Developing a geo-data frame using dasymetric mapping principles to facilitate data integration

Gerbrand Mans[1]

[1]Planning Support Systems, Built Environment, CSIR, South Africa, gmans@csir.co.za

Abstract

Integrating socio-economic and ecological or environmental data spatially, for the purpose of planning processes, allows for more effective and targeted decision making. The integration of data from different sources is however an issue which is not easy to resolve. What makes the integration of data difficult is the widely differing scale and nature of analysis units. They are commonly dictated by academic discipline, nature of the study, survey methodology, administrative / service considerations or physiographic characteristics.

This paper discusses the development of a disaggregation procedure for socio-economic data based on the principles of dasymetric mapping and areal interpolation in order to develop a flexible geo-data frame which allows the data to be assigned to different demarcations seamlessly. The geo-data frame is based on the spot building count points-dataset from ESKOM. The reason for selecting this point dataset is that it represents all formal and informal built up areas in South Africa. It is therefore a good stand-in indicator of human or socio-economic activity.

The geo-data frame was used to determine the total population per census sub-places. The accuracy was tested against the recent population updates of eThekwini and City of Cape Town by calculating the intraclass correlation coefficient (ICC). The cities' data was used as control totals. The ICC for both cities shows a significant correlation (very strong) between the cities' vs. the geo-data frame's outputs.

1. Introduction

1.1 The need for data integration

The need to work with data in an integrated fashion is a topic that has enjoyed much attention in the past two decades, especially with the rise in popularity in the use of Geographical information systems (GIS). Walker and Young (1997) made some of the earliest and most compelling arguments regarding this saying that:

> '*Area-based analysis of social and environmental factors taken together should aid understanding of the range and variety of needs for policy intervention. It would enable a more customized approach to be taken to resource allocation for environmental protection, social development and well-being, and the maintenance or improvement of economies*… in the end: *ensuring that economists and ecologists are more aware of the spatial and environmental consequences of their recommendations*'.

People with possibly the greatest need to work with data in an integrated fashion are policy and decision makers within government. Also, integration of data is necessary on primarily the social, economic and ecological level in order to ensure sustainability (Costanza 1989; Costanza et al. 1996). An area-based analysis of social and environmental factors taken together, aids in understanding the range of needs on which to base policy intervention. It enables a more customized approach to be taken to resource allocation for environmental protection, social development and well-being, and the continuance or improvement of economies.

Despite many calls for the integration of socio-economic and ecological data for decision making, there has been little real progress. In the South African government context economic and ecological advice arrives from different departments or sections and is then integrated subjectively, for example in the integrated development plans (IDPs)[1] of municipalities, and thus does not pertinently address the question of data integration.

The first solutions pertaining to data integration centered on the use of geographic information systems (GIS). This was however in the early days, but soon people realized that using GIS brought along many issues of its own. The first of which was that the construction of geographical information system databases and the licensing of proprietary software are expensive and are criticized by some

---

[1] An IDP is a participatory approach to integrate economic, sectoral, spatial, social, institutional, environmental and fiscal strategies in order to support the optimal allocation of scarce resources between sectors and geographical areas and across the population in a manner that provides sustainable growth, equity and the empowerment of the poor and the marginalised. An IDP is therefore a plan that guides the activities and decisions of a municipality for the next five years in terms of Chapter 5 of the Municipal Structures Act, 2000.

government administrators responsible for strategic policy development as investments unable to provide adequate returns (Daly 1990; Young 1992). Recently this has started to change as the initial development of databases is usually the most expensive aspect and the databases are now established and cheaper to maintain. In terms of the software, the market for GIS has become competitive, extensive and settled enough such that licensing and maintenance is not such an issue any more and there are also freeware and open source GIS-packages available.

1.2 Obstacles to data integration: boundaries and scale

The biggest issue with respect to the integration of socio-economic and environmental data arises from the fact that often one has datasets with differing scales and differing demarcations. These differences are commonly dictated by academic discipline, nature of the study, survey methodology, administrative / service considerations or physiographic characteristics (Huby et al. 2007).

Typically, environmental and socio-economic studies adopt different types of spatial areas as their basic units. Data structures usually reflect the specific interests of data collectors and reporters (Sang et al. 2005) and different units are selected and tailored to meet specific needs, often for organizational as well as for technical reasons. Socio-economic studies, for example, tend to be based on administrative units such as local authority districts, wards or primary care trusts (Martin 1996). The boundaries of such units are abstract creations with no particular relation to physical realities on the ground and may be subject to periodic change (Martin & Bracken 1993). In contrast, environmental data is mainly collected on the basis of units related to either physical attributes of the environment such as watersheds, or regular sampling grids related to the technology or methodology used to acquire or store the data.

In attempts to overcome the problems related to scale and demarcation, basic spatial units have been created to which the attribute data of the different layers is then assigned. An example of a study where basic spatial units (BSUs) were used is the SECRA-study by Huby et al. (2007). In this study they were concerned with the integration of spatial data sets designed to characterize rural England in terms of what is there, what it is like, the living and working conditions, and the political and economic context from a sustainable development point of view. The CSIR developed the mesozones demarcation for a similar purpose in mind, i.e. integration of socio-economic and environmental data (Naudé et al. 2007).

There are a host of considerations when defining a BSU. The problem one ends up with is that the demarcation only suits the purpose of the specific project at hand and therefore it is not generic enough, i.e. just another arbitrary demarcation. The other issue is - with the fact that socioeconomic studies' boundaries are dictated by administrative demarcations (e.g. South African census demarcations) - that when a BSU takes these administrative boundaries into consideration it become superfluous once the administrative demarcation changes, which is the case in point for the CSIR mesozones (BSUs). Especially in a South African context this is a major issue with local municipality boundaries changing frequently.

A number of researchers have inverted the problem of the BSU identification and rather advocate the development of base units derived from the distribution of the underlying dataset (Eagleson et al. 2003). Openshaw and Rao (1995) argue that census users should abandon official zones and re-engineer zoning systems dependent on the data and relationship they are investigating. Whilst this would result in the optimal solution with regard to data representation and integration, from a more pragmatic policy-making point of view it is not the optimal solution either. This approach assumes full competency in working with GIS-based data in different formats (vector and grid based) and integration of the data in a platform suitable for the question at hand. A person tasked with decision making based on integrated datasets will not have the time and, most probably, not the skill to do this.

2. Methodology

2.1 Purpose of the study

What we wish to address in this paper is the need to integrate data in an appropriate way in order to allow for two things: firstly, integrated decision making on socio-economic and ecological or environmental data; and, secondly, to ensure that the data is portrayed in such a way that it overcomes the MAUP effect. This will be done by developing a flexible socio-economic frame using the sport building count (SBC) dataset (point dataset) and assigning socio-economic characteristics to each point using the principles of dasymetric mapping. The SBC was produced by the CSIR and ESKOM in 2008 and is a geo-referenced building frame developed using Spot 5 satellite imagery. The inventory concerned contains all classifiable building structures within the borders of South Africa (Breytenbach 2010). The accuracy of this process will be tested and discussed herein.

The reasons why we suggest putting the socio-economic data set in a flexible format are the following:

1. Environmental data is dictated by physiographic boundaries (for example plant biomes, river catchments, mountainous areas, etc.) which do not adhere to any administrative or human activity related boundaries;
2. Administrative boundaries are dictated by many influences and subject to constant change. It is neither practical nor implementable to adjust these boundaries according to environmental based demarcations, or to try and make it fit socio-economic trends accurately.

The SBC-points dataset is therefore considered as the static spatial units to which one will assign the socio-economic data.

2.2 Dasymetric mapping

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). Thus, the process of 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. 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 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).

In the examples taken from the literature, a dasymetric map is the result of intersecting polygon layers which predicts where the actual concentration of variability would be within the data source layer (Eicher & Brewer 2001). We propose to move away from a polygon based dataset which represents the underlying statistical surface to using a point dataset by using the spot building count (SBC) data as an ancillary source. This makes it a novel approach from a dasymetric principle point of view. The argument is that the SBC-points are an accurate ancillary source for human activity and therefore for all socio-economic related activities. The inverse of the argument is that one would not find any socio-economic activity where there is not any type of building present, whether formal or informal.

3. Establishing a flexible geo-data frame

In order to test the viability of using the SBC data set as a flexible frame to which socio-economic characteristics can be assigned, we only focused on classifying the dataset in order to do an accurate population distribution classification. The SBC is only a building count and no other information is available regarding the size or type of structure it represents. It was therefore necessary to develop a method by which the points (buildings) could be classified.

To use the SBC as a proxy layer for population distribution, a weight had to be assigned to each point indicating the relative contribution of that point to the total population. In other words the weight of the point would represent the probable household size of the building (household) in question. It was, however, not possible to do a regression analysis[2] of the 2001 StatsSA census data and the SBC data regarding the ratio between SBC points and the number of households in, for instance, a census sub-place as the SBC was only done from 2006 to 2008. If the SBC data was available for 2001, the growth from 2001 to 2006 could be used to predict what the growth or expansion of a certain area was as well as to predict new growth areas beyond the boundaries of more densely populated areas. The characteristics of the new growth area could then be related to the 2001 area of which it is a natural extension of and then regression analysis used to determine population growth. The incompatibility of the datasets arising from the different time lines of the datasets, however, rendered this impossible.

A five-step hierarchical approach – incorporating a process of elimination – was followed in order to get a more detailed distribution of the current population figures for South Africa. This process is based on the principles of dasymetric mapping which is reliant on underlying ancillary data sources to determine the characteristics of the SBC-points. The process was as follows:

1) Assign weights (potential household size) to the SBC-points

The average household size per Census small area layer (SAL) was calculated (population divided by households) and this figure was then assigned the SBC-points falling inside each SAL. The assumption underlying this process is that each SBC is a potential household, but we know that this is not always the case. Hence, a hierarchical exclusion process is followed to get a more accurate representation (Steps 2 to 5).

2) Identify new growth areas

The household sizes determined in Step 1 are dependent on the 2001 census data. From 2001 to 2008 there has been development in urban areas which extends into the SALs which have previously been classified as rural. These new growth areas will therefore have similar characteristics to that of the urban areas that they extend from and the SBC-points here must therefore inherit the characteristics of these urban areas. The following process was followed to identify these areas:

- Convert SBC-points to a 100 m by 100 m grid;
- Pass a 3x3 focal majority filter over the grid;
- Convert the grid to a polygon layer;
- Link new growth areas with the 2001 urban SALs of which they are natural extensions by undertaking to:
    - Select only SALs classified as urban
    - Amalgamate the urban SALs with the polygons from the majority filter process
    - Select the polygons which were part of a polygon that intersects with an urban SAL – these are the potential growth areas;
- Assign the household size of the closest neighbouring urban SAL to the SBC-points situated

[2] Regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed (Wikipedia 2010).

in the new growth areas.

3) Identify the growth of informal areas

In the SBC-dataset there are polygons which indicate the boundaries of informal areas for which it was not possible to identify individual houses in high density informal settlements. There is thus no indication of the number of dwellings, and hence households, situated in these areas. The approximate number of households was determined and the representative amount of points added to the existing SBC-points in order for the points to more accurately represent population distribution by the following process:

- Amalgamate informal area polygons with SALs;
- Select the most dense SAL (2001) – based on density of households per square metre – overlapping the informal polygon (2008);
- Assign the same density to the rest of the informal area polygon;
- Based on this density, calculate the approximate amount of households;
- Create random points equal to the amount of households within the polygon;
- Use the household size of the densest SAL overlapping the informal area as the average household size of the informal area.

The underlying assumption of the above mentioned process is that the density and household size of an informal area are the same as those of the densest SAL (2001) it overlaps.

4) Identify SBC-points situated in commercial and industrial areas

Census SALs are determined by the demarcation of enumerator areas (EAs). Enumerator areas are functional demarcations to ensure that a representative sample of the population is taken during the census survey; therefore EAs also include land-uses other than just residential land. In Step One the assumption was made that each SBC-point is a residential house and therefore represents a household but it may actually be a different type of building. Hence, all commercial and industrial areas had to be identified as these are usually high density built-up areas and the ratio between buildings and households may differ significantly for these two land-use types. Points which are part of the informal areas and which are situated in industrial or commercial areas were excluded from the process as the informal areas are structures used for housing and therefore the weights calculated in the previous process are appropriate for these cases. Step 4 thus entailed the following:

- Identify industrial and commercial areas from the 2001 land-use dataset;
- Calculate the potential 2001 population of these areas using the area weighting method;
- Calculate the ratio between the 2001 total population and the amount of SBC-points to obtain the relative weight of the point as a contributor to the total population.

This process therefore assumes that the same ratio between the amount of buildings and amount of people will apply as was the case in 2001.

5) Calculate weights for SBC-points situated in rural areas

Unlike urban residential areas where the amount of buildings strongly correlates with the amount of households in the area, rural areas have a lot of buildings which may either be uninhabited or used for a different purpose, for example storage facilities on farms. All SBC-points situated in SALs classified as rural were selected while SBC-points which had already been part of any of the above mentioned processes were omitted. The procedure followed was to:

- Identify SBC-points situated in 2001 census SALs classified as rural;
- Exclude SBC-points which have been part of any of the previous processes (Steps 1 to 4);
- Calculate the ratio between total population and SBC for each SAL in order to assign a weight to the point for the dasymetric process.

Assumption: The ratio between the amount of buildings and the total population stayed the same from 2001 to 2008 (population divided by SBC). Exceptions: All points which were part of Steps 2 to 4.

4. Results and accuracy

In order to determine the accuracy of the analysis the results must be tested against a reputable source which gives a detailed distribution of population figures for the same time period as that covered by the analysis. The City of Cape Town and eThekwini Municipality both did recent updates of their population figures down to a very detailed level. eThekwini identified all structures used as households (formal and informal) in their municipality using aerial images and combined with household surveys derived the household size for every housing structure in the municipality (Aiello 2011). The City of Cape Town identified all new developments (including infill developments) based on information from their planning department, digitized all informal housing developments and identified backyard dwellings through surveys. This information was used to calculate the current population totals per census sub-place (Sinclair-Smith 2010). The accuracy of these processes can be debated, but for the purpose of the paper the results from the cities were accepted as the controls against which the accuracy of the SBC-data frame was tested.

To measure the accuracy of the SBC-data frame it was decided to assign the population figures to the census sub-place demarcation (StatsSA 2001) as the City of Cape Town did its population update based on the sub-place demarcations. The eThekwini dataset is also a point-dataset, like the SBC-data frame. The difference is, however, that each point in this dataset is definitely a housing structure, i.e. there are people living in the structure represented by a point. The point locations were assigned to the sub-place in which they are situated and the total population per sub-place calculated accordingly. The **final dataset** produced was therefore of total population based on the two cities' population counts versus the population counts produced through the study's methodology per census sub-place.

A correlation analysis was run of the estimates derived using the SBC-data frame and of the results received from the City of Cape Town and eThekwini Municipality. The cities' derived population totals per sub-place were used as the control variables.

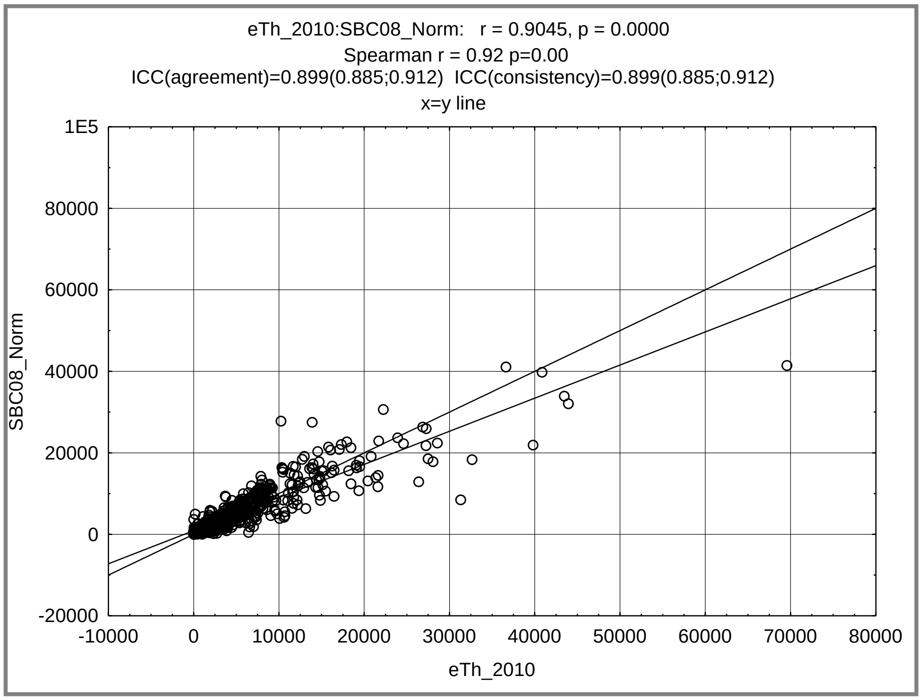

Figure 1: Correlation tests for eThekwini Municipality

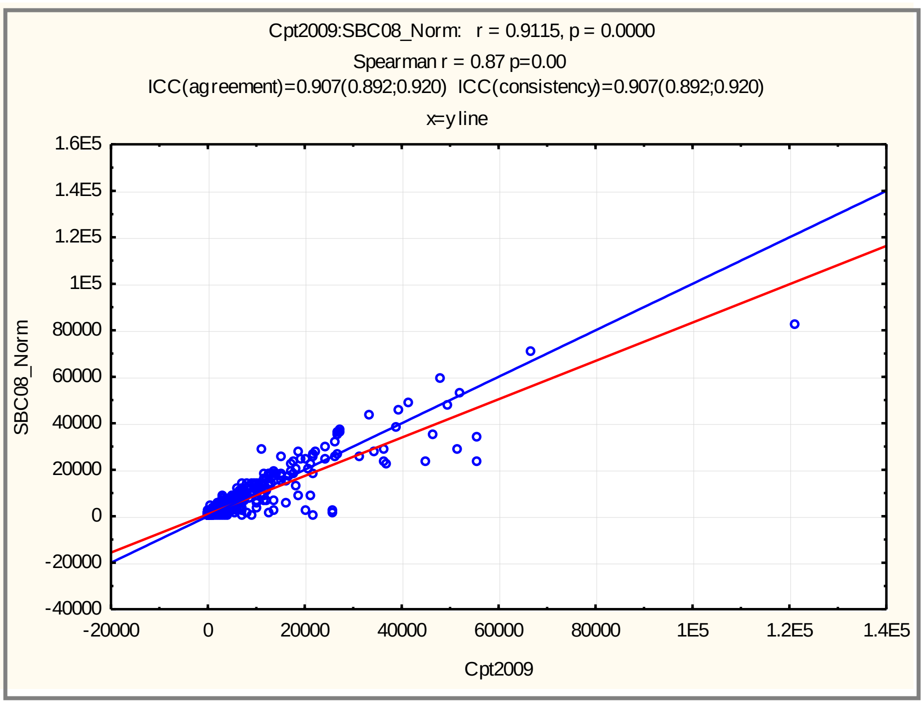

Figure 2: Correlation tests for the City of Cape Town

There is a strong positive Pearson's (r =0.9045) correlation, which is significant for the sample size, between the results per sub-place on city wide level for eThekwini Municipality (see Figure 1). The same significant correlation results were obtained for the City of Cape Town, see Figure 2, with Pearson's being 0.9115. The correlation for the Cape Town output is slightly weaker than that for eThekwini.

The problem with measuring correlation is that it only shows that the Y-values follow the same trend of the X-values, i.e. if dataset Y is 1;200;50;10 it will still be strongly correlated with X if its values are 2;400;100;20. In order for the methodology by which the SBC-data frame was developed to be reliable, the allocations must not only correlate but also correspond or agree internally. In other words the values calculated using the SBC-data frame must be as close as possible to the control value (the population totals as determined by the city), despite following the same trend in general.

The intraclass correlation coefficient (ICC) can be used to measure the consistency or reproducibility of quantitative measure made by two different observers, or in the instance of this study where the same aspect (total population) was determined using two different methods. 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.

Figures 1 and 2 show that there is a significant intraclass correlation between the measures of the cities and that of the method used to prepare the SBC-data frame. The closer the ICC value is to 1 the stronger the correlation. For the City of Cape Town the ICC is 0.907 and for eThekwini it is 0.899. 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). The blue line on the figures indicates a perfect correlation and the red line the deviation based on the SBC-data frame dataset. It is only for the areas with an exceptional high population figure that SBC-data frame underestimates the total population, but only slightly, for both the study areas.

The City of Cape Town and eThekwini Municipality have identified areas they deem to be "rural" due to lower population densities. eThekwini does however have traditional tribal areas which are not formal built up areas, but still have high population densities. These areas are called dense rural areas. The correlation analysis was also done for the **rural areas** of the City of Cape Town and eThekwini and lastly the **dense rural areas** of the latter. Figure 3 shows rural-urban classifications of eThekwini and City of Cape Town.

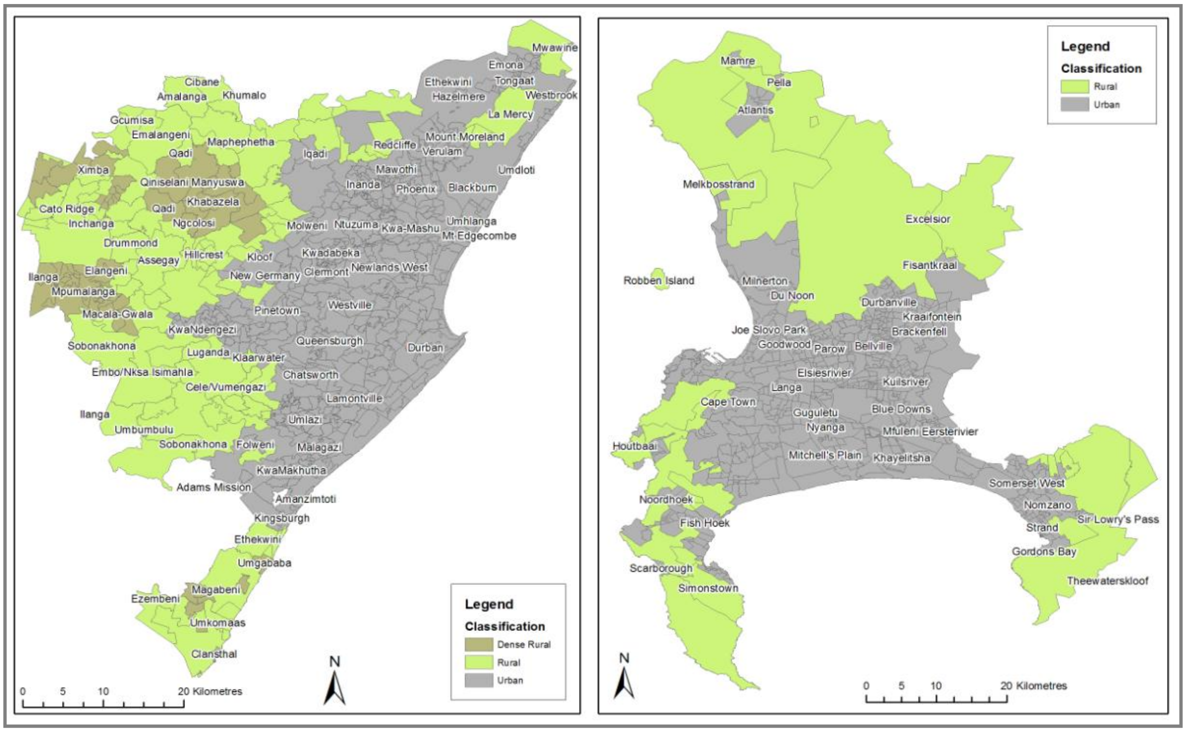

Figure 3: Rural-urban classification of eThekwini and City of Cape Town

The reason for this is that in these areas there are sharper changes in population distribution over time due to: (1) informal areas being established on the fringes of the city; and, (2) city growth expanding into these areas over time as space becomes restricted in the formal build up areas. It is important that the SBC-method also picks up these changes which may be more subtle as the relative impact on total population is not that big.

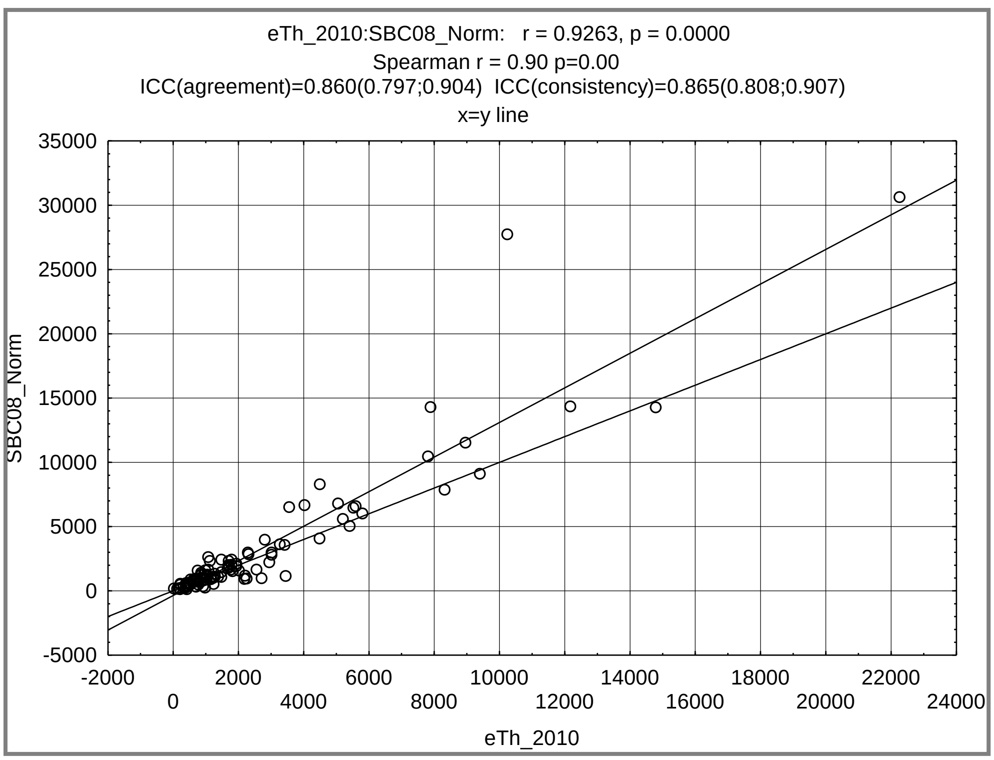

Figure 4: Correlation tests for eThekwini Municipality rural areas

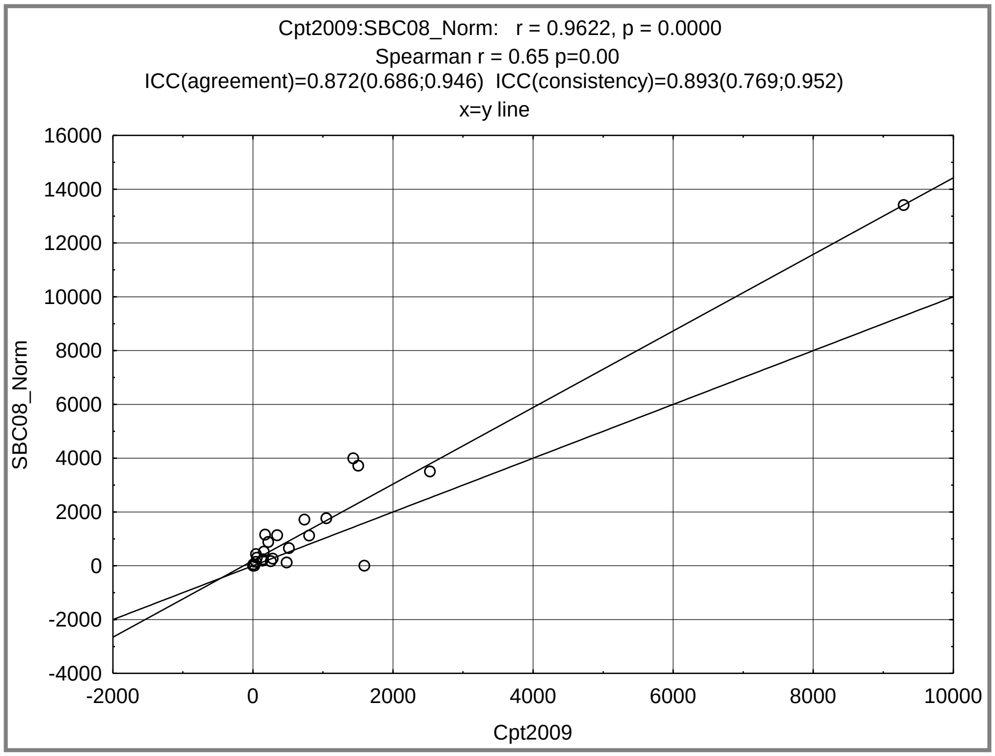

Figure 5: Correlation tests for the City of Cape Town rural areas

As the red line in Figures 4 and 5 shows, there is an overestimation of the population in areas with the higher population totals in the areas the cities classified rural. These are however only a few isolated cases and would not have much of an impact on the population distribution on a city wide level. This is evident in the ICC value of 0.806 for the City of Cape Town and 0.8720 for eThekwini which again is very significant based on the sample size, i.e. the SBC-data frame produces to a significant degree the same result as the figures estimated by the two cities respectively.

The last type of area the correlation was tested for was that of the dense rural areas of eThekwini Municipality. (The City of Cape Town does not have areas that fit into this category.) These are mainly former tribal areas with high population densities and not much in terms of municipal services and infrastructure. The dense rural areas of eThekwini host about 8% of the total population. For the dense rural areas the correlation between the SBC-data frame and the city’s figures are almost identical, see Figure 6. The ICC is 0.965. The results based on the SBC-data frame for eThekwini’s dense rural areas are the strongest for all the areas, which is a positive result for the SBC-data frame owing to the informal / complex nature of these areas and the SBC-data frame capturing the total population accurately.

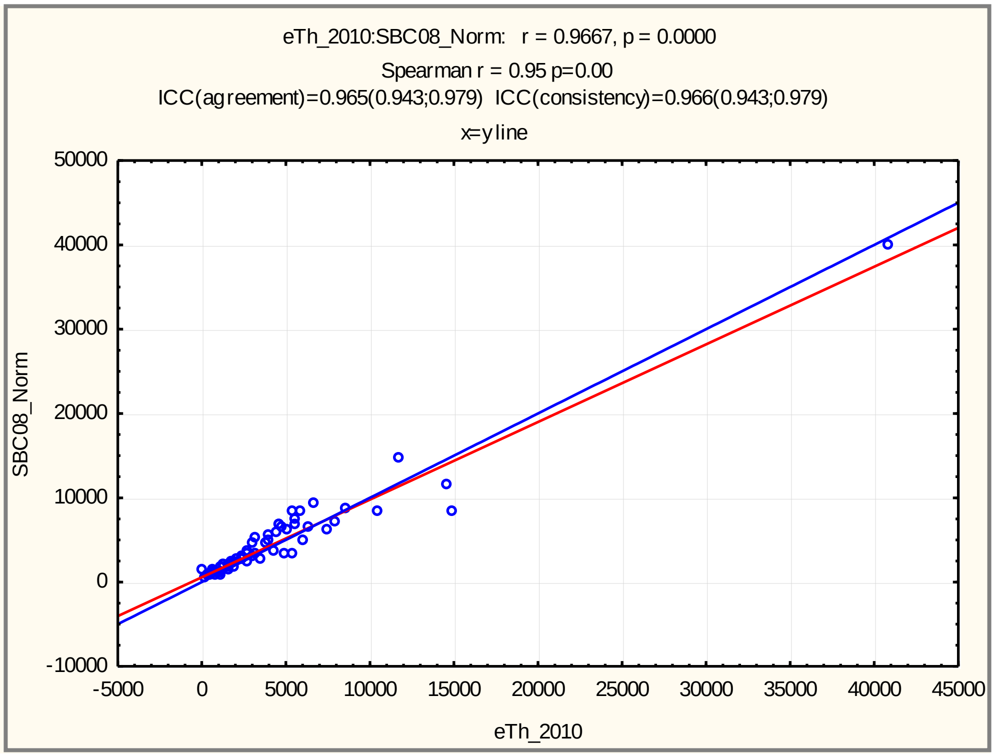

Figure 6: Correlation tests for eThekwini Municipality dense rural areas

Looking at the bigger picture and reminding oneself that the SBC-data frame was prepared on a national level, it is good to know that the dense rural areas are being captured accurately because it is these areas (think Eastern Cape, Limpopo, Mpumalanga) for which is difficult to get accurate and up-to-date population data.

5. Synthesis and way forward

This paper started by reiterating the need for the integration of different spatial datasets (ecological as well as socio-economic) in order to support decision making, especially from a government perspective. This has been a topic of discussion for literally decades, which is an indication of the complexity of finding an easy or one size fits all solution. GIS has in the past ten years gained popularity as well as accessibility (becoming more user friendly and freeware packages becoming more available) which makes the collection, analysis and use of spatial information more commonplace in government at different levels as well as in the private sector. The issue regarding data integration is not ameliorated however as people sit with a host of different datasets with varying demarcations and scales and it is difficult to make decisions on, or to integrate them.

Different attempts have been made to develop platforms for integration which involved the development of basic spatial units (BSUs). These did not solve the problem, however, as even BSUs are dictated by a certain purpose in mind and one ends up with just another demarcation. This lead to people arguing that the units of analysis must be dictated by each analysis or study on its own, which is noble, but does not solve the problems of decision makers who have to base their decisions on existing administrative boundaries, or where fixed physiographic boundaries (like water catchments) are the units of analysis and socio-economic data (with completely different demarcations) forms part of the analysis.

A novel approach based on dasymetric mapping principles was used in order to classify a point dataset of all buildings (spot building count) in South Africa based on socio-economic characteristics. The reasons behind using a socio-economic classified point dataset to develop a flexible data integration frame were that:

- ecological / environmental data is bound by hard physiographic data (inflexible boundaries like water catchments, plant biomes, or mountain ranges);
- point data can be assigned easily to any demarcation;
- SBC-points are an accurate ancillary source for human activity.

The approach is novel in that, firstly, data integration processes and studies usually focus on using polygon data (a demarcation) and, secondly, dasymetric mapping principles have only been applied to processes which involve assigning the result to polygons. For the purposes of the paper and for testing the accuracy of the method only total population was generated as an output.

A hierarchical exclusion process based on dasymetric mapping principles was used to in the classification of the SBC. 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 analysis results were produced for the whole of South Africa. However, to test the results an up-to-date data set on population distribution on a detailed level was needed. The City of Cape Town and eThekwini Municipality had recently undertaken updates of their population figures. The cities' respective data was thus used as a control total and a correlation analysis on census sub-place level was executed. The SBC-data frame results were assigned to the sub-place polygons for the analysis. A regular correlation analysis would not yield optimal results in this instance as we did not want to only test whether one variable fluctuates according to another, but the two variables must also be as close as possible in value. The intra class coefficient (ICC) was therefore determined. ICC 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. The closer the ICC value is to 1 the stronger the correlation. For the City of Cape Town the ICC is 0.907 and 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).

The method therefore yielded a very reliable result in terms of population distribution prediction. Based on these results the dasymetric mapping based methodology for developing the SBC-data frame can be deemed as successful. The way forward would be to refine the variables in the SBC-data frame and include other socio-economic variables. Some of these variables may need a more detailed classification according to the dasymetric principles, but the base (which has been created in this study) will stay the same. It might also be interesting to do the same kind of test regarding accuracy for other areas rather than the two metros concerned. The changes in character of metros are however quite sharp within short distances and, despite this, the method (done on a national level) still yielded a significant result statistically. If a host of socio-economic related variables can be linked to the SBC-data frame through the use of dasymetric principles, a powerful and flexible socio-economic dataset can be developed which can be assigned to any ecological or administrative demarcation. This will enhance, streamline and simplify the process of working with data in an integrated fashion. As illustrated in the introduction, the advantages of this are limitless.

References

Aiello A 2011, email, 26 January.

Bloom LM, Pedler PJ & Wragg GE 1996, 'Implementation of enhanced areal interpolation using MapInfo', *Computers and Geosciences*, vol. 22, pp. 459-466.

Breytenbach A 2010, 'A national inventory using earth observation and GIS', *ScienceScope*, vol. 5, no. 1, September 2010, pp. 66-68.

Costanza R 1989, 'What is ecological economics?', *Ecological Economics*, vol. 1, pp. 1-7.

Costanza R, Segura O & Martinez-Alier J 1996, *Getting Down to Earth: Practical Applications of Ecological Economics*, Island Press, Washington DC.

Daly HE 1990, 'Toward some operational principles of sustainable development', *Ecological Economics*, vol. 2, pp. 1-6.

Eagleson SF, Escobar I & Williamson I 2003, 'Automating the administration boundary design process using hierarchical spatial reasoning theory and geographical information systems', *International Journal of Geographical Information Science*, vol. 17, pp. 99-118.

Eicher CL & Brewer CA, 2001, 'Dasymetric mapping and areal interpolation: Implementation and evaluation', *Cartography and Geographic Information Science*, vol. 28, pp. 125-138.

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.

Goodchild MF & Lam NS 1980, 'Areal interpolation: A variant of the traditional spatial problem', *Geo-processing*, 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.

Huby M, Owen A & Cinderby S 2007, 'Reconciling socio-economic and environmental data in a GIS context: An example from rural England', *Applied Geography,* vol. 27, no. 1, pp. 1-13.

Martin D 1996, *Geographic information systems: Socioeconomic applications*, Routledge, London.

Martin D & Bracken J 1993, 'The integration of socioeconomic and physical resource data for applied land management information systems', *Applied Geography*, vol. 13, no. 1, pp. 45-53.

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.

Naudé AH, Badenhorst W, Zietsman HL, Van Huyssteen E and Maritz J 2007, Technical overview of the mesoframe methodology and South African Geospatial Analysis Platform, Technical Report for the CSIR, Pretoria.

Openshaw, S 1983, *The modifiable area unit problem (Concepts and techniques in Modern Geography)*, Geo Books, Norwick.

Openshaw S & Rao L 1995, ‘Algorithms for re-engineering 1991 census geography’, *Environment and Planning A*, vol. 27, pp. 425-446.

Sang N, Birnie RV, Geddes A, Bayfield NG, Midgley JL, Shucksmith DM, et al. 2005, ‘Improving the rural data infrastructure: The problem of addressable spatial units in a rural context’, *Land Use Policy*, vol. 22, pp. 175-186.

Sinclair-Smith K 2010, email, 21 April.

Walker P & Young M 1997, ‘Using integrated economic and ecological information to improve government policy’, *International Journal of Geographical Information Science*, vol.11, no.7, pp. 619-632.

Wikipedia, 2010, Regression Analysis, Viewed 05 April 2011, http://en.wikipedia.org/wiki/Regression_analysis.

Young MD 1992, *Sustainable Investment and Resource Use: Equity Environmental Integrity and Economic Efficiency*, Parthenon Press, Carnforth and UNESCO, Paris.