Finding The Jobless: An Initial Spatio-Temporal Modelling Of Unemployment In South Africa

Gina Weir-Smith

Human Sciences Research Council, Private Bag X41
Pretoria, 0001, South Africa
Tel: +27 12 302 2529
Fax: +27 12 302 2525
Email: gweir-smith@hsrc.ac.za

Abstract

Unemployment globally has been constantly on the rise in the past few years. In South Africa the figure for unemployment (strict definition) has risen from 21.0% in 1996 to 25.8% in 2001. Besides the identification of poverty and unemployment as priorities by NEPAD (New partnership for African Development), the Growth and Development Summit (GDS) of South Africa has also set the ambitious task in 2003 to reduce unemployment by half over the next few years.

To date some reference has been given to the spatial aspect of unemployment research in South Africa, but mostly very generalised. Very little research has been done in comparing the patterns spatially over time. A temporal comparison of unemployment patterns is important to predict future occurrences.

The primary aim of the research is to understand the spatio-temporal structure in unemployment in South Africa. The first phase of the research is focused on obtaining all the relevant data sets for the same spatial extent. These include poverty, unemployment, income, gross geographic product (GGP) and other relevant socio-economic indicators.

Results from three consecutive censuses were compared at a small area level in GIS and assisted in identifying and describing patterns of unemployment – especially spatially. Other data sets like GGP was used to enhance the analysis where required.

The findings showed a marked increase in unemployment over time as well as in a spatial extent. Correlations between unemployment and low levels of education and the number of African people were recorded. The coincidence of unemployment and certain economic sectors were observed.

The work presented here contains the initial findings made in one of the provinces in South Africa.

1. Introduction

Unemployment globally has been constantly on the rise in the past few years. To this end this study aims to develop a GIS model of unemployment for South Africa that will assist in reaching the Millennium Development Goals (MDG) as agreed to by the G8 countries.

However, numerous representational challenges face a study of this nature since the display of socio-economic data over time is difficult and the geographical units have changed and are therefore affected by the modifiable areal unit problem (Openshaw 1984). The presentation of change over time in GIS remains a research challenge and this study will aim to address the issue.

In South Africa the figure for unemployment (strict definition [1]) has risen from 21.0% in 1996 to 25.8% in 2001 (Statistics South Africa 1996 and 2001). Unemployment and poverty have been identified as critical issues by the New Partnership for African Development (NEPAD) and other international organisations. The Growth and Development Summit (GDS) of South Africa has set the ambitious task in 2003 to reduce unemployment by half over the next few years (Sunday Times 2003). In order to fulfill this and other international objectives like the reduction of poverty and creating jobs, data is needed to identify where interventions, investment and upliftment is needed.

This research is initiated out of the need to understand the thus far neglected spatial aspects of unemployment and to build a spatial model around it. This will help to improve macro-economic planning, resource allocation and governance. Spatial disparities between regional employment levels and poverty have been persistent in South Africa – as in many African countries - over the past decades. This persistence of the regional structure of unemployment (Arrufat 1998) needs to be investigated in comparison to inequality and poverty.

Although unemployment is experienced at a macro economic level, it has a specific micro level manifestation (Kingdon 2003). In order to solve the macro economic phenomena it is important to take note of the micro level characteristics of unemployment. By analysing and predicting unemployment at a local level, the research will aim to address this issue. Although many unemployment trends have been identified in recent years in South Africa very little research has been done in comparing its spatio-temporal patterns. A temporal comparison of these patterns is important to predict future occurrences.

In South Africa, both the censuses (1991, 1996, 2001) and a Labour Force Survey (LFS) collect statistical data on unemployment regularly. Although the LFS is regarded as the official source of labour market statistics, the census data was preferred for this analysis due to its longitudinal aspect and the ability to present this data spatially with ease. Many macro-economists argue that the unemployment figure is over-represented in the census [2], and this might be the case, but for the purpose of this study it is important to establish trends and patterns over time and the census data was the only data source which fulfilled the requirements of being able to compare data over significant time periods. The census data also covers the whole country and it is therefore preferred for purposes of prediction based on geographical units.

2. Background and methodology

The first phase of the project is aimed at analysing provincial unemployment patterns at the most local spatial level in Gauteng. This will assist in identifying underlying issues and problems related to unemployment. It also makes the task more manageable, because it is

[1] The unemployed are those people within the economically active population who (a) did not work in the seven days prior to census night, (b) wanted to work and were available to start work within a week of census night, and (c) had taken active steps to look for work or to start some form of self-employment in the four weeks prior to census night.

[2] For example in 2001 the comparative statistics for unemployment between the two sources were: 24.0% (Census) and 16.5% (LFS).

broken down into smaller units of analysis. It was decided to break the analysis down into the nine provinces of South Africa, because they have distinct socio-economic characteristics and units of analysis are easier to handle. These are therefore treated as regions of economic activity and it is expected that each region will have a core-periphery structure (Suedekum 2003). The core being the area where unemployment is low and population and income is high. The periphery refers to areas of peripheral economic activity and high unemployment. The first province to be analysed is Gauteng. This is the economic heartland of South Africa with 20 % of the total population residing on 1.4% of the country's land. It is highly urbanised with a strong emphasis on service occupations. The province also includes non-metropolitan areas on the periphery.

When comparing the change in unemployment data over time, considerations should be given in terms of spatial and non-spatial (attribute) changes. These two issues will be considered separately in the following paragraphs. It is also necessary to regard the changes in methodology in order to make inter-censal population change (Martin & Gascoigne 1994). Initial data capture was done at Enumerator Area (EA) level - the smallest spatial units used during a census.

3. Census data used

3.1 Spatial data

Up to now census data is collected every five years in South Africa. The censuses compared here will include the last three censuses, namely those conducted in 1991, 1996 and 2001. The primary purpose of an Enumerator Area (EA) is to demarcate are where one enumerator must distribute and collect census forms. It takes into consideration the change in neighbourhoods, political boundaries, urban development, etc. It is therefore possible that the spatial extent of EAs can vary quite significantly between consecutive censuses. Besides this South Africa also underwent a major polititial change in 1994 which affected most statutory boundaries at that stage. The provincial boundaries of South Africa also changed in 1994 and the EA boundaries were affected by this change. The percentage change between EA boundaries was difficult to assess. Since no official record keeping was done about changes a visual comparison yielded the following result for the three censuses.

The main spatial difference in Gauteng between the 1991 and 1996 censuses was that townships were not demarcated in detail in 1991. Traditionally townships are characterised by high-density population concentrations with some informal settlements in-between. Due to political unrest at the time of the census, townships were demarcated as one spatial entity and the census data was provided for the entire entity. Therefore, no census data was provided for smaller spatial units within the townships. The previous homelands (like GaRankuwa and KwaNdebele) were also excluded in the 1991 census, but a survey was done by the HSRC to collect this data. This was then merged with the EA census data to obtain a complete picture for the province. The total number of EA's in Gauteng for the 1991 census was 7550, compared to 15 962 in 1996 and 13 177 in 2001.

Figure 1. Comparison of spatial boundaries from various censuses.

The spatial differences between 1996 and 2001 censuses included the expansion of suburban neighbourhoods and townships alike. Detailed spatial data at an EA level was available in 1996. Although the spatial boundaries for EA's were updated during the 2001 census, Statistics SA decided not to release attribute data at an EA level for confidentiality reasons.

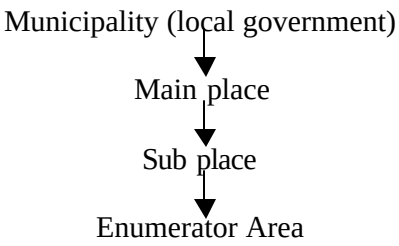
3.2 Attribute data

In the 1991 census unemployment statistics was not directly calculated. To generate these statistics the number of employed people had to be subtracted from the economically active population. In the 1996 census the number of unemployed people per EA was provided. The statistics was therefore in comparable formats.

The 2001 census attribute data was not released at an EA level and this consequently made comparisons with the previous two censuses very difficult. The spatial boundaries for the EAs were made available and statistical modelling techniques could be used to compute unemployment statistics for these boundaries. Attribute data was released at a sub place level and this was used for the analysis done here.

3.2.1 Data comparison based on individual years

Due to the difficulty of incomparable spatial boundaries, it was decided to aggregate the unemployment statistics to a common boundary. This was the most uncomplicated way to overcome the modifiable area unit problem (MAUP). It was decided to use the 2001 sub place boundaries as the common denominator. Sub places are part of a spatial hierarchy developed by Statistics SA. This hierarchy includes the following levels:



The number of polygons for sub places is considerably less that those of EAs and a certain degree of generalism will therefore occur. For example the 2001 EA polygons for Gauteng equalled 13 177 while the sub places were 2 215. The latter dataset tend to consist of larger polygons in the peri-urban areas of the province, but smaller ones in the higher populated areas. It is therefore foreseen that it will still be able to portray enough detail at a local level.

The attribute data of EA for 1991 and 1996 censuses was then proportionally aggregated to the sub place boundaries of 2001. The following paragraphs compare the statistics resulting from this aggregation.

The aggregated census data for Gauteng is displayed in the following graph.

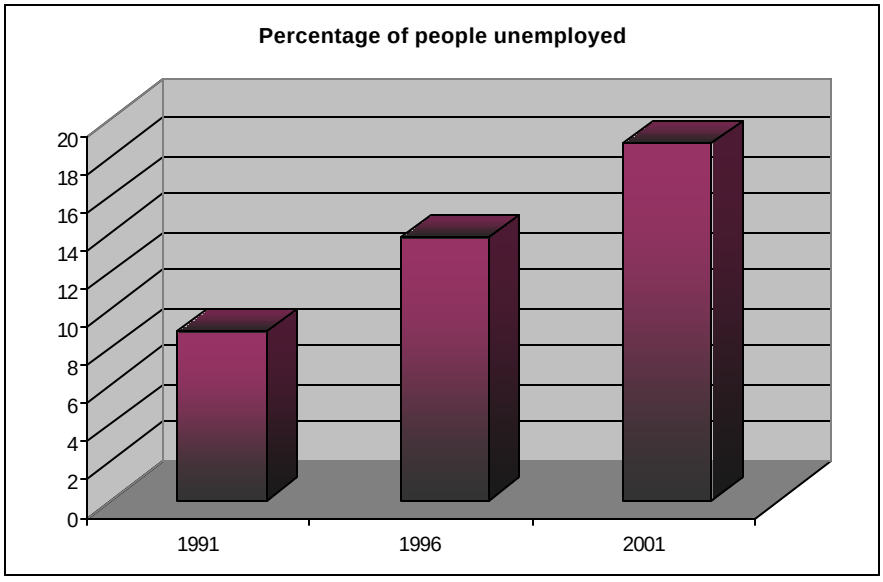


Figure 2. Percentage of unemployed people in Gauteng per census

To calculate the unemployment statistics for Gauteng in 1991, data from the neighbouring province, Mpumalanga had to be incorporated since the provincial boundaries changed between 1991 and 1996.

Figure 3 indicates the distribution of unemployment in the province in 1991 and 1996. Extremely high unemployment (more than 30% of the economic active population) was recorded in almost 390 places (18%) including townships in the south, south east, south west, other areas on the West and East Rand and areas in the northeast of Gauteng.

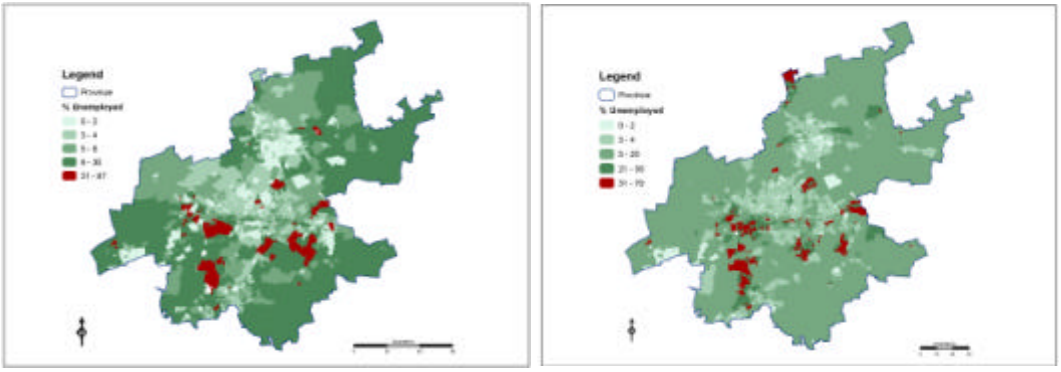


Figure 3 (a) and (b). Percentage of unemployed people 1991 and 1996 in Gauteng per sub place

In 1996 the number of areas with an unemployment rate higher than 30% amounted to 324 (15%). The distribution of unemployment extended/shifted in the following direction: areas in the north west corner of the province were now included, reduction in central north, inclusion of small areas in north east, addition based on new townships in central parts of the province (Diepsloot), expansion in non-urban and townships in the central east (Benoni), reduction in the southeastern and southwestern parts.

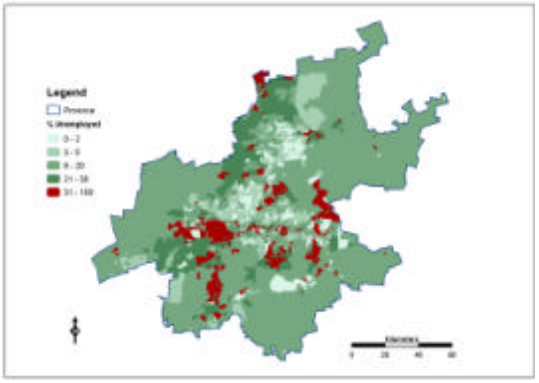


Figure 4. Percentage of unemployed people (2001) in Gauteng per sub place

In 2001 high levels of unemployment had expanded quite extensively as shown in Figure 4. In comparison to 1991 the areas of high unemployment had increased to 588 places (27%) and covers a wider area of the province. The expansion includes more areas in Soweto, the south east (Ekurhuleni), the east west corridor between Benoni and Kagiso, Kempton Park and east of Pretoria.

The core area of Gauteng in terms of low unemployment is visible in Figure 4. The highest unemployment (periphery) in the province occurs in a north south corridor from Sharpeville to Kagiso and from Kagiso eastward to Benoni. Other areas of high unemployment are mostly outside the metropolitan areas of Pretoria and Johannesburg.

A decrease in the number of areas was recorded that had an extremely high unemployment between 1991 and 1996, while in 1996 to 2001 there was an increase. This means that unemployment intensified in the latter period. Several socio-economic factors could be at play here – namely increased migration into Gauteng or natural population growth.

3.2.2 Data comparison between years

To calculate the difference between unemployment rates in 1991, 1996 and 2001 the percentage of unemployment of the previous year was subtracted from the statistics of the subsequent year. A negative result therefore indicated a decrease in unemployment while a positive result indicated an increase. In a number of cases (n=35) the unemployment value for 2001 was missing and it was therefore not possible to calculate the change for these records. Table 1 indicates the descriptive statistics for the change in unemployment.

	N	Minimum	Maximum	Mean	Std. Deviation
% Unemployment change between 1991 and 1996	2215	-69.91	48.29	1.52	9.21
% Unemployment change between 1996 and 2001	2180	-40.83	90.9	4.86	9.19

Table 1: Descriptive statistics for the change in unemployment.

Between the 1991 and 1996 census 55% of all polygons showed an increase in employment levels. The same figure between the 1996 and 2001 census was 76%. The spatial representation of these changes are shown in Figure 5(a) and (b).

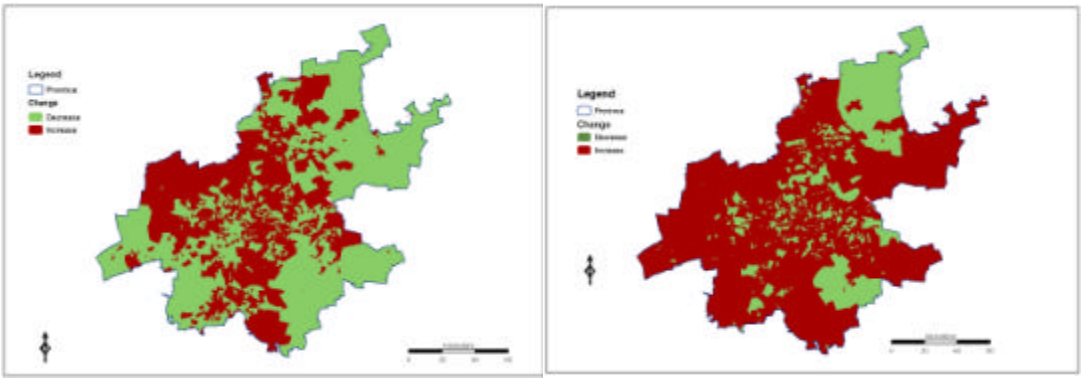


Figure 5 (a) and (b). Absolute change in unemployment rate between 1991 and 1996, 1996 and 2001.

An increase of 20% in the number of areas that experienced an increase in unemployment is quite significant. If this trend continues one would expect that the during the next five year period (after Census 2001) almost all places in Gauteng would show an increase in unemployment.

A statistical analysis of the socio-economic data from the different years rendered the following result.

In 1991 the strongest correlation of unemployment was with African females, African people in general, highest education level Grade 12, highest education level Grade 10 and African males. It is interesting to note that there was a higher correlation among young working age groups (e.g. younger than 40) and unemployment. A high correlation among primary and secondary education with unemployment was recorded, but the correlation drops significantly as the highest level of education becomes tertiary education. In conclusion it was predominantly African people and people without completed schooling that was unemployed.

Table 2: Pearson's correlation of unemployment in 1991 with other socio-economic variables

		African females
Number of unemployed	Pearson Correlation	.960**
	Sig. (2-tailed)	.000
	N	2213
		Africans
Number of unemployed	Pearson Correlation	.953**
	Sig. (2-tailed)	.000
	N	2213
		Highest education Grade 12
Number of unemployed	Pearson Correlation	.945**
	Sig. (2-tailed)	.000
	N	2213
		Highest education Grade 10
Number of unemployed	Pearson Correlation	.945**
	Sig. (2-tailed)	.000
	N	2213
		African males
Number of unemployed	Pearson Correlation	.921**
	Sig. (2-tailed)	.000
	N	2213

**Correlation is significant at the 0.01 level (2-tailed).

In 1996 the variables with the highest correlation to unemployment were education levels of Grade 8, 6, 9, 11 and Africans. Table 3 displays the result of Pearson's correlation for these variables. A high correlation was recorded among young working age groups (e.g. younger than 40) and unemployment. High values of correlation also occurred among those with secondary education as highest education level (Grades 8, 9 and 11).

Table 3: Pearson's correlation of unemployment in 1996 with other socio-economic variables

		Highest education Grade 8
Number of unemployed	Pearson Correlation	.974**
	Sig. (2-tailed)	.000
	N	2215
		Africans
Number of unemployed	Pearson Correlation	.970**
	Sig. (2-tailed)	.000
	N	2215
		Highest education Grade 6

Number of unemployed	Pearson Correlation	.969**
	Sig. (2-tailed)	.000
	N	2215
		Highest education Grade 9
Number of unemployed	Pearson Correlation	.965**
	Sig. (2-tailed)	.000
	N	2215
		Highest education Grade 11
Number of unemployed	Pearson Correlation	.959**
	Sig. (2-tailed)	.000
	N	2215

**Correlation is significant at the 0.01 level (2-tailed).

In 2001 the socio-economic variables which correlated highly with unemployment were, Africans, highest education Grade 7 and 11, males and females (see Table 4). Grade 7 is part of primary education. It was also interesting to note an increase in age related to unemployment. For example in 1991 there was no age group with a value of higher than $r^2 = 0.8$ to unemployment. In 1996 the age groups 15-19, 30-34, 25-29 and 35-39 all recorded a correlation of more than 0.8 to unemployment. In 2001 all the above age groups as well as the 20-24 and 40-44 year olds recorded an r^2 value of more than 0.80. It would therefore seem that an increase in the real age of age groups related to unemployment is taking place.

Table 4: Pearson's correlation of unemployment in 2001 with other socio-economic variables

		Africans
Number of unemployed	Pearson Correlation	.974**
	Sig. (2-tailed)	.000
	N	2222
		Highest education Grade 7
Number of unemployed	Pearson Correlation	.971**
	Sig. (2-tailed)	.000
	N	2222
		Highest education Grade 11
Number of unemployed	Pearson Correlation	.971**
	Sig. (2-tailed)	.000
	N	2222
		Males
Number of unemployed	Pearson Correlation	.955**
	Sig. (2-tailed)	.000
	N	2222

		Females
Number of unemployed	Pearson Correlation	.941**
	Sig. (2-tailed)	.000
	N	2222

**Correlation is significant at the 0.01 level (2-tailed).

From the above comparison, the variable representing Africans occurred in each analysis. The other variable that is related to unemployment is level of education. The lower the level of education completed, the higher the relation to unemployment.

4. Analysis of secondary and supportive data

An analysis of supportive data was done to aid the understanding of unemployment. The first data set analysed, was the Gross Geographic Product (GGP) for 1996 and 2001. This data is presented at the spatial level of magisterial district (1996) and municipality (2001) and is therefore firstly difficult to compare and secondly too generalised obtain specific answers with regards to unemployment at a sub place name level. It is however a useful source of background information.

The 1996 GGP data was spatially overlayed with the 1996 unemployment data and an analysis rendered the following insights. Several magisterial districts had a majority of sub places with an unemployment figure of more than 30%. These districts were Alberton, Brakpan, Nigel and Soweto. The first three districts are located in the east of the province, while the latter is the former township south of Johannesburg. The dominant GGP sector in the three districts in the east of Gauteng, was manufacturing (ranging from 28% - 53% of the total GGP). In Soweto the dominant GGP sector was government services (31% of total). Soweto was also the only district which recorded a negative economic growth between 1991 and 1996. The rate was –0.26.

One other magisterial district, Westonaria, recorded comparatively low economic growth between 1991 and 1996 (0.04). This district is located in the south west of the province and was heavily dependent on mining (77% of total GGP). A number of magisterial districts were highly dependent on one economic activity as part of their GGP. Besides Westonaria, these included: Oberholzer (mining 89%) and Springs (manufacturing 63%). The first two districts are located in the south west of the province.

The high concentration of unemployment went hand in hand with dominance in manufacturing as the primary contributor to the GGP. Trade was the second highest contributor to GGP in these districts in1996.

In 2001 GGP was presented at the municipality level. This spatial level is a more aggregate level than magisterial district, but became the accepted form of spatial representation at the local government level. The smaller, underlying 1996 magisterial district boundaries could be used to add up to the 2001 municipality boundaries.

High unemployment was recorded in the majority of sub places in the Ekurhuleni metro (previously the magisterial districts of Germiston, Alberton, Kempton Park, Boksburg, Brakpan, Benoni and Springs). In comparison the sub places in Brakpan, Benoni and Alberton recorded high values of unemployment in 1996 as well. In 2001 manufacturing was still the dominant contributor to GGP, but at a lower percentage (e.g. compare an average 49% then to 30% now). It is also interesting to note that community services (which included public services in 2001) had exceeded trade as the second highest contributor to GGP in these municipalities.

The metro of Johannesburg (consisting of Soweto. Johannesburg and Randburg from 1996) also recorded high levels of unemployment in the majority of sub places. Almost 80% of all sub places in this municipality had an unemployment figure of more than 30%. The dominant economic sector was finance (28%), followed by community services (25%). In 1996 the dominant economic sector in Johannesburg and Randburg was trade, while in Soweto government services recorded as the highest economic sector.

The Tshwane metro (constituting of Pretoria and part of Wonderboom) and the West Rand were the only municipalities where low unemployment was recorded in the majority of sub places (i.e. 52% of sub places had an unemployment rate of less than 6% in Tshwane). The dominant economic sectors in these municipalities were community services (40%) and trade (33%).

Based on the above analysis, high unemployment was recorded in municipalities where manufacturing and financial services were the dominant economic sectors. Low unemployment was recorded where community services and trade were dominant.

5. Final statistical analysis and forecasting

The ultimate aim of this work is to create a model which will enable the forecasting of unemployment at a localised level (like sub places). The diffusion of unemployment over the Gauteng landscape was clear in Figures 5a and b. It will be useful to determine what were the contributing factors in the expansion of unemployment over time.

A spectral analysis will analyse the combination of spatial and time factors contributing to the phenomena of unemployment (Robinson 1998). Therefore the next step will be to create an analysis for each time series represented in the data (e.g. 1991, 1996 and 2001) at the Gauteng locations. Once an understanding has been gathered about the change in unemployment patterns over time, this knowledge will be used to predict what the future occurrence of unemployment will be.

An interim analysis based on linear regression was applied to the various time series in the data. The 1991 data yielded the following result:

An r^2- value of 0.9 was recorded after the first model using a stepwise linear regression. An r^2- value of 0.999 was realised after 59 iterations. The variables best predicting unemployment in 1991 were: people employed in other economic sectors (which were not pre-specified), average number of inhabitants per household, number of households with an unspecified income, number of households with unspecified ownership and number of unspecified dwellings. One hundred and ninety two variables were used as input to the model.

The same analysis was done for 1996 data and yielded the following predictors: number of people with population group unspecified, number of people with derived household income of R360 001 or more per annum, the total number of Indian/Asian people, number of people with derived household income not applicable – institution or hostel and number of people with occupation skilled agricultural and fishery workers. An r^2- value of 1.00 was realised after the first model was completed. One hundred and eleven variables were used as input for the model.

In 2001 the predictors for unemployment were: number of people who are not economically active, total population, number of people who are employed, number of people aged 5-9 years and number of people aged 75-79 years. An r^2- value of 1.00 was reached after the 15th model was completed. A total of 89 variables were used as input to the model.

When comparing the above predictors, it is clear that there is no pattern between predictors for the various years. It is therefore necessary to analyse each time series independently in order to obtain a complete understanding of unemployment.

6. Conclusion

Based on the recorded global increase in unemployment, this article aimed to build a GIS model of unemployment at a detailed level in the South African context. This model would serve as a tool not only to predict future occurrences of unemployment, but also to intervene to reduce unemployment.

The findings highlighted that unemployment is a complicated issue and requires the interpretation of various factors that influences it. In this article no analysis of macro-economic issues (like policy and foreign direct investment) were investigated. These have, however, an influence on economic health, but time constraints put a limit to including these in the analysis. It would therefore be beneficial to exploit the findings made here in the broader economic context of the country.

Comparison of unemployment statistics over time indicated that unemployment has increased (in real terms) and also in spatial extent. The expectation would therefore be that this pattern will continue in future. The significant part of the spatial expansion of unemployment, is that 75% of sub places in Gauteng experienced an increase between 1996 and 2001. It would therefore be fair to expect that this percentage will increase in future.

Further results showed that unemployment correlated with Africans and lower levels of education, that a dominance of the manufacturing sector and unemployment coincided and that different predictors were used to forecast unemployment at different time frames.

Although considerable progress has been made in understanding unemployment, no conclusive model has been formulated. The next steps would be to calibrate the unemployment models developed through 1991 to 2001 to the recorded statistics. This will verify the accuracy of the model and the reliability with which it can be used. Hereafter the GIS modelling of unemployment should be attainable.

References

Arrufat, J.L., Diaz Cafferata A.M., Figueras, A.J. 1998. Unit-roots in spatial unemployment in Argentina. Testing in the presence of structural breaks.

Kingdon, G. and Knight, J. 2003. Unemployment in South Africa: a microeconomic approach. Available at: http://www.csae.ox.ac.uk/resprogs/usam/default.html

Martin, D.J. and Gascoigne, R. 1994. Change and change again: geographical implications for intercensal analysis. *Area 26*: 133-141.

Openshaw, S. 1984. The modifiable area unit problem. Concepts and Techniques in *Modern Geography* 38. Norwich: GeoBooks.

Robinson, G.M. 1998. Methods and techniques in human geography. John Wiley:and Sons, Chichester.

Stats SA. 1996. October Household Survey (OHS). Pretoria: Statistics South Africa

Stats SA. 2000. Labour Force Survey (LFS). Pretoria: Statistics South Africa.

Suedekum. J. 2003. Increasing returns and spatial unemployment disparities. Paper presented at European Regional Science Association (ERSA). (www.geocomputation.org). Accessed December 2003.

Sunday Times, 2003. Summit vows to halve jobless rate, June 6. Johannesburg: Times Media.