Unemployment in South Africa: Building a Spatio-temporal Understanding

Gina Weir-Smith[1], Fethi Ahmed[2]

[1]Human Sciences Research Council, gweir-smith@hsrc.ac.za
[2]University of Johannesburg

Abstract

The spatial understanding of unemployment in South Africa is often limited to provinces. This paper explores ways to integrate unemployment data from 1991 to 2007 to obtain a spatially more detailed understanding of the phenomenon. Creating such a temporal GIS is challenging because of the changing administrative boundaries and this phenomenon is part of the modifiable areal unit problem (MAUP). The MAUP was overcome by using proportional aggregation to the 2005 municipal boundaries. The unemployment data were analysed spatially through mapping standard deviation intervals and spatial autocorrelation. Unemployment showed a spatial pattern over time in South Africa and municipalities with high unemployment figures over time were spatially concentrated. Furthermore, it is recommended that census data should be released at a small area level to allow more accurate comparison of socio-economic data over time.

1. Introduction

Unemployment globally has been on the rise till around 2004 and increased again since 2008 due to the recent global recession. According to the annual report on global employment (ILO, 2012), the world faces the "urgent challenge" of creating 600 million productive jobs over the next decade in order to generate sustainable growth and maintain social cohesion. After three years of continuous crisis conditions in global labour markets and against the prospect of a further deterioration of economic activity, there is a backlog of global unemployment of 200 million (ILO, 2012). Moreover, the report indicates that more than 400 million new jobs will be needed over the next decade to absorb the estimated 40 million growth of the labour force each year.

In South Africa the figure for unemployment (strict/official definition - excludes discouraged workers) has risen from 21.0% in 1996 to 25.8% in 2001 and slightly decreased to 25.2% in 2010 (Statistics South Africa, 1996; Statistics South Africa, 2003; Statistics South Africa, 2010). Unemployment in South Africa is regularly measured at different spatial levels and includes national or sometimes provincial (e.g. Labour Force Survey – LFS) statistics. Since data at provincial level is not expected to provide meaningful variation in unemployment over time, this paper will use census data which is presented at municipal level and is collected less frequently. The census data is the only data source which includes all citizens and is spatially comprehensive in terms of unemployment in the country. The census definition of unemployment focuses on the narrow (official) definition of unemployment, because it excludes discouraged workers. Many economists believe that the broad definition is a more accurate reflection of unemployment (Kingdon and Knight, 2005; Kingdon and Knight, 2000; Klasen and Woolard, 2008). Although it is acknowledged that there are differences in the measurement, for the purposes of this analysis the narrow definition will be used to ensure a spatially comprehensive, sub-provincial comparison.

1.1 Displaying socio-economic data spatially

The first time census data was captured in a GIS format in South Africa was in 1991. Prior to 1991 data was available at an aggregate level and usually inaccessible to the general public. Since the 1991 census, socio-economic data has not only been available electronically, but has also progressively become more available in the public domain. The Community Survey (CS) of 2007 from Statistics South Africa (Stats SA) was released electronically and can be downloaded from the internet.

Census data should be presented at spatial levels that are small enough to enable the analyst to see the variations within an area of analysis (e.g. municipality or sub-place). Metropolitan areas and municipalities would, in most cases, require census information at a ward, sub-place or Enumeration Area level to enable users to understand the variations in demographic patterns within their areas (Statistics South Africa, 2007).

The comparison of socio-economic trends over time also requires that the definition of such concepts remain comparable and that the spatial unit of analysis remain consistent. If not, both of these become a challenge. In South Africa the definition of some variables collected during census periods have changed, however the statistics attached to labour trends have remained the same. The greatest challenge is however the comparison of data for spatial units (e.g. for Enumeration Areas or municipalities). Enumeration Area (EA) boundaries not only changed between 1991, 1996 and 2001, but there were also several cross-provincial local municipalities between 1996 and 2005. Such cross-boundary local municipalities, district councils, district management areas and metropolitan areas have to be taken into account when aggregations are calculated from EA data.

Statistics South Africa (2007) reports that the following factors complicate the work of researchers and academics who wish to do comparative studies using data collected in South African censuses:

a) the continuous and comprehensive changing of administrative boundaries
b) the revision of the set of enumeration areas that was used in Census 1996
c) the decision not to release the Census 2001 data at enumeration area level.

1.2 Displaying change over time

In order to compare change in data over time it is important to ensure the measuring units (i.e. polygons of analysis) are the same. Since the delineation of EAs is primarily concerned with the management of enumerator workloads and has to take residential development, demolishing, change in density and residential structure into consideration, it is expected that these boundaries will change over time (Martin et al., 2002).

There are three ways to overcome the above-mentioned problem: either remodel the data to some underlying surface-based representation (Bracken and Martin, 1995); use areal interpolation to transfer data from one set to another (Goodchild et al., 1993); or use lookup tables to make best-fit assignments of one set of areal units into another (Martin et al., 2002).

The first option refers to the aggregation of old enumeration areas to new boundaries and this task is relatively simple if there is a set of boundaries that are identical in both censuses - that is, a "lowest common denominator". For instance, in the most recent census, only the EA boundaries may have changed, but not the sub-place or main place boundaries. Population data can then be compared for sub-places simply by aggregating the EA data. However, EAs may have been reassigned to different sub-places between the censuses without actually changing the boundaries. In such a case one would have to determine into which sub-place in census 1 each census 2 EA falls. GIS polygon overlay operations can assist in this task (United Nations, 2000).

Option two refers to instances where the boundaries of reporting units for the two censuses are not nested at some geographic level of aggregation then some form of areal interpolation is required to obtain compatible census data. Areal interpolation is the process of transferring data - for example, population totals - from one set of areal units to another incompatible set of units.

Areal interpolation refer to the process of using data published for differing administrative units and estimate their values for a single set of standardized units (also referred to as target units) (Langford et al., 1991). In its simplest form areal interpolation is based on the assumption data is evenly distributed across the source units. According to Gregory the accuracy of areal interpolation will vary according to the nature of the variable being interpolated, the nature of the ancillary data, and the shape and size of both the source and target units (Gregory and Ell, 2006).

Option three requires a numbering system which indicates corresponding numbers for each small area that fits into a larger area. This process could require more manual input than the previous options.

Where researchers are concerned with the analysis of change over very small geographical areas, the options one or two may be the only alternative, but for larger areas, option three has considerable utility. For the purpose of analysing inter-censal change, higher-level geographical units (e.g. local districts) often form the most appropriate units for analysis (Martin et al., 2002), (Gregory and Ell, 2006).

It is important to note that no interpolation method can provide error-free estimates of target zone socio-economic indicators. In fact, the errors may often be unacceptably large for applications requiring high accuracy. Areal interpolation should thus be seen as a method of last resort, where more accurate options - such as re-aggregation of small data collection units - are unavailable (United Nations, 2000).

Comparative work on solving the MAUP was done by Martin et al. (2002) and Gregory and Ell (2006). Martin et al linked the census data of 1971, 1981 and 1991 for England, Scotland and Wales by aggregating the 1991 data to 1981 Enumeration Districts using a point in polygon technique (Martin et al., 2002). Gregory and Ell (2006) interpolated British census data from 1851 to 1930 to the parish boundaries of 1951 (the target units) and created a historical GIS of census data. No literature could be found which indicates solutions for South African census data or other time series data of a spatial nature.

This article will address the following objectives:

a) Attempt to overcome the modifiable areal unit problem (MAUP)
b) Identify spatial patterns of unemployment over time
c) Briefly analyse trends of spatio-temporal unemployment variation at a sub-provincial geographical level.

2. Methodology

In using unemployment data from censuses, the following assumptions are made:

a) Census data are adjusted statistics based on enumeration status where rates of unemployment are projected.
b) The number of unemployed people and other information relevant to unemployment are correctly reported.
c) All unemployed people reside in the respective (households) in the Enumerator Areas.
d) All citizens are covered by the census.

The following are limitations of using unemployment data as extracted from census statistics: no distinction is made between different forms of unemployment, census does not explore options of non-economic activities and unemployment data does not account for migration patterns. Furthermore, public perceptions could influence the response rate obtained from censuses and require different levels of imputation.

The data sources used for this study originates from Stats SA, the Human Sciences Research Council and Global Insight. Stats SA, the official producer of statistics in South Africa, produces unemployment data at a municipality level, but only during censuses and the data used here refer to Census ’96 and Census 2001. Censuses are conducted across the whole country and supposedly collect data from all citizens in the country. In 2007 Stats SA conducted the Community Survey (CS) which presents data on unemployment at municipality level. The CS was a very large sample survey which was conducted in the absence of a census.

In 1991 the Human Sciences Research Council (HSRC) produced census data spatially for magisterial districts and EAs. Global Insight, a commercial company, produces demographic data for South Africa and this data was used for comparison purposes of data in 1996, 2001 and 2007. Global Insight has created their own labour market model which estimates a time series of employment, unemployment and the economically active population by drawing on selected, reliable, points from the various official data sources and interpolating for the rest (Global Insight, 2011).

2.1 Overcoming the MAUP

Considering the different ways of overcoming the census MAUP in South Africa, the ideal solution would be to use the boundaries of existing small area features like EAs or sub-places. The challenge of not having data for all EAs in 1991 and 1996, made this difficult to attain. The same challenge would apply to small area grids and automated zoning, since the lack of underlying data would make it difficult to achieve results at small area levels. Automated zoning could be used at higher level geographies like municipalities, but since data existed for magisterial districts in 1991 and 1996, it was more opportune to aggregate these to the 2005 municipalities.

Different areal interpolation methods were tested at an EA level for the 1991 data and areas with small populations were more prone to errors. A difference of up to 75% in the results was experienced in areas of small populations while differences in areas with larger populations were less than 1%. This finding underlies the fact that there will always be some form of error associated with areal interpolation (Gregory et al., 2008).

Using the 1991 EA data, the option to aggregate data by centroid to the nearest containing feature resulted in 14.7% of the municipalities not having data. To overcome the problem of not being able to do successful areal interpolation from an EA level and to ensure the comparability of unemployment data from the various official data sources, the census data was aggregated proportionally from the magisterial districts of 1991 to the municipal boundaries of 2005. The aggregation was proportional to the size of the area of a magisterial district that fell within a municipality. Such an aggregation is based on the assumption that the population is evenly distributed across space and the detailed methodology is referred to in a different article. The 2005 municipal boundaries were used as the common denominator and was the most recent source of municipal boundaries at the onset of this research. For Census 2001, data fashioned for the 2005 boundaries was downloaded from Stats SA website, while Community Survey 2007 data was produced at this spatial level and did not require any further re-calculation.

2.2 Spatial autocorrelation and standard deviation mapping

Spatial autocorrelation measures feature similarity based on both feature locations and attribute values simultaneously. It establishes the spatial randomness of a trend. Positive spatial autocorrelation refers to a map pattern when geographic features of similar value tend to cluster, whereas a negative spatial autocorrelation indicates a map pattern in which geographic units of similar values scatter throughout the map. Where no statistically spatial autocorrelation exists, the pattern of spatial distribution is considered to be random (Cliff and Ord, 1973; Griffith, circa 2002).

Spatial autocorrelation was calculated for unemployment rates at a municipality level for the whole country. A separate autocorrelation was run for each year under investigation. The ESRI ArcMap software tool was used and it calculates the Moran's I Index value and both a Z score and p-value evaluating the significance of that index. The Z score indicates measures of standard deviation. For example, if a value of +2.5 is calculated, it is interpreted as "+2.5 standard deviations away from the mean". Besides this a Moran's I value is also calculated and it indicates whether a pattern is random, clustered or dispersed. A value close to +1.0 indicates clustering while a value close to -1.0 indicates dispersion (ESRI, 2009). Zero spatial autocorrelation means geographically random phenomena and chaotic landscapes (ESRI, 2009).

In order to calculate the effect of neighbouring municipalities, a distance calculation is used. The distance method used in this case was "inverse distance" which means that all municipalities impact/influence all other municipalities, but the farther away municipalities are, the smaller the impact it has. An Euclidean distance counting method was used and this means distance was calculated "as the crow flies" – the shortest distance between two points (which would be the centroid of the municipality).

The standard deviation provides a measure of dispersion relative to the mean (Pagano, 2004). Class breaks were created by using standard deviations and this classification is based on the fact that in the normal distribution for example 72 % of the data were within the arithmetic mean (or one standard deviation interval) and 91% of the data were within the arithmetic mean (or two standard deviation intervals). The example used here refers to the percentage unemployed people in 1991 and Figure 1 shows the distribution of the data as well as the standard deviation intervals (thin dotted lines). The dot-dash line between 15.8 and 27.1 indicates the arithmetic mean while the blue lines indicate the class breaks used to map the data.

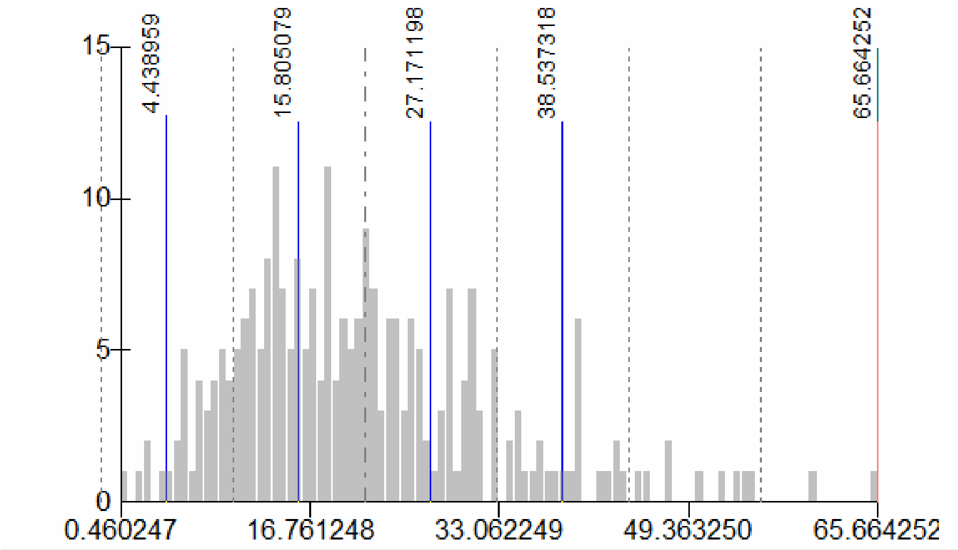

Figure 1: Histogram of 1991 unemployment data distribution

The figure shows that many municipalities have unemployment values that fall within one standard deviation on both sides of the mean. It also indicates that the majority of municipalities has an unemployment figure of between 4% and 40% and that municipalities that fall outside this range, can probably be considered as outliers in the data.

3. Discussion of Findings

This section considers results from the MAUP aggregation, the standard deviation analysis and spatial autocorrelation of the unemployment data. The first part focuses on overcoming the modifiable areal unit problem.

3.1 Results of solving the modifiable areal unit problem

Results from the areal interpolation of the 1991 and 1996 magisterial districts used in this article yielded the statistics shown in Figure 2 and only selected municipalities are discussed here since there are too many municipalities to analyse in one figure. The figure shows results for two municipalities in each of the nine provinces and these municipalities reflect different types of entities like urban, metropolitan and rural areas. The municipality with the highest unemployment in all years was Nongoma in KwaZulu-Natal.

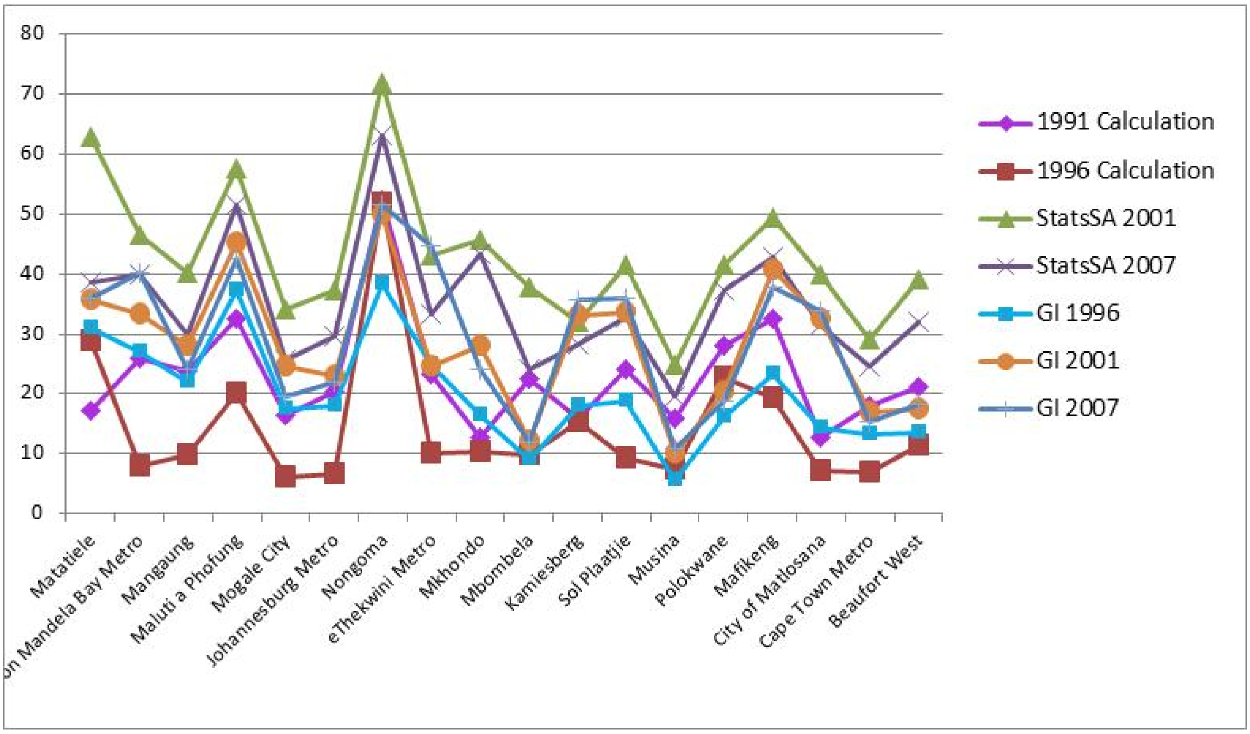

Figure 2: Comparison of unemployment statistics

This municipality showed continuous high levels of unemployment (above 31%) in 1991, 1996, 2001 and 2007. There was an increase of about 20% in unemployment between 1996 and 2001 while the statistics dropped somewhat in 2007. Data from Global Insight (2011) in comparison, show somewhat lower unemployment statistics of 38.4% in 1996 and a steady increase to 51.4% in 2007. Large differences are seen between the two sets of 1996 and 2001 statistics from Stats SA, especially in the metros. For example, City of Cape Town recorded 6.9% unemployment in 1996 using the areal interpolation method, while Global Insight (GI) (2011) indicated 13.2%. In some metros, e.g. Cape Town and Johannesburg the differences between Stats SA results and GI data prevailed for 2001 and 2007 data. Since the areal interpolation was based on Stats SA data, the results are closer to the Stats SA data than the GI data. These differences in results, underscore the complexity of labour data (Illeris, 1995).

Municipalities throughout the country showed a steep increase in unemployment in 2001. Stats SA (2012) cautioned analysts about the 2001 data since it might reflect a disproportionate percentage of unemployed people. The differences in the results over the years may be partly attributable to questionnaire changes rather than to actual developments in the labour market (Stats SA, 2012).

3.2 Spatial analysis

The spatial analysis was done using standard deviation class breaks as shown in Figures 3 and 4. Yellow and light green colours show unemployment values close to the mean while brown and dark green colours show unemployment values far from the mean with the darkest green being a positive difference. The standard deviation classification does not show actual values for polygons, but only which ones are below or above the mean. In 1991 few municipalities were grouped in the dark green category (highest above the mean). This category indicates the municipalities which had the highest deviation from the mean, i.e. worse off, and included municipalities like Intsika Yhetu (Eastern Cape), Umzimkhulu (KwaZulu-Natal) and Makhuduthamaga (Limpopo). The latter municipality is predominantly rural and situated in a mountainous area with limited access roads. Intsika Yhetu (Cofimvaba) is a sparsely populated, mountainous municipality while Umzimkhulu is has large areas of forestry and is also sparsely populated. All metropolitan municipalities were in the middle or second best off category (-0.50 to 0.50).

In 1996 many municipalities in the Eastern Cape, Limpopo, Mpumalanga and North West were in the worse off category and included large parts of the previous Transkei and Ciskei areas in the Eastern Cape. Other municipalities in the worse off category were Moshaweng (Northern Cape), Ratlou and Greater Taung (North West), Blouberg, Aganang, Mutale, Greater Tubatse and Makhuduthamaga (Limpopo). All metros were in the two better off categories. In 2001 more municipalities in the Eastern Cape became part of the worse off group, while Makhuduthamaga (Limpopo) remained in this group. The eThekwini metro (KwaZulu-Natal) fell into the category of the second worst group in 2001. In 2007 the Kareeberg municipality (Northern Cape) was grouped in the worst off category for the first time.

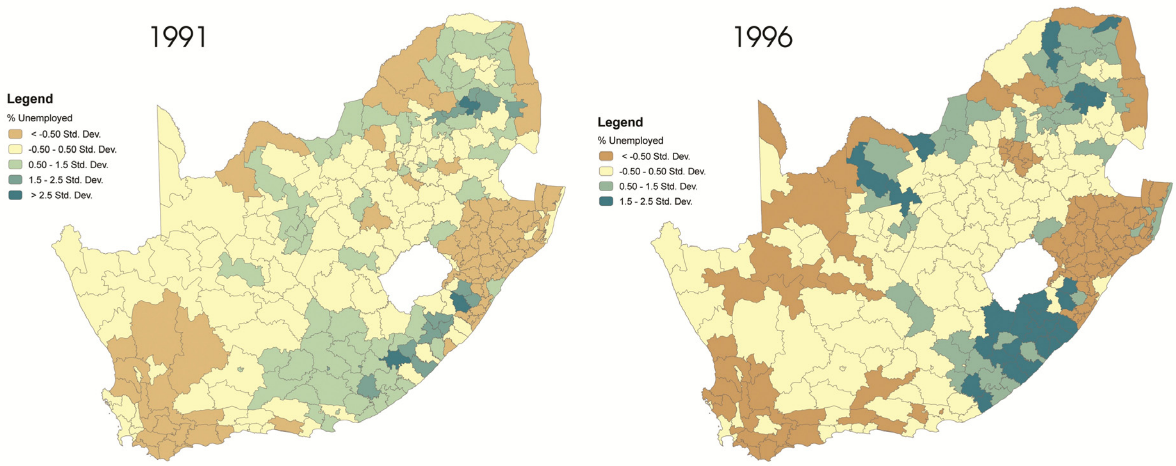

Figure 3: Unemployment trends for 1991 and 1996 using standard deviation intervals

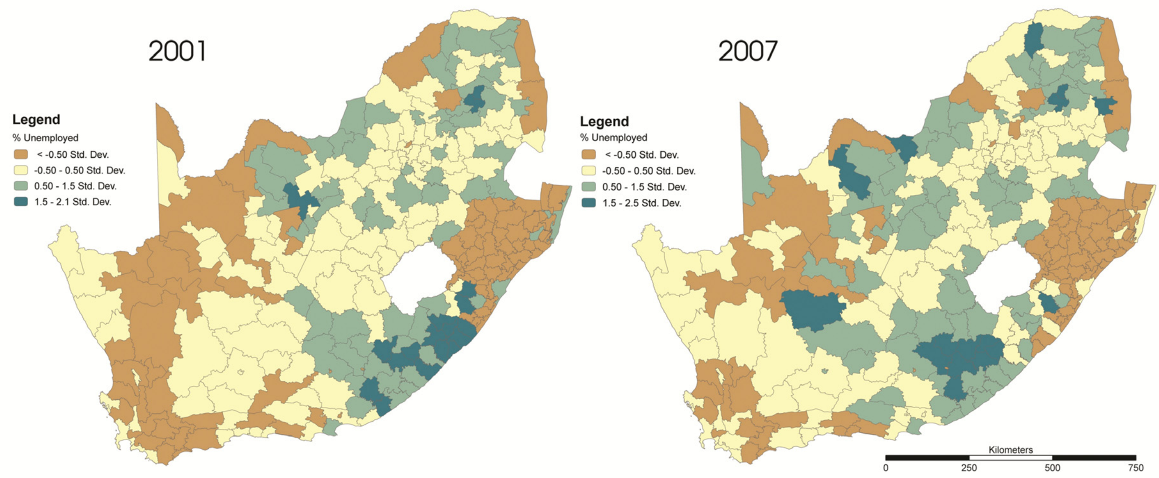

Figure 4: Unemployment trends for 2001 and 2007 using standard deviation intervals

Municipalities in the Western Cape, Northern Cape, Free State, Gauteng and most of KwaZulu-Natal were consistently in the standard deviation class that is best off in terms of unemployment. High unemployment rates were concentrated in the southern parts of KwaZulu-Natal, the north eastern parts of the Eastern Cape and selected municipalities in Limpopo and North West. In terms of the period in time with the highest number of municipalities, the 1996 data showed that 28 municipalities were grouped in the worst-off category while 2001 had the second highest number (i.e. 18).

This type of analysis helped to identify spatial trends of unemployment over time and showed that increasingly more municipalities (from 3 in 1991 to 16 in 2007) were grouped in the worse off unemployment category over time period under investigation. Unemployment was predominant in municipalities that were sparsely populated and often remote.

3.3 Spatial autocorrelation of unemployment data

The spatial autocorrelation results showed that in 1991 the Moran's I index was 0.61 for the percentage unemployed people. The Z-score was 25.87 standard deviations and the spatial distribution of unemployment was not random (values were calculated at a 99% confidence level). Similar statistics were found for subsequent years (see Table 1).

Table 1: Spatial clustering of unemployment

	Moran's I	Z-score	Spatial Concentration
1991	0.61	25.87	Clustered
1996	0.50	21.25	Clustered
2001	0.71	30.08	Clustered
2007	0.71	29.92	Clustered

The results of the spatial autocorrelation showed that unemployment was not randomly distributed, but was clustered in all years. In fact the spatial clustering of unemployment increased between 1991 and 2007 and the Moran's I value was 0.71 for the 2007 data. There was a slight decrease in 1996 which indicates that the spatial distribution of the percentage unemployed people was less clustered than in other years. In fact, in all years except in 1996 when the unemployment statistics were slightly lower, the Z-score was high and fell outside the normal distribution range. This indicates a statistically significant spatial pattern.

Furthermore, unemployment seemed to have a regional persistence, because provinces with high unemployment rates remained high over time. Table 2 indicates for example that KwaZulu-Natal, Limpopo, North West and Eastern Cape recorded unemployment rates of over 40% in 2001. Slight declines were recorded in 2007 and the data excludes current unemployment trends which would reflect the impacts of the post 2008 recession.

Table 2: Percentage unemployed people per province between 1991 and 2007

Province	1991	1996	2001	2007
Eastern Cape	26.1	29.5	52.4	40.4
Free State	16.6	12.3	38.5	35.2
Gauteng	18.0	7.2	32.1	26.0
KwaZulu-Natal	29.9	25.5	53.3	38.4
Mpumalanga	17.7	11.6	40.3	32.2
Northern Cape	18.1	16.2	30.9	28.9
Limpopo	23.1	26.5	43.8	37.0
North West	18.7	17.1	41.9	37.9
Western Cape	11.1	8.6	20.2	17.0

Unemployment rates below 25% were recorded in the Western Cape during all years.

4. Conclusion and Recommendations

The data analysis of this article showed that unemployment has a spatial pattern over time in South Africa. It can therefore not be solved without considering the spatial nature of the phenomena. Since the spatial pattern of unemployment was persistent along provincial boundaries and showed an increased spatial clustering, unemployment interventions should be focused at local areas like municipalities or at least at a provincial level instead of nationally.

The MAUP was overcome by proportionally aggregating data from 1991 and 1996 magisterial districts to 2005 municipal boundaries. Future work could consider calculating a statistical error related to each interpolation or identifying which individual interpolated data values are suspected of containing error. The latter is the only way to completely eliminate areal interpolation errors and to explain the large difference in the findings.

In order to conduct detailed spatial analysis of socio-economic phenomena over time, Stats SA should release data at small area level for research purposes and design small areas that fit within existing administrative boundaries so that the percentage error in areal interpolation could be minimized. This will also allow easy comparison of socio-economic trends over time.

Large differences in the areal interpolated unemployment data and other data sources, emphasizes the importance of using official statistical sources. It also underscores the need for official statistics to be released at small area levels so that reliable datasets can be built from that. Even though it is an official dataset, Stats SA (2012) cautioned users to take care when interpreting labour statistics from 1996.

5. References

Bracken, I and Martin, D 1995, 'Linkages of the 1981 and 1991 censuses using surface modelling concepts' *Environment and Planning A,* vol. 27, pp. 379-390.

Cliff, AD and Ord, JK, 1973. *Spatial Autocorrelation,* Pion, London.

ESRI, 2009, *ArcGIS desktop help manual*, Redlands, Environmental Systems Research Institute.

Global Insight, 2011, *Regional eXplorer encyclopedia*, Global Insight, Pretoria.

Goodchild, MF, Anselin, L and Deichmann, D, 1993, A framework for the areal interpolation of socio-economic data, *Environment and Planning A,* vol. 25, pp. 383-397.

Gregory, IN, Marti-Henneberg, J and Tapiador, F, 2008, A GIS reconstruction of the population of Europe, 1870 to 2000, viewed 19 July 2011, <http://web.udl.es/dept/geosoc/europa/cas/img/GIS%20Approaches%20-%20resubmission.6.06.pdf>.

Gregory, IN and Ell, PS, 2006, 'Error-sensitive historical GIS: identifying areal interpolation errors in time-series data', *International Journal of Geographical Information Science,* vol. 20, no. 2, pp. 135-152.

Griffith, DA, circa 2002, Spatial Autocorrelation, Syracuse University, New York.

International Labour Organisation (ILO), 2012, *Global employment trends 2012: preventing a deeper jobs crisis*. Employment Trends unit of the ILO, Geneva.

Illeris, S, 1985, The geography of unemployment: the case of Denmark. *Espace, populations, sociétés*, 1985-2. Chômage, non-emploi, sous-emploi, pp. 285-297.

Kingdon, GG and Knight, J, 2000, *Are searching and non-searching unemployment distinct states when unemployment is high? The case of South Africa*. Oxford: University of Oxford, Centre for the Study of African Economies.

Kingdon, GG and Knight, J, 2005, *Unemployment in South Africa, 1995 – 2003: causes, problems and policies,* Stellenbosch, 28-29 October, Paper presented at the '10 year review of the South African economy' conference.

Klasen, S, and Woolard, I, 2008, 'Surviving unemployment without state support: unemployment and household formation in South Africa', *Journal of African Economies,* vol. 18 no. 1, pp. 1–51.

Martin, D, Dorling, D and Mitchell, R, 2002, 'Linking censuses through time: problems and solutions', *Area,* vol. 34, pp. 82-91.

Langford, M, Maguire, D and Unwin, DJ, 1991, 'The areal interpolation problem: estimating population using remote sensing in a GIS framework', in I Masser and M Blakemore (eds.), *Handling Geographical Information: Methodology and Potential Applications,* Longman, Harlow.

Pagano, RR, 2004, *Understanding statistics in the behavioral sciences,* 7th edition, Wadsworth, Belmont.

Statistics South Africa, 1996, Census '96, Dataset.

Statistics South Africa, 2003, *Census in Brief, 2001*, Statistics South Africa, Pretoria.

Statistics South Africa, 2010, *Key indicators*, viewed 11 July 2010, <http://www.statssa.gov.za/keyindicators/keyindicators.asp>.

Statistics South Africa, 2007, *Using the 2001 Census: approaches to analysing data,* viewed 22 August 2010, <http://www.statssa.gov.za/Publications/CensusHandBook/CensusHandbook.pdf>.

Statistics South Africa, 2012, *Census 2011: provinces at a glance*, Statistics South Africa, Pretoria.

United Nations, 2000, *Handbook on geographic information systems and digital mapping*, United Nations, New York.