

# Using a hybrid methodology of dasymetric mapping and data interpolation techniques to undertake population data (dis)aggregation in South Africa

Mawande Ngidi, Gerbrand Mans, David McKelly, Zukisa Sogoni

Urban and Regional Planning, Spatial Planning and Systems, Built Environment, Council for Scientific and Industrial Research (CSIR), Stellenbosch, South Africa. [mngidi@csir.co.za](mailto:mngidi@csir.co.za)

DOI: <http://dx.doi.org/10.4314/sajg.v6i2.8>

## Abstract

*The ability of GIS to produce accurate analysis results is dependent on the accuracy and the resolution of the data. In many instances the resolution of census enumerator tract data is too coarse and therefore inefficient in conducting fine grained spatial analysis. Dasymetric techniques can increase the spatial resolution of data by incorporating related high resolution ancillary data layers allowing the primary data to be represented at finer resolutions. Areal interpolation relates to a geostatistical process of transferring data from one set of polygons to another. This paper proposes the application of a hybrid technique using dasymetric mapping and areal interpolation principals to overcome the issues of transferring data from arbitrary spatial units to fit for purpose analysis zones on demand. As a consequence the technique also overcomes the problems of coarse scale population data as well as issues relating to the modifiable areal unit problem (MAUP). The data used to illustrate the value and accuracy of the developed methodology is that of the 2011 census population data and ESKOM's SPOT building count. The final outcome is an algorithm allowing the disaggregation and aggregation of population data to any spatial unit with a high level of accuracy.*

*Key Words: Dasymetric Mapping; MAUP; Population; Census; GIS*

## 1 Introduction

The uptake of Geographic Information Systems (GIS), advances in computing power as well as advances in software development has led to an increase in the use of spatial data in decision making. The ability of GIS to produce accurate analysis results is dependent on the data that is available for analysis. Having accurate geographic data at the right spatial resolution is paramount to analysing phenomena such as population shifts and distributions. High resolution data depicting people and place dynamics as close to reality as possible is paramount to inform localised policy and planning processes.

Globally it has been noted that in many instances the resolution of census enumerator tract data is too coarse and therefore inefficient in conducting fine grained settlement level spatial analysis

that clearly represents the distribution of population, as it seldom clearly represents the underlying variation in the distribution of phenomena such as population in the areas of enumeration. The reason for this is that the demarcation of an enumerator area is undertaken for the purposes of obtaining a relevant sample and does not necessarily adhere to only populated areas. Spatial resolution (especially finer grained population distribution data) plays an important role for planning purposes in many fields. For example: accessibility analysis for facility location planning; location allocation modelling; spatial interaction modelling; infrastructure development; urban growth studies; public services; and resource allocation optimization (Yao Yao *et al.*, 2017, Barrozo *et al.*, 2016, Green *et al.*, 2012, Mennis and Hultgren, 2005; Páez and Scott, 2004).

In South Africa the population census is conducted by the national statistics agency, Statistics South Africa (StatsSA). This census data is collected using varying (area) sized enumeration areas (EAs) which are fundamentally discrete work units for enumerators during the census and generally comprise between 100 and 250 households (Mokhele *et al.*, 2016). The fundamental principle for this sub-division is that an “EA should be within reach of a fieldworker and all households in that EA must be covered within the allocated number of days” (StatsSA, 2011). Therefore one can conclude that it is created for logistical reasons. After enumeration, the results are aggregated and made available to the public at demarcations that decrease the spatial resolution of the data for confidentiality reasons.

In the case of the StatsSA data, EAs are the foundation of sub-places which can be aggregated into bigger zones known as main-places, then into local municipalities, districts and provinces (The HDA, 2013). Having data in these particular spatial units creates issues when this data needs to be used in advanced spatial analysis procedures where different spatial units may be more appropriate. For example, the use of hexagons when conducting geographical accessibility analysis. The reason for using hexagons for such analysis is that the units can nest to create a continuous surface, while being fairly isotropic; that is having the least variation in the internal distances or direction from the centroid (Mokgalaka *et al.*, 2014). Advanced spatial analysis techniques need to be employed in order to overcome the above problems. Dasymetric mapping and areal interpolation are two such techniques.

This paper proposes the application of a combination of dasymetric mapping and areal interpolation techniques to overcome the issues of coarse scale population data. The proposed methodology provides one with the capability to transfer spatial data from arbitrary units to fit for purpose units seamlessly and accurately. By default this technique then also empowers the user to negotiate the age old foe, the modifiable areal unit problem (MAUP). The MAUP is “a problem arising from the imposition of artificial units of spatial reporting on continuous geographical phenomena resulting in the generation of artificial spatial patterns” (Heywood, 1988). Thus, there is a level of generalisation when you put an arbitrary boundary over a continuous dataset. How the boundary is defined will thus influence the subsequent results of any analysis.

## **2 Literature Review**

As mentioned in the introduction, population data in South Africa is collected by Statistics South Africa (StatsSA) and the 2011 Census results are disseminated to the public using their SuperWEB or SuperCROSS software. The introduction outlined the challenges created by the method of enumeration and the subsequent upward aggregation of the data according to administrative boundaries.

To ensure confidentiality, the 2011 census was not released at the EA-level, but was aggregated to what StatsSA calls the small area layer (SALs) and from there followed the upward aggregation to Sub Places (SPs), Main Places (MPs), Local Municipalities (LMs) and higher administrative units. The SAL is “a spatial layer that corresponds as much as possible to the EA layer, but within confidentiality limits” (Verhof and Grobbelaar, 2005 pp.1); the SP “was created by combining all EAs with a population of less than 500 with adjacent EAs within the same sub-place” (StatsSA, 2017). The 2011 data was also released per ward. It must be noted that the wards do not align with the SPs or MPs, but nest into the LMs and upward administrative boundaries. Wards are delimited to be roughly equal in size in terms of registered voters in each ward; their purpose is to assign voters to various geographic areas which periodically change as a result of population shifts (Kanyane, 2016).

All of these data zones mentioned above are created to be ‘fit for purpose’, that is to be used for logistical (wards and EAs), reporting (SPs and MPs) and administrative (LMs and higher order administrative units) purposes. Although appropriate with regards to what they are intended for, as they all adhere and nest into municipal and provincial boundaries which are created for administrative purposes, they however make settlement or neighbourhood level population analyses difficult. These polygonal units portray an inaccurate notion of homogenous population distribution and density that may lead to analytical and cartographic problems (Barrozo *et al.*, 2016 and Weichselbaum *et al.*, 2005) such as attributing population to uninhabited or uninhabitable areas. They also make it difficult to create customized analysis units for specific analyses that may need data at finer scales than what is available.

Figure 1 is a representation of this issue; part A showing the real world scenario in terms of the distribution of households in relation to natural features in an enumeration area; and part B of the figure illustrating how the population would be represented in an EA dataset. There are clear areas that are not inhabited such as those of the water body and forest areas, as well as a clear non-uniform distribution of the population settlement; however, the data is represented as a homogenous areal spatial unit. This makes it necessary to employ techniques that can better represent the real distribution of population within these zones.

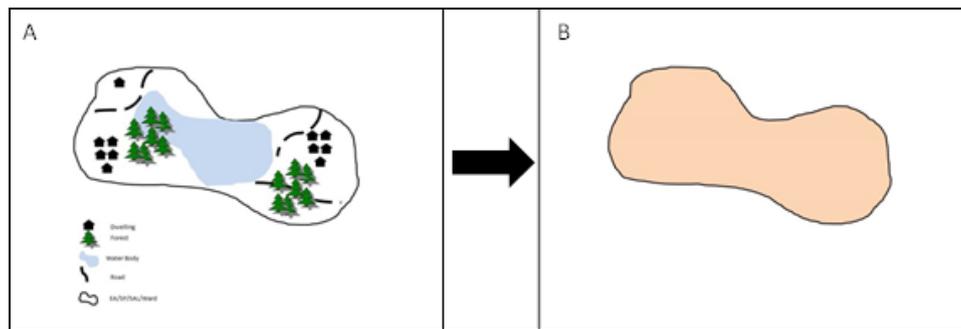


Figure 1: Graphical representation of real world vs population data representation in EA data

## 2.1 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. A dasymetric map involves transforming data from the arbitrary zones of the aggregate dataset to recover (or try to recover) and depict the underlying statistical surface. The aggregate dataset's demarcation is however not based on variation in the underlying statistical surface, but rather relates to the convenience of enumeration (Eicher and Brewer, 2001 and Mennis and Hultgren, 2005). This transformation process incorporates the use of an ancillary dataset that is separate from, but related to, the variation in the statistical surface (Eicher and Brewer, 2001). Dasymetric mapping has a close relationship with areal interpolation, which is a process of transforming data from a set of source zones to a set of target zones with different geometry (Bloom *et al.*, 1996 and Fisher and Langford, 1995 and Goodchild and Lam, 1980).

It is imperative for dasymetric mapping to have an appropriate ancillary dataset that depicts the underlying statistical surface of the area under investigation in order to disaggregate the data accurately. Datasets such as a building count (a points' dataset) would significantly enhance the accuracy, as (if classified) these building points can act as a proxy for human settlements. That is, building or household points in that dataset would not be found in areas that are inhabitable such as cliffs, roads, mines, waterbodies and the like, as well as offering the ability to exclude those points which are not residential or commercial.

Areal interpolation is 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 and Langford, 1995 and Hay *et al.*, 2005). In the examples taken from the literature, a dasymetric map is the result of intersecting polygon layers which predict where the actual concentration of variability would be within the data source layer (Eicher and Brewer, 2001).

The question is thus: how can we move from arbitrary zones to relevant analysis zones as accurately as possible using a combination of these two spatial analysis techniques?

### **3 Methodology and Data**

As mentioned earlier this paper proposes the use of dasymetric mapping and areal interpolation techniques to disaggregate and re-aggregate population data into desired analysis units in order to undertake several analyses relating to population distribution and overcome the MAUP in the process. It is proposed to move away from a polygon based dataset which does not represent the underlying statistical surface to a point dataset by using the Spot Building Count (SBC) data as an ancillary source.

The SBC was originally produced by the CSIR and Eskom in 2008 and is a geo-referenced building frame developed using SPOT satellite imagery. The dataset has subsequently been maintained and updated by ESKOM. The inventory concerned contains all classifiable building structures within the borders of South Africa (Breytenbach, 2010).

The thinking behind using the SBC as an ancillary dataset is that each point represents a potential household; however a generic 'household size' cannot be applied equally to all points in the country as these points are situated in areas with differing household sizes and not all the points represent inhabited structures. These points therefore needed to be classified in order to assign the necessary weight to each point representing the potential number of people who will be occupying that structure. The way in which this was done was by assigning the average household size, of the sub-place the point is situated in, to that point. A four-step hierarchical classification approach (a process of elimination) was followed in order to classify the points dataset in more detail in order to differentiate between points representing inhabited versus uninhabited buildings. This process involved:

- Identifying new growth areas, thus points representing buildings which are most likely residential due to the morphology and proximity to existing residential areas, but which were not present in the census year from which the household sizes were inherited
- As part of the new growth areas, differentiation between most likely informal dwelling structures versus formal dwelling structures also based on the morphology of the 'settlement' observable from the continuous form of a cluster of points
- Differentiation between points of structures being used for commercial or industrial purposes based on the underlying land use
- Identifying points in commercial farming and other sparsely populated areas where the use of buildings clustered together can vary significantly from one building to the next, e.g. a house and barn on a commercial farm.

For a more detailed discussion on the classification of the points please refer to Mans (2011), but it would suffice to say that the classification process was critical to ensure that points do not give a skewed representation of the actual distribution of population.

This makes it a novel approach from a dasymetric principle point of view. The argument is that the classified SBC-points are an accurate ancillary source for human activity and therefore for other socio-economic and population related activities. The inverse of the argument is that it is unlikely that socio-economic activity will be found where there is not any type of building present, whether formal or informal (Mans 2011).

The method will be presented through algorithms in a two-step process that has a number of sub sets of step-wise processes. The first process is the disaggregation of the data, and the second process is the re-aggregation of the data into new analysis zones.

### 3.1 Data

The datasets used to undertake the initial disaggregation process are:

- The original census tract data, using the sub-place as the unit in which the data is obtained in this instance (source: StatsSA Census 2011 interactive data in SuperCross and Demarcation Board)
- Updated Spot Building Count (original source: ESKOM)

### 3.2 Input

$m$  = number of original census tract data (e.g. sub-places)

$m'$  = number of newly tessellated zones (e.g. grid or mesozones)

$n$  = Number of the set of the ancillary dataset depicting the underlying statistical surface (e.g. SBC points)

In the following example:

- Let  $T = \{t_1, \dots, t_m\}$  be the set of population totals per the census tract data, (n sub-places)
- Let  $S = \{s_1, \dots, s_m\}$  be South Africa split into  $m$  e.g. sub-places
- Let  $P = \{p_1, \dots, p_n\}$  be the weight representing the potential household size of each point

In the first part of this process the following algorithms represent the process. Process 1– step 1 is expressed as:

$$\text{Let } A = \{S_i\}_{i=1}^m : \bigcup_{i=1}^m S_i = P \text{ be a partition of } P \text{ into } m \text{ sub-places} \quad [1]$$

Let  $A$  be the set consisting of elements  $S_1, \dots, S_2, \dots, S_i, \dots, S_n$  (represented by  $\{S_i\}_{i=1}^m$ ) given that the union for all  $S_i$  (represented by  $\bigcup_{i=1}^m S_i$ ) is  $P$  which is the total number of points. This simply represents a spatial join, where each point then gets the unique identifier for the sub-place that it falls within.

Process 1– step 2 is expressed as:

[2]

$$W = \{w_i : w_i = \sum_{j=1}^m p_j \forall p_j \in S_i\}^m$$

All the weights ( $p_j$ ) per sub-place ( $S_i$ ) are summed to give the sum of weights ( $w_i$ ) being an element of  $W$ . This represents the sum of weights of the points belonging to each sub-place.

Process 1– step 3 is expressed as;

$$A' = \{S'_i\}_{i=1}^m \text{ with } S'_i = \{s'_j : s'_j = s_j/w_i \forall s_j \in S_i\} \quad [3]$$

Let  $A'$  be the set consisting of elements  $S'_i$  with  $S'_i$  being the set consisting of elements  $s'_j$  given that  $s'_j$  is equal to  $s_j/w_i$  for all  $s_j$  in each sub-place ( $S_i$ ) as calculated in step 5. This equation represents the division of the weight of each point belonging to that sub-place by the sum of weights in that sub-place.

Process 1– step 4 is expressed as;

$$\hat{P} = \{\hat{p}_i : \hat{p}_i = s'_i \times t_j \forall s'_i \in S'_j\}_{j=1}^m \quad [4]$$

Where  $\hat{P}$  is the set consisting of elements  $\hat{p}_i$  given that  $\hat{p}_i$  is the multiplication / product of  $s'_i$  ( $s'_i = s_j/w_i \forall s_j \in S_i$ ) and total population of the sub-place ( $t_j$ ) for all  $s'_i$  in  $S'_j$ . In essence this process then multiplies the proportional contribution of each point with the population of the sub-place that the point belongs to; therefore, redistributing the population per sub-place proportionally to the points inside the sub-place based on the relative weight of the point, producing the population per point.

This first process then produces a set of points with the potential household size of each point. The next process then transfers the point's data to a new demarcation. For that reason Process 2– step 1 is expressed as;

$$\text{Let } \Theta = \{M_i\}_{i=1}^{m'} : \bigcup_{i=1}^{m'} M_i = \hat{P} \text{ be a partition of } \hat{P} \text{ into } m' \text{ new analysis unit} \quad [5]$$

$\Theta$  is the set of elements consisting of the new tessellations / desired analysis units (e.g. grid or mesozone tessellation of South Africa represented by  $M_i$ ) given that the union of all new tessellations ( $\bigcup_{i=1}^{m'} M_i$ ) is the set of the total number of weighted population points ( $\hat{P}$ ). Hence linking or spatially joining the points to the new tessellation or analysis unit and thereby transferring the unique identifier of the new analysis unit it falls inside of.

And finally Process 2– step 2 is expressed as;

$$\hat{W} = \{\hat{w}_i : \hat{w}_i = \sum_{j=1}^{m'} \hat{p}_j \forall \hat{p}_j \in M_i\}_{i=1}^{m'} \quad [6]$$

Thus, summing the population per point for the new polygon a group of points belong to. Where,  $\hat{W}$  represents the set of elements  $\hat{w}_i$  given that  $\hat{w}_i$  is the sum for all of the weighted population per point ( $\hat{p}_j$ ) in  $M_i$  in the new tessellation (e.g. the grid or mesozones).

Undertaking this method allows for the distribution of census population data to be better represented within the data zones, this will be presented in the next section.

#### 4 Results

In short, the process discussed above produces a points' dataset with a potential population size assigned to each point. These points can now be linked to any demarcation seamlessly, whilst representing the underlying statistical service accurately, and then summed per unit for the new demarcation.

Using the set of points that has been produced through the above method (where the original population data that was used was originally on a sub-place level), the disaggregation result can be compared to the population in a ward. The reason for using a ward is that wards are populated with the same census year's population data, but the extent of a ward is completely different to that of a SP, thus, SPs do not nest into wards. If the SP data is used as the main data source, and the SP data is disaggregate it to the points; and sum of the points' based population compares to the original population StatsSA assigned to the ward based on the original source, it illustrates that the ancillary dataset is an accurate representation of the underlying statistical surface.

The example used to illustrate this output is of ward 7 (2011 ward boundaries) of the Ratlou Local Municipality in the North West province, an area known as the Madibogo Pan. This ward had a population of 8 014 (StatsSA Census SuperCross data) in 2011 and is 90km<sup>2</sup> in area. The dominant land cover in this ward is low to medium subsistence cultivation, shrub land and settled villages (Geoterra Image, 2015).

Figure 2A shows Ward 7 and the land cover (LC). It is clear that in this ward there are two distinct villages with the dominant one to the south east and a smaller one to the north west of the ward. Figure 2B is the land cover with the SBC points overlaid on it. It shows that points correspond strongly with the urban village class. In a regular ward level choropleth map showing ward population in this region, this intra ward variation of population distinction would not be visible, instead population would be represented using a mono-colour polygon.

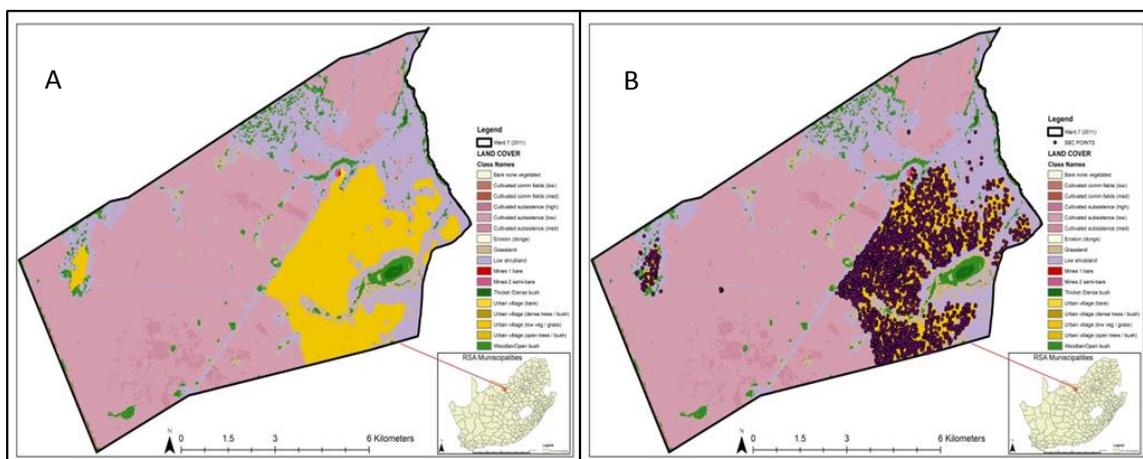


Figure 2: Land Cover of Madibogo Pan

After undertaking the method described in the previous section, a 30mx30m (900m<sup>2</sup>, at the same resolution as the LC layer) grid was created for the ward (see figure 3 below).

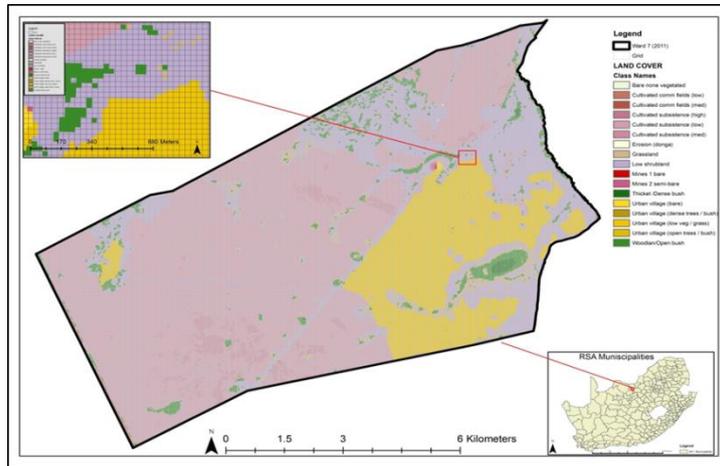


Figure 3: Illustration of Madibogo Pan with LC and grid created overlaid

The SBC points that now have the population attributed to it were joined to the grid and the population per point was summed per grid cell producing a grid that had population transferred from the points dataset. Only grid cells that cover areas with population are populated whilst those that do not have any SBC building points (such as those that are classified as not being in one of the urban village categories) are not. This procedure produced a grid showing the population distribution. Figure 4 graphically demonstrates the result of this process.

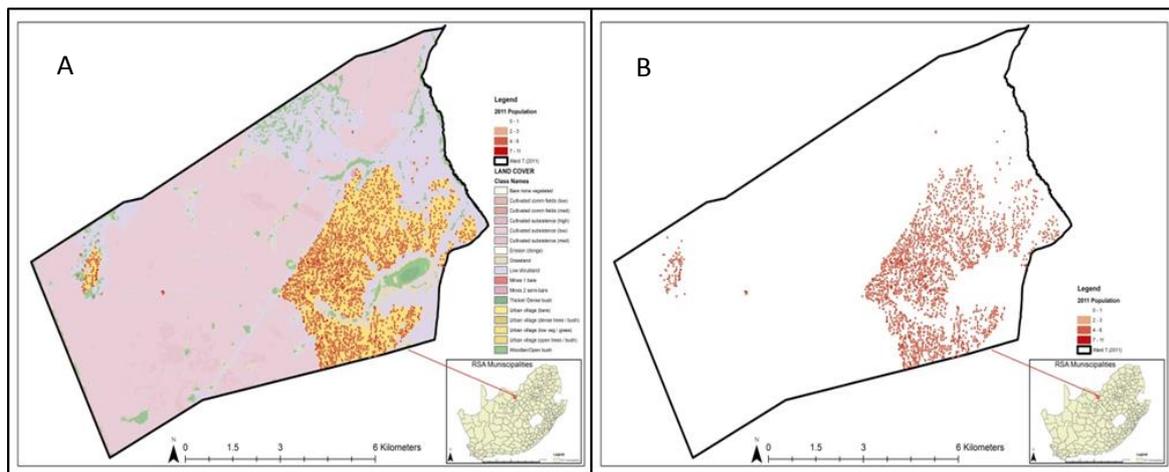


Figure 4: Population Distribution in Madibogo Pan

Figure 4A shows the population distribution as per the grid overlaid on the LC set to 50% transparency to emphasise the population distribution in the grid. Figure 4B shows the population grid without the LC layer in order to clearly indicate the areas that have population and those that do not, this shows the internal variation and distribution of population within the ward.

After undertaking the proposed method; the total population in the ward was calculated to be 8 019; a discrepancy of 5 people or 0.05% between the original ward population as per StatsSA and the redistributed population which originated from the sub-place dataset. It important to note here that (as mentioned earlier in the paper) the boundaries for wards and sub-places do not align and that sub-places do not nest into ward boundaries.

This process can be replicated at a larger scale or used to create custom, fit for purpose tessellations with a relatively accurate population distribution, better depicting the settlement in an area. By being able to disaggregate and aggregate the population in this fashion it also overcomes the problems associated with the MAUP by default.

## **5 Limitations**

For an accurate result, this method can only be undertaken when data is available on scales below municipal level. As community surveys (such as the 2016 community survey) do not offer data at scales such as sub place level, a process such as this can only be undertaken following the dissemination of data from the census which is conducted every 5 to 10 years in South Africa.

## **6 Application of Results**

This methodology has been applied over a long period to a range of research projects analysing data ranging from local to regional and national scales. It has been fundamentally used to develop customised and project specific analysis units that best suit the research question.

For example at a local level it has been used in a range of analyses to test for geographic access levels. As part of a project for the Department of Public Service and Administration it was used to test for geographic access levels to government service points in the municipalities of eThekweni and Johannesburg to evaluate how best to address the provision of social services facilities within the metropolitan areas and identified facility backlogs where applicable (Green et al., 2012). In the project this method was utilised to create a customised population based service demand profile in the form of a hexagonal grid from which population distribution and access to facilities was calculated. The results of the project produced multi-departmental integrated facility plans which covered facilities from all three tiers of government in each of the metropolitan areas and thereby supporting the goals of more equitable and affordable access to a range of services in all parts of the selected cities and the clustering of facilities (Green et al., 2012).

At a national level, this method has recently been used to generate population profiles for a customised national set of catchment areas as part of a research project for the Department of Rural Development and Land Reform. This projects aim was to create differentiated norms and standards for social facility provision in rural areas. The approach taken was to create and then profile a set of functional service catchment areas. These profiles could then be linked to a defined minimum basket of services for each level of catchment for a full spectrum of services.

As these catchments did not align to any administrative boundaries apart from the provincial boundaries; the process of disaggregating the population using dasyametric principals referred to earlier had to be employed to then allocate population to the catchment areas (DRDLR, 2016).

The principal demonstrates how the proposed method in this paper was used to create a customised analysis unit for particular analysis at a national level (DRDLR, 2016).

## **7 Conclusions**

Improvements in GIS and spatial analysis methods have permitted increased sophistication and accuracy of geospatial studies that may have enormous impacts on how spatial information and data can be studied. However; the ability of GIS to do this is fundamentally dependent on the available data for analysis. The increased spatial resolution of datasets, in this case population data, can lead to producing better plans to inform policies and decisions that impact the lives of the people in South Africa.

The problem presented by having coarse population data (collected at different and coarse resolutions) is that it impedes the ability to conduct settlement level analyses. Knowing the internal variations of the distribution of population is an important input to plans. These plans could be very local or require analyses that transcend administrative boundaries or predefined reporting zones; overcome issues presented by frequent boundary shifts; as well as those that offer the ability to create appropriate and/or standardised analyses units that offer a view of subtle changes in population differences.

This paper has presented the application of a hybrid technique of dasymetric mapping and areal interpolation in the South African context and shown how it can be applied in order to represent this data at finer scales than currently readily available. This technique can be applied to data for a wide range of research applications to offer finer or refined scaled population data that can be applied at different scales thereby overcoming the problems that occur when data is not available at suitable resolution.

The next step in this research is to apply this technique to different temporal datasets that will allow for temporal spatial data alignment in order to analyse trends at fine scales and then using those outputs to inform policy at local, regional or national scales.

## **8 References**

- Barrozo LV, Pérez-Machado RP, Small C & Cabral-Miranda W 2016, 'Changing spatial perception: dasymetric mapping to improve analysis of health outcomes in a megacity', *Journal of Maps*, vol. 12, no.5. pp. 1242-1247.
- Breytenbach, A 2010, 'A national inventory using earth observation and GIS', *Science Scope*, vol. 5, no. 1. pp. 66-68,
- Bloom LM, Pedler PJ & Wragg GE 1996, 'Implementation of enhanced areal interpolation using MapInfo', *Computers and Geosciences*, vol. 22, pp. 459-466.

- Department of rural development Land Reform (DRDLR) 2016, 'The Social Facility Provision Toolkit', Department of rural development Land Reform, Pretoria, 15 May 2017, <https://www.socialfacilityprovisiontoolkit.co.za/#/>
- Green CA, Mans G, Le Roux A, Schmitz P, Mokgalaka M, McKelly D, Ngidi M, Badenhorst W, Zietsman HL 2012, 'Geographic accessibility study of social facility and government service points for the metropolitan cities of Johannesburg and eThekweni 2011/12'. Report No. CSIR/BE/SPS/ER/2012/0061/B. ISBN 978-0-7988-5608-9. CSIR, Pretoria.
- Fisher PF, & Langford M 1995, 'Modelling errors in areal interpolation between zonal systems by Monte Carlo simulation', *Environment and Planning A*, vol. 27, pp. 211-224.
- Geoterra Image 2015, '2013-2014 South African National Land-Cover Dataset', Geoterra Image, South Africa.
- Goodchild MF & Lam NS 1980, 'Areal interpolation: A variant of the traditional spatial problem', *Geoprocessing*, vol. 1, pp. 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*, vol. 10, pp. 1073-1086.
- Heywood DI, Cornelius S & Carver S 1998, 'An introduction to geographical information systems', Addison Wesley Longman, New York.
- Kanyane M 2016, 'Ward delimitation: impact on municipal service delivery in South Africa', Human Sciences Research Council, Pretoria, 2 May 2017, [http://www.demarcation.org.za/site/wp-content/uploads/2016/07/WARD-DELIMITATION\\_-IMPACT-ON-MUNICIPAL-SERVICE-DELIVERY\\_BARWA-KANYANE-HSRC.pdf](http://www.demarcation.org.za/site/wp-content/uploads/2016/07/WARD-DELIMITATION_-IMPACT-ON-MUNICIPAL-SERVICE-DELIVERY_BARWA-KANYANE-HSRC.pdf)
- Mans G, Green C, Ngidi M, Sogoni Z & McKelly, D 2016, "Development and prioritisation of catchments", Technical report commissioned by the Department of Rural Development and Land Reform, Pretoria, viewed 15 May 2017 [https://www.socialfacilityprovisiontoolkit.co.za/attachments/Development\\_and\\_prioritisation\\_of\\_catchments\\_technical\\_report\\_web.pdf?etag=true](https://www.socialfacilityprovisiontoolkit.co.za/attachments/Development_and_prioritisation_of_catchments_technical_report_web.pdf?etag=true)
- Mans G 2011, 'Developing a geo-data frame using dasymetric mapping principles to facilitate data integration' in *AfriGEO Conference: Developing Geomatics for Africa*, Cape Town, May-June 2011.
- Mennis J & Hultgren T 2006, 'Intelligent dasymetric mapping and its comparison to other areal interpolation techniques', *Cartography and Geographic Information Science*, Vol. 33, No. 3, pp. 179-194.
- Mennis J and Hultgren (2005). Dasymetric mapping for disaggregating coarse resolution population data, *proceedings of the 22nd International Cartographic Conference*, July 9-16, A Coruna, Spain.
- Mokgalaka H, Mans G, Smit J, & McKelly D 2014, 'Validating the accuracy of GIS based accessibility analysis in determining public primary health care demand in metropolitan areas' in J Whittal and S Motala, *Second AfricaGEO Conference*, Cape Town, 2014 viewed 5 May 2017, [http://www.africageoceedings.org.za/?page\\_id=403](http://www.africageoceedings.org.za/?page_id=403)
- Mokhele T, Mutanga O, Ahmed F 2016, 'Development of census output areas with AZTool in South Africa', *South African Journal of Science*, vol. 112 no.7/8, viewed 18 May 2017, [http://www.sajs.co.za/sites/default/files/publications/pdf/SAJS%20112\\_7-8\\_Mokhele\\_Research%20Article.pdf](http://www.sajs.co.za/sites/default/files/publications/pdf/SAJS%20112_7-8_Mokhele_Research%20Article.pdf).
- Páez A & Scott DM 2004, 'Spatial statistics for urban analysis: A review of techniques with examples' *GeoJournal*, Vol. 61, pp. 53-67.
- Statistics South Africa (StatsSA) 2012, 'Census 2011 Statistical Release (Revised) P0301.4', Statistics South Africa, Pretoria, South Africa, viewed 11 May 2017, <https://www.statssa.gov.za/publications/P03014/P030142011.pdf>
- Statistics South Africa (StatsSA) 2017, 'Small area statistics', Statistics South Africa, Pretoria, South Africa, viewed, [http://www.statssa.gov.za/?page\\_id=4086](http://www.statssa.gov.za/?page_id=4086).

- The Housing Development Agency (HDA) 2013, 'Western Cape: Informal settlements Status 2013', Research Series Published by The Housing Development Agency, Johannesburg, South Africa, viewed 18 May 2017, [https://www.thehda.co.za/uploads/files/HDA\\_Western\\_Cape\\_Report.pdf](https://www.thehda.co.za/uploads/files/HDA_Western_Cape_Report.pdf)
- Weichselbaum J, Petrini-Monteferri F, Papathoma M, Wagner W and Hackner N 2005, 'Sharpening census information in GIS to meet real-world conditions: the case for Earth Observation' *Sustainable Development and Planning*, vol. 2, no. 1 pp. 143 – 152.
- Yao Y, Liu X, Li X, Zhang J, Liang Z, Mai K & Zhang Y 2017, 'Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data', *International Journal of Geographical Information Science*, vol. 31, no. 6, pp. 1220-1244.