DataFirst

DataFirst Technical Papers

Throwing light on rural development: using nightlight data to map rural electrification in South Africa

by

Takwanisa Machemedze, Taryn Dinkelman, Mark Collinson, Wayne Twine and Martin Wittenberg

Technical Paper Series
Number 38

About the Author(s)

Takwanisa Machemedze: DataFirst, University of Cape Town, South Africa
Taryn Dinkelman: Economics Department, Dartmouth College, United States of America
Mark Collinson: MRC/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, South Africa. INDEPTH Network, Ghana. DST/SAMRC South African Population Research Infrastructure Network (SAPRIN)
Wayne Twine: School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, South Africa
Martin Wittenberg: DataFirst and School of Economics, University of Cape Town, South Africa

Acknowledgments

This work is based on research supported by the National Research Foundation, South Africa (ESRC-NRF International Centre Partnership, Ref ICPC150423117553). Any opinion, finding and conclusion or recommendation expressed in this material is that of the author(s) and the NRF does not accept any liability in this regard. We wish to thank Nathan Cunningham, Louise Corti and the UK Data Archive for assistance with the work on the satellite data. Suleiman Afolabi from the Agincourt Unit extracted the Agincourt HDSS data and supplied the shape files used to extract the satellite data.

Recommended citation

Machemedze, T., Dinkelman, T., Collinson, M., Twine, W., Wittenberg, M. (2017). Throwing light on rural development: using nightlight data to map rural electrification in South Africa. A DataFirst Technical Paper 38. Cape Town: DataFirst, University of Cape Town.

DataFirst, University of Cape Town, Private Bag, Rondebosch, 7701, Tel: (021) 650 5708,
Email: info@data1st.org/support@data1st.org

Throwing light on rural development: using nightlight data to map rural electrification in South Africa

Takwanisa Machemedze, Taryn Dinkelman, Mark Collinson, Wayne Twine and Martin Wittenberg

DataFirst Technical Paper 38
University of Cape Town
September 2017

1. Introduction

How do you measure development in the absence of high quality social survey information? Several authors have argued that nightlights data captured by satellites can provide a reliable measure of development in contexts where other measures do not exist or where the quality of the data is such that the statistics cannot be trusted. The results of these studies look plausible, but thus far there has been no independent verification that the satellite data reliably pick up local development patterns.

In this paper we investigate parts of this question in the context of rural electrification in South Africa. Since the end of apartheid there has been a massive roll-out of electricity connections, raising the proportion of rural households with connections from under 30% to around 65%[1] by 2008. Furthermore this process did not occur evenly – different parts were electrified at different times. In this paper we investigate to what extent the satellite data can pick up this process. We look at these questions in the context of Bushbuckridge, a rural municipality in the east of South Africa, adjoining the Kruger National Park and close to the Mozambican border.

Bushbuckridge was in one of the "homeland" areas in apartheid South Africa and like most of these areas lacked basic infrastructure. After the advent of democracy in 1994 the national electricity roll-out also reached this area. Different parts of the area were, however, electrified at different times. What distinguishes this municipality from many other locations, however, is that since 1992 it has hosted the Agincourt Health and Demographic Surveillance site (AHDSS), which implies that higher quality information is available for this area than elsewhere. In particular since 2001 we have information for every second year on the number of households in each of the 31 AHDSS village that use electricity for lighting. In this paper we set out to investigate three questions:

a) Does the satellite data pick up the temporal patterns of rural electrification?
b) Does the satellite data pick up the spatial patterns of electrification? Can it pick up the difference between developed and undeveloped areas? And at what spatial resolution?
c) What is the correlation between the satellite data and the household electrification data?

Our research highlights both the possibilities and the limitations of satellite data for measuring local development. It also raises interesting questions about measurement of the delivery of household services via standard survey instruments.

[1] Own calculations using the 1993 Project for Living Standards and Development and the 2008 National Income Dynamics Study.

2. Literature Review

2.1 Satellite data

There is a burgeoning literature on the use of nightlight data in measuring economic development. The two main ways that nightlights data have been used is (1) in measuring income growth or economic growth over time and (2) as a way to examine the impacts of new technology in less developed countries.

In the first class of papers, Henderson *et al* (2012) argue that the use of satellite data significantly improves the measures of economic growth obtained by more conventional means. More controversially Pinkovskiy and Sala-i-Martin (2016) argue that satellite data suggest that survey data are less reliable in measuring poverty and economic well-being in Africa than GDP. Michaelopolous and Pappaioanou (2014) have used nightlights as a measure of economic success, linking variations in this success to variations in early institutional factors (like ethnicity) across Africa. In the second class of papers, researchers have used nightlights data to look at the impacts of access to electricity e.g. Burlig and Preonas (2016), and have found – at least in India – surprisingly little relationship between nightlights and measurable economic outcomes.

Much of this literature has used the nightlight data at the level of countries or major cities (Burlig and Preonas 2016 is an exception). Little research has been done on how well these data perform in more rural contexts. Furthermore there has been little attention paid to validating the nightlight data against external benchmarks. We aim to close this research gap in this paper.

2.2 Rural electrification in South Africa

South Africa provides a good setting for investigating some of these issues, as it has had a good economic infrastructure, from which large portions of the population were excluded in the past due to the policy of apartheid. We can therefore exploit variation in electricity access both within the country and over time.

The scale and process of rural electrification after the advent of democracy is discussed in Bekker *et al* 2008 and Dinkelman 2011. Nationwide, household electrification rates climbed from 35% of households in 1990 to 84% in 2011 (StatsSa 2012). Since 1994, over 7 million households have received new, heavily subsidized, connections to the grid (DOE website). Much of this expansion involved low capacity domestic connections. As noted earlier, it led to a major increase in the proportion of rural households using electricity for lighting between 1993 and 2008. In the Bushbuckridge Local Municipality only three households in a sample of 356 surveyed across five villages (including two of the ADHSS study villages) had electricity in 1991 (Madubansi and Shackleton 2006). By 2002, almost all households in four of the five villages were electrified. In two villages immediately north of our study site, 32% and 11% of households respectively were not yet connected to the grid in 2009 (Matiska et al. 2013). By 2011, only 6% of all households in the ADHSS study site lacked an electricity connection (Wittenberg *et al* 2017). Indeed the number of new connections exceeded the initial backlog, since there was a strong increase in the number of households, not only in Agincourt but nationally. Given the scale of the roll-out can one capture it by satellite data? One limitation might be that the affected households may very well have had other forms of domestic lighting beforehand. Furthermore Harris *et al* (2017) note that the roll-out didn't proceed linearly – alongside new connections there were also disconnections. These could be for a variety of reasons, including economic (lack of affordability), physical (faults in the infrastructure) and spatial (migration from rural areas that were electrified to new settlements on the urban fringe but lacking infrastructure). Nevertheless the aggregate patterns suggest a strong increase in the availability and use of electricity for lighting. Furthermore Dinkelman (2011) has noted that electricity access promotes economic activity which may also show up in the nightlight data.

3. Data and data processing

3.1 The satellite data

Our data comes from the DMSP-OLS Nighttime Lights Time Series (http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html). These data are annual cloud-free composites of average digital brightness values for the detected lights, filtered to remove ephemeral lights (e.g. fires) and background noise (Elvidge et al 2009). The data come from six satellites over 21 years. We extracted the data for the Agincourt study site and for the villages within the site using shape files supplied by the MRC/Wits University Rural Public Health and Health Transitions Research Unit. Each pixel within the area represents around a square kilometer on the ground. The data per pixel gives the annual average brightness level with digital numbers (DN) ranging from 0 to 63. The top value is saturated light. Increases above that level will not be measured (that is, however, not a problem for our study area). Up to the saturation point the values represent a scale with 0 representing the absolute absence of light.

Elvidge *et al* (2014) recommend that users perform an intercalibration before direct comparison of the digital values (DN) across the time series. The reasons are that: the original instrument (the OLS) had no on-board calibration, there are differences in the performance of instruments, different sensors had different detection limits and saturation radiances. As a result, the intercalibration converts data values from individual satellite products into a common range defined by a reference year. We apply the calibration coefficients in Elvidge et al. (2014) before comparing data over the years. Negative values are treated as zero or no light. After calibration, we derived the "sum-of-lights" (SOL) as the sum of the digital values for the Agincourt area for each year and satellite. The raw and calibrated results for the Agincourt are shown in Figure 1. It is evident that the results from different satellites are closely aligned after the intercalibration. Where we have two readings for a year, we average the intercalibrated values, to produce one estimate per year.

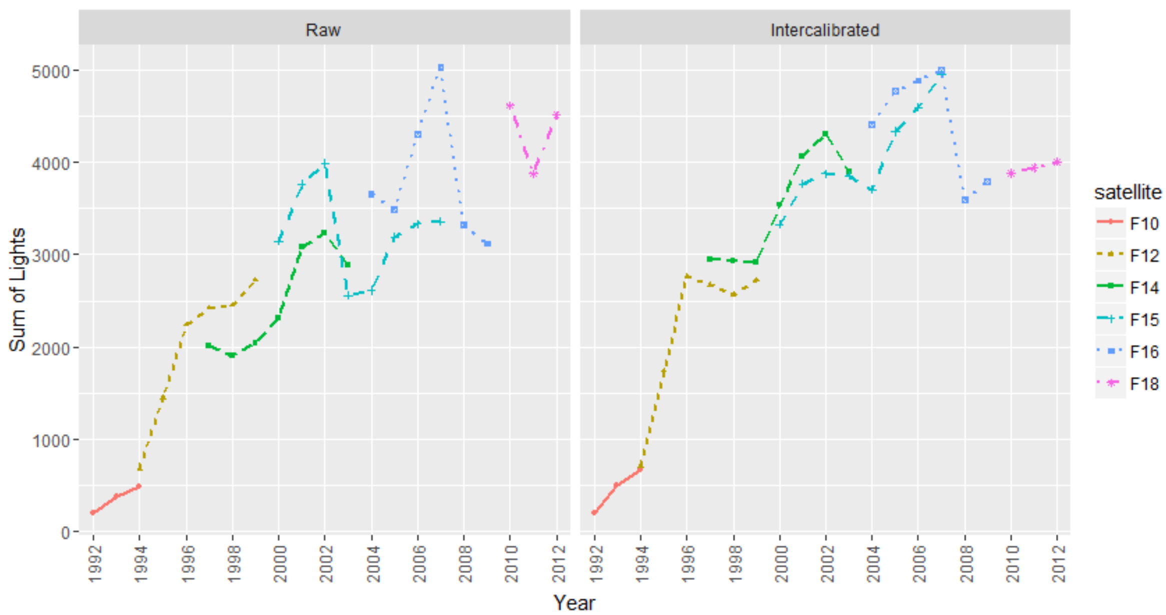

Figure 1: Comparison of the raw and intercalibrated nightlight data for Agincourt

One of the limitations of the "sum-of-lights" measure is that it is sensitive to the size of the area over which the sum is calculated. We want to compare the results for different villages, so for most of our analyses we use the "average sum-of-lights" (i.e. sum-of-lights divided by total pixels) which ranges from 0 (complete absence of light) to 63 (fully saturated).

One immediate issue is that our villages are irregularly shaped and do not fit neatly into the square pixel grid of the satellite data. As a result we calculate the proportion of the village that fits into a pixel and add that proportion of its light measure to the "sum-of-lights" score for the village. The process is shown in Figure 2.

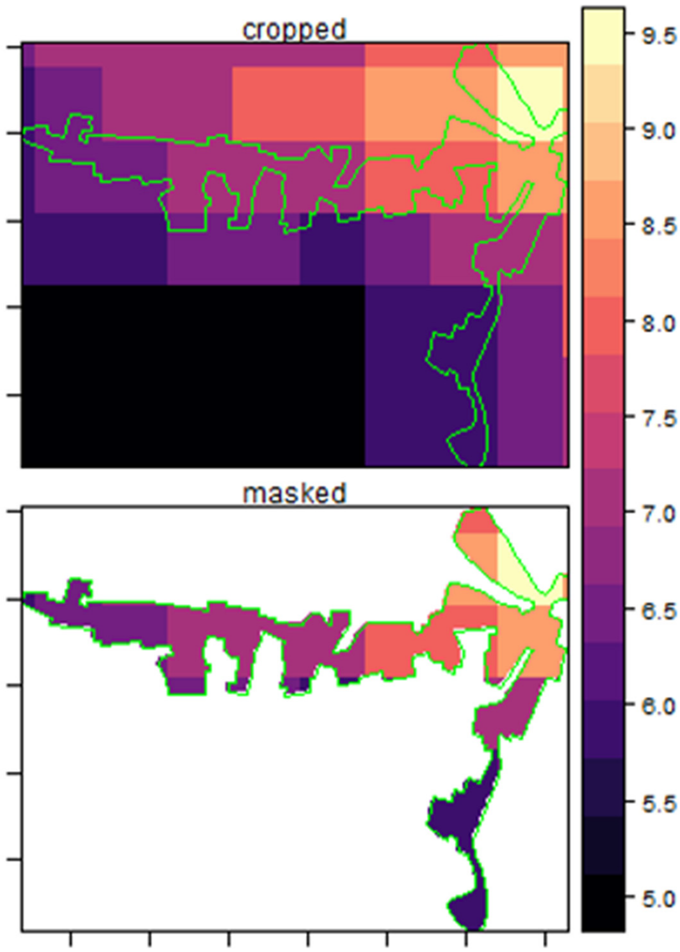

Figure 2: The top panel shows how one of the villages is positioned relative to the pixel grid. The bottom panel shows how the light measures are allocated to that village.

It is evident that there will be considerable measurement error at the scale that we are interested in, i.e. the village level. Some of the light that emanates from a village will be lost to adjoining areas. In fact the measurement error is worse than this, given how the original data are constructed (Elvidge et al. 2007; Doll 2008). From an altitude of 830 km, the DMSP-OLS satellite captures images at a fine nominal resolution of 0.56km. These are then smoothed on-board by averaging 5x5 pixel blocks of fine data to produce data with a ground sampling distance (GSD) of 2.7 km. According to Doll (2008) this is done to reduce the amount of memory required on board the satellite. This data again is re-mapped with 1 km x 1 km spatial grid (Elvidge et al. 2007; Doll 2008). So although the data arrives with a nominal resolution of a square kilometer, the pre-averaging that has been done means that the resolution is not as accurate as that.

3.2 The Agincourt data

The MRC/Wits University Rural Public Health and Health Transitions Research Unit has been doing demographic and health surveillance in the Agincourt area in the east of South Africa since 1992. Its purpose was to provide accurate information for health planning (Tollman 1999, Tollman, Herbst, Garenne, Gear and Kahn 1999) and to investigate the delivery of health services in a deprived rural area. The Agincourt sub-district was selected as location in part because it reflects many of the key developmental challenges. It formed part of the previous Gazankulu homeland and exhibited many of the characteristics of these areas: a lack of infrastructure and a population subjected to forced removals, betterment planning and migrant labour (see Niehaus 2001).

Since 1992 trained fieldworkers have been conducting annual census rounds in which births, deaths, in- and out-migrations, and household rosters are captured. Since 2000 the Unit has also fielded specialised modules dealing with other topics of interest, such as receipt of grants or labour market status. We will be using the household assets module that was fielded every second year since 2001. This module asks what type of energy is used for lighting. The count of households indicating that they mainly use electricity for lighting will be our measure of the number of households with electricity access.

By 2001, the first year in which we can measure household electrification, much of the area had already been electrified. According to the HDSS data 69% of households had access in that year. Nevertheless over the ten year period for which we have both HDSS and satellite data there was still considerable change in the level of electrification. In our dataset we observe that seven of the twenty one villages in the original study site (it was enlarged in 2006) did not yet have electricity in 2001. Furthermore villages with electricity saw the number of connections increase, in part due to new household formation in those areas, so we might expect that the brightness of the area should increase even in the period since 2001.

Besides our "objective" measure of electricity connections we also have the results of an exercise categorising the study site villages in terms of their development status which was conducted in 2000, i.e. a year before our first connection data from the HDSS (Hargreaves 2000). The method involved collecting infrastructure data as well as socio-economic information on all of the villages in the study site. A workshop of community representatives ended up developing a four-fold classification: a) "Central communities" b) "Established communities" c) "Undeveloped villages" and d) "Refugee settlements". The first two categories had electricity in 2000, while the latter two did not. Electricity access was, however, not the only basis for differentiating between them. The former two also had better road infrastructure and were deemed to be more affluent. The refugee settlements were largely on the eastern fringe of the study site (with one notable exception) and housed people who had fled the Mozambican civil war in the 1980s.

4. Methods

We outlined three research questions above. We organise this discussion around methods designed to establish a) an increase in light over time, due to rural electrification b) variation across space, due to differences in access and c) variation across time and space due to changes in access.

4.1 Temporal variation in the nightlight data

Our first cut at establishing whether the nightlight data has picked up the rural electrification programme is to look at the temporal trends. An increase in the level of light from Agincourt does not, however, establish that this is due to the electrification. It could be an artefact of measurement changes in the satellite instruments that the intercalibration did not properly deal with. To that end we use a ready-made counterfactual: the Agincourt area immediately adjoins the Kruger National Park. The park, of course, was not electrified (except for isolated camp sites which had received electricity prior to 1992). Our strategy can be thought of as a difference-in-difference estimation strategy, as expressed by the equation

$$sol/pixel = \beta_1 + \beta_2 t + \gamma Agincourt + \delta t * Agincourt + \varepsilon \qquad (1)$$

Here t is a time trend, Agincourt is a dummy variable for the "treated" area and the parameter of interest is δ which measures the difference in the rate at which Agincourt has increased in brightness compared to the counterfactual, i.e. the Kruger Park.

There are two potential problems with conventional tests of significance in regression 1: the errors ε are almost definitely heteroscedastic. Furthermore it is possible that errors for the same year are correlated between Agincourt and the Kruger Park. Consequently we used robust standard errors with clustering by year. Conventional standard errors were uniformly smaller than the ones we report.

Of course the Agincourt area could increase in brightness due to changes in lighting that were not brought about the government's rural electrification programme, e.g. if those households and businesses that did have access used more lights. There is a second way of trying to think about the temporal pattern in the nightlight data. Established urban areas in the traditionally "White" part of South Africa were not the primary targets of the electricity roll-out (although some settlements on the outskirts of cities did get electrified early in the electrification programme). We use Nelspruit, the closest big city to Agincourt, as another counterfactual. This gives a second "difference-in-difference" estimate. Finally we test to see if the Nelspruit and Kruger trends are statistically equal and then compare the Agincourt pattern to that pooled trend estimate. In all we therefore have three difference-in-difference estimates:

- Agincourt vs Kruger
- Agincourt vs Nelspruit
- Agincourt vs Kruger/Nelspruit

4.2 Spatial variation in the nightlight data

The comparisons with the Kruger Park and Nelspruit provide circumstantial evidence that electrification can be picked up by the satellites, but it isn't direct evidence of the importance of new connections. As noted earlier, the roll-out within the Agincourt area did not proceed evenly across space. Our second set of methods are designed to check whether the satellite data can pick up differences between villages that have been electrified versus those that haven't.

We begin with purely descriptive analyses, checking whether the places that light up within Agincourt correspond to places that are electrified and more developed. We test the light data against the village typology developed in Hargreaves (2000), i.e. we run the regression

$$sol/pixel = \beta_1 + \beta_2 type2 + \beta_3 type3 + \beta_4 type4 + \varepsilon \qquad (2)$$

The base category is the "central" communities as listed in Hargreaves, type2 is the "established" ones, type3 the "undeveloped" and type4 the "refugee" ones. We test for differences in the year 2000 (when the typology was created) but we also run this pooling over all the years to see whether these differences are a persistent feature. Again we use robust standard errors and (in the pooling regression) allow for correlation of errors within years.

We then test to see whether our direct measure of household electricity connections (available since 2001) is informative by regressing the average brightness of a village on the number of households using electricity for lighting (per square kilometre, i.e. per pixel). The cross-sectional regressions can be written as

$$sol/pixel = \beta_1 + \beta_2 connections/pixel + \varepsilon \qquad (3)$$

Again we correct the standard errors for heteroscedasticity.

4.3 Variation in space and time

Our preferred approach is to use both the cross-sectional and the temporal variation. One possibility is to run regression 3 over the pooled village-year panel. We do that. We also run that regression with a quartic in pixel size, to control for the fact that there may be more noise in measurements made on smaller villages. The final specification (and our preferred one), however, is

$$sol/pixel_{it} = \beta_1 + \beta_2 connections/pixel_{it} + \theta_i + \eta_t + \varepsilon_{it} \qquad (4)$$

where *i* subscripts village and *t* year. The θ_i terms are village fixed effects, η_t year fixed effects and ε_{it} is an idiosyncratic error.

5. Results

5.1 Variation in time

The extent of temporal variation is already shown in Figure 1. The comparison with the Kruger Park and Nelspruit is given in Figure 3. More detailed summary statistics are provided in the appendix. It is interesting to note that in 1992 the average brightness of Agincourt was not much different to that of Kruger – and both were close to zero. The contrast with the city is quite stark. The difference between the Agincourt trajectory and the Kruger Park one is a nonparameteric "difference-in-difference" estimate of the impact of rural electrification. Interestingly the Kruger Park line is not completely flat but also shows an increase, although the value of brightness at the end is still below three and so still very dark. The Nelspruit trajectory shows some growth in brightness, which is not surprising, given that the city became the capital of Mpumalanga province in 1994 and has been a hub for development of the Lowveld area.

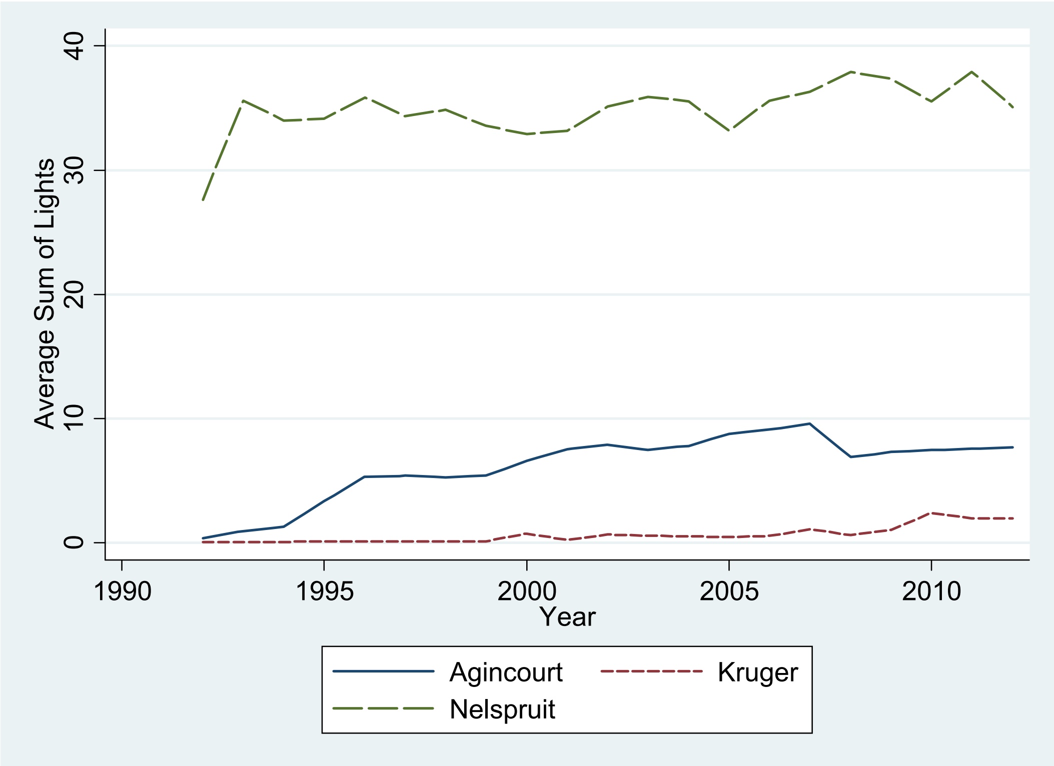

Figure 3: Average Sum of Lights (SOL) by year and location.

Notes: SOL is the average luminosity per pixel averaged over all pixels in each location. The range of SOL is 0 (no light) to 63 (fully saturated light).

One aspect of the data that deserves comment, although we cannot investigate it fully in this paper, is the dip in the Agincourt brightness in 2008, from which it never properly recovers. The timing coincides with South Africa's electricity supply crisis which led to extensive "load shedding" in 2008. This episode was followed by major tariff increases, which might account for the lack of a rebound after 2008. We leave this question for future research, although we think it is a potentially important feature of the data. We will make some speculative comments about its significance in the discussion.

Table 1 provides parametric estimates of the average increase in brightness over the period 1992 to 2012. Columns 1 to 3 present the estimates for Agincourt, Kruger and Nelspruit respectively. It is clear that brightness increased in all areas, but the increase is biggest in the Agincourt area. The "treatment effects" regression in column 4 shows that the difference in trend between Agincourt and Kruger (given by the coefficient on t*Agincourt) is big and statistically significant. The point estimate of .255 would suggest that the roll-out achieved a brightness increase of five units over this period (over and above what would have happened anyhow). Given that the levels were less than one to start with, this is a huge increase. The results in column 4 also allow us to test to see whether the trends in the Kruger Park and Nelspruit are statistically different. The point estimate for the difference of .118 is sizable, but is not big enough to conclude that the trends are different.

The difference in the trends between Agincourt and Nelspruit (i.e. between the estimates in columns 1 and 3) is .137, which is our second "difference-in-difference" estimate of the average annual change in brightness induced by rural electrification. This is still a sizable effect. Nevertheless this effect is statistically not significant. The 95% confidence interval for the true difference goes from -0.064 to 0.338. Nevertheless there are good reasons for suspecting that the point estimate is actually below the true value. With agglomeration economies we would expect cities to grow faster (and therefore become brighter) than the rural areas over time, even if there is no additional residential electrification.

In column 5 we obtain our third difference-in-difference estimate. In this regression we impose a common trend on the Kruger Park and Nelspruit data (which was not rejected by the regression in column 4). This trend coefficient is .15 and is related, presumably, to a general increase in economic activity after the end of apartheid. Nevertheless the data suggest again that the Agincourt area became brighter more rapidly than this. The point estimate for the "treatment effect" is .196 which is again strong and is statistically different from zero.

Table 1: Trends in brightness, Sum of lights 1992-2012

VARIABLES	(1) Agincourt	(2) Kruger	(3) Nelspruit	(4) All areas Pooled	(5) All areas Pooled
Time trend	0.348**	0.0932**	0.211*	0.0932**	0.152**
	(0.0649)	(0.0164)	(0.0889)	(0.0167)	(0.0429)
Agincourt				2.690**	3.336**
				(0.923)	(0.899)
Nelspruit				32.88**	34.18**
				(1.312)	(0.453)
Time trend*Agincourt				0.255**	0.196*
				(0.0776)	(0.0722)
Time trend*Nelspruit				0.118	
				(0.0970)	
Constant	2.322**	-0.368*	32.52**	-0.368*	-1.014*
	(0.799)	(0.140)	(1.240)	(0.142)	(0.481)
Observations	21	21	21	63	63
R-squared	0.670	0.696	0.355	0.992	0.992

Robust standard errors in parentheses

** $p<0.01$, * $p<0.05$

Notes: Table shows regression output from equation 1

5.2 Variation in space

An initial look at the spatial patterns in brightness over time is given in Figure 4. Here we show the pattern of nightlights for the Agincourt area as a whole in five year intervals from 1992 to 2007 (the peak of brightness as shown in Figure 1). It is evident that initially the whole area was dark, but that an area near the centre of the site and an area in the North-West were the first to light up. By 2007 the lights had diffused fairly widely, but there were still some darker patches along the eastern border. These last areas to be illuminated were the refugee settlements mentioned in the village typology. These were also among the last places to be electrified. The 2012 distribution is given in Figure 5. The area as a whole is darker than in 2007 (which is to be expected given that the aggregate brightness in 2012 is below that in 2007, as shown in Figure 1). Nevertheless there is a marked difference in brightness between different parts of the site.

There are two noteworthy "bright spots" in that picture (detectable also in the earlier years). One in the extreme North-West and one in the central part. The bright area in the North-West is a small administrative town with a shopping centre, police station, municipal offices, and urban residential stands which were electrified in the late 1990s. The second bright spot in the centre of the area is similarly characterised by a variety of economic activities. Figure 6 provides a close-up of this area, which proves to be informative. Village X lies at the intersection of several transportation routes – the railway line goes through the area in a North-South direction. A train siding is indicated with the black dot. A major East-West road meets the NW to SE road from Agincourt. At the intersection there is taxi rank and nearby there is a shopping centre (indicated with the cross). Given this economic activity it is not surprising that this intersection "lit up" early and remains bright.

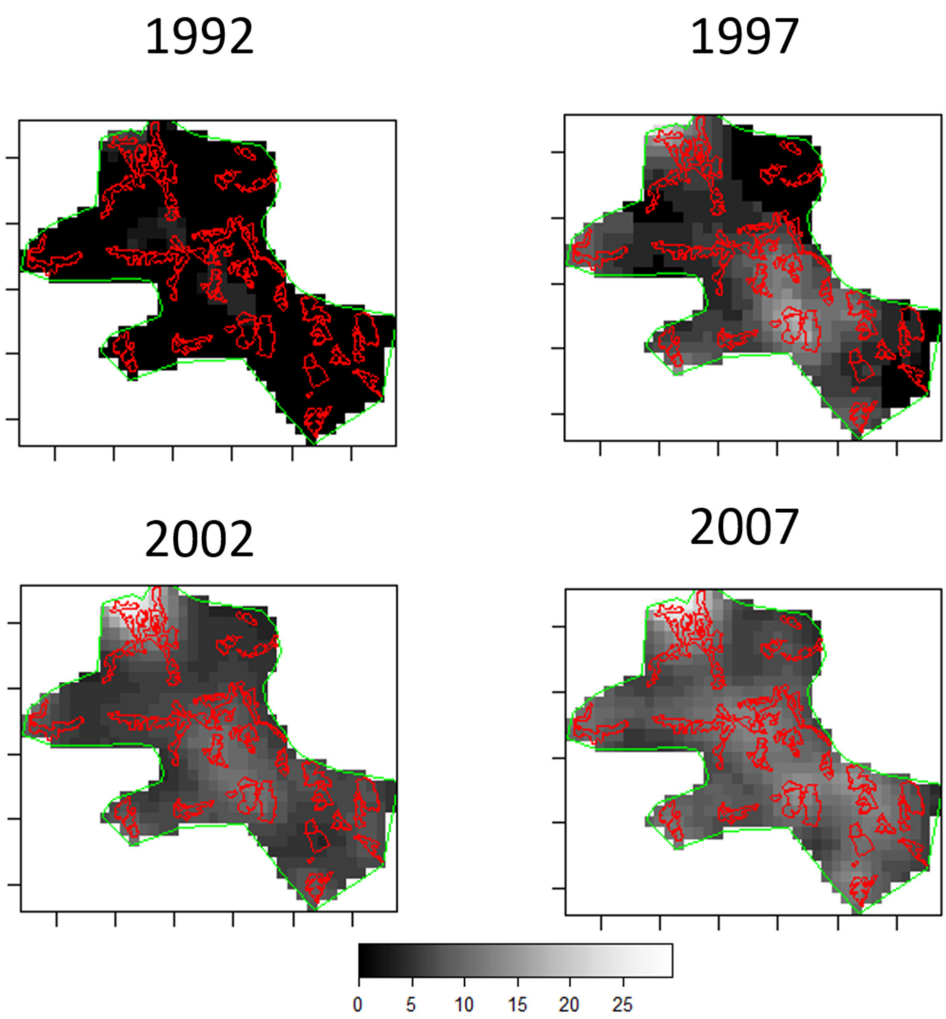

Figure 4: Differences in nightlight intensity in the Agincourt study site 1992 to 2007

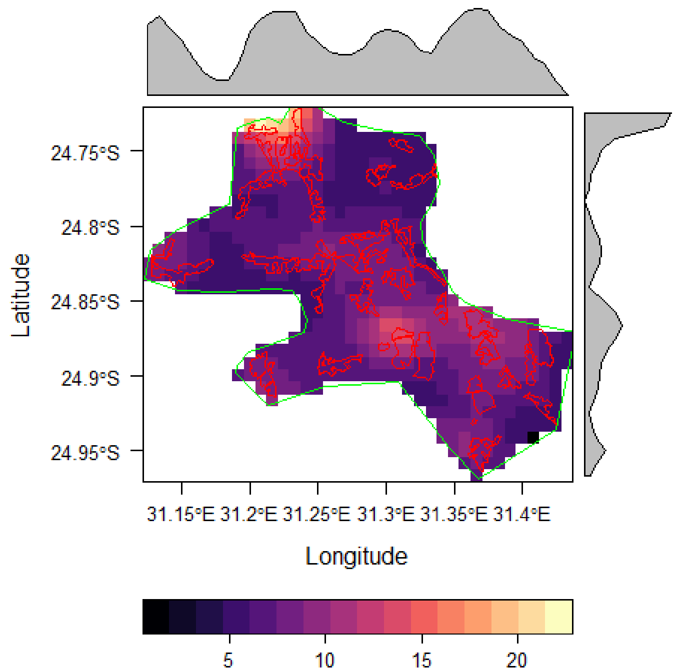

Figure 5: Distribution of nightlights in three villages in the AHDSS study area 2012

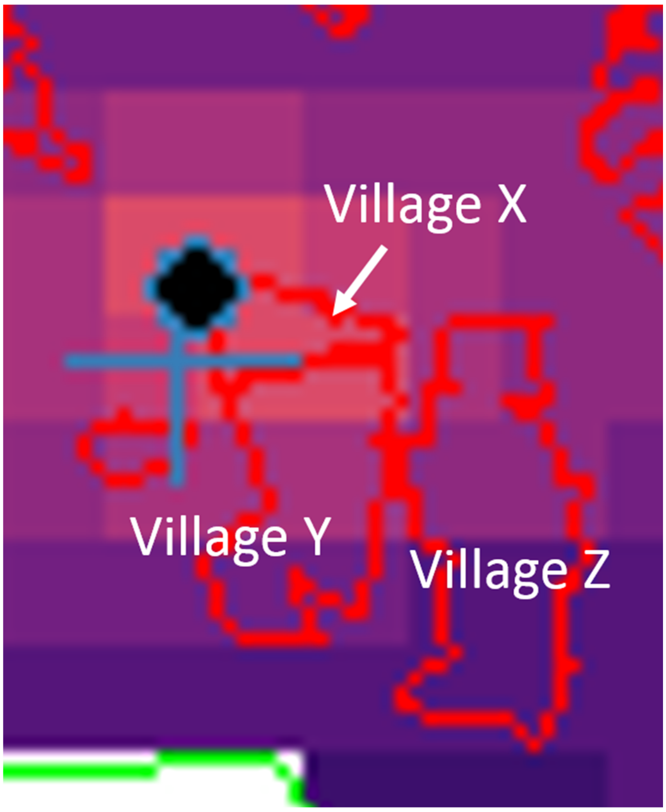

Figure 6: Detail from the 2012 nightlight distribution

Nevertheless the detailed picture in Figure 6 also raises several interesting issues about the data. Firstly, the brightest part of the area is Village X. This also happens to be a "refugee" settlement and one of the last to be electrified (it is the only such settlement that is not around the eastern fringe of the site). By contrast Village Z was classified as a "central" community and Village Y an "established" one. It seems likely that the economic facilities next to Village X (in particular shopping centre and taxi rank) are generating the brightness that is incorrectly attributed to the residential area of Village X. Indeed much of that activity also seems to be lighting up the surrounding farmland. Furthermore it needs to be remembered that the data was pre-averaged at a coarser scale and then rendered back down to square kilometre blocks. This seems a case where we are testing the limits of (and possibly going beyond) the resolution of the satellite data.

In summary there is little doubt that the satellite data is showing spatial variation in brightness. It also seems that this spatial variation is at a broad level correlated with the pattern of local development and electrification. Given the measurement issues, however, it remains to be established whether we can show a firm link between local electricity connections and village brightness.

Our first "hard" evidence is shown in Table 2. The first column shows that "undeveloped" communities were three points darker on the brightness scale compared to "central" and, indeed, to "established" communities in the year 2000. Given the average brightness in central communities (8.8) this is almost a 40% reduction. The point estimate for the refugee villages is also negative, although this is estimated with a lot of noise. Given the fact that Village X was actually one of the "bright spots" on the map, we see (in column 3) that once we eliminate its influence the other refugee settlements were also three points darker than the central settlements. Indeed the results of that column suggest a clear two-way split in brightness, with "central" and "established" communities having a brightness level around 8.8 and "undeveloped" and "refugee" settlements three points below that.

Table 2: Regression of nightlights (average SOL) on type of village

VARIABLES	(1) 2000	(2) Pooled	(3) Without Village X 2000	(4) Without Village X Pooled
Established villages	-0.0646	-0.0949	-0.0646	-0.0949
	(0.991)	(0.454)	(0.998)	(0.454)
Undeveloped villages	-3.356*	-1.774**	-3.356*	-1.774**
	(1.156)	(0.455)	(1.164)	(0.455)
Refugee villages	-1.829	-0.296	-3.086**	-1.219*
	(1.545)	(0.506)	(0.957)	(0.511)
Constant	8.842**	7.544**	8.842**	7.544**
	(0.797)	(0.321)	(0.802)	(0.321)
Observations	20	420	19	399
R-squared	0.360	0.041	0.568	0.049

Columns 2 and 4 provide the results when we pool over all the time periods. These results show that over the period 1992 to 2012 as a whole there was a noticeable difference in brightness between the central and the established communities on the one hand and the undeveloped and refugee ones (without Village X). These differences were not as stark as for the single cross-section, given that all villages started off fairly dark at the beginning and electricity had diffused fairly widely by the end.

From 2001 we have village level connection data from the HDSS. If we project average village brightness on electricity connections per pixel, we get the results in Table 3. In the top panel we have included all villages, including the anomalous data for Village X. Several points are evident. Firstly, the point estimates on the connections variable are positive in all years, although statistically different from zero at the 5% level only in 2003. The size of the coefficient in 2001 and 2003 suggests that 200 new connections[2] in a village (per square km) would increase the brightness level by 1.7 units, which given the baseline average brightness is a marked increase. Secondly, the explanatory power of the regressions (and the point estimate) goes down markedly after 2003. The main reason for this is that (as shown in the summary statistics in the appendix) the number of un-electrified villages drops from six in 2003 to just two in 2005 – and one of those is Village X. This means that the variation in the connections variable is driven by too few observations.

[2] It is useful to note that 200 connections per pixel is the average number of connections for electrified villages. In all cases villages that become electrified go from zero (or close to it) connections to more or less fully electrified, i.e. hundreds of connections. This value is therefore a reasonable benchmark against which to consider the size of the coefficient.

Table 3: Relationship between average village brightness and number of connections/square km

Panel A: All villages	(1) 2001	(2) 2003	(3) 2005	(4) 2007	(5) 2009	(6) 2011
Connections/pixel	0.00875	0.00961*	0.00113	0.00475	0.00150	0.00231+
	(0.00564)	(0.00424)	(0.00260)	(0.00389)	(0.00123)	(0.00122)
Constant	7.249**	6.557**	9.766**	9.439**	7.581**	7.601**
	(0.928)	(0.985)	(0.638)	(0.952)	(0.375)	(0.451)
Observations	21	21	21	25	27	27
R-squared	0.124	0.233	0.007	0.068	0.016	0.045
Panel B: Without Village X	(1) 2001	(2) 2003	(3) 2005	(4) 2007	(5) 2009	(6) 2011
Connections/pixel	0.0126*	0.0124**	0.00326	0.00750*	0.00144	0.00243+
	(0.00443)	(0.00334)	(0.00209)	(0.00303)	(0.00136)	(0.00126)
Constant	6.476**	5.865**	9.220**	8.732**	7.495**	7.440**
	(0.585)	(0.761)	(0.546)	(0.759)	(0.390)	(0.440)
Observations	20	20	20	24	26	26
R-squared	0.322	0.422	0.061	0.169	0.017	0.060

Robust standard errors in parentheses
** p<0.01, * p<0.05, + p<0.1

The bottom panel shows that the strength of the relationship is increased in the 2001 to 2007 period if Village X is excluded from the regression.

5.3 Using both the temporal and spatial variation

Our last set of estimates regress village brightness on household connections over time and space. The results are in Table 4. Column 1 includes year fixed effects but no other controls. We have kept Village X in all of the regressions. The estimate in that column suggests that an increase in 200 connections (per square km) would increase average brightness by 0.8 units. This amounts to a 10% increase on the baseline brightness, as reflected in the intercept. The coefficient is intermediate between the 2001-2003 estimates in Panel A of Table 3 and the estimates from 2005 and 2009-2011 in that Table. Interestingly the year effects provide strong evidence that the post-2008 data does not revert to trend. In column 2 we add a control for a quartic in the size of the area, on the assumption that there may be systematic measurement differences given how the village measures were constructed. These variables are jointly significant, but the coefficients other than the intercept are not materially affected. Adding village fixed effects does, however, lead to a much larger coefficient on the connection variable. The size of this coefficient now suggests that 200 new connections would lead to a 1.4 unit increase in brightness or a 20% increase on the baseline in 2001.

Table 4: Relationship of average village brightness to household electricity connections, 2001-2011

	(1) Pooled	(2) Pooled	(3) Pooled
HH connections/pixel	0.00413**	0.00419**	0.00720**
	(0.00128)	(0.00128)	(0.00189)
Year 2003	-0.429	-0.431	-0.521
	(0.703)	(0.673)	(0.493)
Year 2005	1.422*	1.417*	1.193**
	(0.607)	(0.598)	(0.426)
Year 2007	1.777**	1.797**	1.890**
	(0.657)	(0.640)	(0.407)
Year 2009	-0.764	-0.898	-0.837
	(0.577)	(0.566)	(0.426)
Year 2011	-0.642	-0.778	-0.841
	(0.600)	(0.588)	(0.429)
Village Effects	N	N	Y
Quartic in pixels	N	Y	N.A.
Constant	7.779**	11.00**	7.151**
	(0.542)	(1.281)	(0.397)
Observations	142	142	142
R-squared	0.272	0.319	0.761

Robust standard errors in parentheses
** p<0.01, * p<0.05

2001 is the base year

6. Discussion

We began by outlining a set of simple research questions. The first of these was whether the nightlight data is capable of detecting a rural electrification problem. The answer is clearly in the affirmative. The difference-in-difference estimates suggest that the Agincourt area brightened somewhere between .2 and .3 units per year over twenty years, leading to a net increase in brightness of around 4 to 6 units. The regressions reported in Table 4 suggest that increases in household connections help to predict the rise in brightness.

Nevertheless there is a mismatch between the magnitude of the increase suggested by the difference-in-difference estimates and the regression ones. The latter one would predict only around a one-and-a-half unit increase in brightness if a village were electrified. Indeed most of the explanatory power in the final regression is due to the year and village fixed effects. There are at least three reasons for the gap between the two sets of estimates:

- A connection is not sufficient for usage
  This is particularly obvious in relation to the drop in brightness in 2008 and the failure of brightness to recover after that. Nevertheless this was a period in which more connections continued to be rolled out. Interruptions in power (load shedding) and reduction in use (perhaps due to costs) will reduce the correlation between brightness and the number of connections.
- Non-residential electricity use
  We noted that the bright spot in Figure 6 was more likely to have come from commercial activity than from residential use. Of course rural electrification will have spin-offs for commerce, so those should feature in the treatment effect even if commercial use is not captured through household surveys.

- Measurement error
  It is clear that there is at least some measurement error in the brightness data as shown by the case of Village X and the farmland that is lit up in Figure 6. This is non-classical measurement error, i.e. the brightness of unserviced areas can only be overestimated. This reduces the explanatory power of the connection data.

Our second question asked whether the satellite data could pick up the difference between connected and unconnected areas. The answer to this is: somewhat. At a broad level, the evidence in Table 2 suggests that it can. Specifically, the results in this table show that the villages classified as less developed ("underdeveloped" and "refugee" villages) are consistently less bright. However, once we break the data down to the finer resolution of individual villages, the satellite data do less well at discriminating between places getting connected and those not. The case of Village X makes that clear. Nevertheless that example is not only about the spatial resolution of the data, but also about the connection between commercial activities and individual well-being. It is possible for the light data to pick up economic activity even when individuals living in the same area are deprived. It is therefore not clear how satellite data could settle debates about levels of deprivation and poverty. This naturally leads to the next phase in our research: looking at the economic impacts of connections to electricity.

As far as our third question is concerned, the data suggest that there is a clear correlation between local survey estimates of access and the nightlight data. This raises the possibility that with due attention to the measurement issues raised above, the light data could be used as a proxy measure of connection and development – even in rural areas and at a spatial resolution where this has hitherto not been attempted.

This raises the question what one might gain from establishing this relationship? Firstly one can now derive imputed measures of access in years and for contexts where there is no corroborating survey evidence. Secondly, one may be able to track usage patterns (e.g. load shedding) where the connection data is uninformative.

7. Conclusion

We have explored the reliability of satellite data in a local context where we have a lot of ancillary information. We find that on the whole the nightlight data seems to have captured the electricity roll-out in the Agincourt study site. It shows marked increases in brightness over time and captures the broad differences between "developed" and "undeveloped" parts of the site. Nevertheless we have also shown that there are some measurement issues which contaminate the relationship.

Acknowledgements

This work is based on research supported by the National Research Foundation, South Africa (ESRC-NRF International Centre Partnership, Ref ICPC150423117553). Any opinion, finding and conclusion or recommendation expressed in this material is that of the author(s) and the NRF does not accept any liability in this regard. We wish to thank Nathan Cunningham, Louise Corti and the UK Data Archive for assistance with the work on the satellite data. Suleiman Afolabi from the Agincourt Unit extracted the Agincourt HDSS data and supplied the shape files used to extract the satellite data.

References

Bekker, Bernard, Anton Eberhard, Trevor Gaunt and Andrew Marquard (2008) "South Africa's rapid electrification programme: Policy, institutional, planning, financing and technical innovations", ***Energy Policy***, 36:3125-3137.

Burlig, Fiona and Louis Preonas (2016) "Out of the darkness and into the light: Development effects of rural electrification", Energy Institute at Haas Working Paper 268

Dinkelman, Taryn (2011) "The Effects of Rural Electrification on Employment: New Evidence from South Africa", ***American Economic Review***, 101 (7): 3078-3108

Doll, C. N. (2008). CIESIN thematic guide to night-time light remote sensing and its applications. Center for International Earth Science Information Network of Columbia University, Palisades, NY.

Elvidge, C. D., P. Cinzano, D. R. Pettit, J. Arvesen, P. Sutton, C. Small, R. Nemani, T. Longcore, C. Rich, J. Safran, J. Weeks, and S. Ebener, (2007) "The Nightsat Mission Concept." ***International Journal of Remote Sensing*** 28: 2645–2670.

Elvidge, C. D., et al. (2009). "A fifteen year record of global natural gas flaring derived from satellite data." ***Energies* 2**(3): 595-622.

Elvidge, C. D., et al. (2014). "National Trends in Satellite-Observed Lighting." Global urban monitoring and assessment through Earth observation**:** 97-118.

Hargreaves, James (2000) "Village Typology : The Agincourt Field Site. Descriptive results and preliminary discussion", unpublished report.

Harris, Tom, Mark Collinson and Martin Wittenberg (2017), "Aiming for a Moving Target: The Dynamics of Household Electricity Connections in a Developing Context", ***World Development***, 97:14-26.

Henderson, Vernon J., Adam Storeygard and David N. Weil (2012) "Economic growth from outer space", ***American Economic Review,*** 102(2):994-1028.

Madubansi, M. and Shackleton, C.M. (2006) Changing energy profiles and consumption patterns following electrification in five rural villages, South Africa. ***Energy Policy***, 34, 4081–4092.

Matsika, R., Erasmus, B.F.N. and Twine, W.C. (2013) Double jeopardy: The dichotomy of fuelwood use in rural South Africa. ***Energy Policy***, 52, 716–725.

Michaelopolous, Stelios and Elias Pappaioanou (2014) "National Institutions and Subnational Development in Africa", ***Quarterly Journal of Economics***

Niehaus, I., (2001), ***Witchcraft, power and politics: Exploring the occult in the South African Lowveld***, Pluto Press, London.

Pinkovskiy, Maxim and Xavier Sala-i-Martin (2016) "Lights, Camera…Income! Illuminating the National Accounts-Household Surveys Debate", ***Quarterly Journal of Economics***, 579-631.

StatsSA. (2012) ***Census 2011 Census in Brief***. Pretoria: Statistics South Africa.

Tollman, S., (1999), "The Agincourt field site - evolution and current status", ***South African Journal of Medicine*** 89(8), 855—857.

Tollman, S., Herbst, K., Garenne, M., Gear, J. and Kahn, K., (1999), "The Agincourt demographic and health study - site description, baseline findings and implications", ***South African Journal of Medicine*** 89(8), 858—864.

Wittenberg, Martin, Mark Collinson and Tom Harris (2017) "Decomposing changes in household measures: Household size and services in South Africa 1994-2012", ***Demographic Research*** (in press).

Appendix: Summary Statistics

Table 5: Average Sum of Light 1992-2012

Area	Obs	Mean	Std. Dev.	Min	Max
Agincourt	21	6.149488	2.636918	0.377561	9.578964
Kruger	21	0.657972	0.693745	0.070729	2.415774
Nelspruit	21	34.8343	2.195828	27.59448	37.92947

Table 6: Average Sum of Light for Agincourt villages, by year

Year	Mean	Sd	Min	Max
1992	0.448365	1.006462	0	4.006763
1993	1.460189	2.820189	0	10.5396
1994	2.063257	2.352458	0	7.554447
1995	3.738389	2.776086	0	9.543436
1996	6.561656	3.550006	0	13.97618
1997	6.486015	4.056231	0	15.00769
1998	6.453361	3.656762	0	12.73489
1999	6.607349	3.707718	0	14.07452
2000	7.43928	3.397259	0.6121	14.72466
2001	8.648458	3.240899	3.976309	18.14873
2002	8.874296	3.29558	4.852222	19.26156
2003	8.413579	3.313221	2.378953	17.76361
2004	8.906009	2.896607	5.03021	17.58694
2005	9.992118	2.687678	5.133927	17.00544
2006	10.55441	2.774435	5.410313	18.33564
2007	10.98146	2.998985	6.10351	19.44133
2008	7.86658	2.410584	2.510348	15.40788
2009	8.210897	1.835965	5.105967	13.60118
2010	8.391726	2.01587	5.127581	14.10517
2011	8.710425	2.543336	5.220101	17.4213
2012	8.738582	2.441951	5.465878	16.45724

Table 7: Number of villages electrified, by year

	electrified	
year	No	Yes
2001	8	13
2003	6	15
2005	2	19
2007	2	23
2009	2	25
2011	1	26

Table 8: Number of household connections per square km, by year

year	mean	sd	min	max
2001	114.5676	97.73179	0	279.5402
2003	144.7195	120.5369	0	474.0741
2005	188.8342	112.3502	1.20075	519.9161
2007	185.9727	115.3923	1.20075	542.2781
2009	215.3624	116.6797	0	590.3564
2011	256.4579	144.8974	0	688.7491

About DataFirst

DataFirst is a data service dedicated to making South African and other African survey and administrative microdata available to researchers and policy analysts.

We promote high quality research by providing the essential research infrastructure for discovering and accessing data and by developing skills among prospective users, particularly in South Africa.

We undertake research on the quality and usability of national data and encourage data usage and data sharing.

DataFirst

www.datafirst.uct.ac.za
Level 3, School of Economics Building, Middle Campus, University of Cape Town
Private Bag, Rondebosch 7701, Cape Town, South Africa
Tel: +27 (0)21 650 5708
info@data1st.org / support@data1st.org

University of Cape Town · iYunivesithi yaseKapa · Universiteit van Kaapstad

SPES BONA