An Introduction To The Demographics Model

Development planning and budget projections hinge on accurately anticipating how the population of an area will change. For metropolitan cities these changes have to be anticipated for the city as a whole as well as for its various sub-regions. In South Africa electoral wards are at the heart of budget allocations, development initiatives and revenue generation making it desirable that population projections are available st this level at the very least.

This project seeks to improve the utility and accuracy with which ward level projections have been made. To date, projection methods have reflected those used for countries and population is treated as a product of births and deaths in each area as well as the level of migration to and from these areas. Unfortunately, for South African cities the level of migration between wards is unknown and possibly even unknowable. Instead of modelling ward population on this basis the method employed here rests on the assumption that wards and cities are part of an open system. This means that an increase in, say, life expectancy in a ward may very well result in an increase in population in another ward or city. At the core of this method is projecting past population trends at enumerator area level then recombining these areas to estimate the aggregate impact on wards.

Verification of the resulting projections was sought by examining supplemental data. In addition the supplemental information was examined to see if it could contribute to improving the method. Supplemental data included satellite imagery, voting returns, measures of social media activity and cellphone connectivity. Of all these new sources of data two additional sources proved to be of particular value.

1. The aerial brightness of cities at night is a good reflection of economic activity, traffic densities, level of electrification and thus the presence of populations. By measuring the nighttime light intensity of cities changes in population distribution should, theoretically, be discernible. Ultimately the luminosity of cities proved to be of little value but major changes in nightlight brightness was, in certain circumstances, a reliable indicator of increasing populations. As the nighttime luminosity of cities is constantly measured by satellites this may prove to to be a useful supplement to the method. As the satellite telemetry will be routinely updated it can be used on an ongoing basis to measure population changes. As the 2017 data has not yet been released the 2016 nightlight data was incorporated into this facility and is used to modify the enumerator area level projections.

2. It was anticipated that a strong indication of population changes would come from cellphone activity as would the extent to which people adopted social media. In the absence of the co-operation of the main cellular service providers alternative sources of information reflecting this connectivity were examined. These included the location registration of web browsers, tweets as recorded by twitter users with GPS-enabled phones, cloud-sourced indicators of WiFi activity and cellphone tower location. Each of these sources had their value and all suffer flaws that indicate they are not fit for the purpose intended here. For example, the nighlights data indicated marked population increases in several residential developments of Johannesburg. These developments were confirmed by satellite imagery. Twitter activity, web-browser registrations and measures of wifi and cellphone activity failed to identify these population changes. This failure can be attributed to many factors, however the main issue is that areas become "connected" some time after being populated. Unlike the nightlight data, these measures reflect infrastructure development (like installing cellphone towers or provision of ADSL/fibre services) that occurs only after the area is already occupied. In other words they better reflect infrastructure than they do population changes per se.

The quest for ways of improving the method (and of verifying the projections) were not confined to innovative sources. An analysis of changes in voter registration patterns proves to be a robust measure of population changes. In this instance voter registration patterns were used to verify the projections methodology rather than to modify the projections themselves. However, as the projection period increases, voter registration data will become an increasingly important way to recalibrate the projections made available here. After the next election in 2019 the voter returns can be used to recalibrate the existing model.

Two facilities are presented here. The main facility is the set of enumerator area projections to 2030, aggregated to ward level. An additional facility has this data aggregated to the resolution of the nightlight telemetry data. Users can then select the threshold at which these projections are over-ridden by nightlight-derived values.

How Not To Project Population

At country level population projections have proved to be relatively reliable and have afforded planners clear indications of what to expect. This is because the main determinants of population change, fertility and mortality rates and the levels of immigration and emigration can be reliably modelled to provide estimates of trends. Implicit in these models is the assumption that, within these parameters, the population of the system is closed. This means, for example, that the fertility rate of a country does not affected the population of neighbours except through immigration and emigration. The national population is depleted by death and emigration and supplemented by births and immigration. It is axiomatic that those born in the country will live there until they die or emigrate.

At sub-national level these assumption no longer hold. Estimates of fertility and migration rates, for example, get progressively less accurate at provincial, municipal and suburb levels. While it is possible to derive estimates of fertility and mortality from demographic profiles (socio-economic characteristics of households, gender and age profiles etc.) migration present a particular problem for modellers. To measure international counts of movements into and out of the country are required. To understand migration between municipalities the flows between every municipality in the country must be quantified. South Africa has 278 municipalities, meaning that over 77 000 different flows have to be measured. Moreover the focus of this study are electoral wards and with 4 392 wards the number of flow permutations rises to almost 20-million.

Although there are ways to decrease the demands on the information the central point is that modelling migration flows at ward level makes near impossible demands on those collating the data. The information required to treat cities or wards as closed systems do not exist and there is little point other than to treat the systems as open. In open systems changes in fertility and mortality in one area can readily impact on adjacent areas or other cities. A change in fertility in ward A can prompt a change in the population of ward B. Consequently knowing the enumerator areas fertility and mortality rates of any part of an open system will reveals little about population changes in that area.

In open systems changes in mortality and fertility rates can result in an increase in the population of the ward population or prompt migration from that ward. Take the example of a household where an aged parent dies. The death may prompt a decline in the population of the ward in which they live or may prompt the household to move to a smaller dwelling which, quite possibly, is in another ward. Another example is given by the desire of many middle class Gautengers to retire on the south coast of kwaZulu-Natal. Increased longevity in Gauteng then results in population increase in kwaZulu-Natal.

Ultimately the need to treat sub-national administrative areas as open ensures that the methods used to make projections in closed systems fail.

Have Data Or Treat The System As "Open"

It is tempting to employ closed system methodologies to make ward and city projections but the information requirements are not met. At sub-national level it cannot be assumed that those born in a city will continue to live there and the information demands for quantifying flows between small spatial entities enumerator areas are not met. Without adequate information on migration patterns knowing the fertility and mortality trends lends little insight into population changes. There are three main constraints on modelling migration at city or ward level:

1. The frequency with which censi are conducted
2. The ability of the census to capture movement between wards as these are unnamed entities which change frequently
3. The need to classify migration patterns where a significant proportion of migration can be typified as 'oscillating' with people moving to-and-from between cities and homesteads or a regular basis.

South Africa finds itself at a peculiar junction which further emphasise the centrality of migration in population estimates. The UN indicates that in 2016 the number of which births in the country matched the number of deaths for the first time. This means that there is currently no natural population growth in the country and all increases are due to immigration. It is expected that from this point on births will increasingly lag behind deaths. Consequently the SA population will gradually age and, putting the impact of immigration aside, start declining. The age pyramid of SA reflects this trend and displays little difference between the number of people in the youngest age cohort and those aged 25 to 29.
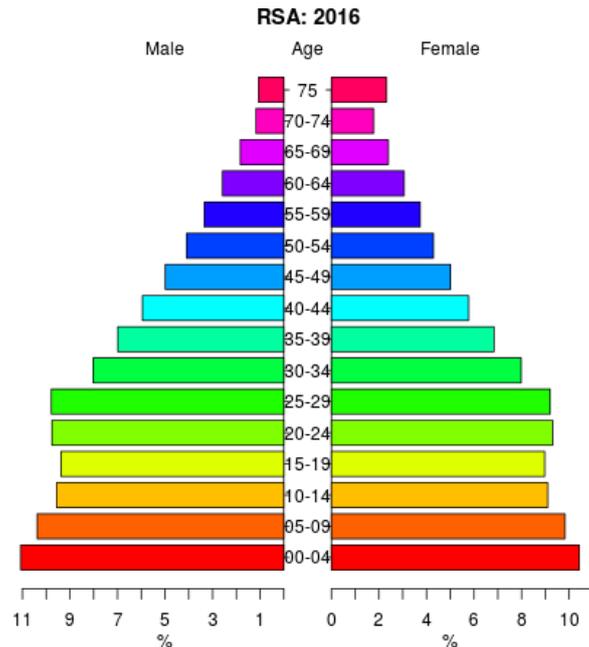
Urban areas tend have led the transition towards declining fertility and increased longevity (i.e declining mortality rates). These factors cast the role of migration at the centre of determining population trends. As indicated above it is precisely at this point that the data is the least systematic.

In other countries migration trends have been derived from the address registers associated with tax records, utility accounts or social security benefits. However South Africa joins that majority group where such data is inadequate. Essentially sub-national migration patterns can only be inferred from population changes post-facto.

Sub-national projections can now only be made on the basis of established or hypothesized trends. For example, if a key assumption is made that past growth patterns of any administrative area is informative of what is will come then population projections can be made. The value of these projections then rest on the veracity of that key assumption and the how well those trends are extrapolated. More importantly these projections can be made without reference to fertility, mortality or migration rates.

Metro Projections

At its simplest level the projections of city populations require knowledge of the population at two distinct periods (like the latest two censi). With this information the population growth can be projected linearly forward. This method of projection assumes that the observed growth rate continues unchanged for the period over which projections are made.

Improved population projections can however be made by obtained additional data points and by basing the projections on smaller spatial entities. For cities two additional data points can be derived for post-1994 South Africa. These are for 1996 and 2016.

- 1996 population estimates can be obtained by combining enumerator areas from the 1996 census to reconstitute the 2016 city boundaries.
- Recently estimates of the 2016 population for each municipality were provided by StatsSA. These estimates are derived from a national survey of 10% of dwellings units. While that survey did not involve a re-listing of all the dwellings in the sampled areas the number of dwellings was adjusted on the basis of satellite and other imagery. That survey is taken as a reliable reflection of the population in each municipality.

Collectively the data thus provides four data points (1996, 2001, 2011 and 2016) for each metropolitan area. Projecting the population of each metro then becomes a question of distilling the underlying trend and extrapolating that trend to 2021 and beyond. Deriving the underlying trend is essentially a question of fitting a curve to the observed points. That curve is then extrapolated forward to make projections.

Curve-fitting And Metro Projections

The growth rate of metropoles has, in general, declined markedly since 1994. Between 1994 and 2001 the annual growth rate of metros was 3.4%. During the 2011-2016 period this had halved to 1.7% per annum. Although the rate of growth is declining the absolute increase in population is substantial - particularly in Gauteng. Thus while the cities are experiencing massive population growth the tempo at which that population is growing is abating. These changes are shown below.

Metro	1996-2001	2001-2011	2011-2016
Johannesburg	4.7	3.7	2.3
Cape Town	3.5	2.9	1.4
eThekwini	2.7	2.3	1.3
Tshwane	4.1	3.6	2.4
Ekuruhleni	4.7	2.8	1.3
Nelson Mandela Bay	1.0	1.5	1.9
Buffalo City	0.7	0.7	1.5

There are significant differences in the growth rate of the larger cities and the two smaller metros (NMA and BUF). The smaller cities enumerator areas currently show increasing rates of growth while the larger metros show declining growth rates. In Tshwane and Johannesburg the growth rate is declining more slowly than in other cities. In Cape Town and Ekuruhleni the drop in the growth rate has been more marked. The growth rate of enumerator areas Ekuruhleni the growth rate is currently one-quarter what it was during the 1996-2001 period. In Cape Town and Ekuruhleni the drop in population growth has been so rapid that these cities may see population stagnation before 2030.

The trend to population stabilisation conforms to the evidence that the South African population is no longer growing "naturally" and will, within the next decade, enter a phase (but for immigration) of decline. Internal migration has a markedly different dynamic and there is remains significant scope for ongoing urbanisation for some time to come.

Identifying the equation that best fits the observed population for each city in 1996, 2001, 2011 and 2016 is a the centre of making projections. In each instance a substantial improvement on a linear projection is obtained by using a linear projection based on a "second order polynomial". These curves correspond to the form $y=ax + bx2 +c$ where y is the population in the year for which a projection is to be made, 'x' is the population observed in the past and parameters 'a' and 'b' show how the curve changes over time. For each of the cities that curve that a) best fits the four observed points and b) projects that trend to 2030 is shown in the graphic below.
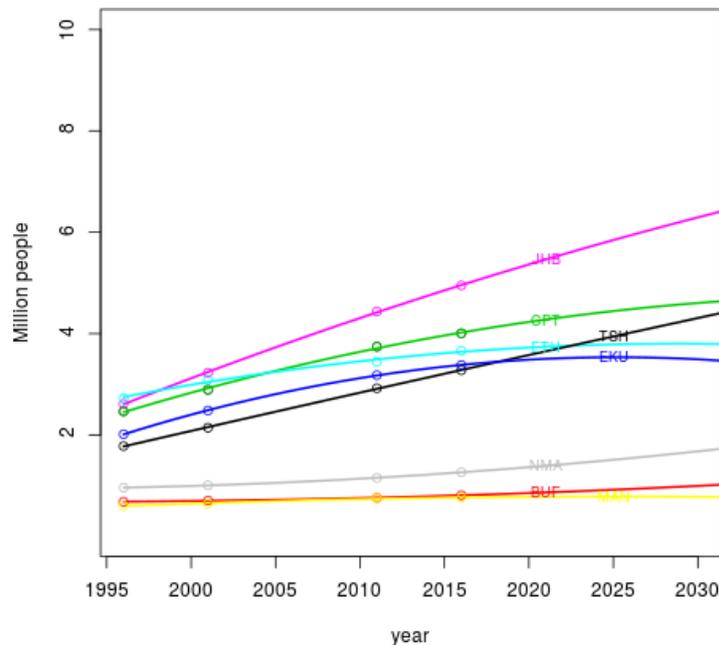
Metro Populations

The city population projections are robust and reliably indicate population trends - without calling for details of fertility, mortality and migration rates. The core role of the city level projections is to provide a target for the ward estimates. For any given year the sum of the population in a city's wards has to equal the estimate for the city as a whole.

Given that the city projections are unknown the key question then becomes how the ward population is distributed in each city. The simplest method is to project trends in each ward and calibrate these to ensure they sum to the total of that city.

The objective of this study was to make projections of ward population for the metropolitan areas. The simplest way to do this is to make a linear projection of ward population form the last two censi. Unfortunately this rapidly gives rise to implausible results. For example, some wards become entirely denuded of population and others show massive increases in population and unrealistic population densities. The situation is partly ameliorated by basing the projections on the last three censi (1996, 2001 and 2011). Fitting curves to the three data points (1996, 2001 and 2011) and projecting that trend does not entirely eliminate the tendency to implausible populations. Despite the large sample size the 2016 Community Survey results were not released for sub-metro entities like wards or enumerator areas.

The shortcoming described above can be addressed by treating the wards as aggregations of smaller components, fitting curves to each of the components and enumerator areas aggregating them back to reconstitute ward profiles. The 'components in this instance can be taken to be enumerator areas (a term used here as synonymous with "small area level"). This method results in ward populations being derived from a complex interplay of the constituent components. By doing this no ward can be totally denuded of population unless every enumerator area in that ward has a declining population. Conversely implausibly large ward populations are undermined by rapid growth in some enumerator areas being offset by declining populations in other enumerator areas. The reasonableness of population growth is enhanced when enumerator area populations are themselves limited to reasonable population densities.

The Method

The methodology adopted here rests on the ability to project population for sub-national entities for three points in time - namely the 1996, 2001 and 2011 censi. To do this it was necessary to derive reliable 1996 and 2001 population estimates for each enumerator area as they were defined in 2011. Once again "second order polynomial curves" were fit to these points. This curve was then projected and used to derive population estimates for each enumerator area for the period 2012 to 2030.

The projections (inevitably) resulted in some implausible estimates for some, as ward downward sloping projections eventually result in population estimates that are below zero (negative). Conversely the projections based on an increasing trend could result in rapid population growth and, by implication, implausible population densities. For this reason limits are placed on projections. The two basic rules applied were:

1. Negative populations are assumed to reflect empty enumerator areas and negative populations were set to zero.
2. The population of any area was not allowed to exceed the maximum density of enumerator areas observed elsewhere in that metro. The maximum population density was determined for the main (largest) category of housing type in every ea. The population density for that enumerator area was restricted to the maximum density observed for that housing type elsewhere in that metro. Areas that, for example, were dominated be informal settlement housing were not allowed to exceed the highest density of other enumerator areas dominated by informal areas in that metro. A threshold was similarly set for formal dwellings, flats and other categories of housing.

    Once the projected population reached the threshold area in terms of density the population for the enumerator area was held at that level. The resulting projections were enumerator areas re-weighed to ensure that the sum of the enumerator areas equalled the population projected for the metro. This was simply a question of reweighing each enumerator area population pro-rata.

Like wards the projections of enumerator area populations gave rise to implausible or, at least, questionable results. However the essence of this method is to aggregate the enumerator areas to wards (or similar level). To determine the ward population the over five thousand enumerator areas in Johannesburg are combined to reflect trends in 130 wards - with an average of 43 enumerator areas in each ward.

Although estimates are made for every enumerator area these are, on their own, unreliable and often misleading. By contrast the aggregated result are treated as reliable and should be verified as such. Unfortunately projections, almost by definition cannot be verified timeously. The next census will be in 2021. Only a year or so after that will new population estimates at ward level become available. Verification of the projections has to be sought elsewhere. One candidate for verification are the returns for voting districts in each city.

Verification

As a rule projections require the use of the most recent data available and are to used to reflect a situation some point in the future. Consequently it is extremely to difficult to verify projections. It will only be possible to ascertain how accurate the projections were some time after the next census. By that time the utility of the projections will have passed. This said, the veracity of the projections may be given by supplementary data like the registration statistics for local government elections. Local government elections (which use the wards as the basis for choosing representatives) were last held in 2011 and 2016. It should be possible to use this data to what extent population projections for 2016 correspond to the population profile as inferred from voter registration statistics.

In a classical 'gotcha' there proves to be a poor correlation between the projected ward population and the ward registration figures within any city. This is because, by law, wards have the same number of registered voters thereby ensuring that the proportion of voters in a ward generally approximates the proportion of the population of the ward. Differences between these two proportions arise from age profile differences, the proportion of immigrants that were not entitled to vote and differing propensities of adult citizens to register. Consequently the wards all have similar populations with the differences between wards reflecting these factors rather than discrepancies in population size. The upshot is that the ward population projections cannot be verified by correlating the number of registered voters in each ward in 2016 to the population projected for that ward.

In Johannesburg, for example, there is a statistically significant correlation between registered voters and the projected population estimates. However only a very small proportion of the difference in the number of registered voters can be attributed to changes in the population (population accounts for 6% of the variation in registered voters). Using the unprojected (and thus 'reliable') enumerator areas 2011 statistics 8% of the variance in the number of registered voters could be attributed to differences in the total ward population. This means that the projection model cannot be verified within a municipality when using the 2016 Local Government Election data or, for that matter returns from any other election

Voting districts by contrast differ. Unlike wards, voting districts are defined in a way that facilitates the management of elections and are not required to be the same size. Unfortunately the average voting district covers a small number of enumerator areas and, given that the design rests on combining an adequate number of enumerator areas, cannot be used to verify the projections.

An alternative verification process is to create artificial ward-sized agglomerations of enumerator areas. These pseudo-wards are not required to be the same size in terms of registered voter populations yet approximate wards in terms of the number of enumerator areas they contain. Using the 2011 data from Johannesburg,

as an example, clusters of 32 enumerator areas were arbitrarily combined. The population of these pseudo-wards were then correlated to estimates of the number of registered voters they contained. The relationship was as expected: positive (greater 2011 populations corresponded to more voters registered for the 2011 election), statistically significant (the correlation was not random). Most importantly 69 percent of the variation in the number of registered voters could be explained by the population size in pseudo-wards. The remaining 32% of variation would be accounted for by differences in age profiles, the proportion of non-citizens in the population and differing propensities of adult citizens to register to vote. The 2011 estimates were drawn directly from the census and IEC data and did not, to any extent, rely on projections. The r2 value of 0.69 (i.e. the 69% of variation) sets a benchmark against which the projected estimates could be evaluated.

In 2016 the number of both wards and enumerator areas. The pseudo-wards were accordingly based on agglomerated clusters of 43 enumerator areas. Estimates of the number of registered voters and the 2016 population were made for each pseudo-wards. The 2016 populations were derived from the projections made earlier. Unfortunately a small number of anomalies in the aggregation of registered voters were evident. These anomalies were a product of the aggregation method which resulted in sparsely inhabited industrial areas being allocated registered voters. The anomaly lies with the allocation of registered voters and not in the projected population. When comparing the 2016 values against the 2011 benchmark of 68% the most anomalous outliers were excluded.

The correlation of the registered voters against the projected 2016 population was, once again, to be in the expected direction and statistically significant. Given this, the most important factor was how much of the variation in the number of registered voters in 2016 could be explained by the projected population of 2016. Given the nature of projections it was expected that the projected population would have less explanatory power than the 69% observed in 2011. Reductions in the r2 value would suggest that the projections were departing from the trend implied by voting registration. In fact the r2 enumerator areas value increased slightly to 0.7 (70 percent of the variation in registered voters could be explained by the 2016 population projection). While the increase in not thought to be statistically significant it does affirm that the underlying method appears reliable and, at the very least, conforms to what the IEC data implies.

Aggregating EAs

The creation of the wards offers further insight into how the enumerator area data can be used. In the above example the wards were agglomerated into equal sized chunks containing 43 enumerator areas to reveal a r2 value of 0.7. Creating a larger number of pseudo-wards (by agglomerating fewer enumerator areas) reduces the explanatory value - the r2 value drops. By examining the impact of reducing enumerator areas the number of enumerators so combined guidance is given to those wishing to create their own 'pseudo-wards'.

The plot below compares the explanatory value of the projections in a given year (the r2 value shown as percentage on the graph) to the bin size (the number of enumerator areas in each pseudo-ward). Essentially reducing the bin size undermines the explanatory value of the population estimate. If bin size is dropped from the current 43 enumerator areas to 26 then the proportion of unexplained variation in the registered voters rises from 30% to 37%. A drop to 13 results in over half (54%) of the variation being unexplained.
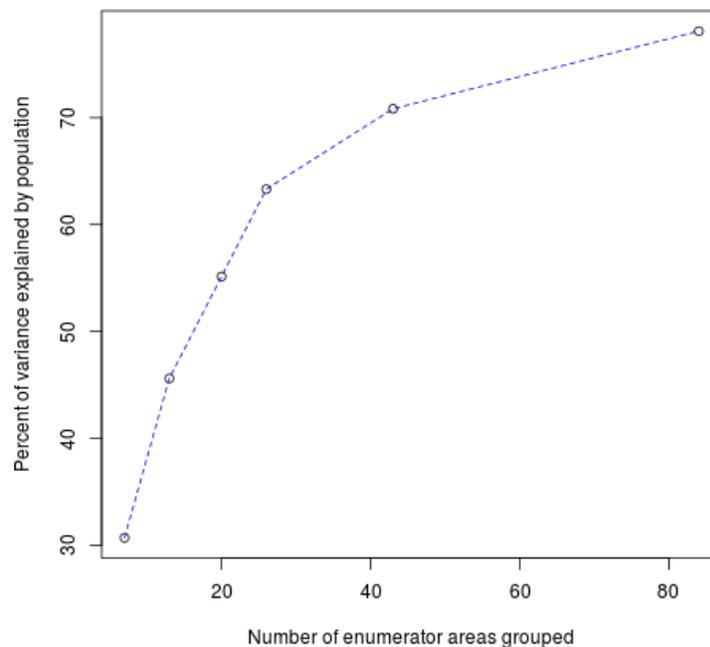
The projections for individual enumerator areas are made available on the understanding that these may be recombined to create elements akin to the pseudo-wards. However the binning of too few enumerator areas is clearly problematic. The size of the resulting problem is shown in terms of the declining explanatory value.

The above model examines the impact of bin size on the utility of the estimates. The explanatory value of the projections will also invariably decline as they reach ever further into the future. However knowing to what extent this occurs can only be known after the fact - possibly in 2022 or 2023.

Modifying Projections

The projection model rests on the assumption that recent changes in the population density of enumerator areas will continue until some threshold level is reached. Once that density has been realised the population is fixed until the end of the projections period. There is obviously no certainty that past trends will continue indefinitely and the model should change when new information comes to light. One potential source of information is satellite telemetry showing how luminous areas are at night. At city level there is, internationally, a clear correlation between night time luminosity and the level of economic activity taking place. There is also a general correlation between the level of economic activity and the population of the city. Brighter cities clearly denote higher populations(at least in developed economies). Unfortunately, as an examination of South African cities shows, the correlation between population density and night time luminosity

breaks down.

Statistically there is very little correlation between luminosity and population. With heavy traffic, more regular street lighting and wide-spread use of lit windows by shops business districts tend to be very well lit. These areas also tend to have small residential populations. Townships and suburbs, on the other hand, tend to be poorly lit (at least as seen from the air) yet contain higher densities of residents. These interactions result is a poor correlation between luminosity and population density.

The graphic below reflects the luminosity of Johannesburg at nigh in June 2016. Those familiar with the areas will be able to identify the main business districts and filaments indicating major arterial routes.

Johannesburg: Nightlight values
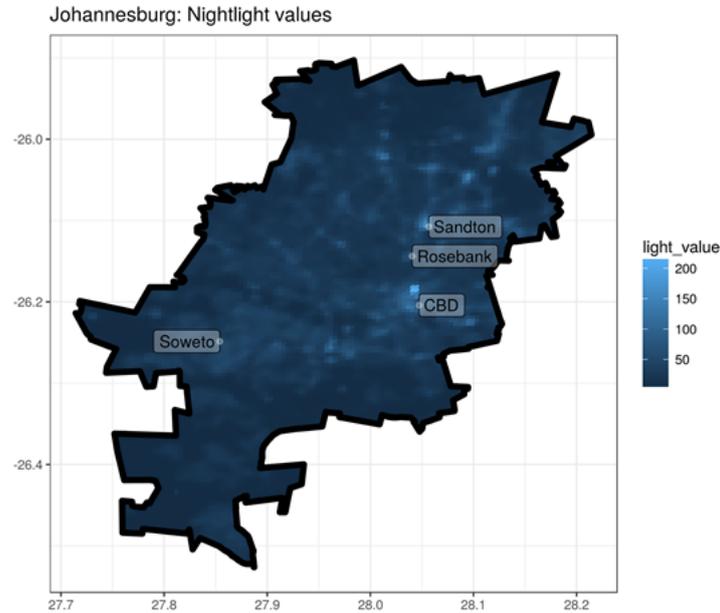
The absence of a clear correlation between population density and luminosity does not render the nightlights data irrelevant. Of particular value is the fact that changes in luminosity are informative of population density. In particular increased luminosity indicates rising economic activity and increasing population density.

Areas that 'brightened' significantly after 2011 were typically residential areas that had become more densely populated after 2011. These areas are illustrated by the yellow dots on the map below.
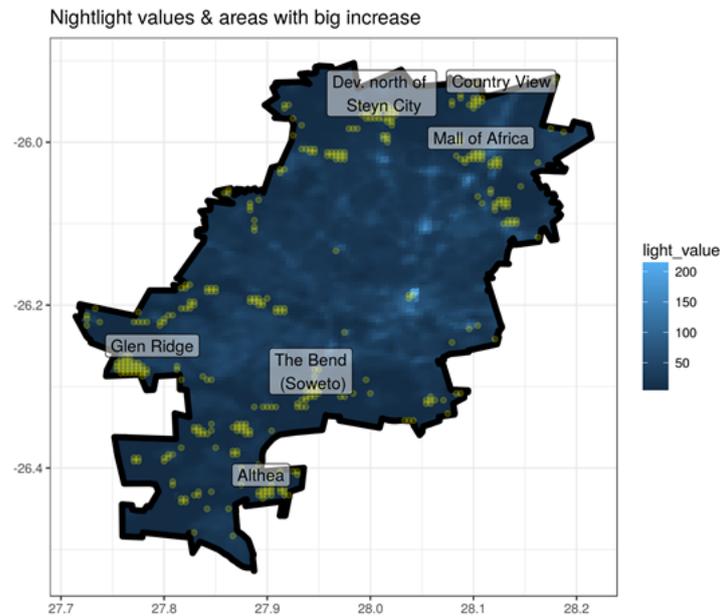
Nightlight values & areas with big increase

The above model examines the impact of bin size on the utility of the estimates. The explanatory value of the projections will also invariably decline as they reach ever further into the future. However knowing to what extent this occurs can only be known after the fact - possibly in 2022 or 2023.

The map below shows those areas where luminosity changed by more than 0.5 points between 2011 and 2016. Change in luminosity is measured on a scale of 0 to 2 with 2 indicating the most dramatic change in night time luminosity between June 2012 and June 2016. June values are used as this corresponds to winter - the time of year where, in Gauteng at least, the enumerator areas tree canopy in residential areas is at its seasonal minimum. Using winter values reduces the canopy differential between older suburbs and newer areas where trees are poorly established. Unfortunately nightlight values for June enumerator areas are not available prior to enumerator areas 2012 - hence the use of the 2012 values.

The change in luminosity is measured, technically, by the difference between the 2016 and 2012 z scores of the logarithm of observed values. A luminosity value of 2 corresponds to the 2016 value (logged) being two standard deviations more than the 2012 value (logged).

The system operates by the user first identifying how big a difference between the 2011 and 2016 values is called for before the projected value is over-ridden. This is a reflection of the preference the user has for using population estimates based on luminosity rather than the projections. The luminosity value of all the areas above the selected threshold are used to derive an estimate of population density using a simple formula. Most of the areas in which there has been marked increases in luminosity are new townships. By examining the nightlight and population values enumerator areas of a cross selection of townships a relationship between density and luminosity was derived. This formula was then used to estimate population figures for all areas falling above the selected threshold. Those new values were then used in lieu of the projected value.

The projections for 2016 may thus be 'enhanced' by identifying those areas which had significant increases in night time luminosity. Those luminosity value of those areas falling above the selected threshold were used to derive a new population estimate. As indicated the method is unproven and rests on confidence in the nightlight enumerator areas satellite telemetry.

The nightlight telemetry does not conform to administrative boundaries and is published at a fairly coarse resolution: squares of approximately 500m by 500m. These pixel are thus larger than many enumerator areas. Consequently it is necessary to provide the nightlights data as a grid with the coarser resolution.

Users are presented with current projections for 2016. They then decide at what level (the threshold) they want the population estimates to over-ride the projections. This is expressed as a number between 0.5 and 2 (with 2 indicating the biggest increase in luminosity enumerator areas over the period. The new values for each grid reference is then presented. The new population calculations are based on enumerator area values and are aggregated to grid level. For enumerator areas above the change threshold projections are based on the actual luminosity value and not the original projection. For enumerator areas enumerator areas below the enumerator areas threshold the original projections are used. These estimates are then re balanced to make sure they the total matches the projected city population.

The nightlights component is used to adjust measurements for 2016 (the latest year for which the telemetry is available). The intention is for the model to be updated each year after the June nighlights values become available. The new telemetry will the n be used to adjust that years projection based on the differences in luminosity between the last census and the most recent telemetry.