G Mans
Planning Support Systems
CSIR
e-mail: gmans@csir.co.za
tel: 021 888 2546

Population grid for the WC: 2011 census population totals

Creating the variable sized grid

In order to know what the potential demand for a service would be, it is important to know as accurately as possible the population distribution of the Western Cape. In other words, one needs to know on a detailed geographical level where people are living and the density of the population in these inhabited places.

The base data to establish the population distribution is a dataset developed for ESKOM by CSIR. It is a dataset consisting of points digitized in GIS for each building in South Africa. The digitizing was done using SPOT satellite images to identify the buildings. This means that the starting base data is a point dataset with potential weights assigned to the building relating to the potential amount of people living in that structure.

The goal for this project was to create a population dataset that can distinguish the distribution of the population on a detailed level. So in the case of Stellenbosch, this means that Kayamandi and Stellenbosch Central are not in the same raster so people's accessibility to facilities can be determined as accurately as possible. After some tests in different towns and with different raster sizes, it was decided that a 1 X 1km raster would offer the needed detail in built-up areas for the analysis. It is, however, not necessary to have the same level of detail in lower density areas or rural areas as the change in socio-economic profile and contribution to the total demand is not that dramatic over the same distances. In other words, the lower the density of an area the bigger the size of the raster can be in order to yield results similar in accuracy to that of the more densely populated areas. This approach makes the dataset also more manageable from a computational perspective; that is, the computer processing time to undertake an analysis based on the dataset is more manageable.

The result is a nested raster with different grid-cell sizes in one layer. The layer has been vectorised. One squared kilometre for densely populated areas (>250 households in that area), 2km^2 for less densely populated areas (1-250 households), and 4km^2 for areas with no population (0 households). This allows the grid to be fine where needed but saves computation power in the less densely populated areas. The process is described in Figure 1.

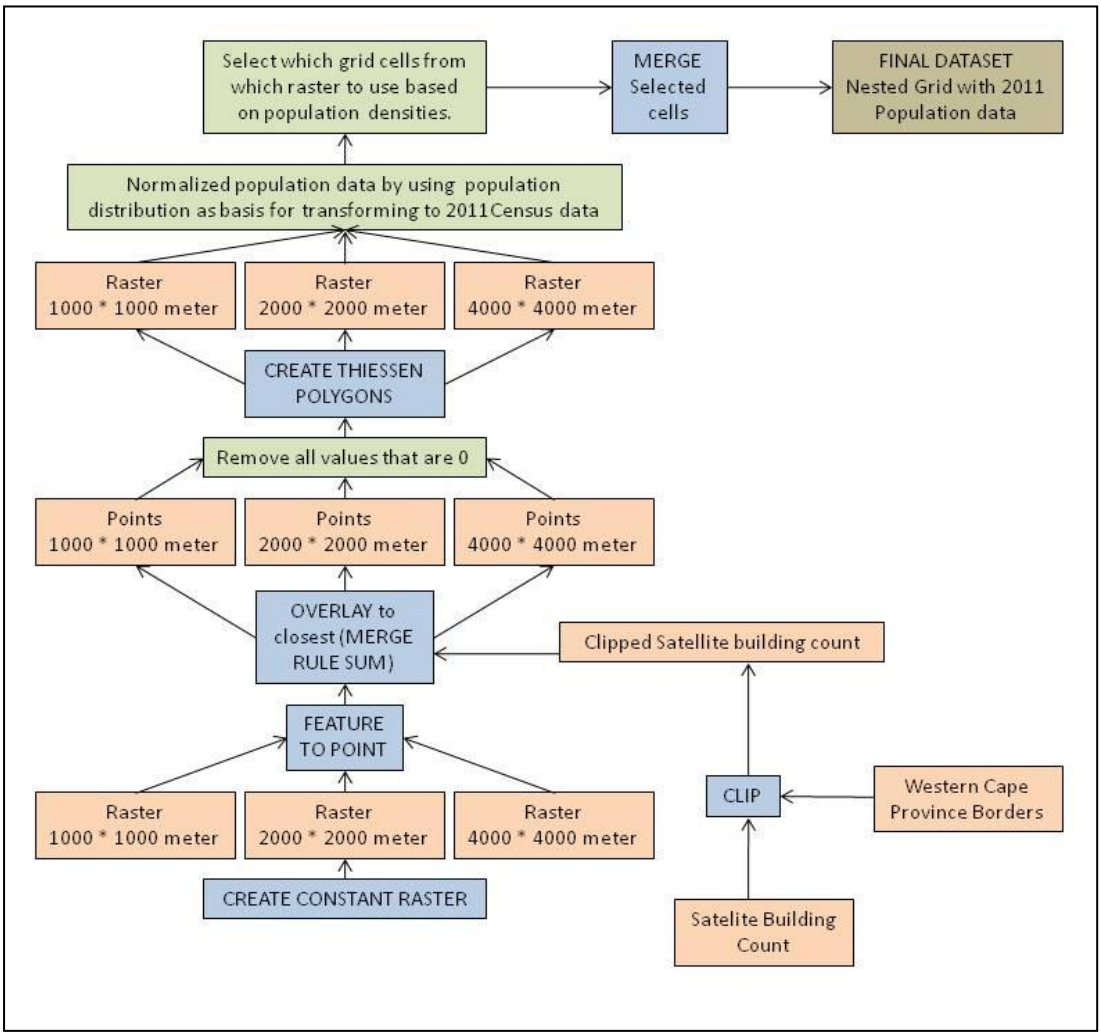

Figure 1: *Creation of a nested grid dataset*

Assigning population totals to the grid

Dasymetric mapping and areal interpolation

A dasymetric map is the result of a procedure applied to a spatial dataset for which the underlying statistical surface is unknown, but for which the aggregate data already exists. The aggregate dataset's demarcation is however not based on variation in the underlying statistical surface, but rather the result of convenience of enumeration (Eicher & Brewer 2001; Mennis & Hultgren 2005). The process of a dasymetric map therefore involves transforming data from the arbitrary zones of the aggregate dataset to recover (or try to recover) and depict the underlying statistical surface. This transformation process incorporates the use of an ancillary dataset that is separate from, but related to, the variation in the statistical surface (Eicher & Brewer 2001). Dasymetric mapping therefore has a close relationship with areal interpolation – the transformation of data from a set of source zones to a set of target zones with different geometry (Bloom et al. 1996; Fisher & Langford 1995; Goodchild & Lam 1980). Areal interpolation is mostly an areal weighting procedure and does not take ancillary sources into consideration when the spatial distribution of data is refined. Many areal

interpolation methods can be incorporated into dasymetric mapping methods to improve the detail of a choropleth map below the level of the enumeration unit (Fisher & Langford 1995, Hay et al 2005).

As can be deduced from the principles of dasymetric mapping as a method of areal interpolation, the accuracy of the depiction of the data is heavily dependent on the quality of the ancillary data used to predict the variation in the spatial distribution of the variable in question. Another consideration is also that the ancillary data used must be updated regularly. The ancillary data used for the population grid update was the SBC done for ESKOM.

Preparing the SBC as ancillary dataset

A hierarchical exclusion process based on dasymetric mapping principles was used do the classification of the SPOT building count – a count done for ESKOM of all buildings in South Africa. According to this classification, each point inherited a weight representing the potential contribution (household size) of the point in question. The following factors were taken into consideration in order to undertake the classification:

- residential areas which have not changed since the last census;
- new urban growth areas;
- informal areas;
- commercial and industrial areas;
- rural areas (agricultural, nature reserves and other sparsely populated areas).

The results of this process were tested against detailed population updates of the City of Cape Town and eThekwini in 2011. The intraclass correlation coefficient (ICC) per census sub-place was calculated. The intraclass correlation coefficient is a descriptive statistic that can be used when quantitative measurements are made on units that are organized into groups. It describes how strongly units in the same group resemble each other. The groups in this instance would be each sub-place with two measures indicating the total population. It differs from Pearson correlation in that it shows not only if there is a correlation in trend but also intra (internal) correlation between the values obtained (Mans, 2011). The results were an ICC of:

- 0.907 for City of Cape Town
- 0.899 for eThekwini

In other words, the values produced using the SBC-data frame method is consistent with that of the control totals (total population as derived by the two cities respectively). Please note that this test was done to show the relevance of the SBC after the classification process as an appropriate ancillary set to assign population data to the grid.

References:

Bloom LM, Pedler PJ & Wragg GE (1996) Implementation of enhanced areal interpolation using MapInfo. Computers and Geosciences 22:459-466

Eicher CL & Brewer CA (2001) Dasymetric mapping and areal interpolation: Implementation and evaluation. Cartography and Geographic Information Science 28:125-138.

Fisher PF & Langford M (1995) Modelling errors in areal interpolation between zonal systems by Monte Carlo simulation. Environment and Planning A 27:211-224.

Goodchild MF & Lam NS (1980) Areal interpolation: A variant of the traditional spatial problem. Geoprocessing 1:297-312.

Hay SI, Noor AM, Nelson A & Tatem AJ (2005) The accuracy of human population maps for public health application. Tropical Medicine and International Health 10:1073-1086.

Mans GG (2011) Developing a geo-data frame using dasymetric mapping principles to facilitate data integration. AfriGEO Conference: Developing Geomatics for Africa, May – June 2011, Cape Town.

Mennis J & Hultgren T (2005) Dasymetric mapping for disaggregating coarse resolution population data. In: Proceedings of the International Cartographic Conference, July 9-15, A Coruna, Spain.