14	58 ± 15
		5	0.82 ± 0.08	60 ± 57	0.76 ± 0.10	26 ± 4	0.84 ± 0.07	41 ± 17	0.81 ± 0.08	60 ± 28	0.98 ± 0.03	22 ± 17	0.51 ± 0.26	56 ± 23
		10	0.80 ± 0.13	59 ± 85	0.74 ± 0.17	25 ± 5	0.84 ± 0.11	41 ± 25	0.8 ± 0.13	59 ± 42	0.99 ± 0.02	22 ± 25	0.59 ± 0.34	56 ± 34
2014	2015	2	0.95 ± 0.01	30 ± 4	0.94 ± 0.01	29 ± 4	0.94 ± 0.01	39 ± 5	0.92 ± 0.02	37 ± 6	0.99 ± 0.0	13 ± 2	0.59 ± 0.15	46 ± 16
		5	0.96 ± 0.01	25 ± 5	0.96 ± 0.02	24 ± 6	0.96 ± 0.02	33 ± 7	0.94 ± 0.02	31 ± 8	1.0 ± 0.0	11 ± 2	0.64 ± 0.2	44 ± 19
		10	0.96 ± 0.02	24 ± 7	0.96 ± 0.03	23 ± 8	0.96 ± 0.03	31 ± 10	0.94 ± 0.03	30 ± 11	1.0 ± 0.0	10 ± 3	0.69 ± 0.28	45 ± 28
2015	2016	2	0.95 ± 0.01	25 ± 4	0.95 ± 0.01	25 ± 5	0.94 ± 0.01	38 ± 8	0.9 ± 0.03	29 ± 4	0.99 ± 0.0	11 ± 1	0.70 ± 0.08	39 ± 8
		5	0.96 ± 0.02	22 ± 8	0.96 ± 0.02	22 ± 9	0.95 ± 0.03	34 ± 15	0.91 ± 0.05	25 ± 8	1.0 ± 0.0	10 ± 2	0.76 ± 0.13	37 ± 13
		10	0.95 ± 0.04	22 ± 12	0.96 ± 0.03	21 ± 14	0.95 ± 0.04	34 ± 23	0.90 ± 0.09	24 ± 12	1.0 ± 0.0	10 ± 3	0.81 ± 0.16	36 ± 20

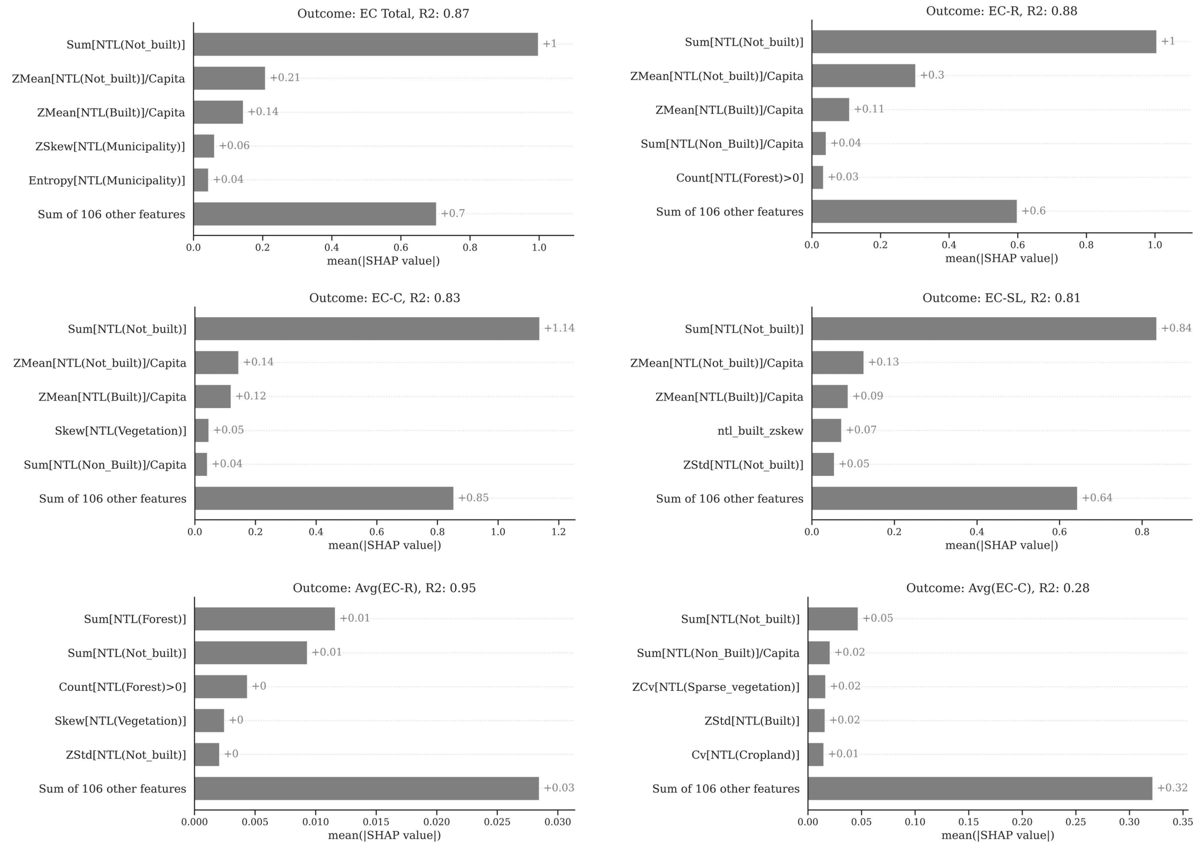

Figure 7. Most important NTL-derived variables for the prediction of 2016 EC values using 2015 values as training data. The prediction models include NTL-derived features.

actual and predicted 2016 values for these two types of EC, with the distributions for all other EC types illustrated in Figure 1 in the Appendix. We observe that the average commercial EC is largely overestimated both in the far north and the far south of the country. An important difference between average residential and average commercial EC is that residential ECs increase with increasing numbers of users and with the size of the municipality, while the relationship between commercial EC and the number of users is more susceptible to change between years, as illustrated in Figure 2 in the Appendix.

Concerning temporal consistency, there is a discernible improvement in predicting EC since 2014, compared with the 2012–2013 period. This can be explained by increased electricity coverage and smaller variability in EC in the same municipality between years. Pooling data from multiple previous years for creating a richer training data improves the prediction accuracy across all outcomes, as illustrated in Table 6. This is particularly useful for models based on NTL-derived variables and fixed municipality characteristics, and for predicting street lightning, for which the prediction improves from a MAPE of 48% to 27%.

When analyzing the prediction intervals of the EC values of individual municipalities, as shown in Figure 9, we observe a moderately high correlation between the uncertainty of prediction and municipality population size for outcomes that represent total electricity consumption. No relation was observed between the size of the prediction interval and other observed municipality characteristics, such as the degree of urbanization or economic or energy poverty. For a few EC outcomes, uncertainties have a geographic location aspect. Over all types of EC outcomes, higher uncertainties are observed for the prediction of average commercial EC, especially for municipalities in the Oruro and Pando departments. For street lightning, significantly large prediction intervals are associated with the EC values of municipalities in the Potosi department. For average residential EC, higher uncertainties are observed for municipalities in the Beni, Santa Cruz, and Pando departments.

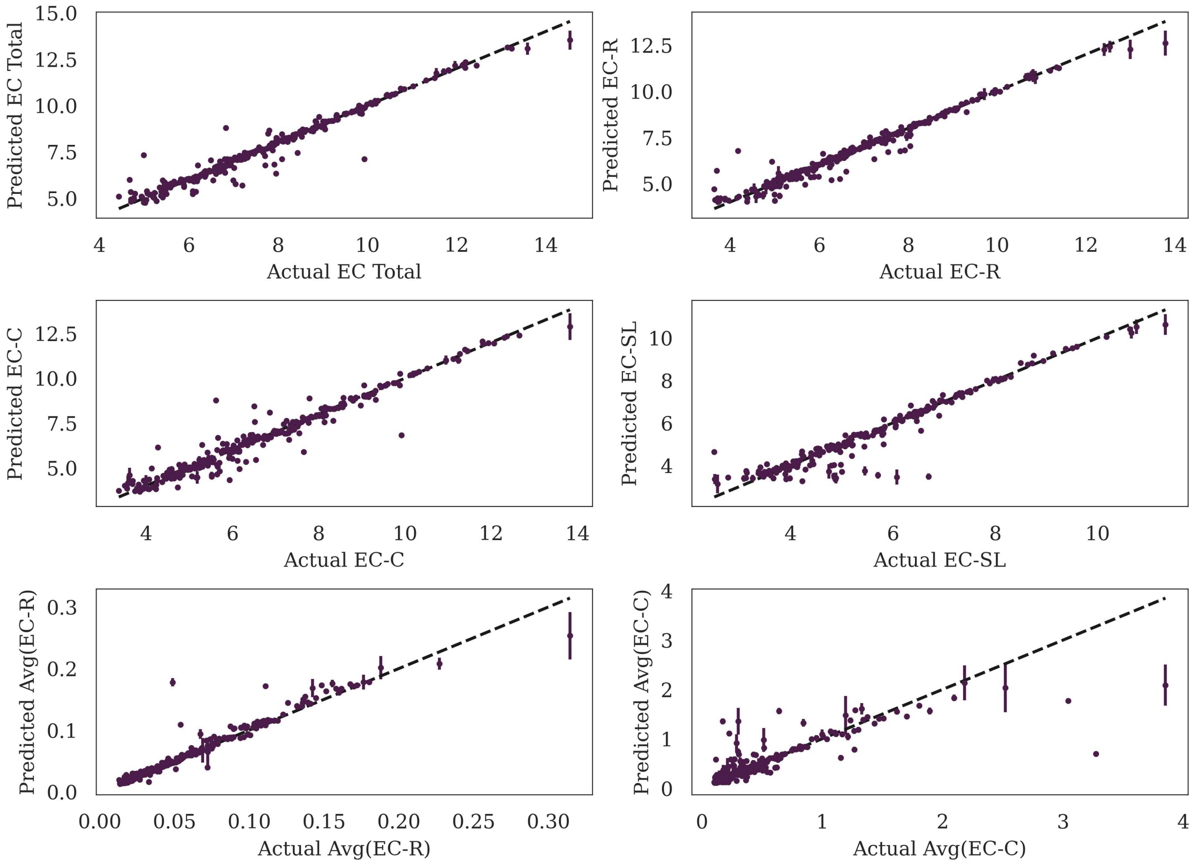

Figure 8. Prediction intervals versus actual values for various types of EC in 2016. The predicted annual values for 2016 are obtained by learning from 2015 data with a model containing all types of relevant features.

Table 6. Model goodness of fit (R^2) and mean absolute percentage error (MAPE) for the prediction of annual values of different types of electricity consumption using pooled data from all years previous to the year for which predictions are made. Regression models include NTL-derived variables, extended by municipality attributes (model 1) and past values of electricity consumption (model 2). Including/excluding in the model each additional group of variables is marked by 1/0.

			Outcome											
Model			EC Total		EC-R		EC-C		EC-SL		Avg(EC-R)		Avg(EC-C)	
(1)	(2)	Predict	R2	MAPE	R2	MAPE	R2	MAPE	R2	MAPE	R2	MAPE	R2	MAPE
0	0	2014	0.80	72	0.83	56	0.77	87	0.75	99	0.89	51	0.28	134
		2015	0.89	51	0.91	41	0.87	66	0.87	56	0.97	27	0.35	115
		2016	0.91	37	0.92	45	0.89	49	0.89	48	0.98	24	0.36	87
	1	2014	0.90	39	0.90	36	0.89	45	0.84	70	0.9	34	0.34	125
		2015	0.93	33	0.93	29	0.92	41	0.91	40	0.98	14	0.40	100
		2016	0.94	26	0.95	26	0.93	32	0.92	34	0.98	13	0.55	59
1	0	2014	0.82	68	0.79	28	0.86	44	0.82	66	0.93	24	0.42	63
		2015	0.97	22	0.94	21	0.95	31	0.95	27	0.99	11	0.65	50
		2016	0.96	24	0.96	26	0.96	31	0.94	27	0.99	11	0.72	45
	1	2014	0.82	59	0.77	25	0.85	39	0.82	57	0.92	22	0.40	58
		2015	0.97	18	0.96	18	0.96	26	0.96	24	0.99	9	0.67	44
		2016	0.96	21	0.97	21	0.96	31	0.94	23	0.99	11	0.70	45

In all years available for analysis, there is a significantly high variability in electricity consumption between municipalities. A diverse pool of examples, while highly generalizable, yields low estimation accuracies. For practical purposes, we aimed to identify the optimal subset of municipalities for which yearly forecasts, based on NTL variables and fixed municipality characteristics, would be most accurate. Figure 10 illustrates the results of this process when predicting total EC for 2016 using 2015 data. The results for the other outcome variables are included in the Appendix. Most criteria were found to be important for selecting successful subsets, except for the urbanization degree. This is due to the fact that

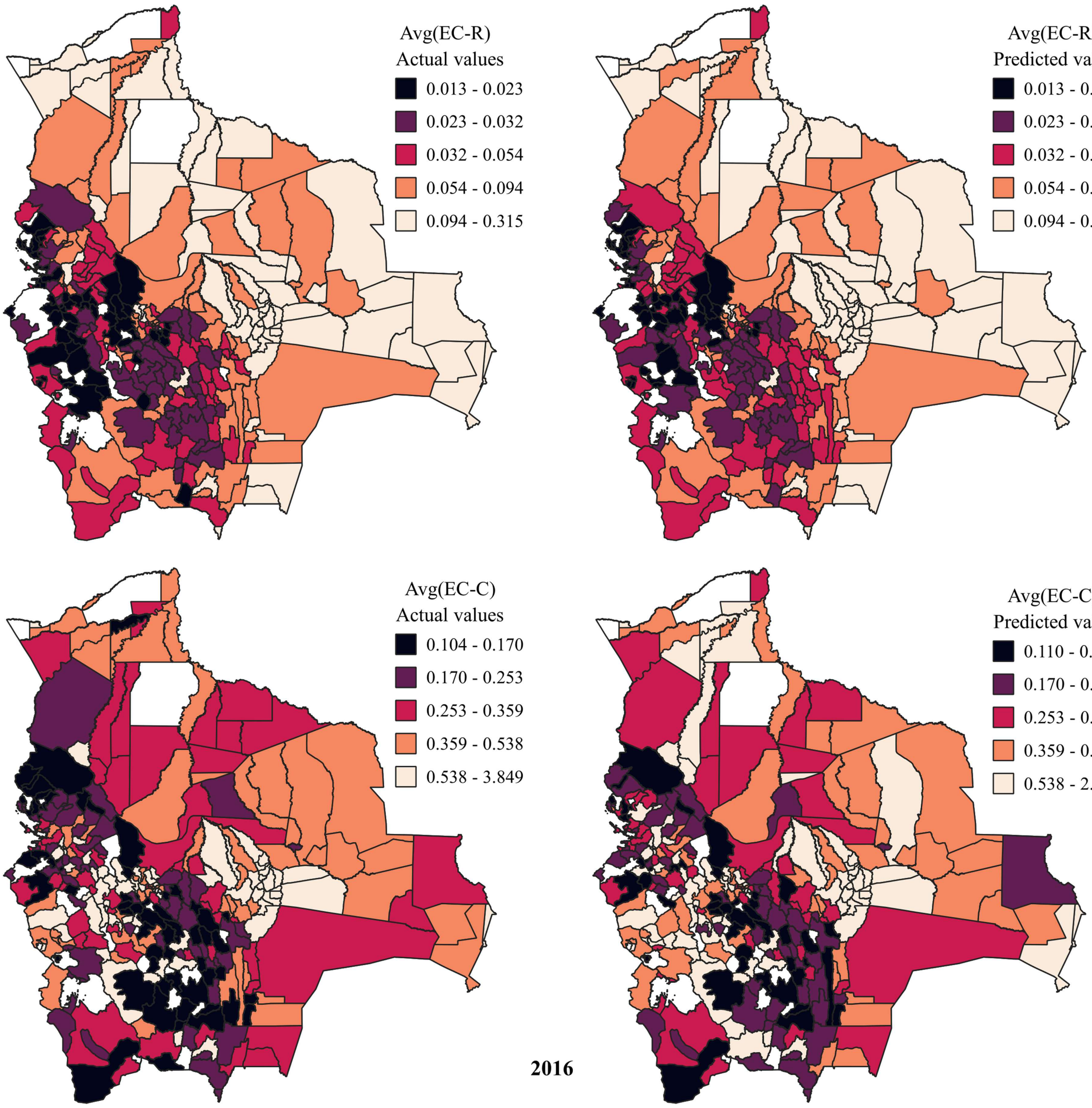

Figure 9. Actual and predicted values for average residential and commercial energy consumption in 2016. The predicted annual values for 2016 are obtained by learning from 2015 data with a model containing all types of relevant features.

most types of electricity consumption increase with population size, and the urbanization degree is the municipality attribute the least correlated with population size, as illustrated in Figure 5. For all EC types, we can find multiple subsets of municipalities for which prediction is improved compared with the whole set, as illustrated in Table 7. The best results across multiple prediction tasks are obtained for municipalities that are either medium in size, have average to higher-than-average degrees of sustainable development, or are among the least energy poor. We also identify subsets where prediction is improved on singular tasks. For example, total residential and total EC are estimated more correctly for larger than average municipalities while average residential EC per user is estimated more correctly for municipalities that experience higher than average degrees of energy poverty.

The subset analysis has shown that in all cases, eliminating municipalities with the most extreme values on either one of the four dimensions improves prediction accuracy. This confirms expectations: on the one hand, it is difficult to predict EC in small communities with an unpredictable pattern of consumption, and an irregular pattern of NTL luminosity; on the other hand, municipalities with more homogeneous

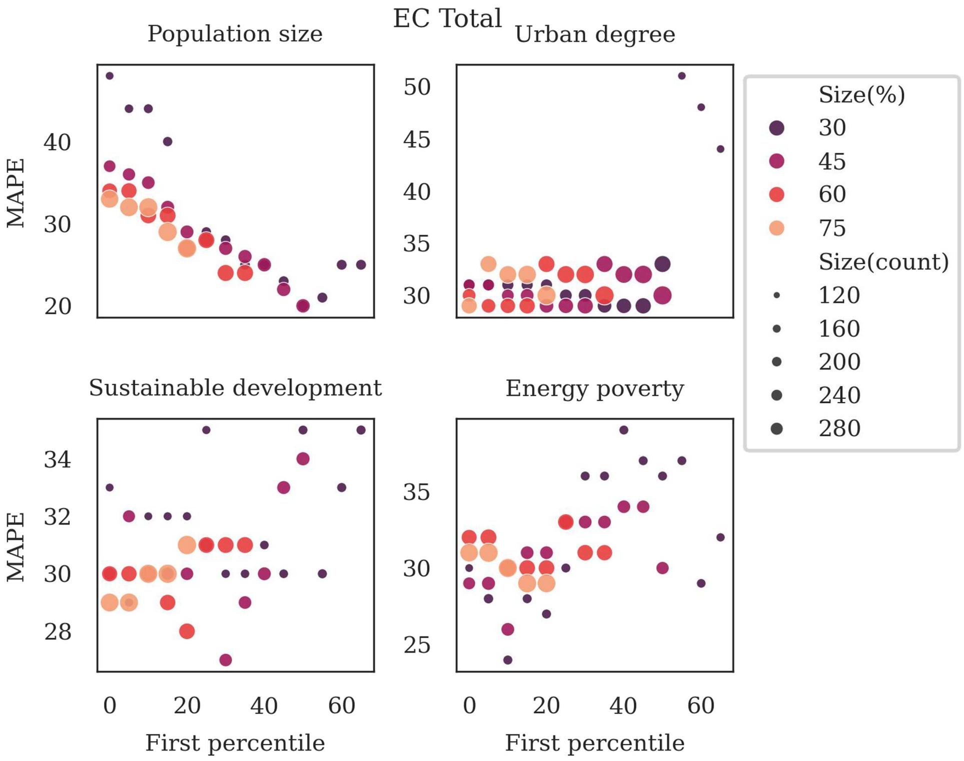

Figure 10. Mean absolute percentage error (MAPE) for the prediction of 2016 annual values of total electricity consumption. Predictions are performed for subsets of municipalities of various sizes, selected based on specific values in the distribution of a municipality characteristic. Selections are defined by the first percentile in the distribution of values of the criteria variable, and by the size of the subset: 30%, 40%, 60% and 75% of the total number of municipalities. Prediction models include NTL-derived features and municipality attributes.

Table 7. Mean absolute percentage error (MAPE) for the prediction of 2016 annual values of different types of electricity consumption. Predictions are done for subsets of municipalities, of various sizes, selected based on specific values in the distribution of a municipality characteristic (criteria). Selections are defined by the first and last percentiles in the distribution of values of the criteria variable. Only optimal subsets in terms of MAPE are included.

	Subset attributes					Prediction	
Outcome	Criteria	Subset size (count)	Subset size (%)	First percentile	Last percentile	MAPE (1)	MAPE (2)
EC Total	Population size	91	30	40	70	25	20
	Population size	87	30	35	65	25	22
	Population size	96	30	60	90	25	15
EC-R	Energy poverty	94	30	10	40	18	10
	Energy poverty	94	30	20	50	18	11
	Population size	90	30	40	70	18	13
	Population size	96	30	60	90	19	12
	Energy poverty	140	45	10	55	20	14
	Sustainable development	125	45	30	75	21	14
	Population size	129	45	25	70	22	14
	Energy poverty	184	60	20	80	22	20
	Energy poverty	227	75	10	85	23	21
EC-C	Population size	91	30	40	70	35	40
	Sustainable development	78	30	5	35	40	44
EC-SL	Sustainable development	61	30	45	75	30	18
	Sustainable development	95	45	40	85	31	16
	Sustainable development	127	60	25	85	39	17
Avg(EC-R)	Energy poverty	91	30	50	80	12	8
	Energy poverty	138	45	35	80	11	8
Avg(EC-C)	Energy poverty	93	30	15	45	32	17
	Energy poverty	137	45	15	60	37	19

economic activities are more likely to share patterns of NTL luminosity. Conversely, it is most difficult to estimate total and residential EC in municipalities that have a higher-than-average urbanization degree and in smaller-than-average municipalities. The electricity for street lighting is more difficult to estimate both in municipalities that are very small or have lower degrees of development, and in municipalities that have higher than average urbanization degrees. Average commercial EC per user is more difficult to estimate in municipalities that are either poorer than average in terms of energy consumption or that have lower than average degrees of development.

Finally, in our forecast exercise, we observe that when the entire set of observations and a regression model with all available predictors and data from 2016, the forecasts underestimate future residential energy consumption: between −18% in 2019 and −25% in 2021. When focusing on the best subset of observations for prediction, forecasts are found approximately in trend with the values reported at the national level (International Energy Agency 2025). These results are listed in Table 8. The evolution of residential EC in large municipalities (population sizes between the 60th and 90th percentile of the population size distribution) is the most likely to mirror the increasing trend in residential EC. This holds true, albeit in a less marked manner, for two other groups of municipalities–average in size and average in degree of sustainable development–groups comprising almost half of all municipalities. The 2019 residential EC forecasts for these two subsets are displayed in Figure 11, expressed as a percentage change with respect to the 2016 values. Further tests show that combining forecasts resulting from the two subsets, by averaging the common values, leads to an estimated overall decrease in residential EC, which deviates from the expected national trend. We cannot, however, firmly (in)validate our results without official statistics at sub-national levels. Of note is the fact that across all subsets of observations tested, models based on NTL-derived data estimate that EC has an upward trend until 2020 and a pronounced downward trend in 2021. This attests to the fact that the NTL data captured a considerable decrease in economic activity in the first year of the COVID-19 pandemic, although the overall electricity consumption in the country did not decrease, according to official reports.

3.3. Monthly models

Using unsupervised learning, we detected four municipality clusters based on 2015–2016 monthly total EC, labelled based on increasing average urbanization degree of municipalities within the clusters, as illustrated in Figure 12. Two profiles are consistent in terms of the degree of urbanization, poverty levels, and population sizes of the municipalities they contain: the first profile, with mostly poor and rural municipalities, and the last profile, with more affluent municipalities, which are mostly localized in the eastern part of the country. The third profile shows a decreasing trend in EC between 2015 and 2016, while the other profiles show increasing EC between 2015 and 2016 and similar seasonal patterns between the two years. The seasonality pattern is profile-specific. The mostly rural and poorer municipalities of the first profile show a significant peak in EC in October and November. Municipalities in the fourth profile show increased EC several times throughout the year, in January, April, May and November. In comparison, municipalities of the second profile register less variation in EC throughout the year.

Table 8. Percentage change in predicted values of annual residential electricity consumption for the years 2017–2021, with respect to 2016 values. Predictions are done for subsets of municipalities, of various sizes, selected based on specific values in the distribution of a municipality characteristic (criteria). Selections are defined by the first and last percentile in the distribution of values of the criteria variable. Only optimal subsets in terms of prediction MAPE for 2016 are included.

	Subset attributes					EC-R Forecasts using 2016 data				
Criteria	Subset size (count)	Subset size (%)	First percentile	Last percentile	MAPE	2017	2018	2019	2020	2021
Population size	96	30	60	90	12	3.2	2.9	3.5	4.0	3.1
Sustainable development	125	45	30	75	14	2.1	2.5	2.3	4.1	1.4
Population size	129	45	25	70	14	1.7	1.6	2.9	2.2	1.1
Energy poverty	94	30	20	50	11	−0.1	0.0	0.8	0.9	−0.7
Energy poverty	185	60	15	75	20	−0.1	0.5	0.0	0.5	−0.1
Energy poverty	184	60	20	80	20	−1.0	−1.5	−0.8	−2.1	−1.3
Energy poverty	224	75	15	90	21	−1.0	−0.9	−0.6	−0.6	−1.3

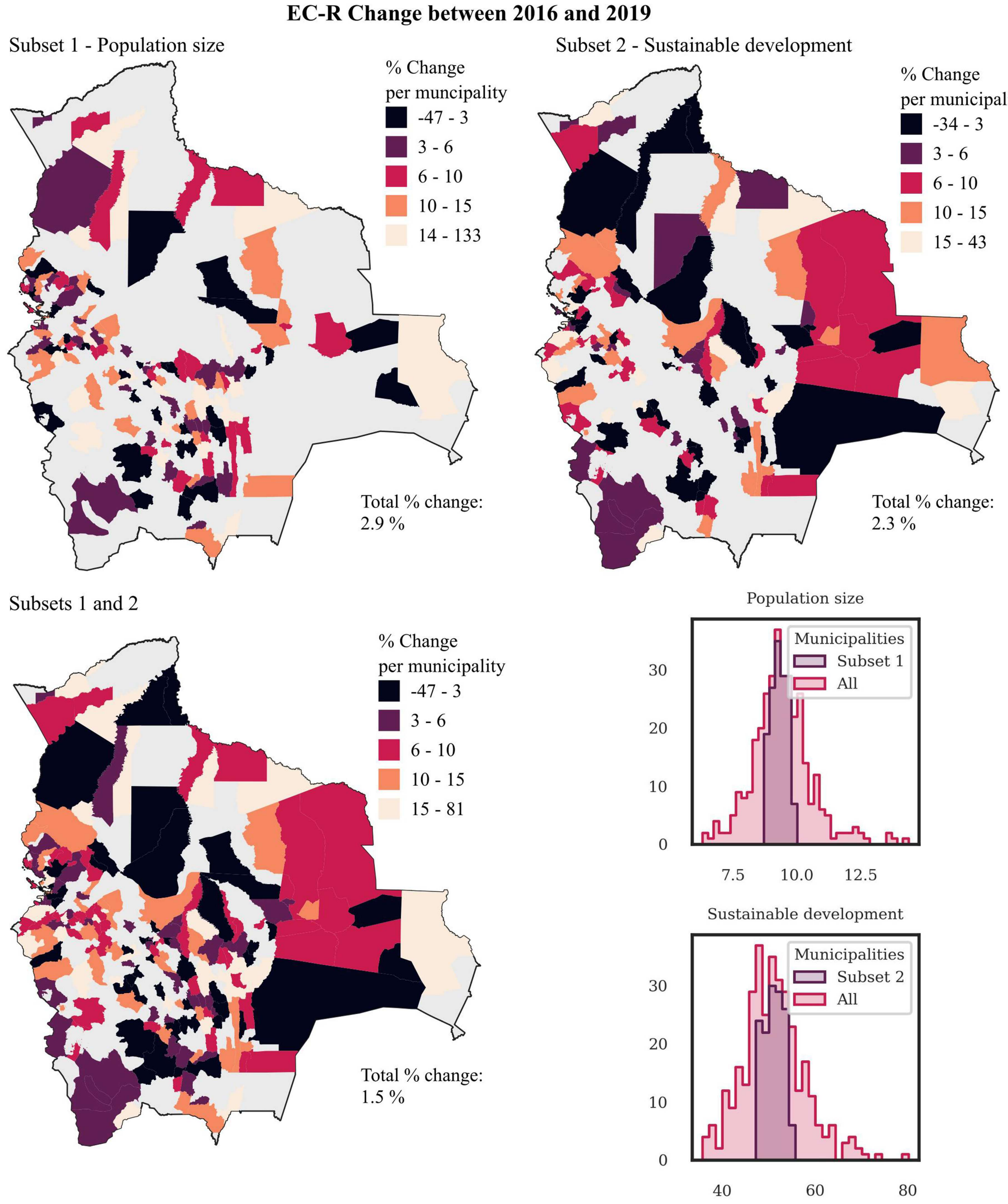

Figure 11. Percentage change in predicted values of annual residential electricity consumption for the year 2019, with respect to 2016 values. Comparison of two subsets of municipalities for which prediction errors are minimized (top maps). The bottom map aggregates results from the prediction models applied to each subset.

The temporal profiles of the NTL and EC are distinctly different. While EC monthly consumption is consistently high for most municipalities, with seasonal variations, the corresponding NTL values remain low with a small number of peaks, generally around the months of February—March and September—October. The lack of NTL radiance data for many months makes it difficult to directly correlate the monthly NTL and EC values for most months. This is reflected when predicting monthly EC values from the NTL-derived variables extracted from the same month's data. First, models that rely only on NTL-derived variables yield a prediction error for total EC between 54% for the month of March and 91% for the month of October. We

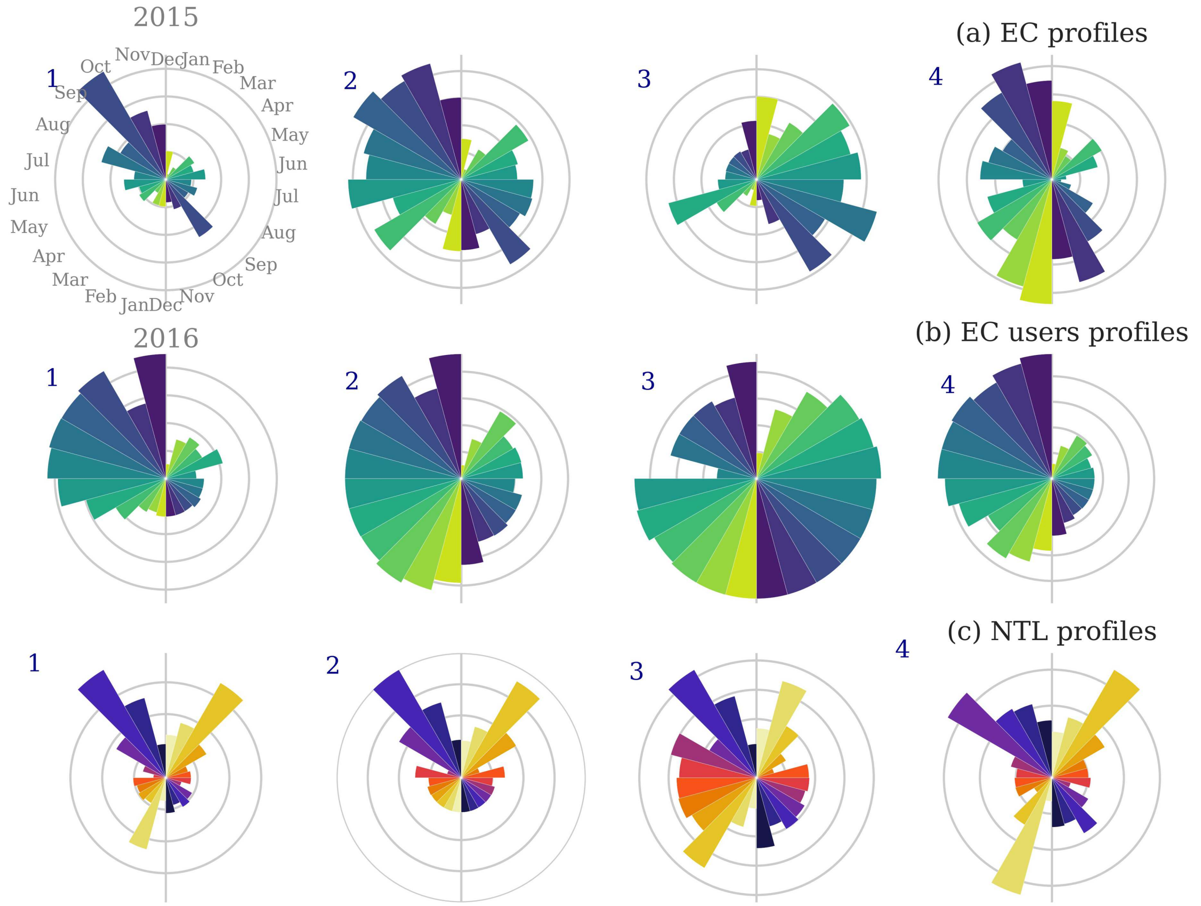

Figure 12. Summary profiles of monthly total electricity consumption (a), total electricity users (b) in Bolivian municipalities between 2015 and 2016, obtained by time series clustering, and the equivalent NTL profile for each EC cluster (c).

observe no improvement between using NTL information from 1 month versus using it for several months. Then, including municipality characteristics improves the prediction significantly, with errors for total EC prediction ranging between 27% in July and 39% in October, as illustrated in Table 9, Appendix. Finally, when using simple EC-derived variables such as the EC of the previous year, we obtain errors between 19% in April and 26% in October. Finally, the addition of the information of the EC for the same month of the previous year does not improve predictions. Errors are similar between all months, except for October, for which the highest errors are recorded. As observed with the annual models, the monthly predictions for the later years are improved compared to the beginning of the period under analysis.

We expect that, similarly to the annual models, the predictions can be improved when filtering observations. Figure 13 illustrates the differences between estimations performed using the entire set of observations versus optimal subsets of various sizes. The predictions are significantly improved, especially for total and average commercial EC. Predictions respect the average monthly trend for all EC types, and for the various subsets tested. Figure 14 illustrates the temporal patterns of actual and predicted monthly values of electricity consumption when using the subset of observations that is optimal for each outcome. Figures 8–10 in the Appendix illustrate the temporal patterns for the entire set of municipalities and the other various subsets considered. Across the different EC types, we observe that the predictions tend to underestimate the range of values and to reduce the number of statistically extreme values. As already identified with the help of temporal clusters, the average monthly trend over different subsets of municipalities is not the same. This attests to the variety of paths to electrification and of patterns of electricity usage, shaped by socioeconomic and geographic conditions, that many Bolivian municipalities are undergoing.

Table 9. Mean absolute percentage error (MAPE) for the prediction of monthly values of different types of electricity consumption, expressed as the mean and standard deviation aggregated over all months. Regression models include NTL-derived variables and municipality attributes, extended by the past annual value of electricity consumption (model 1) and the past monthly value of electricity consumption (model 2). Including/excluding in the model each additional group of variables is marked by 1/0.

Model				Outcome					
(1)	(2)	Train	Predict	EC Total	EC-R	EC-C	EC-SL	Avg (EC-C)	Avg (EC-R)
0	0	2013	2014	41 ± 13	36 ± 13	66 ± 25	73 ± 11	183 ± 103	26 ± 8
		2014	2015	34 ± 1	29 ± 2	54 ± 5	55 ± 7	96 ± 20	20 ± 2
		2015	2016	30 ± 3	26 ± 5	52 ± 5	49 ± 4	80 ± 9	19 ± 2
1	0	2013	2014	50 ± 22	24 ± 2	60 ± 38	55 ± 6	95 ± 81	17 ± 4
		2014	2015	23 ± 4	21 ± 3	33 ± 10	31 ± 3	54 ± 9	13 ± 2
		2015	2016	21 ± 2	18 ± 3	34 ± 3	22 ± 5	41 ± 8	12 ± 1
	1	2013	2014	50 ± 22	24 ± 2	60 ± 38	55 ± 6	95 ± 81	17 ± 4
		2014	2015	22 ± 3	20 ± 3	32 ± 10	30 ± 3	51 ± 10	13 ± 1
		2015	2016	21 ± 2	18 ± 3	34 ± 4	21 ± 4	43 ± 16	12 ± 1

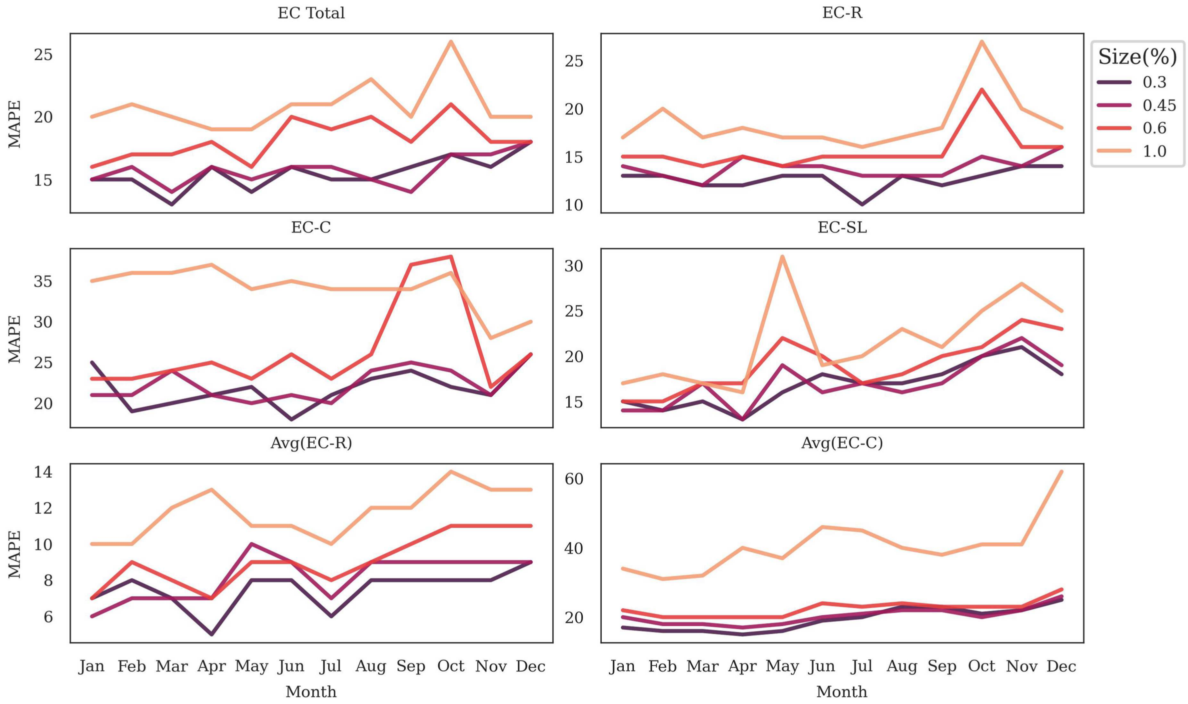

Figure 13. Mean absolute percentage error (MAPE) for the prediction of 2016 monthly values of different types of electricity consumption. Predictions are performed for the entire municipality set and for subsets of municipalities of various sizes, which minimize prediction errors. The prediction models include NTL-derived features, municipality attributes, and annual electricity consumption in the previous year.

4. Discussion

4.1. Prediction accuracy

Estimating electricity consumption using proxies such as nighttime lights is complex owing to the influence of population structure, land use, and seasons, among other issues. This is related to the general understanding that standalone nighttime lights data are unable to fully capture the different facets of electricity consumption, which are largely due to diurnal activities, but merely approximate them. However, meaningful relationships between nighttime lights information and electricity consumption can still be established, as our analysis and the larger related literature have documented.

In this study, we show that NTL data can successfully proxy multiple types of electricity consumption at the municipality level in Bolivia, both at the annual and monthly temporal scales. Using models based on features derived from NTL data and municipality attributes such as population size, urbanization degree,

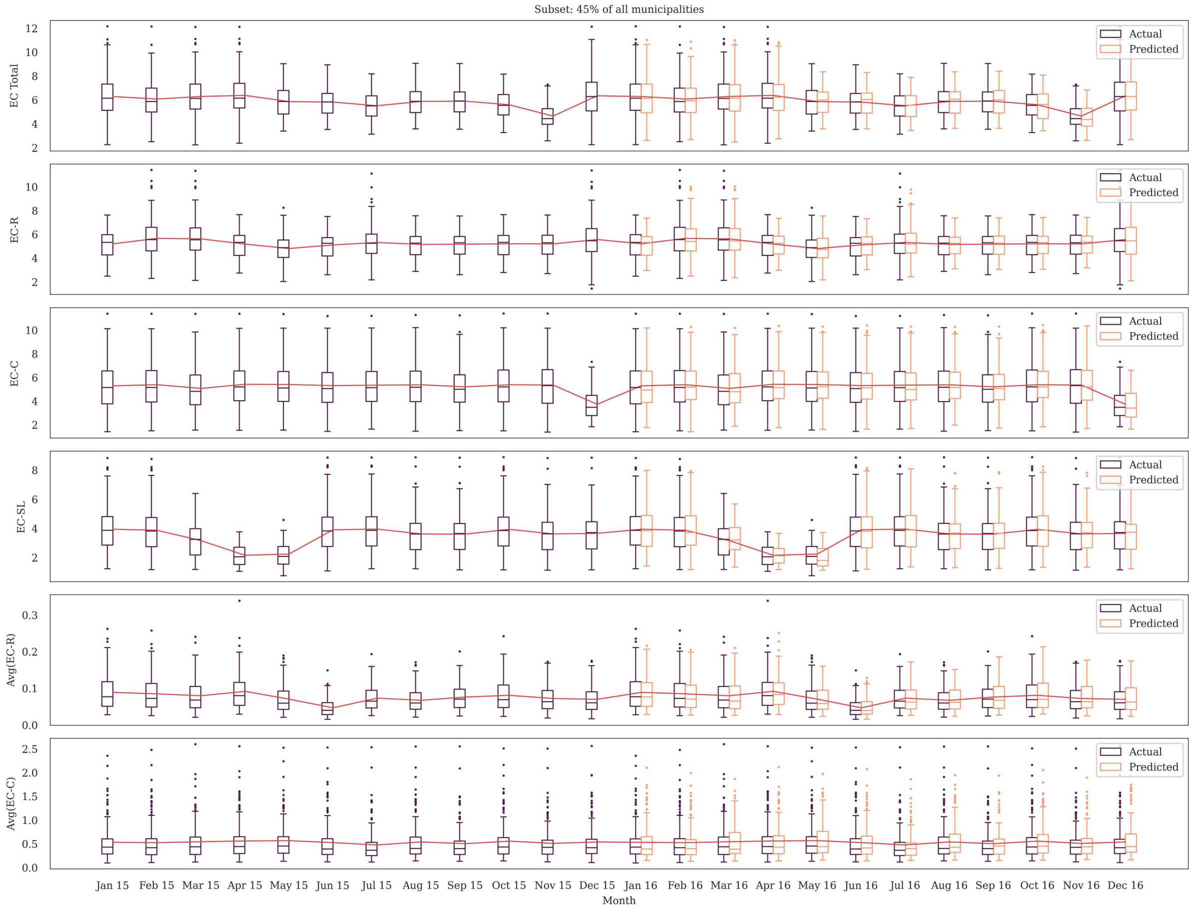

Figure 14. Actual monthly EC values in 2015–2016 for all municipalities considered in a learning scenario using 45% of all municipalities. The predicted monthly values for 2016 are obtained by learning from the 2015 data. The red line marks for each month the average EC value over all municipalities.

and poverty levels, we can estimate annual total electricity consumption with a mean absolute percentage error of 26%, residential electricity consumption with 26%, commercial electricity consumption with 31%, electricity used for street lightning with 34%, average residential consumption with 13%, and average commercial consumption with 59% MAPE. On the one hand, our results for Bolivia are consistent with similar studies for different areas across the globe that have reported high and very high model goodness-of-fit metrics (Falchetta et al. 2019; Wang and Lu 2021). On the other hand, our results clearly show the importance of reporting estimation errors and align with similar studies in this sense (Shi et al. 2016), since within a small range of very high values of model performance, there is a great variability in prediction errors.

For monthly electricity consumption, our results show that the combination of NTL-derived features, municipality attributes, and past values of annual electricity consumption is sufficient to obtain estimates that replicate the accuracy of annual predictions, for all types of electricity consumption, and with small variations between the different month estimates. This is an important finding for energy policy implementation and monitoring whenever only scarce official data are available. More generally, increasing the ability to nowcast or forecast temporally disaggregated EC at sub-national levels, while minimizing auxiliary data needs, is an essential part of planning long-term energy demand, alongside the development of scenarios that assess planning feasibility, required investments, and environmental impacts (Fernandez Vazquez et al. 2024).

As an indicator of light intensity during the night, NTL data can only imperfectly approximate economic activities with a strong daily component. This is reflected in lower estimation accuracies for total and average commercial residential consumption. This can be partially explained by significant heterogeneity in commercial activities across the different municipalities. In many rural areas in Bolivia, the main income-generating activities are not commerce, but agriculture (Sanchez Solis et al. 2023). Different geographic regions have specific profiles of commercial activity: in the lowlands, the predominant businesses are entertainment businesses, workshops, grocery stores, and restaurants, while in the highlands, the predominant businesses are grocery stores, dairy shops and restaurants (Sanchez Solis et al. 2022). With respect to average commercial consumption per user, we also attribute lower accuracies to the heterogeneity of activities within the municipality. We consider a worthy direction of further research, the investigation of various metrics that summarize the variability in electricity use for commercial activities at local levels, in relation to higher spatial resolution NTL data (Lu et al. 2025).

We can highlight municipality groups for which prediction errors can be further reduced. While varying the purpose of the prediction task, population size, energy poverty, and sustainable development levels can all be successful criteria for selecting subsets of municipalities that optimize prediction. Among them, population size produces the most consistent results as a municipality selection criterion. These municipality attributes, alongside the urbanization degree, mediate the regional heterogeneity in NTL data and significantly improve mapping with electricity consumption data. This is expected since attributes of this type are essential parameters in global scenarios for forecasting future energy demand (Moksnes and Rozenberg, 2019).

The data under analysis covers a period (2012–2013) when a significant number of municipalities were in the process of gaining access to electricity, either for residential or commercial purposes. For this reason, models that use either past electricity consumption or NTL-derived information do not perform well on prediction tasks in earlier years. This is likely replicated in other areas in the Global South where the electrification process is still ongoing. With increasing coverage, we believe that the accuracy of estimation can only increase. Future energy demand is difficult to estimate for developing countries because of uncertainty in economic activities and industries (Cheng et al. 2022). The increased population density in urban areas, growth of road networks, or development of industrial activities would all translate into higher NTL radiances, which would better correlated with the increased economic activity. Overall, for further prediction and forecast scenarios, we expect that errors will be higher for smaller municipalities or municipalities that still experience high levels of energy poverty.

4.2. Important predictors

A large set of NTL-derived indicators has been generated, including both total aggregates and measures of variance in pixel intensity, with the goal of mapping the spatial distribution of human and economic activities, in line with recent studies (Qin et al. 2025). Methodologically, complex variance indicators have the purpose of enriching the NTL information associated with rural areas that are predominantly unlit. Conceptually, these indicators are meaningful in view of the significant spread and fragmentation of human settlements in many rural municipalities in Bolivia. The variance in nighttime lights intensity in such areas stands in contrast with that observed in large consolidated urban municipalities.

While indicators describing the variance in NTL radiance were found to be important for the estimation of all types of electricity consumption, the aggregation of NTL radiance over the area of the municipality outside of what was labelled as *built* or *settlement* had the largest positive impact on prediction accuracy. This is an important result that first points to the spread of economic and human activity outside of conventionally defined settlement boundaries. Second, it attests to the variability of urban expansion in many areas in developing countries, and the difficulty in precisely estimating them in a manner that is both accurate and consistent over time (Parés-Ramos et al. 2013; Duque et al. 2015). Further methodological refinements, such as the definition of intermediate boundaries to delineate the increasing density of human and economic activity, will likely result in more accurate estimates, especially in terms of commercial electricity use or electricity use for street lighting. Then, the aggregation of NTL radiance over areas of distinct land use—forests, vegetation, or cropland—was also deemed important for prediction. Accurate mapping of land use classes pertaining to industrial purposes would further enhance the ability to predict electricity consumption generated by sources other than residential sources.

4.3. Data limitations

Despite the extensive preprocessing applied to the Black Marble NTL products, NTL data still suffers from inherent technical limitations. First, cloud contamination reduces the number of valid observations and represents a major source of outliers. In this analysis, we relied on the cloud mask provided with the Black Marble product, which is reported to have 94% accuracy for areas between 60°S and 60°N, Bolivia included (Wang et al. 2021). However, residual cloud misclassification may persist in regions with consistently high cloud frequencies, potentially leading to an underestimation of artificial light radiance. Second, light scattering and overglow are common issues in which the radiance emitted from brightly lit urban centres spreads into adjacent pixels because of atmospheric scattering along the satellite scan direction. This effect can lead to inflated radiance values in peri-urban areas (Wang et al. 2021). With a majority of rural and mid-sized urban areas under analysis, we postulate that this issue affects our results to a lesser measure. Third, angular effects introduce uncertainties, as repetitive observations of the same area may report changes that are not related to actual NTL brightness variations, but rather to the viewing angle (Tan et al. 2022). According to Wang et al. (2021), nadir observations in urban centres could be up to three times higher than off-nadir observations. Even though we considered images with different viewing angles for our composites to cover the different light sources, this uncertainty poses a challenge when comparing composites of different years and months. Additionally, given the diverse range of landscapes within the study area, the angular effect can exhibit variations influenced by landscape factors such as vegetation and buildings, thereby introducing additional uncertainty into the analysis (Li et al. 2019; Tan et al. 2022).

The annual Black Marble composites used for this study generally have more than 97% high-quality pixels. However, the sensitivity of the sensors is insufficient for detecting low NTL emissions in low-density rural areas of the Andean and Amazonian regions. The detection of electrification in rural areas using nighttime lights is difficult because of low luminosity (Min and Gaba 2014), and our results reflected this issue. Errors in forecasting annual EC were greater for rural areas, and we encountered both over-estimation and under-estimation errors for these types of areas. This limitation becomes evident when comparing month-to-month EC and NTL values.

Seasonality plays an important role in capturing accurate information in a consistent manner using NTL images. On the one hand, Bolivia is a country with low solar seasonality, which minimizes variations in night length throughout the year. On the other hand, the long duration of the rainy season, from December to March, hinders sensor activity. The region East of the Andes is especially characterized by more intense precipitation during the nighttime (Giles et al. 2020). Vegetation phenology and snow presence are other key sources of error that are related to seasonality. Snow amplifies the scattering of the NTL, while the variations in the vegetation canopy cause the seasonal occlusion effect, influencing the amount of surface light that penetrates the canopy during the leaf-on and leaf-off periods (Wang et al. 2021).

While the spatial resolution of the Black Marble data is generally sufficient for electricity consumption estimations at regional or national scales, it limits intra-urban small-scale analyses. Additionally, in topographically complex and remote regions such as the Andean highlands and the Amazon rainforest, the 500 m resolution may lead to mixed-pixel effects and increased sensitivity to terrain obstruction and persistent cloud cover. These factors can reduce the precision of NTL measurements. While preprocessing mitigates several of these challenges, future research could benefit from higher-resolution NTL measurements; for example, the SDGSAT-1 data provide fine spatial detail and have shown potential for fine-scale socioeconomic analysis (Chen et al. 2025).

4.4. Policy implications

Tracking residential energy consumption is instrumental in the implementation of appropriate energy subsidy policies. A key reason for which the Tarif Dignidad policy had a positive welfare effect was its targeted design, with the monthly 70 kWh threshold ensuring that subsidies reached predominantly low-income households (Rubin de Celis Cedro and Espinoza Vasquez 2025). Moreover, predicting future energy consumption can enable long-term policy planning. The variation in energy demand is an essential input for energy planning and energy system models (Sanchez Solis et al. 2022). Measuring the effects of different subsidy and pricing interventions on residential consumption is critical for designing energy conservation measures (Rubin de Celis Cedro and Espinoza Vasquez 2025). Furthermore, energy policies and investment initiatives need to take into consideration local realities for optimal effect. Interruptions in the electricity network are more likely to occur and take longer in rural areas (Delavechia et al. 2023). Very small rural communities generate a demand that renders the installation of micro-grids feasible economically and technically, due to the dispersed nature of such communities (Sanchez Solis et al. 2023).

The availability of annual and monthly EC data at subnational levels has opened the path towards a high-resolution analysis of energy demand. Households in different types of regions – such as the Bolivian highlands versus the lowlands – have differing types of electrical appliances and needs for electricity systems. The country's location in the tropics makes it a prime candidate for the development of renewable energy sources, solar energy and biomass (Lopez et al. 2021; Cheng et al. 2022). Combining energy needs with the local potential for renewable sources or with region-specific infrastructure requirements is one of multiple ways to fulfil the potential of the electricity database.

5. Conclusion

The available data and analysis captured a shift in electricity coverage and electricity consumption in a country on the path of growing energy demand, coupled with increasing urbanization. This trend is expressed as high spatial variation in both NTL-radiance and electricity consumption patterns, and as considerable mismatch between the two sets of indicators at fine temporal scales. Despite these limitations, the data-driven approach in our study has shown that a comprehensive exploration of various forms of NTL-derived information, in combination with other spatial data sources and municipality attributes such as population size, energy poverty levels, or sustainable development, can produce reliable estimates of various types of electricity consumption. Furthermore, documenting the inherent uncertainty in accuracy and identifying optimal prediction scenarios facilitates the further use of electricity forecasts and nowcasts for planning and policy modelling. Finally, with continuous efforts in the production of consistent time series of spatial data, such as nighttime lights, but also land cover, the ability to proxy important socioeconomic developments will only improve.

Endnotes

1. The total number of municipalities after eliminating those where the EC is either null or within the lowest 5% of non-null values.

Acknowledgements

All authors agree to be accountable for all aspects of the work.

Author contributions

CRediT: **Oana M. Garbasevschi:** Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft; **Andrea Sofia Garcia de León:** Data curation, Formal analysis, Visualization, Writing – original draft; **Lykke E. Andersen:** Conceptualization, Funding acquisition, Project administration, Resources, Writing – review & editing; **Guillermo Guzmán Prudencio:** Data curation, Validation, Writing – review & editing; **Michael Wurm:** Supervision, Validation, Writing – review & editing; **Hannes Taubenböck:** Conceptualization, Funding acquisition, Resources, Writing – review & editing.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This work was supported by the Compagnia di San Paolo Foundation and the Volkswagen Foundation, through the international cooperation project INEQUALITREES, under Grant number A130902.

Data availability statement

The electricity consumption data that support the findings of this study are openly available in a public Google Drive folder, accessible from the page of the associated technical report: https://sdsnbolivia.org/working-paper-*n*-2-2023-elbol-an-anonymized-research-databse-with-monthly-electricity-consumption-in-bolivia-2012-2016/ (Andersen et al. 2023). While the repository does not provide a persistent identifier, the authors commit to maintaining long-term accessibility of the files.

References

Alabi O, Abubakar A, Werkmeister A, Sule SD. 2022. Keeping the lights on or off: tracking the progress of access to electricity for sustainable development in Nigeria. GeoJournal. 88(2):1535–1558. https://doi.org/10.1007/s10708-022-10689-2

Andersen LE et al. 2023. ELBOL: An Anonymized Research Database with Monthly Electricity Consumption in Bolivia: 2012-2016. 02--23Universidad Privada Boliviana. https://sdsnbolivia.org/working-paper-n-2-2023-elbol-an-anonymized-research-databse-with-monthly-electricity-consumption-in-bolivia-2012-2016/

Andersen LE, Branisa B, Guzmán Prudencio G. 2023. Midiendo la pobreza y la desigualdad energética a nivel municipal en Bolivia. SDSN Bolivia. https://www.sdsnbolivia.org/en/midiendo-la-pobreza-y-la-desigualdad-energetica-a-nivel-municipal-en-bolivia/

Andersen LE, Branisa B, Calderón F. 2019. Estimaciones Del PIB per Cápita y de La Actividad Económica a Nivel Municipal En Bolivia En Base a Datos de Consumo de Electricidad. Centro de Investigaciones Sociales (CIS) de la Vicepresidencia del Estado Plurinacional de Bolivia.

Andersen LE, Canelas S, Gonzales A, Peñaranda L. 2020. Atlas Municipal de Los Objetivos de Desarrollo Sostenible En Bolivia 2020. Universidad Privada Boliviana, SDSN Bolivia: La Paz https://www.sdsnbolivia.org/atlas/

Andreano M, Simona R, Benedetti F, Piersimoni G, Savio. 2021. Mapping poverty of Latin American and Caribbean countries from heaven through night-light satellite images. Soc Indic Res. 156(2–3):533–562. https://doi.org/10.1007/s11205-020-02267-1

Breiman L. 2001. Random forests. Mach Learn. 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Buchhorn M et al. 2020. Copernicus Global Land Service: Land Cover 100m: version 3 Globe 2015-2019. Product User Manual (Dataset v3.0, doc issue 3.3). Zenodo. https://doi.org/10.5281/ZENODO.3938963

Cao Z, Gao F, Wu Z, Zhang Q, Zheng Z, Li Z, Hai X. 2025. Quantifying the spatiotemporal mismatches between energy consumption and population: a comparative study in Guangzhou and Shenzhen, China. Cities. 156:105522. https://doi.org/10.1016/j.cities.2024.105522

CEPAL. 2022. Statistical Yearbook for Latin America and the Caribbean, 2021. Comisión Económica Para América Latina y El Caribe UNITED NATIONS PUBLICATION: NEW YORK.

Chakraborty S, Alexander O, Kittner N, Wang Z. 2025. A global stocktake on electricity access and gaps from NASA black marble nighttime lights. Earth's Future. 13(6):e2024EF005916. https://doi.org/10.1029/2024EF005916

Chen Z, Luo H, Li M, Lin J, Zhang X, Li S. 2025. Fine-scale poverty estimation by integrating SDGSAT-1 glimmer images and urban functional zoning data. Remote Sens Environ. 329:114925. https://doi.org/10.1016/j.rse.2025.114925

Cheng C, Gutierrez NP, Blakers A, Stocks M. 2022. GIS-Based solar and wind resource assessment and least-cost renewable electricity modelling for Bolivia. Energy Sustainable Dev. 69:134–149. https://doi.org/10.1016/j.esd.2022.06.008

Chindarkar N, Goyal N. 2023. Did it increase energy consumption? A difference-in-differences evaluation of a rural electrification policy in Gujarat, India using night-time lights data. Energ Policy. 183:113814. https://doi.org/10.1016/j.enpol.2023.113814

Davies DL, Bouldin DW. 1979. A cluster separation measure. ITPAM. PAMI-1(2):224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Delavechia R, et al. 2023. Electricity supply regulations in South America: a review of regulatory aspects. Energies. 16(2):915. https://doi.org/10.3390/en16020915

Deutsches Zentrum für Luft- und Raumfahrt (DLR). 2016. Global Urban Footprint (GUF) v1. https://geoservice.dlr.de/web/datasets/guf

Duque JC, Patino JE, Ruiz LA, Pardo-Pascual JE. 2015. Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. Landsc Urban Plan. 135:11–21. https://doi.org/10.1016/j.landurbplan.2014.11.009

Elvidge CD, Zhizhin M, Ghosh T, Hsu F-C, Taneja J. 2021. Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. Remote Sens. 13(5):922. https://doi.org/10.3390/rs13050922

Eras-Almeida AA, Fernández M, Eisman J, Martín JG, Caamaño E, Egido-Aguilera MA. 2019. Lessons learned from rural electrification experiences with third generation solar home systems in Latin America: case studies in Peru, Mexico, and Bolivia. Sust. 11(24):7139. https://doi.org/10.3390/su11247139

Esch T, et al. 2012. TanDEM-X Mission—New perspectives for the inventory and monitoring of global settlement patterns. J Appl Remote Sens. 6(1):061702–1. https://doi.org/10.1117/1.JRS.6.061702

Falchetta G, Pachauri S, Parkinson S, Byers E. 2019. A high-resolution gridded dataset to assess electrification in Sub-saharan Africa. Sci Data. 6(1):110. https://doi.org/10.1038/s41597-019-0122-6

Fehrer D, Krarti M. 2018. Spatial distribution of building energy use in the United States through satellite imagery of the earth at night. Build Environ. 142:252–264. https://doi.org/10.1016/j.buildenv.2018.06.033

Fernandez Fuentes M, et al. 2021. Evaluación del costo de Electrificación rural en Bolivia para alcanzar el ODS 7. Journal Boliviano de Ciencias. 17(50):32–55. https://doi.org/10.52428/20758944.v17i50.2

Fernandez Vazquez CAA, Vansighen T, Fernandez Fuentes MH, Quoilin S. 2024. Energy transition implications for Bolivia. Long-term modelling with short-term assessment of future scenarios. Renew Sustain Energy Rev. 189:113946. https://doi.org/10.1016/j.rser.2023.113946

Georg I, Blaschke T, Taubenböck H. 2023. Spatial delineation of urban corridors in North America: an approach incorporating fuzziness based on multi-source geospatial data. Cities. 133:104129. https://doi.org/10.1016/j.cities.2022.104129

Giles JA, Ruscica RC, Menéndez CG. 2020. The diurnal cycle of precipitation over South America represented by five gridded datasets. IJCli. 40(2):668–686. https://doi.org/10.1002/joc.6229

INE. 2020. Geographical Aspects of Bolivia. https://www.INE.gob.bo/index.php/bolivia/aspectos-geograficos/

INE. 2021. Producto Interno Bruto por Departamento https://www.INE.gob.bo/index.php/estadisticas-economicas/pib-y-cuentas-nacionales/producto-interno-bruto-departamental/producto-interno-bruto-departamental/

International Energy Agency. 2025. Residential energy use by source. https://www.iea.org/countries/bolivia/efficiency-demand. Source: IEA, World Energy Balances, IEA, Paris. https://www.iea.org/data-and-statistics/data-product/world-energy-balances, Licence: Terms of Use for Non-CC Material.

Ivan K, Holobâcă I-H, Benedek József, Török I. 2020. Potential of night-time lights to measure regional inequality. Remote Sens. 12(1):33. https://doi.org/10.3390/rs12010033

Jasiński T. 2019. Modeling electricity consumption using nighttime light images and artificial neural networks. Energy. 179:831–842. https://doi.org/10.1016/j.energy.2019.04.221

Jimenez Zabalaga P, et al. 2025. Towards a sustainable bolivian energy system in 2050: the pathway for decarbonization under high renewable potential. Energy 337:138335. https://doi.org/10.1016/j.energy.2025.138335

JRC. 2019. GHSL Data Package 2019. LU: Publications Office.

Kuffer M, Karin Pfeffer R, Sliuzas H, Taubenböck I, Baud M, van Maarseveen. 2018. Capturing the urban divide in nighttime light images from the international space station. IJSTA. 11(8):2578–2586. https://doi.org/10.1109/JSTARS.2018.2828340

Li X, et al. 2019. Anisotropic characteristic of artificial light at night – systematic investigation with VIIRS DNB multi-temporal observations. Remote Sens Environ. 233:111357. https://doi.org/10.1016/j.rse.2019.111357

Lin B, Raza MY. 2021. Analysis of electricity consumption in Pakistan using index decomposition and decoupling approach. Energy. 214:118888. https://doi.org/10.1016/j.energy.2020.118888

Lopez G, et al. 2021. Pathway to a fully sustainable energy system for Bolivia across power, heat, and transport sectors by 2050. J Clean Prod. 293:126195. https://doi.org/10.1016/j.jclepro.2021.126195

Lu S, Xiao Y, Lu Y, Lin J. 2025. Spatialization of electricity consumption by combining high-resolution nighttime light remote sensing and urban functional zoning information. Geo-Spatial Information Science. 28(2):527–540. https://doi.org/10.1080/10095020.2024.2356757

Lundberg SM, et al. 2020. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence. 2(1):56–67. https://doi.org/10.1038/s42256-019-0138-9

MacQueen JB. 1967. "Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. Vol 1. University of California Press: Berkeley. p 281–297.

Min B, et al. 2024. Lost in the dark: a survey of energy poverty from space. Joule. 8(7):1982–1998. https://doi.org/10.1016/j.joule.2024.05.001

Min B, Gaba K. 2014. Tracking electrification in Vietnam using nighttime lights. Remote Sens. 6(10):9511–9529. https://doi.org/10.3390/rs6109511

Min B, Gaba KM, Sarr OF, Agalassou A. 2013. Detection of rural electrification in Africa using DMSP-OLS night lights imagery. Int J Remote Sens. 34(22):8118–8141. https://doi.org/10.1080/01431161.2013.833358

Moksnes N, Rozenberg J, Broad O, Taliotis C, Howells M, Rogner H. 2019. Determinants of energy futures—a scenario discovery method applied to cost and carbon emission futures for South American electricity infrastructure. ERCom. 1(2):025001. https://doi.org/10.1088/2515-7620/ab06de

NASA. 2021. VNP46A4 - VIIRS/NPP Lunar BRDF-Adjusted Nighttime Lights Yearly L3 Global 15 Arc Second Linear Lat Lon Grid. NASA VIIRS Land Science Investigator-Led Processing System. https://doi.org/10.5067/VIIRS/VNP46A4.001

Navia M, Orellana R, Zaráte S, Villazón M, Balderrama S, Quoilin S. 2022. Energy transition planning with high penetration of variable renewable energy in developing countries: the case of the bolivian interconnected power system. Energies. 15(3):968. https://doi.org/10.3390/en15030968

OSM. 2023. Open Street Map. https://www.geofabrik.de/

Paparrizos J, Gravano L. 2015. K-Shape: efficient and accurate clustering of time series. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. ACM: Melbourne Victoria Australia. p 1855–1870. https://doi.org/10.1145/2723372.2737793

Parés-Ramos I, Álvarez-Berríos N, Aide T. 2013. Mapping urbanization dynamics in major cities of Colombia, Ecuador, Perú, and Bolivia using night-time satellite imagery. Land. 2(1):37–59. https://doi.org/10.3390/land2010037

Petitjean F, Ketterlin A, Gançarski P. 2011. A global averaging method for dynamic time warping, with applications to clustering. Pattern Recogn. 44(3):678–693. https://doi.org/10.1016/j.patcog.2010.09.013

Polimis K, Rokem A, Hazelton B. 2017. Confidence intervals for random forests in python. JOSS. 2(19):124. https://doi.org/10.21105/joss.00124

Qin Z et al. 2025. Fine-scale mapping and spatiotemporal heterogeneity analysis of industrial CO2 emissions using nighttime light remote sensing in mainland China. Environment, Development and Sustainability. 1–22. https://doi.org/10.1007/s10668-025-06444-5

Rodríguez López F, Guzmán Prudencio G, Marchi Moyano BD, Escalante Pacheco D. 2020. Efectos de La Minería En El Desarrollo Económico, Social y Ambiental Del Estado Plurinacional de Bolivia. CEPAL: Santiago.

Román MO, et al. 2018. NASA's black marble nighttime lights product suite. Remote Sens Environ. 210:113–143. https://doi.org/10.1016/j.rse.2018.03.017

Román MO et al. 2019. Satellite-based assessment of electricity restoration efforts in Puerto Rico after hurricane Maria. In: McFadden J, editor. PLoS One. 14, p e0218883. https://doi.org/10.1371/journal.pone.0218883

Rubin de Celis Cedro R, Espinoza Vasquez L. 2025. Welfare Analysis of Electricity Distribution: The Dignity Tariff in Bolivia. http://dx.doi.org/10.2139/ssrn.5382909. Available at SSRN 5382909.

Sakoe H, Chiba S. 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process. 26(1):43–49. https://doi.org/10.1109/TASSP.1978.1163055

Sanchez Solis C et al. 2023. Electricity demand forecasting for rural communities in developing countries: calibrating a stochastic model for the bolivian case. In: Proceedings of ECOS 2023 - 36th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems. ULPGC: Las Palmas de Gran Canaria, Spain. p 1935–1946. https://doi.org/10.52202/069564-0176http://www.proceedings.com/069564-0176.html

Santillán OS, Cedano KG, Martínez M. 2020. Analysis of energy poverty in 7 Latin American countries using multidimensional energy poverty index. Energies. 13(7):1608. https://doi.org/10.3390/en13071608

Shi K, et al. 2016. Detecting spatiotemporal dynamics of global electric power consumption using DMSP-OLS nighttime stable light data. ApEn. 184:450–463. https://doi.org/10.1016/j.apenergy.2016.10.032

Siksnelyte-Butkiene I, Streimikiene D, Lekavicius V, Balezentis T. 2021. Energy poverty indicators: a systematic literature review and comprehensive analysis of integrity. Sustainable Cities and Society. 67:102756. https://doi.org/10.1016/j.scs.2021.102756

Solis S, et al. 2022. The energy sufficiency concept and its impact on energy demand estimation in rural communities from developing countries. In: 3rd Latin American Conference on Sustainable Development of Energy, Water and Environment Systems. SDEWES Centre: Sao Paulo, Brazil.

Takaku J, Tadono T, Tsutsui K. 2014. Generation of high resolution global DSM from ALOS PRISM. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL–4. 243–248. https://doi.org/10.5194/isprsarchives-XL-4-243-2014

Tan X, Zhu X, Chen J, Chen R. 2022. Modeling the direction and magnitude of angular effects in nighttime light remote sensing. Remote Sens Environ. 269:112834. https://doi.org/10.1016/j.rse.2021.112834

Tavenard R, et al. 2020. Tslearn, A machine learning toolkit for time series data. J Mach Learn Res. 21(118):1–6.

Wager S, Hastie T, Efron B. 2014. Confidence intervals for random forests: the jackknife and the infinitesimal jackknife. The Journal of Machine Learning Research. 15(1):1625–1651.

Wang J, Lu F. 2021. Modeling the electricity consumption by combining land use types and landscape patterns with nighttime light imagery. Energy. 234:121305. https://doi.org/10.1016/j.energy.2021.121305

Wang H, Maruejols L, Yu X. 2021. Predicting energy poverty with combinations of remote-sensing and socio-economic survey data in India: evidence from machine learning. Energ Econ. 102:105510. https://doi.org/10.1016/j.eneco.2021.105510

Wang Z, Shrestha RM, Roman MO, Kalb VL. 2022. NASA's black marble multiangle nighttime lights temporal composites. IEEE Geosci Remote Sens Lett. 19:1–5. https://doi.org/10.1109/LGRS.2022.3176616

Wang Z, Román MO, Sun Q, Molthan AL, Schultz LA, Kalb VL. 2018. Monitoring disaster-related power outages using NASA Black Marble nighttime light product. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII–3. 1853–1856. https://doi.org/10.5194/isprs-archives-XLII-3-1853-2018

Wang Z, Román MO, Kalb VL, Miller SD, Zhang J, Shrestha RM. 2021. Quantifying uncertainties in nighttime light retrievals from Suomi-NPP and NOAA-20 VIIRS Day/Night band data. Remote Sens Environ. 263:112557. https://doi.org/10.1016/j.rse.2021.112557

WorldPop. 2020. Global 1km Population Individual Countries. Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur; Center for International Earth Science Information Network (CIESIN), Columbia University. https://doi.org/10.5258/SOTON/WP00670www.worldpop.org

Wurm M, Weigand M, Stark T, Goebel J, Wagner GG, Taubenböck H. 2019. Modelling the impact of the urban spatial structure on the choice of residential location using 'Big earth data' and machine learning. In: Proceedings of the 2019 Joint Urban Remote Sensing Event (JURSE). IEEE: Vannes, France. p 1–4. 10.1109/JURSE.2019.8808942

Xie Y, Weng Q. 2016. Detecting urban-scale dynamics of electricity consumption at Chinese cities using time-series DMSP-OLS (Defense Meteorological Satellite Program-Operational Linescan System) nighttime light imageries. Energy. 100:177–189. https://doi.org/10.1016/j.energy.2016.01.058