Remote night-time lights sensing: Investigation and econometric application

Orientation: Some recent studies have been published that demonstrated the value of remote sensing night-time lights as descriptors and/or proxies for human activity.

Research purpose: This article investigated the association between night-time light emissions and gross domestic product (GDP) estimates for South Africa.

Motivation for the study: Satellite night-lights data seemed to be a useful proxy for economic activity at temporal and geographic scales for which traditional data are of poor quality, are unavailable or only available with a large time lag.

Research approach/design and method: The article primarily used the remote sensing of night-time light emissions using satellite technologies. The methodology employed in this study involved estimating both a vector error correction modelling (VECM) and autoregressive distributed lag (ARDL) models that map light growth into a proxy for GDP growth.

Main findings: Both the VECM and ARDL models confirmed a long-term co-integrating relationship between GDP (per capita) and night-time lights (total light intensity), a statistically significant short-term error correction term could, however, not be established through the VECM, but indeed through the ARDL model.

Practical/managerial implications: The results of the study suggested that satellite remote sensing technologies held much promise and opportunities in terms of the field of Economics and Development.

Contribution/value-add: This study contributes to our understanding of the spatial and temporal behaviour and trends in economic activity. It also suggested the use of satellite remote sensing technologies as part of official statistical frameworks and methodologies.

Keywords: night lights; light pollution; DMSP/OLS; VIIRS/DNB; International Space Station; urban human activity.

Introduction
This article focuses on examining the economics and economic application of the remote sensing of night-time light (NTL) emissions using satellite technologies within the South African context. Levin et al. (2020) argue that the use of NTL emissions in the visible band offers a unique opportunity to directly observe human activity from space. They further argue that numerous applications of remotely sensed NTLs may be found in the field of economics. Although some studies have been conducted internationally, such a study has not yet been conducted within a South African context. It, therefore, promises to open a new field of study within the South African context. The study also applies several econometric techniques not utilised in previous studies.

Remote sensing products such as NTL emissions are primarily derived from satellite sensors. Two sensors have been collecting NTL emissions from as early as the mid-1970s up to the present. However, the actual data have only been released since 1992. The United States of America (USA) Air Force Defence Meteorological Satellite Program (DMSP) operated the Linescan satellite system (Operational Linescan System [OLS]) originally until 2013. In 2011, the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA) launched the Suomi National Polar Partnership (SNPP) infrared imaging radiometer suite (Visible Infrared Imaging Radiometer Suite [VIIRS]) as the DMSP successor (Jeswani 2018).

Continuous time-series data of global remote sensing now span from 1992 up to the present; however, using two different systems that, to some degree, are not 100% compatible. This is because the VIIRS system is a significant upgrade to the OLS system. Nonetheless, there are also
significant similarities between the two systems to allow for the development of a continuous time series of NTL emissions. The data can be expressed in several different formats, for example, in GeoTiff (raster), shapefile (vector) or statistics file format. This allows for both visual and quantitative analysis and display of data.

Both the OLS and the VIIRS systems can detect artificial lighting, that is, NTL that can be interpreted as a measure of human activity. Both systems have a built-in manipulation system, through calibration and algorithms, for example, to exclude natural light so that the end product only consists of man-made or artificial light. Night-time light now seems to be widely used as a proxy for other more difficult measure variables, for example, gross domestic product (GDP) at a granular level, granular poverty levels, informal economic activity and remittances, human ecological footprint, and electrification rates, to name but a few (Ghosh et al. 2010). Also, see Bauer (2020) and Levin et al. (2020). Applications can be related to environmental factors, human activities and their impacts. According to Jeswani (2018), many studies have shown the possibility of correlations between NTL and several known demographic and economic variables.

The attraction for using the NTL stems from the objectivity, consistency, cost-effectiveness, timeliness and spatial nature of the data. To this end, Goldblatt et al. (2016) argue that the use of satellite remote sensing is now a key methodology in economics and other applied scientific research. This is especially true for countries with poor national accounts data. Bhandari and Roychowdhury (2011) went further stating that NTLs have been used as a proxy to study economic activities for the past 20 odd years.

Unfortunately, as with most, if not all sources, remote sensing NTL data also have some constraints. For example, satellites cannot detect light used inside a high capital intensive production plant, whether mostly empty or running at full capacity. Satellites also cannot distinguish between a light emitted from an office compared to an apartment building. To this end, Mellander et al. (2015) expressed some concern about the extent of the relationship between NTL emissions and economic activity at a micro-level.

With these premises, that is, the possible economic applications of satellite remote sensing NTLs in mind, the remainder of the article is structured as follows. Section ‘Literature review of remote sensing night-time light data’ is devoted to briefly discuss recent theoretical and empirical research regarding the initiation of NTL remote sensing and its progress throughout, especially from an economic perspective. Fairly detailed background on the origins and history of remote sensing of NTL data is supplied in Section ‘Origins and history of remote sensing of NTL data’. Section ‘Mapping NTLs for South Africa’ includes a basic account of the nature and detail of NTL maps as well as the development of NTL maps for South Africa. The focus of Section ‘Inquiry into NTLs for South Africa’ is on a descriptive inquiry into NTLs for South Africa, while Section ‘Theoretical underpinnings and principles’ focuses on the theoretical principles underlying the possible relationship between NTLs and economic performance. Section ‘NTLs and economic performance of South Africa’ endeavours to conduct an empirical investigation into the relationship between NTLs and economic performance within the South African context. Finally, Section ‘Conclusions’ provides the conclusion.

**Literature review of remote sensing night-time light data**

Some studies have been published that attempt to demonstrate relationships that provide insight into the value of using remote sensing NTLs as descriptors and/or proxies for human activity. Given that the release of the global NTL data is fairly recent, the first studies had also just commenced. The initial uptake was very modest but recently experienced a significant increase. Furthermore, over recent years, a proliferation of focus areas of studies has become noticeable. From an economic point of view, studies can be found incorporating global, national and micro-levels.

The discussion herein will only focus on the literature related to the field of economics because the focus of this study will be economic by nature. It is relevant to note that there is two broad field of studies, that is, studies aimed at identification and development of methods to improve the datasets and to extract the information from the datasets, both quantitatively and quantitatively. The second set of studies is application-oriented and it is on the second set of studies that the literature review will focus on.

Croft’s (1978) study seems to be the first one to explore the NTLs economic nexus. The study draws some positive conclusions on the relationship between NTLs and human development and settlement. Doll, Muller and Morley (2006), Ebener et al. (2005), Sutton and Costanza (2002), and Sutton, Elvidge and Ghosh (2007) further support Croft’s (1978) conclusions that show that NTLs reflect human activity. During 1997, Elvidge et al. published an article that found a ‘strong’ relationship between GDP and NTLs. Ghosh et al. (2010) also found that NTLs are very effective in GDP estimations, especially when it comes to countries lacking statistical structures and resources. Elvidge et al. (2017) found a coefficient of determination of 0.97 whilst performing a GDP regression analysis of 21 countries and DMSP NTLs.

Goldblatt et al. (2016) argued that Henderson, Storeygard and Weil (2012) pioneered the use and application of NTLs as a useful proxy for economic activity. Henderson et al. (2012) raised concerns about the inability of many countries, and in particular, developing countries to not or at best poorly measure economic variables such as GDP and economic growth (Kleynians & Coetzee 2017). The study further suggests that economic variables are rarely measured at all for cities or subnational regions. The authors developed a statistical framework incorporating remote sensing NTLs
into the national accounts system to supplement existing economic variables. The results of their empirical work suggest that ‘empirical growth’ no longer needs to be synonymous with ‘national income accounts’.

Omar and Ismal (2019) argue that their article was the first to study the relationship between NTLs and real GDP at the national and subnational levels using sub-national data for Egypt. Indeed, they found a statistically significant positive correlation between NTLs and real GDP at both levels. They state further that NTLs can be a good proxy for GDP (at various levels), especially in the absence of reliable official data. Omar and Ismal (2019) found a great deal of support in the literature, for example, Chen and Nordhaus (2011), Ebener et al. (2005), Elvidge, Chi-Hsu and Ghosh (2014), Henderson et al. (2012), and Kulkarni et al. (2011).

Elvidge et al. (2007) for estimating the density of constructed surfaces; Matsumura et al. (2009) for measuring and spatially mapping and monetising the human ecological footprint, estimating access to electricity and estimating the information and technology development index. The authors explicitly argue that all these measures of well-being can be solely derived from NTLs and population density.

The wide application benefit of NTLs, as suggested by Ghosh et al. (2010), is further supported by Rayner et al. (2010), who looked at fossil fuel carbon emissions; Doll (2008) and Sutton (1997) used it to estimate and map the spatial distribution of the population; Elvidge et al. (2009) for poverty mapping; Elvidge et al. (2007) for estimating the density of constructed surfaces; Matsumura et al. (2009) for measuring and spatially mapping food demand; Zhao et al. (2011) for water use; and Hsu, Elvidge and Matsumo (2013) for stocks of steel and other metals.

### Origins and history of remote sensing of night-time light data

During the 1960s, the Air Force of the United States of America (USA) initiated a new project called the DMSP. The original purpose of the programme was to detect moonlight clouds and weather data to support the USA Department of Defense (DoD). As from 1976, the satellites have included a weather sensor, the OLS, and have been circling the earth 14 times per day recording the intensity of lights as a six-bit (DMSP–OLS system) and 14-bit (SNPP–VIIRS system) digital number (DN), for every 30 arc-second output pixel (approximately 0.86 km² at the equator) spanning -180° to 180° longitude and -65° to 75° latitude (EOG 2020 & NOAA 2020).

The raw data generated from both the DMSP–OLS and the SNPP–VIIRS systems have been made public in 1992. The DMSP–OLS data are available from the NOAA website (https://eogdata.mines.edu/download_dnb_composites.html) from 1992 to 2013 covering their different satellites on an annual basis. The SNPP–VIIRS data are available from the Earth Observations Group (EOG) website (https://eogdata.mines.edu/download_dnb_composites. html) from 2012 to 2019 on both a monthly and annual basis.

The DMSP data are cloud-free composites made using all the available archived DMSP–OLS smooth resolution data for each calendar year and exclude several natural phenomena, including sunlit, glare, moonlit, clouds and lightning. The SNPP–VIIRS data are also filtered to exclude data impacted by stray light, lightning, lunar illumination and cloud cover.

For each of the DMSP–OLS composites, a product called Stable Lights is available. In this product, fires and other ephemerical lights are removed, based on their high brightness and short duration. In the final result, each pixel quantises the 1-year average of stable light in a 6-bit data format. The pixel values, called DN, are integers ranging between 0 and 63. In the case of the SNPP–VIIRS composites, several products are available, that is, VIIRS Cloud Mask, VIIRS Cloud-Mask-NTLs, VIIRS Cloud Mask, Outlier Removed and VIIRS Cloud Mask, Outlier Removed, NTLs, of which the last product is the preferred product.

The final results contain pixels that quantise the monthly or annual average of the VIIRS Cloud Mask, Outlier Removed,
NTLs and other products, in a 14-bit data format, also called DN, but with no upper limit, with integers ranging from zero to infinity. The DN represents the man-made light intensity of the particular pixel with a value of 0 referring to no light (EOG 2020 & NOAA 2020).

Elvidge et al. (2013) found some significant differences between the DMSP–OLS and SNPP–VIIRS composites. Figure 1 shows a side-by-side comparison of the average SNPP versus the average DMSP for the island of Oahu, Hawaii. The SNPP product shows substantially more spatial detail. Furthermore, the DMSP data have saturation (white pixels) centred on the major urban areas. The background areas with no detected lighting appear black on the SNPP product and as a ‘salt and pepper’ noise in the DMSP. Jing et al. (2015) support Elvidge et al. (2013) in that the SNPP products contain significant advantages over the DMSP product; however, not rejecting the significance of the use of the DMSP product in any case.

Henderson et al. (2012) state that the intensity of night lights, as measured by the DN, reflects outdoor and some indoor man-made light and that this has possible significant economic applications.

Mapping night-time lights for South Africa

In Figures 2–5, unlit areas are black, and lights appear with intensity increasing from black, red, yellow to white. Black areas are represented by the DN number increasing from 0 to 63 or infinity from grey to white.

Figures 2 and 3 indicate NTLs for South Africa at two points in time, that is, 1992 and 2013, indicating scales of man-made light from the DMSP–OLS system. The significant increases in man-made light for the Johannesburg–Pretoria region, the Mpumalanga region and the coastal regions are visible. On the contrary, the interior of the country seems to have experienced very little increase in man-made light over the period. Figures 4 and 5 show the NTLs from the SNPP–VIIRS system for 2013 and 2019. Again, the increases in man-made light for the regions, as indicated above, are visible, but not to the same extent.

The DMSP satellites operated until 2013, whereas the SNPP satellites started in 2012 but the SNPP–VIIRS system is not directly compatible with the DMSP–OLS system as highlighted in several studies (Jing et al. 2015). Although there seems to be a continuous series of night-time data, the

FIGURE 1: Suomi National Polar Partnership–Visible Infrared Imaging Radiometer Suite (a) versus Defence Meteorological Satellite Program–Operational Linescan System (b) cloud-free composited average visible band images of Oahu, Hawaii.

FIGURE 2: Defence Meteorological Satellite Program–Operational Linescan System night-time lights South Africa, 1992.

Source: Elvidge, C.D., Baugh, K.E., Zhizhin, M. & Hsu, F.C., 2013, ‘Why VIIRS data are superior to DMSP for mapping night time lights’, Proceedings of the Asia-Pacific Advanced Network 35, 62. https://doi.org/10.7125/APAN.35.7
two series are not directly compatible. Fortunately, 2 years of overlapping datasets are available. Figures 3 and 4 display the South African night-time lights for 2013 for both systems. Some clear differences are observable, mostly related to the saturation characteristic of the DMSP–OLS system.

The presence of man-made light, as displayed in these four figures, suggests that the majority of South Africans live in a few cities and that these cities have expanded significantly over this period, while a large part of South Africa is very sparsely populated. This also seems true for the South African economy. The increasing concentrations of people and economic activity over this period are highlighted through these figures.

Inquiry into night-time lights for South Africa

This section focuses on some preliminary empirical work on the South African NTL view. The first step is to construct an annual series of nightlight data from 1992 to 2018, by linking the annual DMSP–OLS and SNPP–VIIRS data at the national level. According to Beyer et al. (2018), one possible method is to exploit the overlap between the two data sources in 2013; however, in this case, the average overlap between the two data sources in 2012 and 2013 will be used.

The 2012 and 2013 data from the two sources are displayed in Table 1. The first 2012 and 2013 rows display data from the DMSP–OLS system, while, in the second 2012 and 2013 rows, the data from the SNPP–VIIRS system are displayed. The 2012 and 2013 overlap rows display the DMSP–OLS data divided by the SNPP–VIIRS data, while the average overlap row displays the average 2012 and 2013 overlap data. The SNPP–VIIRS data (2014–2018) are then adjusted using the average overlap number as displayed in the table.

This method is possible because both sources generate data in the same frequency, and they describe one phenomenon and there is no time gap between them. A word of caution may be relevant in that this method may not be most optimal because the underlying data generating processes between the two systems are significantly different, that is, the method of measurement has changed between both time series. The most obvious difference is the upper limit characteristic of the DMSP–OLS data. To address some of the shortcomings listed earlier, Jeswani (2018) suggests possibly using logarithmic and/or power function methods for the intercalibration between DMSP–OLS and SNPP–VIIRS.

The average annual DN value of the NTLs (avelight) for South Africa for the period is displayed top left in Figure 6.
Top right and middle left display the average annual number of pixels with DN values greater than 0 (countlight) and the average annual total intensity of all the DN pixels with a value greater than 0 (sunlight). Middle right in the panel displays the average annual total intensity of all the DN pixels with a value greater than 0 (sum) per total area (lightkm). Bottom right in the panel displays the average annual total intensity of all the DN pixels with a value greater than 0 (sum) per person (lightpop). Bottom left displays the average annual total intensity of all the DN pixels with a value greater than 0 (sum) per total national GDP (lightgdp).

Table 1 shows the unit root tests results (probability of the t-statistic) of the various NTL-related variables. Comparing

![Figure 6](https://www.jefjournal.org.za)

**FIGURE 6:** Night-time lights variables for South Africa, 1992–2019. (a), AVELIGHT; (b), COUNTLIGHT; (c), SUMLIGHT; (d), LIGHTKM; (e), LIGHTGDP and (f), LIGHTPOP.
the Augmented Dickey–Fuller (ADF) test statistics of the level and the difference variables with the critical test values at the one per cent level suggests that all the variables or time series are indeed stationary in the differentiated format. The results suggest that the differenced variables are stationary and therefore integrated to the order of 1 or I (1).

Inspecting the graphs of the frequency distribution of the DN pixels with a value greater than 0 for the period 1992–2013 (DMSP–OLS system) and DN pixels with a value greater than 0 for the periods 2012–2019 (SNPP–VIIRS), it is found that the majority of pixels with DN values greater than 0 have DN values of less than 7 and 5. There also seems to be two sets of pixels with DN values greater than 0 in each of the systems, that is, in the range 2–7 and 11–20. Pixels with DN values greater than 40 are very limited. This suggests that there are large parts of the country with very low NTLs, significant parts with fairly high NTLs and very small parts with extreme night-time emissions.

Table 3 displays the frequency distribution of the DN pixels for the periods 2012 and 2013 of the data sourced from the DMSP–OLS system (OLS) and SNPP–VIIRS systems (VIIRS). It shows some significant differences within the lower frequencies (1–3) and higher frequencies above 60. However, the frequency distributions between 5 and 60 seem fairly similar between the two systems.

**Theoretical underpinnings and principles**

Doll et al. (2006) tested the linearity of the log–log relationship between country-level GDP at purchasing power parity and total lit area all over the world, using the data in 1994 and 1995, and obtained a coefficient of determination of 0.85 for the regression. Ghosh et al. (2010) linearly regressed PPP-GDP and sum of lights (SOL) globally in 2006 and found an $R^2$ of 0.73.

Henderson et al. (2012) suggest that the relationship between the growth of lights and the growth of true income is given by

$$x_t = \beta y_t + \epsilon_{t,j}$$

where $y$ is the growth (or log difference) in true real GDP; $x$ is the growth of observed light, with a variance of $\epsilon_{t,j}$.

An important assumption underlying the specification is that there is a simple constant elasticity relationship between total observable lights ($x$) and total income ($y$): $x_t = \phi y_t$, where $\beta$ is the elasticity of lights concerning income.

Wu et al. (2013) also presumed that amount of lights is an increasing function of the corresponding GDP, and define it specifically as a power function, namely, a log-linear relationship between GDP and amount of lights.

$$\text{light} = \phi \text{(GDP)} = k \text{GDP}^\alpha$$

where parameter $k$ is not a constant but is determined by some factors other than GDP. A major concern raised by Wu et al. (2013) is the components of $k$. Gross domestic product per capita would be a probable factor, as a higher income per capita always leads to higher consumption of normal goods, and light is seemingly a kind of normal goods. Latitude per capita would also be a probable factor, as a higher income per capita would also be possible that affects parameter $k$ for its potential influence on residences’ demands for light. Another element that requires close attention is the degree of spatial concentration or dispersion of human activities, which is closely related to the degree of urbanisation.

Beyer et al. (2018) used similar specifications as Henderson et al. (2012), estimating the elasticity of GDP for nighttime intensity in the world and South Asia as

$$\ln(\text{GDP}_{c,t}) = a + b_c + c_t + \delta \ln(\text{intensity}_{c,t}) + \epsilon_{t,j}$$

where $\ln(\text{GDP}_{c,t})$ is the natural logarithm of GDP of country $c$ in year $t$ measured in constant local currency, $\ln(\text{intensity}_{c,t})$ is the natural logarithm of light brightness per km², $b_c$ is a country fixed effect and $c_t$ is a year fixed effect.

Omar and Ismal (2019) state that unlike Chen and Nordhaus (2011), who studied the relationship between GDP growth and NTL growth, they investigate the relationship at the variables’ levels. A very important assumption underlying
the theoretical model is the existence of a structural relationship between the true GDP and observed NTL with constant elasticity, signifying that observable lights increase at the same rate as economic performance. Making the coefficients’ interpretation with elasticity, they used the log–log equation:

\[ \ln(GDP_{it}) = \alpha_0 + \beta \ln(NTL_{it}) + \mu_{it} \]

where \( \mu_{it} \) is the error term of predicting the observed GDP using observed NTL.

Omar and Ismal (2019) use stable light dataset. Three variables are extracted: the SOL, which is the sum of the multiplication of location’s pixels by its digital value of light’s intensity; top-coded cells where the top is the number of pixels that are the highest light with DN 62 and 63; and the unlit-coded cells unlit variable indicates the number of pixels that are dim with DN 0–6. The model is estimated in a panel data structure.

Hu and Yao (2019) suggest that because the correlation between NTLs and GDP is decreasing with income levels, their relationship could be nonlinear. Their study investigates the relationship between night-light intensity (NTL) and real GDP per capita. The relationship is estimated using ordinary least square regressions, including a nonlinear term.

Lopez-Ruiz, Blazquez and Hasanov (2019) regress GDP (GDP\textsubscript{t}) on nationwide night-light intensity (NTL\textsubscript{t}) in logarithmic form, to see how well the latter can explain the former, using:

\[ GDP_{t} = \alpha_0 + \alpha_1 NTL_{t} + \epsilon_{t} \]

From the above, it is relevant to investigate the relationship using a log–log estimation technique considering constant elasticity between economic growth and NTL growth. The estimation can be performed either in level or in growth format using GDP or GDP per capita as the dependent variable. It may also be relevant to test for nonlinearity. In terms of the NTL variable, it is possible to use only one or several derived variables, for example, the SOL on its own or in combination with other derived NTL variables.

**Night-time lights and economic performance of South Africa**

This section investigates the ability to assess the economic performance of South Africa using light emissions recorded by satellite and reported on the Internet. The behaviour of the GDP per capita, total light intensity (total intensity of all the DN pixels with a value greater than 0) and the number of pixels that are dim with DN 0–5 over the period 1992–2019 (in natural logarithm scale) are displayed in Figure 7. Some correlation between the GDP per capita and total light intensity variables seems to be present (Figure 8). The smoothing is done using a Kernel, such as the Kernel K or an Epanechnikov kernel. Estimating the regression fit is done using the Nadaraya–Watson method. The possible nonlinearity of the relationship also seems evident given the flat slope of the relationship at the higher values. This is because of the saturation of DN values at the high bound, remembering that the DN values are capped at 63 using the DMSP–OLS system. Gross domestic product per capita will be used as a proxy for economic performance, whilst total light intensity will be represented by the sum of NTL variable. The unlit-coded cells unlit variable indicating the number of pixels that are dim with DN 0–6 will be used as a control variable because this variable controls for the blooming effect associated with NTLs.

Zou (2018) argues that the use of the vector error correction model (VECM) can be used to establish the relational model amongst economic variables in a non-structural way. The VECM methodology is adopted within this study predominantly because the variables under consideration are not stationary in

---

**FIGURE 7:** Gross domestic product per capita and total light intensity, 1992–2019. (a), LNGDPCAPITA; (b), LNSUMLIGHT and (c), LNBCOUNT.
their levels but are in their differences and the variables are assumed to be cointegrated. The results of the commonly accepted ADF as presented in Table 2 and performed on GDP and bcount (in natural logarithmic format) show that level values of the three variables is nonstationary, and a further test indicates differenced values of the three variables are first-order difference stationary, that is I(1).

The three variables VECM are then integrated into the same ordering of 1:

$$\Delta \ln Y_t = \beta_0 + \sum_{i=1}^{n-1} \beta_i \Delta \ln Y_{t-i} - i + \sum_{i=0}^{n-1} \delta_i \Delta \ln X_{t-i} - i + \varphi_i + \varepsilon_t$$

where $\ln Y_t$ is the real GDP per capita at period $t$, $\ln X_t$ is the total light intensity at period $t$ ($\ln L_t$) and the number of pixels that are dim with DN 0–5 at period $t$ ($\ln C_t$); $\mu_t$ is the stationary error term; $\delta$ is the error correction term and is the ordinary least square residuals from the long-term cointegration regression:

$$\ln Y_t = \beta_0 + \beta_1 \ln X_t + \varepsilon_t$$

and is defined as:

$$\Delta \ln Y_t - \Delta \ln X_t = ECT_{t-1}$$

The Akaike information criterion (AIC) within a vector autoregressive model (VAR) was used to determine the optimal lag period for South Africa. It was found that the optimal lag order for the VAR model is 4 (AIC 4 = -7.623296*, where the * indicates the lag order selected by the criterion). Estimating the VECM, it is important to include $p-1$ lags where $p$ is the lag order selected by the criterion. Testing the stationarities of VAR model via the mod of AR characteristic root reciprocal of the VAR model indicates that the mod of reciprocal of each characteristic root lies inside the unit circle. That implies that the lag order of 4 is appropriate, and the established VAR model is stable after going through a stability test.

The Johansen cointegration test on the three variables (Table 4) shows that, in both trace and maximum eigenvalue tests, under the 5% level (test results are to accept the null hypothesis) positive relationships exist for South Africa’s GDP. This means there are stable and long-term equilibrium relationships amongst the variables. On the premise of the existence of cointegration relationships, VEC modelling can be further conducted.

Estimating the VECM yields the result given in Table 5. The cointegration equation therefore is:

$$\Delta \ln Y_t - 1 = 2.763816 + 0.748619 \ln L_{t-1} - 0.252327 \ln C_{t-1}$$

From this equation, it can be seen that other things equal, each percentage-point increase in total light intensity will cause an increase of 0.75 percentage points in GDP per capita, and each percentage-point increase in the number of pixels that are dim with DN 0–5 will cause a decrease of 0.25 percentage points in GDP per capita.

The results of the VECM yield the following in terms of the error correction mechanism (Table 6). Although the sign of the error correction term is indeed negative as per the economic theory, the term is not statistically significant ($r < 2$). This indicates that there does not seem to be a statistically significant short-term causality between the GDP per capita and total light intensity.
The estimated VECM, therefore, is:

\[
\Delta \ln Y_t = -0.015656 \Delta Y_{t-1} + 0.598921 \Delta \ln Y_{t-1} + 0.185419 \Delta \ln Y_{t-2} - 0.0363115 \Delta \ln Y_{t-3} - 0.029792 \Delta \ln L_{t-1} - 0.095377 \Delta \ln L_{t-2} - 0.073777 \Delta \ln L_{t-3} - 0.003454 \Delta \ln C_{t-1} - 0.024636 \Delta \ln C_{t-2} - 0.017733 \Delta \ln C_{t-3} + 0.006457
\]

The cointegration relationship is displayed in Figure 9. The zero average line represents a stable and long-term equilibrium relationship amongst variables.

The cointegration relationship seems to capture the 1997/1998 Asian financial crisis, the subsequent 2003–2008 economic expansion period, the 2010 financial crises and subsequent sluggish growth period fairly well. Testing the above model for statistical errors indicates that the \( R^2 \) is fairly high at 0.5, the Breusch–Godfrey serial correlation LM test shows no serial correlation with \( p = 0.42 \), the heteroscedasticity test, Breusch–Pagan–Godfrey, indicates that the model does not have heteroscedasticity \( (p = 0.78) \) and the residuals are normally distributed \( (p = 0.14) \). The Cumulative Sum Control Chart (CUSUM) test also suggests model (parameter) stability.

It is also relevant to state that the inclusion of the other night light-related variables within the VECM model yielded insignificant or spurious results. For example, the inclusion of the top-coded cells (i.e. the number of pixels that are the highest light with DN 62 and 63) and/or a quadratic term did not significantly contribute to the performance of the model. On the contrary, there were no statistically significant differences in using total GDP compared to GDP per capita.

The study will make use of the autoregressive distributed lag (ARDL) model methodology in an attempt to verify the results of the VECM model. Ghouse, Khan and Rehman (2018) state that, in an ARDL model, the dependent variable is expressed by the lag and current values of the independent variable and its own lag value. The ARDL model normally starts from a reasonably general and large dynamic model and progressively reducing its mass and altering variable by imposing linear and non-linear restrictions (Charemza & Deadman 1997). The ARDL model is one of the most general dynamic unrestricted models in econometric literature.

A generalised ARDL \((p, q)\) model can be specified as:

\[
Y_t = \gamma_0 + \sum_{i=1}^{p} \delta_i Y_{t-i} + \sum_{j=0}^{q} \beta_j X_{t-j} + \varepsilon_t \quad \text{for } p \geq 1, \, q \geq 0,
\]

for simplicity assume that the lag order \( q \) is the same for all variables in the \( K \times 1 \) vector \( X \). \( b \) and \( d \) are coefficients, \( \delta \) is the constant and \( \varepsilon \) is a vector of the error terms. \( Y \) is the real GDP per capita at period \( t \), \( X \) is the total light intensity at period \( t \) \((L)\) and the number of pixels that are dim with DN 0–5 at period \( t \) \((C)\).

Estimating the generalised ARDL model using the same variables as within the VECM model (and using the logarithm format) suggests the use of an ARDL \((4.0.0)\) model. The optimal lag structure was derived from the Akaike Information Criteria (AIC).

Testing the model for possible cointegration yields an \( F \)-test statistic value (5.41) higher than the upper bound \( I \) (1) up to the 2.5% level. It is, therefore, possible to reject the null hypothesis of no cointegration.

The cointegration of long-term equation (unrestricted constant and no trend) is given as:

\[
EC = \ln Y_t - 0.5897 \ln L_t + 0.0282 \ln C_t
\]

or

\[
\ln Y_{t+1} = 0.5897 \ln L_{t+1} - 0.0282 \ln C_{t+1} + \varepsilon_{t+1}
\]

\((p = 0.000) \quad \text{(ec is error correction term)}\)

The cointegration coefficient (ecm) is estimated at \(-0.185336\) with a \( p \)-value of 0.005, suggesting that the ecm derived from the ARDL model is indeed statistically significant. The cointegration graph is displayed in Figure 10. There seems to be significant similarities between the cointegration relationship between the VECM (Figure 9) and ARDL (Figure 10) models. Testing the model for statistical errors indicates that the \( R^2 \) is fairly high at 0.98, the Breusch–Godfrey serial correlation LM test shows no serial correlation with \( p = 0.09 \), the heteroscedasticity test, Breusch–Pagan–Godfrey, indicates that the model does not have
The literature on the use of remote sensing NTLs seems to support its use in many different fields of study, including economics. A growing number of articles are being published. The use of satellite data seems to be common practice by now and can be a very useful proxy for several economic variables at a global, national and sub-national level, including GDP and population. This is even more relevant for countries where there is a lack of reliable official data and statistics (Coetzee 2019).

There is a continuous series of NTLs data since 1992 available. Data from 1992 to 2013 are generated from the DMSP–OLS system, whilst data from 2012 to present are generated by the new SNPP–VIIRS system. Although not perfectly compatible, it is possible, through some inter-calibration techniques, to link the two datasets. The data are subject to onboard manipulation procedures to ensure the end data only contains man-made or artificial night light.

This article demonstrates how many different applications of remote sensing NTLs may be developed. The data can be visually displayed giving unique insights into the spatial nature and characteristics of a country and/or any sub-national region up to a 1 km² area. The data can also be used in various statistical analyses, for example, comparing the light intensity and consequently human and economic activity of various countries and/or sub-national regions up to a 1 km² area. It is also possible to visually and/or statistically display trend behaviour within a spatial context.

The methodology used in this study involves estimating both a VECM and ARDL model that map light growth into a proxy for GDP growth. The use of VECM and ARDL models is well known and easily applied. Both the VECM and ARDL model point to a long-term cointegrating relationship between GDP (in this case GDP per capita) and total light intensity of NTLs. A statistically significant short-term error correction term could, however, not be established through the VECM. On the contrary, a statistically significant short-term error correction term was derived through an ARDL model. The combined results suggest that indeed a statistically significant long-term relationship exists between GDP and night light emissions, whilst it is plausible that a short-term relationship also exists. Nonetheless, the possible application of NTLs from an economic viewpoint seems both desirable and effective.

The results of the study suggest that satellite remote sensing technologies hold lots of promise and opportunities in terms of the field of economics. Satellite night-light data seem to be a useful proxy for economic activity at temporal and geographic scales for which traditional data are of poor quality or are unavailable. It, therefore, contributes to our understanding of the spatial and temporal behaviour and trends in economic activity. As a policy proposal, it is suggested that the use of satellite remote sensing technologies be included in the official statistical frameworks and methodologies currently being utilised to produce official economic statistics.

Acknowledgements

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this research article.

Authors’ contributions

C.E.C. designed the original concept and analysis. E.P.J.K. assisted in the composition and analysis and wrote the final article.

Ethical considerations

Although an ethical clearance is not considered necessary for this study, ethical clearance was applied for from the Economic and Management Sciences Research Ethics Committee at North-West University, ethical clearance number: NWU-00911-20-A4.
