Remote Sensing Applications: Society and Environment 24 (2021) 100647

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

Are night-time lights a good proxy of economic activity in rural areas in middle and low-income countries? Examining the empirical evidence from Colombia

Xaquín S. Pérez-Sindín [a,*], Tzu-Hsin Karen Chen [b], Alexander V. Prishchepov [a]

[a] Department of Geosciences and Natural Resource Management (IGN), University of Copenhagen, Øster Voldgade 10, 1350 København K, Denmark
[b] School of the Environment, Yale University, New Haven, CT 06511, USA

ARTICLE INFO

Keywords:
Remote Sensing
DMSP
VIIRS
Global south
Economic development

ABSTRACT

The use of satellite imagery, particularly night-time lights, has flourished in the last 20 years for socioeconomic studies. The intensity of the lights captured through remote sensing is frequently used as a proxy of different socioeconomic indicators. While some studies found a high correlation between night-time lights intensity and Gross Domestic Product, there has been an inconclusive debate about the validity of this assumption that night-time lights can serve as a good proxy for the economic development to sub-national level studies, particularly in rural areas of middle and lowincome countries. We test the suitability of night-time lights from publicly available data sources for estimating Regional Domestic Product (RDP) across municipalities with different degrees of urbanization in Colombia. We use a series of cross-sectional regression models to compare correlation between municipality RDP and luminosity from different sources for 2012 (DMSP, VIIRS, harmonized DMSP/VIIRS, and harmonized DMSP/VIIRS masked with Global Urban Footprint), as well as multilevel regression models to estimate RDP time-series from 2011 to 2018. Our findings reveal that all compared night-time light products can serve as a good indicator of municipal RDP patterns, while VIIRS data presents the best model fit. Harmonized data that have extensive temporal night-time light records from 2011 to 2018 were significantly correlated with RDP time-series. For seven population size levels – from big cities (>500,000 inhabitants) to rural areas (<5,000 inhabitants), the results present a comparatively higher model fit for urban than rural areas. The use of Global Urban Footprint further improved model fit for large cities but worsened rural RDP estimates. Therefore, our analysis underscores that the use of night-time lights can be a very valuable method to estimate patterns of socioeconomic change at the municipal level in medium and low-income countries and thus may help to implement better sustainability-oriented policies.

1. Introduction

Achieving a better and more sustainable future for the global population is a major societal challenge in the 21st century. In this context, being able to study the patterns of population dynamics and socioeconomic changes plays a fundamental role in achieving sustainable pathways. However, detailed population census data and socioeconomic surveys are rare and infrequent (United Nations

* Corresponding author.
E-mail addresses: xps@ign.ku.dk (X.S. Pérez-Sindín), thc@au.dk (T.-H.K. Chen), alpr@ign.ku.dk (A.V. Prishchepov).

https://doi.org/10.1016/j.rsase.2021.100647
Received 6 May 2021; Received in revised form 13 September 2021; Accepted 4 October 2021

Available online 11 October 2021

Statistics Division, 2019), and in best, are common in developed countries, for instance in the Global North (Pedersen, 2011; Schunck and Rogge, 2010). At the same time, satellite earth observations represent promising methods to overcome infrequent data availability on socioeconomic development, such as night-time lights (further NTL). The application of NTL satellite imagery for socioeconomic studies has expanded rapidly during the last two decades with a range of topics, including mapping urban expansion and population dynamic (Álvarez-Berríos et al., 2013; Imhoff et al., 1997; Small et al., 2005; Zhou et al., 2015; Zhu et al., 2019), tracking electricity consumption (Min et al., 2013; Min and Gaba, 2014), estimating GDP (C. Elvidge et al., 1997; C. D. Elvidge et al., 1997; Forbes, 2013; Li et al., 2013), poverty (Elvidge et al., 2009; Wang et al., 2012; Yu et al., 2015), monitoring disasters (Chen et al., 2019; Román et al., 2019; Zhao et al., 2018), armed conflicts (Li and Li, 2014) and the environmental impacts of light emissions (light pollution) (Rich and Longcore, 2005), including the impacts on human health (Lunn et al., 2017).

However, despite the progress on the applicability of NTL for socioeconomic studies, there is little progress on understanding the relationship between NTL and socioeconomic transition in rural areas, particularly in middle and low-income countries. Sutton et al.'s (1997) work, which compared NTL satellite imagery and population density for the continental United States in the 1990S, can be named as one of the first applications of NTL for socioeconomic studies. Further, by using Elvidge's (1997) global satellite maps of NTL from The Earth Observations Group (EOG) at The National Oceanic and Atmospheric Administration (NOAA)/National Center for Environmental Information (NCEI), Sutton and his colleagues found that saturated pixels of Defense Meteorological Satellite Program (DMSP) NOAA Operational Linescan System (OLS) NTL imagery predict population with R^2 of 0.63. This drew the attention of the remote sensing community to use NTL as a predictor of population density. Similarly, NTL was found as a good predictor of economic activity. The literature abounds with studies examining the correlation between lights intensity and Gross Domestic Product (GDP) (Elvidge et al., 1997; Forbes, 2013; Henderson et al., 2011, 2012; Levin and Zhang, 2017; Li et al., 2013). Despite of inconsistency of NTL observations due to noise and a lack of calibration of NTL data obtained from NOAA OLS, there has been consensus about the usefulness of DMSP data to measure economic activities and thus could add value to conventional economic statistics like national and regional GDP. Hence, NTL is a "bright idea" (Henderson et al., 2011) to measure income at the country level so that economists can solve the "plague [of] serious measurement errors" (p.2) from available data on GDP growth in middle and low-income countries.

Following Henderson's work (2012), who pioneered the applicability of NTL to study economy, many applied economic studies used NTL as an indicator of economic activity. A recent review found over 150 studies in economics using NTL (Gibson et al., 2020), most of them with DMSP data (1992–2012), but also Visible Infrared Imaging Radiometer Suite (VIIRS), a sensor on board Suomi NPP VIIRS/DNB launched in October 2011 to improve DMSP/OLS sensor in terms of data availability and higher spatial resolution (Miller et al., 2012). Yet, there has been an inconclusive debate about whether NTL is a true proxy of economic activities in a different context than country-level comparison. For instance, Jean et al. (2016) noted that NTL have difficulty distinguishing economic characteristics at a local level, namely, poor and densely populated areas versus wealthy and sparsely populated areas. Compared to cross-sections of GDP, Chen and Nordhaus (2019) found that there was more uncertainty in the lights-based time-series of GDP estimates. Also, Bickenbach et al. (2016), using regional data from Brazil and India, find that the relationship between NTL growth and observed GDP growth varies significantly across regions.

More uncertain is, perhaps, the prediction power of NTL in rural areas. After comparing GDP and luminosity at the 1° latitude × 1° longitude grid-cell level for a sample of 170 countries for the period 1992–2008, Chen and Nordhaus (2011) noted that luminosity data do not provide reliable estimates of economic activity for regions with a GDP density below $8,100/km2. Keola et al., (2015) found that DMSP data have a negative correlation with GDP if the agricultural sector represents a large share in GDP. Chen and Nordhaus (2015) also suggested that VIIRS can potentially improve estimations of economic performance in areas with low population density, including parts of Sub-Saharan Africa. After comparing GDP and luminosity across cities and Kabupaten (rural areas of approximately 279.9 persons per km2) in Indonesia, Gibson et al. (2021), stated that although VIIRS data improve DMSP, neither of them can be considered a useful proxy for GDP outside of cities. Yet, they acknowledged that an important question for external validity is whether evidence from other developing countries can corroborate the results from Indonesia.

Colombia, in South America, is among the low-income countries in the Global South and can be an interesting example of testing the suitability of NTL in various socioeconomic settings, such as municipalities with different GDP. Also, Colombia has detailed socioeconomic censuses as early as 2011, which makes it compelling to contrast with the NTL time series. Therefore, by bringing the example of Colombia in our study, the major objective was to assess the suitability of DMSP and VIIRS NTL products to estimate GDP at small geographies. Specifically, we aimed to answer two research questions:

- Are NTL a good proxy of economic activity at the municipality level, including in rural areas?
- Can yearly variation in luminosity predict yearly variation in economic activity at the municipality level? Do the correlations differ by the municipality's population size?

Our paper is structured as follows. First, we introduce the country, its main biophysical features and socioeconomic characteristics. Following the description of the data sets and sources, we will introduce the analysis, results and culminate the paper with a discussion and conclusions.

2. Study context

The territory of the Republic of Colombia is located in the north-western corner of South America and has a continental area of 1,141,748 km^2 (WorldAtlas, 2021). With an estimated population of 50 million people in 2020, Colombia is the third-most populous country in Latin America, after Brazil and Mexico. The orography of Colombia is quite diverse with mountain systems, plains and valleys (Fig. 1). Through its center run the towering, snow-covered volcanoes and mountains of the Andes. There are tropical beaches in the north and west of the country, deserts in the north and vast grasslands, called Los Llanos, in the east. Dense forests fill Colombia's Amazon Basin, which takes up nearly the country's entire southern half. In northwest Colombia, a warm, wet, jungle-filled area called the Chocó reaches across the Panama border (IGAC, 2011). Colombia has three levels of territorial organization: *veredas*, municipalities and departments (Fig. 1). The first group constitutes the subdivisions of the second, and the second group the subdivisions of the last. There are in total 32 departments and 1,122 municipalities, 39% with less than 10,000 inhabitants and 16% with less than 5, 000 (see Table 1). Although approximately 67% of the territory represents a flat landscape, approximately 70% of the population is concentrated in mountainous areas (Fig. 1) (seeTable 2).

Colombia has experienced economic growth over the last two decades. As of 2019, the GDP per capita increased to over USD 6,645, and GDP increased from USD 99.87 billion in 2000 to nearly USD 355 billion in 2019 (International Monetary Funds, 2019). Poverty levels also decreased from 65% in 1990 to 35% by 2019 (The World Bank, 2019). The departments of Vichada and Vaupes, both located in the Amazon Basin, and the jungle-filled area El Chocó exhibit the lowest GDP per capita with USD 1,737, 1,968 and 2,191, respectively, compared to USD 5,912 at national level and 9,809 in the capital region, Bogotá.

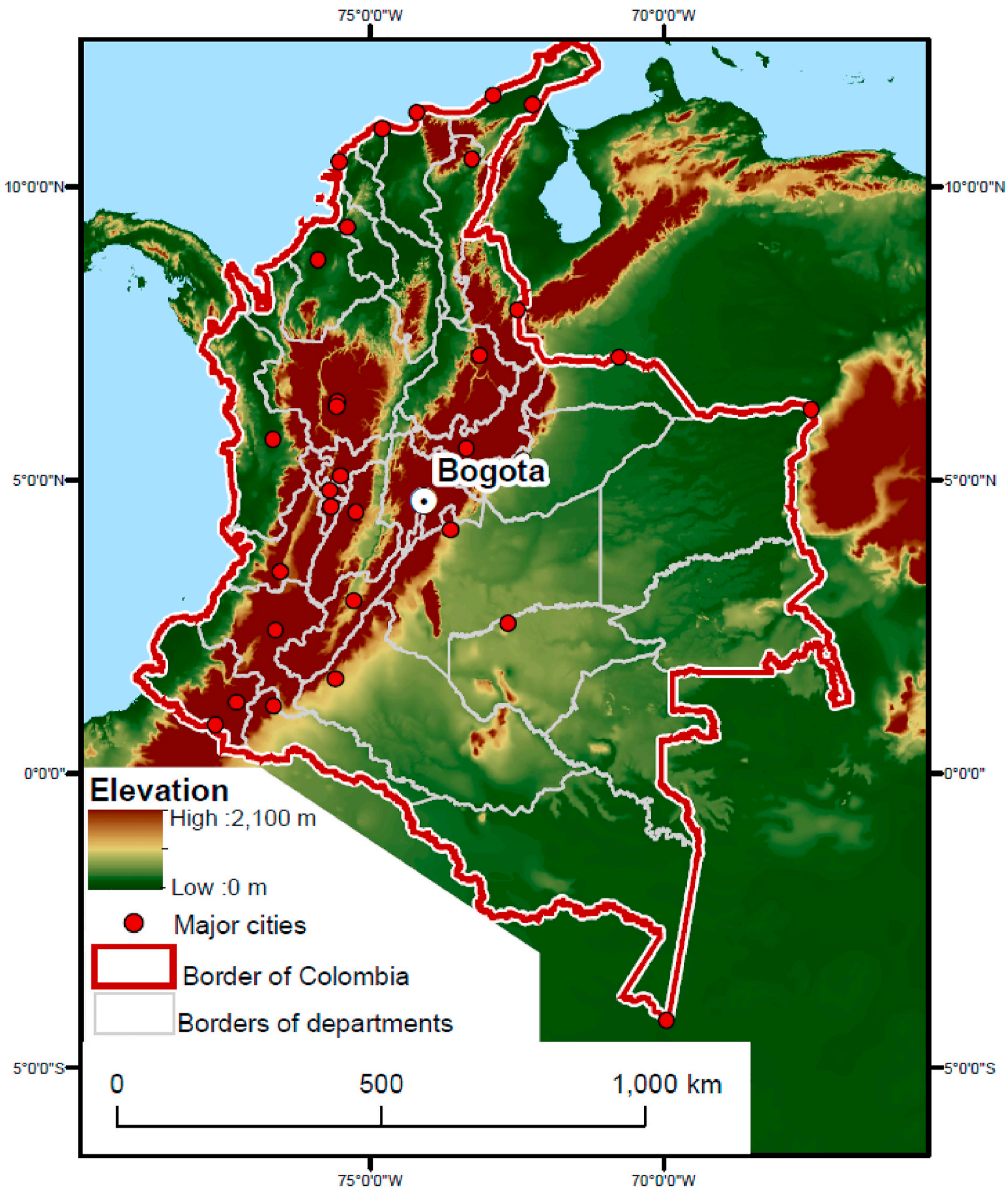

Fig. 1. Physical geography of Colombia, main cities and borders.

Table 1
Descriptive statistics across size-based categories of municipalities.

Number of inhabitants		Pop. (2010) (thousands)	Density (persons per km2)	GDP 2012 (100 Million local currency)	N
500,000 or more	Mean	1,810	3,986	29,035	9
	Std. Dev	2,202	2,764	45,525	
100,000–499,000	Mean	211	932	2,441	51
	Std. Dev	118	2,164	1,804	
50,000–99,999	Mean	67	293	799	58
	Std. Dev	13	461	878	
20,000–49,999	Mean	30	112	336	243
	Std. Dev	8	235	410	
10,000–19,999	Mean	14	58	186	319
	Std. Dev	3	54	909	
5,000–9,999	Mean	7	52	102	259
	Std. Dev	1	47	284	
Less than 5,000	Mean	3	31	37	183
	Std. Dev	1	27	86	
Total	Mean	41	148	540	1,122
	Std. Dev	250	666	4,699	

Source: Colombian National Administrative Department of Statistics https://www.dane.gov.co/

Table 2
Sum of DMSP and VIIRS light across categories of municipalities.

Number of inhabitants	Sum of DMSP lights, 2012	Sum of VIIRS lights, 2012	Number of observations
500,000 or more	18,163	3,902	9
	11,217	2,890	
100,000–499,000	7,566	1,397	51
	5,112	1,252	
50,000–99,999	3,546	645	58
	2,295	746	
20,000–49,999	1,927	645	243
	1,594	746	
10,000–19,999	1,044	275	319
	1,264	400	
5,000–9,999	577	131	259
	636	211	
Less than 5,000	277	125	183
	496	321	
Total	1,566	722	1,122
	2,850	3,184	

Source: NOAA/National Centers for Environmental Information (NCEI)

3. Data and methods

As per the official data, our analysis relies on data on the regional domestic product (further RDP) at the municipal level (In Spanish Valor Agregado Municipal) published by the Colombian National Administrative Department of Statistics (Spanish: Departamento Administrativo Nacional de Estadística; further DANE) (see Fig. 2). RDP accounts are derived from National Accounts of Colombia, which follow the international guidelines and recommendations of the United Nations System of National Accounts (SNA) 2008, the International Monetary Fund and Eurostat. DANE obtains aggregate values for the economic activity by economic branch using either continuous statistics elaborated on a yearly basis and representative on a municipal basis, or non-continuous statistics based on household surveys from which municipal production can be inferred. Then, production is allocated according to national estimates of sectorial technical coefficient. The data is available for the period 2011–2018. Our analysis includes data for 1,122 municipalities grouped in categories according to the degree of urbanization (see Table 1) (seeFig. 3).

3.1. Utilized night-time light data

We used four data types to test the potential relationship between NTL and RDP. Data set#1 is NOAA's VIIRS (further, VIIRS NTL) annual composites for 2012–2019 (Elvidge et al., 2021). Among various versions, we used a masked annual average radiance of VNL V2 data, which is free of biomass burning and other ephemeral events. The data are radiance values in units of nano Watts per square cm per steradian (nanoWatt/cm2/sr) and range from −1.5 to 808.4 for Colombia 2012. Data set#2 is DMSP annual stable light composites for 1992–2012, also from NOAA. The DMSP data are Digital Numbers, ranging from zero to 63, and have no interpretation in terms of radiance values.

Data set#3 is harmonized DMSP and VIIRS NTL product. Despite the valuable records of night-time composites from DMSP (1992–2013) and VIIRS (2012–2018), the inconsistency between both does not allow exploring long-term changes without calibration observations. On these grounds, Li et al. (2020) generated a harmonized NTL data set of both sources that we use in our study to enrich the comparison. The data set is the result of harmonizing the inter-calibrated NTL observations from the DMSP data and, after 2013 –

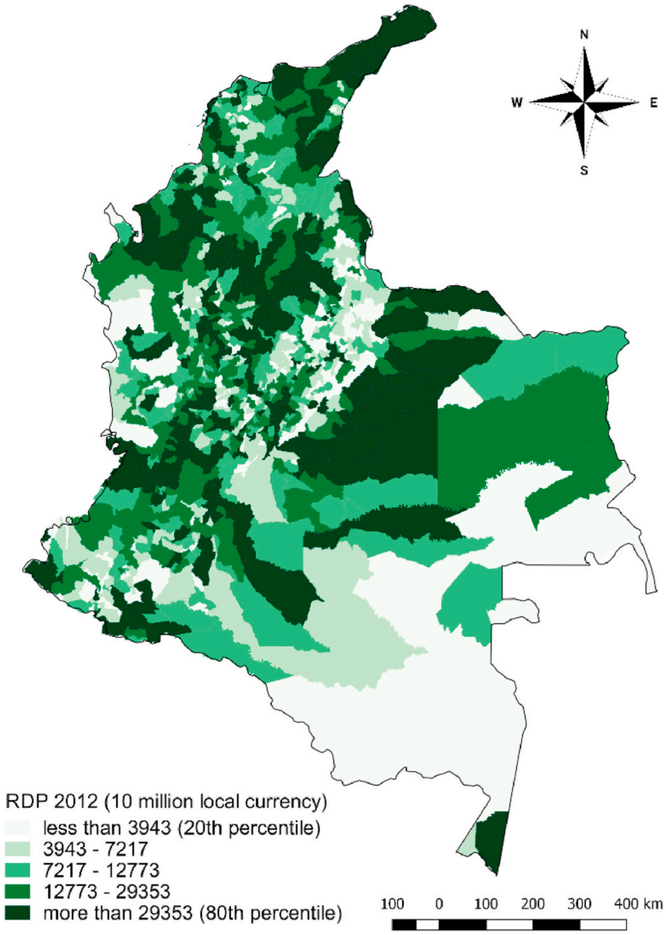

Fig. 2. Map of RDP Municipalities (100 Million Colombian pesos) 2012, top four cities by RDP. Source: Colombian National Administrative Department of Statistics https://www.dane.gov.co/.

the latest year available from DMSP - a simulated DMSP-like night-time lights observation from the VIIRS data.

Finally, with the purpose of improving harmonization DMSP/VIIRS data estimations, we also investigated the correlation between regional GDP and the data resulting from masking the harmonization of DMSP and VIIRS with a layer of residential building (data set#4). The idea behind this layer is to explore whether omitting non-residential lights can help to omit the wrong attribution of light to hinterland areas. For this purpose, we applied the Global Urban Footprint shapefile from the EOC Geoservice of the Earth Observation Center (EOC) of the German Aerospace Center (DLR) (Esch et al., 2017).

3.2. Data processing

We transformed each of the cells of DMSP and VIIRS raster data sets into points as output feature class. Once we processed VIIRS data we replaced 309 cells with negative values for 2012 by zeros. These values were due to the higher resolution of VIIRs composites (0.004167°), compared to DMSP/OLSVIIRS (0.008333°) – which detect moonlit reflected by natural land cover, such as vegetation, water, and desert (Li et al., 2017). We noticed that with DMSP and VIIRS 52 and 61 municipalities were with zero values, respectively (see Tables 3 and 4). Municipalities with zero values are common in Colombia's Amazon Basin (departments of Amazonas, Guainía, Vaupés, Caquetá, in the southeast) and El Chocó (West), which are in areas covered by dense forest and with low population density. We retained zero value-municipalities in our analysis, assuming that zero values reflect absence of economic activity. To be able to work with zero values, we developed correlations using inverse-hyperbolic-sine transformation for the luminosity data (Gibson et al., 2021).

3.3. Data modelling

In the analysis, we use two types of regressions: cross-sectional regression model between municipality's RDP and luminosity with four data sets (DMSP- data set#1, VIIRS- data set#2, DMSP/VIIRS harmonized-data set#3, layer of residential buildings-data set#4), and two panel-data regression models of GDP for the time 2011–2018 with luminosity from DMSP/VIIRS harmonized data-data set#3. Below we provide further details on each of the models.

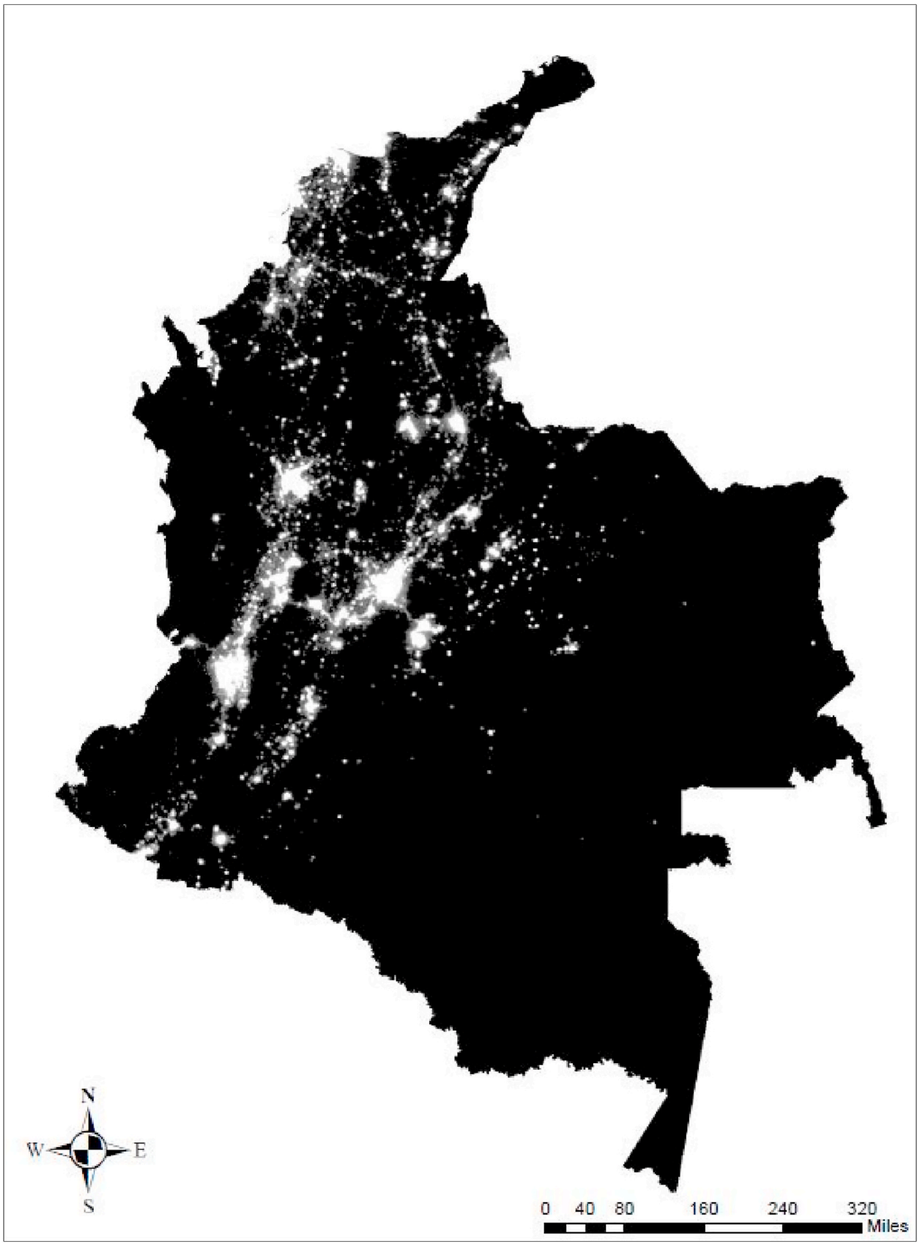

Fig. 3. Night-time lights as detected by satellite (VIIRS annual composite 2012).

3.3.1. Cross-sectional modelling with DMSP and VIIRS

We carried out a bivariate linear regression to test the ability of the selected data sets to predict RDP across the municipalities in Colombia. We used 2012 as a base year, both for DMSP and VIIRS. VIIRS data for 2012 only begin in April; therefore, we used an annual product that includes data from April 2012 to March 2013 to avoid the possible effects of omitting three months in our analysis.

We estimated coefficient and intercept of the linear model with the following equation:

$$\ln(RDP_i) = \alpha + \beta\, NTL_i + \varepsilon_i \tag{1}$$

where $\ln(RDP_i)$ is the dependent variable (natural log values of RDP), i is the unit of analysis (municipalities), α is the intercept, β is the coefficient for NTL, and ε is the error term (residuals).

We estimated the linear model between RDP and light products for each of the population size categories. The results were interpreted in terms of the standardized regression coefficients, as they are expressed in percentage terms and thus omit the effects derived from using NTL products with different units of measurement. A standardized coefficient compares the strength of the effect of each individual independent variable to the dependent variable. The higher the value of the coefficient, the stronger the effect.

Table 3
Description of municipalities with zero value for DMSP 2012

Department	Total municipalities	Number of zero-value municipalities	% over total zero-value municipalities	% cumulated
Chocó	31	11	21.15	
Amazonas	12	8	15.38	36.54
Guainía	10	6	11.54	48.08
Vaupés	7	3	5.77	53.85
Meta	30	1	1.92	55.77
Santander	87	8	15.38	71.15
Antioquia	126	2	3.85	75.00
Cundinamarca	117	2	3.85	78.85
Cauca	43	1	1.92	80.77
Nariño	65	6	11.54	92.31
Boyacán	124	3	5.77	98.08
Casanare	19	1	1.92	100
Total Colombia	1,122	52	100	

Table 4
Description of municipalities with zero value for VIIRS 2012.

Department	Total municipalities	Number of zero-value municipalities	% over total zero-value municipalities	% cumulated
Chocó	31	16	26.23	
Amazonas	12	8	13.11	39,34
Guainía	10	7	11.48	50,82
Vaupés	7	4	6.56	57,38
Caquetá	16	1	1.64	59,02
Guaviare	4	1	1.64	60,66
Meta	30	2	3.28	63.93
Santander	87	2	3.28	67.21
Antioquia	126	4	6.56	73.77
Cundinamarca	117	2	3.28	77.05
Cauca	43	3	4.92	81.97
Nariño	65	7	11.48	93.44
Bolívar	46	1	1.64	95.08
Boyacán	124	3	4.92	100.00
Total Colombia	1,122	61	100	

3.3.2. *Panel data modelling using a multilevel approach*

Further, we used a multilevel approach to examine the trend of level-1 (i.e., year) RDP as a function of NTL (Fig. 4). We also consider the effect of spatial dependency (level-2 municipality-specific) by including random intercepts and random slopes into the model formulation. We analyzed variance partition coefficients by spatial units (i.e., municipality) and temporal units (i.e., year), which showed that 97% of the variance was between municipalities. In contrast, only 1% of the variance was between years. This justified the use of municipalities as the level-2 unit of the multilevel models. The temporal dependency over the years within each municipality was taken into account with a Gaussian serial correlation structure (Diggle et al., 2002).

Thus, the generic structure of the multilevel model with the random intercept effect (multilevel random intercept model) is given by

$$\begin{aligned} &(RDP_{it}|\delta_{0i}) \sim \mathrm{NB}(RDP_{it},\ \theta) \\ &Level1:\ \ln(RDP_{it}) = \beta_{0i} + \beta_1 NTL_{it} \\ &Level2:\ \beta_{0i} = \gamma_{00} + \delta_{0i} \\ &\delta_{0i} \sim N\left(0, \sigma^2\right) \end{aligned} \tag{2}$$

where RDP_{it} represents the RDP in municipality i in year t. β_{0i} refers to the intercept, β_1 is the slope of NTL, δ_{0i} is the level-2 (municipality) random effect for the intercept, and θ is the shape parameter that can control for over-dispersion. The formulation of model

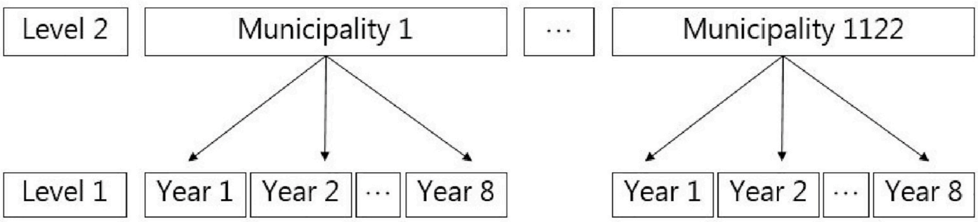

Fig. 4. The multilevel model structure of this study.

Table 5
RDP and luminosity (bivariate linear regression).

	Standardized coefficients	Intercept	R^2	N
DMSP 2012				
All municipalities	0.710***	22.089***	0.504	1,122
	(0.014)	(0.104)		
500,000 or more inhabitants	0.739*	17.613**	0.547	9
	(0.423)	(4.370)		
100,000–499,000	0.654***	21.012***	0.428	51
	(0.127)	(1.204)		
50,000–99,999	0.559***	20.923***	0.313	58
	(0.141)	(1.228)		
20,000–49,999	0.564***	23.525***	0.318	243
	(0.032)	(0.253)		
10,000–19,999	0.531***	23.746***	0.282	319
	(0.021)	(0.149)		
5,000–9,999	0.463***	23.248***	0.214	259
	(0.029)	(0.194)		
Less than 5,000 inhabitants	0.536***	22.959***	0.288	183
	(0.022)	(0.121)		
VIIRS Annual Composites for 2012				
All municipalities	0.827***	23.214***	0.684	1,122
	(0.012)	(0.051)		
500,000 or more inhabitants	0.934***	18.043***	0.873	9
	(0.193)	(1.772)		
100,000–499,000	0.740***	23.705***	0.547	51
	(0.082)	(0.599)		
50,000–99,999	0.612***	23.686***	0.374	58
	(0.097)	(0.597)		
20,000–49,999	0.665***	24.480***	0.443	243
	(0.027)	(0.128)		
10,000–19,999	0.570***	24.323***	0.324	319
	(0.024)	(0.091)		
5,000–9,999	0.494***	23.920***	0.244	259
	(0.035)	(0.109)		
Less than 5,000	0.555***	23.172***	0.309	183
	(0.041)	(0.096)		
Harmonized DMSP and VIIRS for 2012				
All municipalities	0.710***	22.089***	0.504	1,122
	(0.014)	(0.104)		
500,000 or more inhabitants	0.739*	17.613**	0.547	9
	(0.423)	(4.370)		
100,000–499,000	0,654***	21.012***	0.428	51
	(0.127)	(1.204)		
50,000–99,999	0.559***	20.923***	0.313	58
	(0.141)	(1.228)		
20,000–49,999	0.564***	23.525***	0.318	243
	(0.032)	0.253		
10,000–19,999	0.531***	23.746***	0.282	319
	(0.021)	(0.149)		
5,000–9,999	0.618***	24417***	0.382	259
	(0.000)	(0.049)		
Less than 5,000 inhabitants	0.536***	22.959***	0.288	183
	(0.022)	(0.121)		
Harmonized DMSP and VIIRS for 2012 masked with GUF				
All municipalities	0.730***	24.534***	0.533	1122
	(0.011)	(0.038)		
500,000 or more inhabitants	0.940***	18.724***	0.884	9
	(0.176)	(1.587)		
100,000–499,000	0.727***	23.039***	0.528	51
	(0.099)	(0.712)		
50,000–99,999	0.350**	25.779***	0.123	58
	(0.082)	(0.483)		
20,000–49,999	0.491***	25,566***	0.241	243
	(0.020)	(0.082)		
10,000–19,999	0.376***	25.094***	0.141	319
	(0.019)	0.053		
5,000–9,999	0.180***	24.755***	0.032	259

(*continued on next page*)

Table 5 (*continued*)

	Standardized coefficients	Intercept	R^2	N
	(0.032)	(0.054)		
Less than 5,000 inhabitants	0.185***	23.827***	0.034	183
	(0.064)	(0.067)		

Standard errors in ( ), ***, **, and * denote statistical significance at 1%, 5% and 10% levels.

(1) defines the expected value RDP_{it} as being associated with the level-1 NTL predictor NTL_{it} with the corresponding fixed effects β_1 and γ_{00}. Following Diggle et al. (2002) the correlation between pairs of yearly measurements in the same municipality is assumed to have a serial structure which is a Gaussian decreasing function of the time intervals. That is, for two given years t and t' at municipality i, we assume

$$\mathrm{Corr}(RDP_{\mathrm{it}}, RDP_{\mathrm{it'}}) = \tau^2 \exp\left(-\varphi|t-t'|^2\right) \tag{3}$$

for some unknown constant serial variance τ^2 and association parameter φ. This specification considers that count measurements taken close together in time are typically more strongly correlated than those taken further apart in time (Diggle et al., 2002).

To explore whether the marginal effects of NTL across municipalities, we further consider the following slope effect model (multilevel random intercept + random slope model):

$$\begin{aligned}
&(RDP_{it}|\delta_{0i}) \sim \mathrm{NB}(RDP_{it}, \theta) \\
&Level1: \ \ln(RDP_{it}) = \beta_{0i} + \beta_{1i} NTL_{it} \\
&Level2: \ \beta_{0i} = \gamma_{00} + \delta_{0i} \\
&\beta_{1i} = \gamma_{10} + \delta_{1i} \\
&\delta_{0i} \sim N\left(0, \sigma^2\right)
\end{aligned} \tag{4}$$

where γ_{10} denotes the coefficient of NTL and δ_{1i} is random effect for the slope of NTL. In contrast to model (2), model (4) includes the marginal effect of NTL to explore the spatial variability of the NTL correlation with RDP. We used likelihood ratio test to select the model with a better fit.

4. Results

4.1. Cross-sectional results

We found a statistically significant relationship between RDP and different products of NTL for Colombia's municipalities, with an estimated standardized regression coefficient of 0.710 for DMSP (data set #1), 0.827 for VIIRS (data set #2), 0.710 for harmonized DMSP/VIIRS (data set #3) and 0.730 for masked harmonized (data set #4) (Table 5). All coefficients were significantly positive for the analysis aggregating all municipalities. Overall, this finding indicates that light intensity can be a good proxy of RDP.

As to analysis across different population size categories, there was a statistically significant relationship between DMSP, VIIRS,

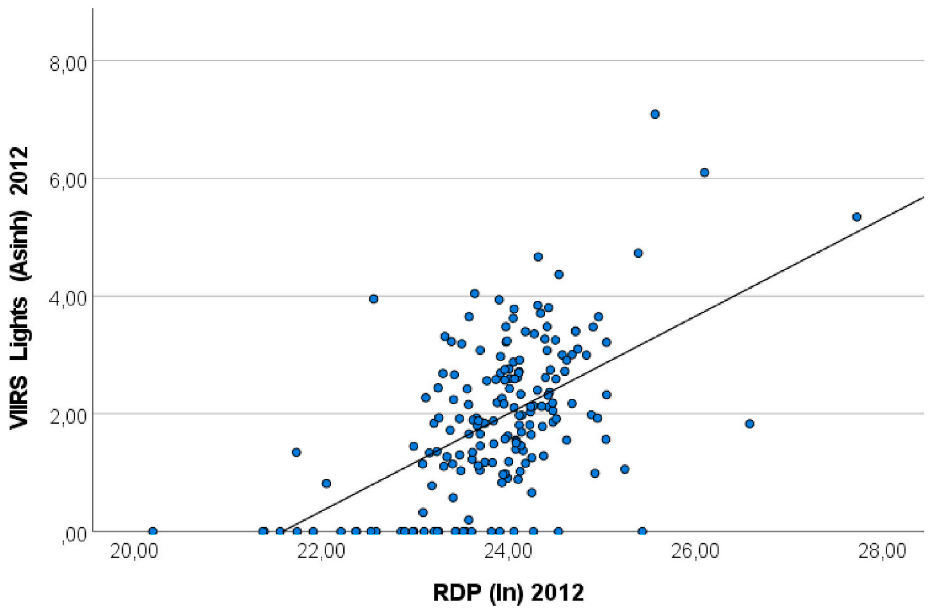

Fig. 5. RDP (Ln, natural logarithm values) luminosity (VIIRS, Asinh, inverse hyperbolic sine function) for municipalities with less than 5,000 inhabitants, 2012

harmonized DMSP/VIIRS and masked harmonized and real RDP for all categories of municipalities. The municipalities with 500,000 or higher population numbers had the highest regression coefficient and model fit in each of data types. Overall, the coefficient decreases as we went down in the population size of municipalities. VIIRS NTL gave the highest coefficient for rural municipalities (less than 5,000 inhabitants), namely 0.555, followed by DMSP NTL and harmonized DMSP/VIIRS (0.536) and harmonized masked with GUF, (0.185). In contrast, harmonized masked with GUF gave the highest coefficient (0.94) for metropolises (more than 500,000 inhabitants), followed by VIIRS NTL (0.934), DMSP NTL and harmonized DMSP/VIIRS (both were 0.739) (Fig. 5).

4.2. Panel-data estimation

The constructed multilevel model with random slopes and intercept fitted significantly better than the model only with random intercepts ($p < 0.001$, log-likelihood ratio test) (Table 6).

The map on the left side of Fig. 6 represents the baseline RDP across municipalities, while the map on the right side indicates which municipalities have higher correlation between NTL and RDP. The marginal slope (right side map, Fig. 6) shows that those municipalities in Colombia's Amazon Basin (southeast), as well as those within the jungle-filled area of El Chocó (West), exhibits a lower correlations between NTL and RDP. Furthermore, the marginal effects for slopes and intercepts were both more significant in highly populated municipalities (see Fig. 6 scatter plots).

As per differences across population size categories, Fig. 7 provides evidence that NTL correlates with RDP time-series at a higher rate in the metropolises. The positive mean random slopes in the municipalities with a population below ten and five thousand suggest that NTL has a positive correlation with RDP time series even in rural areas.

5. Discussion

Overall, DMSP, VIIRS, harmonized DMSP/VIIRS were good proxies of economic activity across the municipalities. The estimated standardized coefficients between night-time light and RDP - of some 0.710 for DMSP, 0.827 for VIIRS, 0.710 for harmonized DMSP/VIIRS and 0.730 for masked harmonized with a layer of human settlements - were relatively high. Our results confirm that the greater spatial resolutions of VIIRS data outperform DMSP and harmonized DMSP/VIIRS data in estimating economic activity at sub-national level, while harmonized DMSP/VIIRS data allow a prediction of economic time series with good performance.

In terms of our first research question ("*Are NTL a good proxy of economic activity at the municipality level, including in rural areas?*"), three among the four data sets (i.e., DMSP, VIIRS, and harmonized DMSP/VIIRS) served as a good proxy of economic activity in rural areas of Colombia, despite the lower model fit compared to that of metropolises. VIIRS exhibited the highest prediction power for rural areas (less than 5,000 inhabitants), with a standardized coefficient of 0.555 and $R2 = 0.309$. Moreover, while early findings in the literature show the low performance of NTL in rural areas with a density mean of 280 persons per km2 (Gibson et al., 2021), our results showed evidence that NTL correlated with RDP in areas with such low population density as 31 persons per km2, that is municipalities with less than 5,000 inhabitants. Consistent with previous studies (e.g., Chen and Nordhaus (2015) study on population and economic output in Africa), our analysis also indicated that VIIRS lights is better than DMSP in estimating economic activity at sub-national level and thus confirmed that this product is a promising supplementary source for estimation of RDP.

We noticed that GUF may help to exclude non-residential areas and, therefore, improve the prediction of RDP for large cities with more than 100,000 inhabitants, but not for rural areas. This is possibly because the human settlement layer is less accurate and complete for rural areas, when sparse settlements and separate households may not be included into GUF (Global Urban Footprint, Mück et al., 2017).

For our second research question "*Can yearly variation in luminosity predict yearly variation in economic activity at municipal level?*", our mixed-effect model indicates that yearly variation in harmonized DMSP/VIIRS luminosity predict yearly variation in RDP. While the statistical relationship is more significant for municipalities with 500,000 or more inhabitants, it can also predict temporal variations in less-populated areas.

This study complements the previous efforts made on correlating NTL and GDP growth that only used the growth rate between two years as the dependent variable. Here, we used observations across eight years and a multilevel model for dealing with the temporal dependency existing in time-series data. Our findings revealed that NTL is a good indicator of inter-annual variance of RDP in a longer period (i.e., eight years). The correlation with RDP time-series was higher in urban municipalities and decreased for rural areas, but remains positive. It is possible to conclude that NTL, particularly the harmonized DMSP/VIIRS product, is a good method to measure long-term patterns of socioeconomic change at very small geographies in Colombia and, potentially, other countries.

One limitation of our study was the uncertainty coming from the use of inter-calibrated NTL products. First, as argued by Gibson et al. (2021), the measurement errors in DMSP data are mean reverting. Brightness variances between places are understated in DMSP data and harmonized DMSP/VIIRS data due to the low spatial resolution and top-coded values. Moreover, if there is a sufficiently strong degree of mean-reversion in a right-hand side variable, the regression coefficient on that variable can be overstated rather than having the normal attenuation bias from random measurement error. These mean-reverting errors - which can potentially bias the regression coefficients here estimated – also make the DMSP/VIIRS calibration difficult. Given the degree of error depending on true luminosity, there is no single adjustment available to align DMSP and VIIRS. Finally, additional uncertainties were added to the analysis of the models using harmonized DMSP/VIIRS data, because it is an estimated variable derived from a statistical model instead of observed values, also known as a generated regressor problem (Pagan, 1984; Sobiech, 2019). The use of generated regressors can potentially bias standard errors toward zero, producing type I errors. Hence, more efforts are needed to reduce the uncertainty and correct the bias associated to the use of inter-calibrated products.

We also notice, despite the significant correlation between NTL and RDP in rural areas, NTL data exhibited the lowest performance

Table 6
Multilevel model results on RDP and harmonized NTL from 2011 to 2018.

	Log (sum of lights)it	R^2	N
	0.006		
Multilevel random intercept model	0.012***	0.903	8976
	0.002	(conditional)	
Multilevel random intercept + random slope model	0.034***	0.912	8976
	0.003	(conditional)	

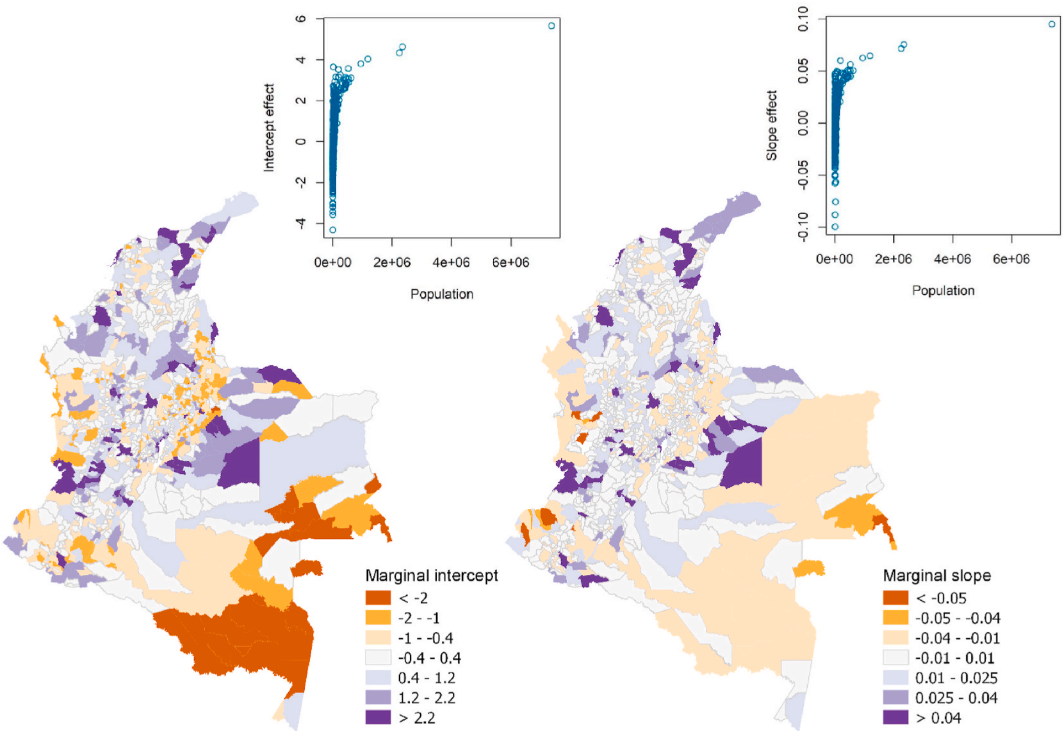

Fig. 6. Mixed-effect model with random slopes and intercept.

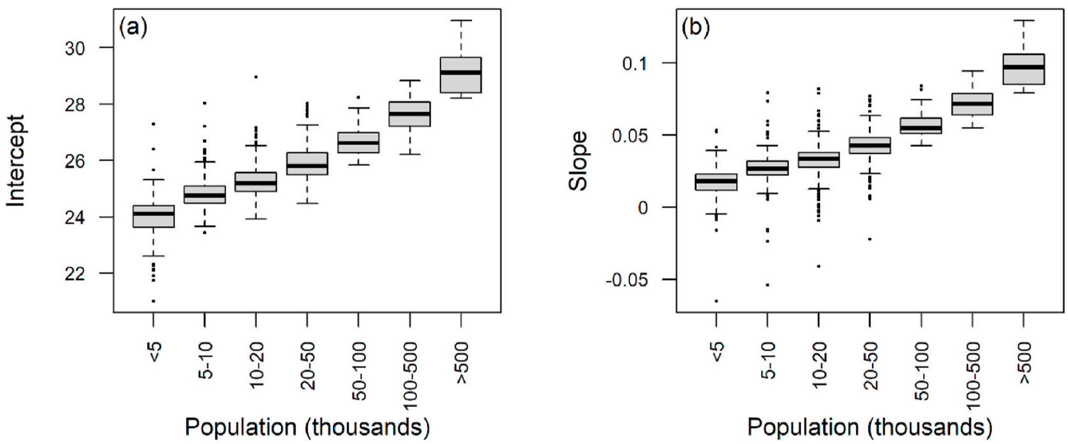

Fig. 7. (a) The intercept and (b) slope values from multilevel random intercept + random slope model by population level. These values were weighted by adding the random effect (municipality-specific coefficient) to the fixed effect (overall coefficient).

for municipalities located in areas densely covered by forest and with low population density, i.e., Colombia's Amazon Basin (southeast) and El Chocó (West). Taking into account that these are the areas with a higher proportion of zero value municipalities – as we observed after visualizing the values extracted from DMSP and VIIRS annual composites (Tables 3 and 4) – our findings confirm that dense vegetation can be a potential limitation for NTL applications.

Finally, the quality of inputs to estimate GDP in Colombia might have influenced the results of this study. It is worth referring Chen and Nordhaus's (2011) study on the utility of luminosity as a proxy of GDP around the world. After comparing the GDP reported by national statistical accounts and luminosity, the authors graded Colombia with a C in a scale from A to E, according to the quality of the statistical systems. A country graded A in this study was, for instance, the United States, where detailed accounts can come from electronic transactions and administrative data sets at fine level. Hence, more efforts are needed to develop transparency and reliability of RDP data sets in the middle and low-income countries.

6. Conclusions

This article explores the suitability of NTL to estimate economic activity at small geographies in low-income country-Colombia. Despite the consensus about the usefulness of NTL to measure GDP in the literature, there has been an inconclusive debate about whether NTL is a true proxy of economic activities at sub-national levels, particularly at local levels and rural areas of middle and low-income countries. In this study, we probed that despite the limited capacity to measure economic activities in regions densely covered by forest and with low population density, overall, NTL is a good proxy of economic activity in rural areas of Colombia. Among publicly available data sets, VIIRS data shows the best fit for country-wide RDP predictions and rural municipality predictions, whilst the mask of Global Urban Footprint on NTL products can enhance the predictions for large cities. The bottom line is that, while the case of Colombia should be compared with a wider set of countries – from different sub-regions of the world and with different geographical characteristics - NTL should not be underestimated as a method to overcome the lack of population census data and socioeconomic surveys at very fine levels and outside cities in middle and low-income countries. Moreover, the results of this research show the potential of NTL as a method to measure long terms patterns of socioeconomic changes, something that can be very useful to better understand the before-after impacts of interventions of different nature (infrastructures, policies etc.), as well as refine and design new sustainability-oriented policies.

Credit author statement

Xaquin S. Perez-Sindin: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization. Tzu-Hsin Karen Chen: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - Review & Editing, Visualization. Alexander Prishchepov: Conceptualization, Methodology, Resources, Writing - Review & Editing, Visualization.

Ethical statement

Hereby, I Xaquín Pérez-Sindín consciously assure that for the manuscript "Are nighttime lights a good proxy of economic activity in rural areas in middle and low-income countries? Examining the empirical evidence from Colombia" the following is fulfilled:

- Ethical practices have been followed in relation to the development, writing, and publication of the article.
- Manuscript has been 'spell checked' and 'grammar checked'
- All references mentioned in the Reference List are cited in the text, and vice versa
- Permission has been obtained for use of copyrighted material from other sources (including the Internet)
- A competing interests statement is provided, even if the authors have no competing interests to declare
- Journal policies detailed in this guide have been reviewed
- Referee suggestions and contact details provided, based on journal requirements

Declaration of competing interest

No competing interests.

Acknowledgement

We would like to thank the reviewers and the editor for their thoughtful comments and efforts towards improving our manuscript. The European Commission's Marie Skłodowska-Curie Individual Fellowship Action supported this work [Grant Agreement number: 795302 — ReCurse].

References

Álvarez-Berríos, N.L., Parés-Ramos, I.K., Aide, T.M., 2013. Contrasting patterns of urban expansion in Colombia, Ecuador, Peru, and Bolivia between 1992 and 2009. Ambio 42, 29–40. https://doi.org/10.1007/s13280-012-0344-8.

Bickenbach, F., Bode, E., Nunnenkamp, P., Söder, M., 2016. Night lights and regional GDP. Rev. World Econ. 152, 425–447. https://doi.org/10.1007/s10290-016-0246-0.

Chen, T.H.K., Prishchepov, A.V., Fensholt, R., Sabel, C.E., 2019. Detecting and monitoring long-term landslides in urbanized areas with nighttime light data and multi-seasonal Landsat imagery across Taiwan from 1998 to 2017. Remote Sens. Environ. 225, 317–327. https://doi.org/10.1016/j.rse.2019.03.013.

Chen, X., Nordhaus, W., 2015. A test of the new VIIRS lights data set: population and economic output in Africa, 2015 Rem. Sens. 7, 4937–4947. https://doi.org/10.3390/RS70404937, 7, 4937–4947.
Chen, X., Nordhaus, W.D., 2019. VIIRS nighttime lights in the estimation of cross-sectional and time-series GDP. Rem. Sens. 11, 1057. https://doi.org/10.3390/rs11091057.
Chen, X., Nordhaus, W.D., 2011. Using luminosity data as a proxy for economic statistics. Proc. Natl. Acad. Sci. U.S.A. 108, 8589–8594. https://doi.org/10.1073/pnas.1017031108.
Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S., 2002. Analysis of Longitudinal Data. Oxford University Press, Oxford.
Elvidge, C., Baugh, K., Kihn, E., Kroehl, H., Davis, E., Davis, C., 1997. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. Int. J. Rem. Sens. 18, 1373–1379. https://doi.org/10.1080/014311697218485.
Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W., Davis, E.R., 1997. Mapping city lights with nighttime data from the DMSP Operational Linescan System. Photogramm. Eng. Rem. Sens. 63, 727–734.
Elvidge, C.D., Sutton, P.C., Ghosh, T., Tuttle, B.T., Baugh, K.E., Bhaduri, B., Bright, E., 2009. A global poverty map derived from satellite data. Comput. Geosci. 35, 1652–1660. https://doi.org/10.1016/j.cageo.2009.01.009.
Elvidge, C.D., 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. Rem. Sens. 13, 922. https://doi.org/10.3390/rs13050922.
Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., Strano, E., 2017. Breaking new ground in mapping human settlements from space – the Global Urban Footprint. ISPRS J. Photogrammetry Remote Sens. 134, 30–42. https://doi.org/10.1016/j.isprsjprs.2017.10.012.
Forbes, D.J., 2013. Multi-scale analysis of the relationship between economic statistics and DMSP-OLS night light images. GIScience Remote Sens. 50, 483–499. https://doi.org/10.1080/15481603.2013.823732.
Gibson, J., Olivia, S., Boe-Gibson, G., Li, C., 2021. Which night lights data should we use in economics, and where? J. Dev. Econ. 149, 102602. https://doi.org/10.1016/j.jdeveco.2020.102602.
Gibson, J., Olivia, S., Boe-Gibson, G., 2020. Night lights in economics: sources and uses. J. Econ. Surv. 34, 955–980. https://doi.org/10.1111/joes.12387.
Henderson, J.V., Storeygard, A., Weil, D.N., 2012. Measuring economic growth from outer space. Am. Econ. Rev. 102, 994–1028. https://doi.org/10.1257/aer.102.2.994.
Henderson, V., Storeygard, A., Weil, D.N., 2011. A bright idea for measuring economic growth. In: American Economic Review, pp. 194–199. https://doi.org/10.1257/aer.101.3.194.
IGAC, 2011. Geografía de Colombia. IGAC. Instituto Geográfico Agustín Codazzi, Bogotá.
Imhoff, M.L., Lawrence, W.T., Stutzer, D.C., Elvidge, C.D., 1997. A technique for using composite DMSP/OLS "city lights" satellite data to map urban area. Remote Sens. Environ. 61, 361–370. https://doi.org/10.1016/S0034-4257(97)00046-1.
International Monetary Funds, 2019. World Economic Outlook Database. Washington, DC.
Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science 84 353, 790–794. https://doi.org/10.1126/science.aaf7894.
Keola, S., Andersson, M., Hall, O., 2015. Monitoring economic development from space: using nighttime light and land cover data to measure economic growth. World Dev. 66, 322–334. https://doi.org/10.1016/j.worlddev.2014.08.017.
Levin, N., Zhang, Q., 2017. A global analysis of factors controlling VIIRS nighttime light levels from densely populated areas. Remote Sens. Environ. 190, 366–382. https://doi.org/10.1016/j.rse.2017.01.006.
Li, X., Li, D., 2014. Can night-time light images play a role in evaluating the Syrian Crisis? Int. J. Rem. Sens. 35, 6648–6661. https://doi.org/10.1080/01431161.2014.971469.
Li, X., Li, D., Xu, H., Wu, C., 2017. Intercalibration between DMSP/OLS and VIIRS night-time light images to evaluate city light dynamics of Syria's major human settlement during Syrian Civil War. Int. J. Rem. Sens. 38, 5934–5951. https://doi.org/10.1080/01431161.2017.1331476.
Li, X., Xu, H., Chen, X., Li, C., 2013. Potential of NPP-VIIRS nighttime light imagery for modeling the regional economy of China. Rem. Sens. 5, 3057–3081. https://doi.org/10.3390/rs5063057.
Li, X., Zhou, Y., Zhao, M., Zhao, X., 2020. A harmonized global nighttime light dataset 1992–2018. Sci. Data 7, 1–9. https://doi.org/10.1038/s41597-020-0510-y.
Lunn, R.M., Blask, D.E., Coogan, A.N., Figueiro, M.G., Gorman, M.R., Hall, J.E., Hansen, J., Nelson, R.J., Panda, S., Smolensky, M.H., Stevens, R.G., Turek, F.W., Vermeulen, R., Carreón, T., Caruso, C.C., Lawson, C.C., Thayer, K.A., Twery, M.J., Ewens, A.D., Garner, S.C., Schwingl, P.J., Boyd, W.A., 2017. Health consequences of electric lighting practices in the modern world: a report on the National Toxicology Program's workshop on shift work at night, artificial light at night, and circadian disruption. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2017.07.056.
Miller, S.D., Mills, S.P., Elvidge, C.D., Lindsey, D.T., Lee, T.F., Hawkins, J.D., 2012. Suomi satellite brings to light a unique frontier of nighttime environmental sensing capabilities. Proc. Natl. Acad. Sci. U.S.A. 109, 15706–15711. https://doi.org/10.1073/pnas.1207034109.
Min, B., Gaba, K., 2014. Tracking electrification in vietnam using nighttime lights. Rem. Sens. 6, 9511–9529. https://doi.org/10.3390/rs6109511.
Min, B., Gaba, K.M., Sarr, O.F., Agalassou, A., 2013. Detection of rural electrification in Africa using DMSP-OLS night lights imagery. Int. J. Rem. Sens. 34, 8118–8141. https://doi.org/10.1080/01431161.2013.833358.
Mück, M., Klotz, M., Taubenbock, H., 2017. Validation of the DLR Global Urban Footprint in rural areas: a case study for Burkina Faso. In: Joint Urban Remote Sensing Event (JURSE). Institute of Electrical and Electronics Engineers Inc., pp. 1–4. https://doi.org/10.1109/JURSE.2017.7924618
Pagan, A., 1984. Econometric issues in the analysis of regressions with generated regressors. Int. Econ. Rev. 25, 221. https://doi.org/10.2307/2648877.
Pedersen, C.B., 2011. The Danish civil registration system. Scand. J. Publ. Health 39, 22–25. https://doi.org/10.1177/1403494810387965.
Rich, C., Longcore, T., 2005. Ecological Consequences of Artificial Night Lighting. Island Press, Washington, DC.
Román, M.O., Stokes, E.C., Shrestha, R., Wang, Z., Schultz, L., Carlo, E.A.S., Sun, Q., Bell, J., Molthan, A., Kalb, V., Ji, C., Seto, K.C., McClain, S.N., Enenkel, M., 2019. Satellite-based assessment of electricity restoration efforts in Puerto Rico after Hurricane Maria. PLoS One 14, e0218883. https://doi.org/10.1371/journal.pone.0218883.
Schunck, R., Rogge, B.G., 2010. Unemployment and its association with health-relevant actions: investigating the role of time perspective with German census data. Int. J. Publ. Health 55, 271–278. https://doi.org/10.1007/s00038-009-0075-1.
Small, C., Pozzi, F., Elvidge, C.D., 2005. Spatial analysis of global urban extent from DMSP-OLS night lights. Remote Sens. Environ. 96, 277–291. https://doi.org/10.1016/j.rse.2005.02.002.
Sobiech, I., 2019. Remittances, finance and growth: does financial development foster the impact of remittances on economic growth? World Dev. 113, 44–59. https://doi.org/10.1016/J.WORLDDEV.2018.08.016.
Sutton, P., Roberts, D., Elvidge, C., Meij, H., 1997. A comparison of nighttime satellite imagery and population density for the continental United States. Photogramm. Eng. Rem. Sens. 63, 1303–1313.
The World Bank, 2019. Poverty headcount ratio at national poverty lines (% of population) - Colombia [WWW Document]. URL. https://data.worldbank.org/indicator/SI.POV.NAHC.
United Nations Statistics Division, 2019. Guidelines on the Use of Electronic Data Collection Technologies in Population and Housing Censuses. New York.
Wang, W., Cheng, H., Zhang, L., 2012. Poverty assessment using DMSP/OLS night-time light satellite imagery at a provincial scale in China. Adv. Space Res. 49, 1253–1264. https://doi.org/10.1016/j.asr.2012.01.025.
WorldAtlas, 2021. Maps Of Colombia [WWW Document]. Orig. Online Geogr. Resour. URL. https://www.worldatlas.com/maps/colombia.
Yu, B., Shi, K., Hu, Y., Huang, C., Chen, Z., Wu, J., 2015. Poverty evaluation using NPP-VIIRS nighttime light composite data at the county level in China. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 1217–1229. https://doi.org/10.1109/JSTARS.2015.2399416.

Zhao, X., Yu, B., Liu, Y., Yao, S., Lian, T., Chen, L., Yang, C., Chen, Z., Wu, J., 2018. NPP-VIIRS DNB daily data in natural disaster assessment: evidence from selected case studies. Rem. Sens. 10, 1526. https://doi.org/10.3390/rs10101526.
Zhou, Y., Smith, S.J., Zhao, K., Imhoff, M., Thomson, A., Bond-Lamberty, B., Asrar, G.R., Zhang, X., He, C., Elvidge, C.D., 2015. A global map of urban extent from nightlights. Environ. Res. Lett. 10, 054011 https://doi.org/10.1088/1748-9326/10/5/054011.
Zhu, Z., Zhou, Y., Seto, K.C., Stokes, E.C., Deng, C., Pickett, S.T.A., Taubenböck, H., 2019. Understanding an urbanizing planet: strategic directions for remote sensing. Remote Sens. Environ. 228, 164–182. https://doi.org/10.1016/j.rse.2019.04.020.