Using Election Registration Data as Proxy for Measuring Population Migration

Johan Maritz*, Pieter Kok**

* Council for Scientific and Industrial Research
**Independent Researcher (retired from the Human Sciences Research Council in 2007)

Abstract. Migration is an issue that remains critical for national as well as regional policy agendas and government planning. Over time migration, in addition to population growth or decline, changes the demographic composition of towns, cities and regions which in turn requires adjustments to service and infrastructure provision. The development of suitable policy responses requires reliable, comparable and timely information which in itself presents a problem as migration-specific surveys at national scale do not occur frequently in South Africa. The most obvious sources of migration data used to be the national census (held every 10 years or so), as well as household and labour surveys (*other surveys mostly extend to particular parts of South Africa*). Although socio-economic data has increased, it has not dealt well with migration. A recent research project conducted at the CSIR entitled the Integrated Planning Development and Modelling (IPDM) project explored the use of voter registration information as an alternative source from which migration proxy data can be extracted. Anonomised voter registration data was provided by the Independent Electoral Commission of South Africa for several consecutive elections covering a 10-year period. The data, once spatialised (and related to voting districts), could then be processed to extract detail movements between different election periods. The results were extremely valuable to identify spatial and other migration trends over various time periods. This paper describes the process applied as well as the initial analyses conducted investigating migration pertaining to South Africa's former homeland territories.

Keywords: migration, spatial processing, GIS, election-registration data.

1. Introduction

Demography is not static, and population figures, distribution and composition change over time and space (Waugh 2000). This is especially prevalent in developing countries including South Africa where urbanisation through migration is still a strong current reality for the foreseeable future affecting the South African settlement landscape (Kok & Collinson 2006). Migration is therefore also an issue that remains critical for South Africa's national as well as regional policy agendas and government planning. Over time migration, in addition to population growth or decline, changes the demographic composition of towns, cities and regions, which in turn requires adjustments to service and infrastructure provision. The development of suitable policy and planning responses requires reliable, comparable and timely information which in itself presents a problem as migration-specific surveys at national scale do not occur frequently in South Africa. According to Polzer (2010:2), data on migration within and into South Africa is at times "poorly collected, weakly analysed, and often misleading". Official data (e.g., the 2001 national census, the 2007 community survey.) does not reliably capture movements between provinces, within them, or within municipalities (Polzer 2010). Migration has many dimensions and therefore cannot be properly analysed without taking into account spatial, economic and social factors – to name but a few.

This paper will not delve into the causes of migration, the focus will instead be on harnessing an alternative data source that can be spatialised and that can contribute to migration and demographic research conducted locally.

1.1. Defining Migration

Often when the term "migration" is mentioned, those who move between countries (international migrants) are considered first, however most migration happens within country borders (Skeldon 2008). In order to conduct comparative and trend analysis it is crucial to define migration clearly. Some concepts frequently emerge in migration studies: these relate to the *origins and destinations* of those migrating. *Table 1* attempts to differentiate between the various basic forms of migration. A second variable is the *duration of migration*. The time-duration also varies between different forms of migration, ranging from weeks for seasonal workers to years for others. For some moving is only temporary while others perceive it as long term or as permanent. The third variable is that of the *spatial extent*. Moving a few street blocks might not be considered by some as migrating as employment and the social facilities and services used might remain largely unchanged. Some definitions also refer particularly to a change of or establishment of a residence (Anon 2008). According to Kok (Kok & Collinson 2006) at its most basic form migration should be defined as *the crossing of the boundary of a predefined spatial unit by persons involved in a change of residence.*

Permanent/definitive	External (international)
	• Voluntary • Forced (refugees)
	Internal
	• Rural – depopulation/ Urbanisation • Urban depopulation • Regional
Semi-permanent	Years
Circular (temporary)	Years/months
Seasonal	Several months/weeks

Table 1: Forms of migration (modified from Waugh 2000).

The definition of migration must also fit the local context – temporary circular migration is a particular form of migration apparent in South Africa where the apartheid system created areas of *forced* settlement in the former homeland territories. Those seeking employment or education opportunities relocate temporarily to urban areas whilst remaining connected to the sending household often sending remittances (Rwelamira & Kirsten 2003). For the purpose of the work conducted and described in this paper migration is described as *the process that results in an individual or household relocating to establish or re-establish residence in a different spatial area (defined here as a voting district) than the prior residence*. The following briefly addresses the associated data challenges that face analysts locally.

1.2. Migration Data Challenges

Statistical data collected nationally does not always suffice when migration is analysed. The recent South African National Development Plan even indicated that data on migration into and within South Africa is poorly collected, weakly analysed and often deceptive. The plan noted that municipalities were often unable to respond effectively because they did not have sufficient data or the necessary skills to make sense of the data they have (Steyn 2013). Skeldon (1990) pointed out that migration analyses should preferably not be restricted by what the available data has to offer, but the reality is that an analysis of migration inevitably has to fit in with the available data, which is mostly census-based. Here the importance of defining migration is re-affirmed otherwise it becomes mere data driven conceptualisations (Kok & Collinson 2006).

1.3. Spatial Unit Used

When considering population movements the spatial extent becomes important as this, in combination with the definition of migration, determines where migration occurs. Migration can only be measured if an administrative or geographical unit for the person or household changes. Often administrative units, defined mostly by bureaucrats or politicians, are used to record migration although they are not always the most suitable units. Afterwards analysts attempt to rationalise such units for demographic analysis (Standing 1984). Information is sometimes also not made available at the more disaggregate level at which it was recorded, but at a more aggregate

level using larger spatial units. Some analysts such as Skeldon (1990) and Standing (1984) have in the past reported on the practical problem associated with migration analysis when working at more aggregate levels and with zone types not always suitable for such analyses. There is also the danger that such limitations can miss out on some migration occurring or even distort observations.

A related problem when dealing with the spatial units of analysis is the varying sizes of the analysis units defined by territorial demarcations which tend to cause arbitrary zone-size distortions of geo-statistical indicators and comparisons (Presidency, 2006). This makes portraying and "reading" information challenging and can even lead to a misinterpretation of the geo-spatial information.

2. The South African Case

In the past one of the main sources of migration data was Statistics South Africa (Stats SA). Such data was collected at the enumerator area(EA)[1] level and made available at different aggregate spatial levels. *Figure* 1 provides a diagrammatic layout of the spatial units used for statistical and election-registration data analysis. The 1996 census provided migration data at an EA level as well as aggregated to magisterial districts and so-called main places (which are named places), but following the 2001 census migration data was provided only at aggregated levels. Comparison of data over time becomes problematic when changes occur in the census questions related to migration (Kok & Collinson 2006). The latest census (2011)[2] will be also be made available at the Small Area Layer (SAL), which is largely at an EA level, with some sparsely populated EAS aggregated into larger zones. In addition the EA demarcation between different censuses differ spatially, which constrains time-series analysis between inter-census periods. Apart from the Stats SA census data there are also other data that contains migration information; these include Household and Labour Force surveys which are conducted regularly. Institutions such as the University of the Witwatersrand (for example) also collect long-term demographic information – at the Agincourt site[3] in the Mpumalanga province. Some municipalities and metros also collect such information but in some cases though they do not have the capacity and resources to appropriately collect, manage and analyse such data.

---

[1] An EA is the smallest geographical unit (piece of land) into which the country is divided for census enumeration purposes, is of a size that can be enumerated by one census fieldworker *(enumerator)* in the allocated period for the census. EAs typically contain between 100 and 250 households (Statistics South Africa 2001).

[2] Not available at time of writing.

[3] A site in the Bushbuckridge area in Mpumalanga.

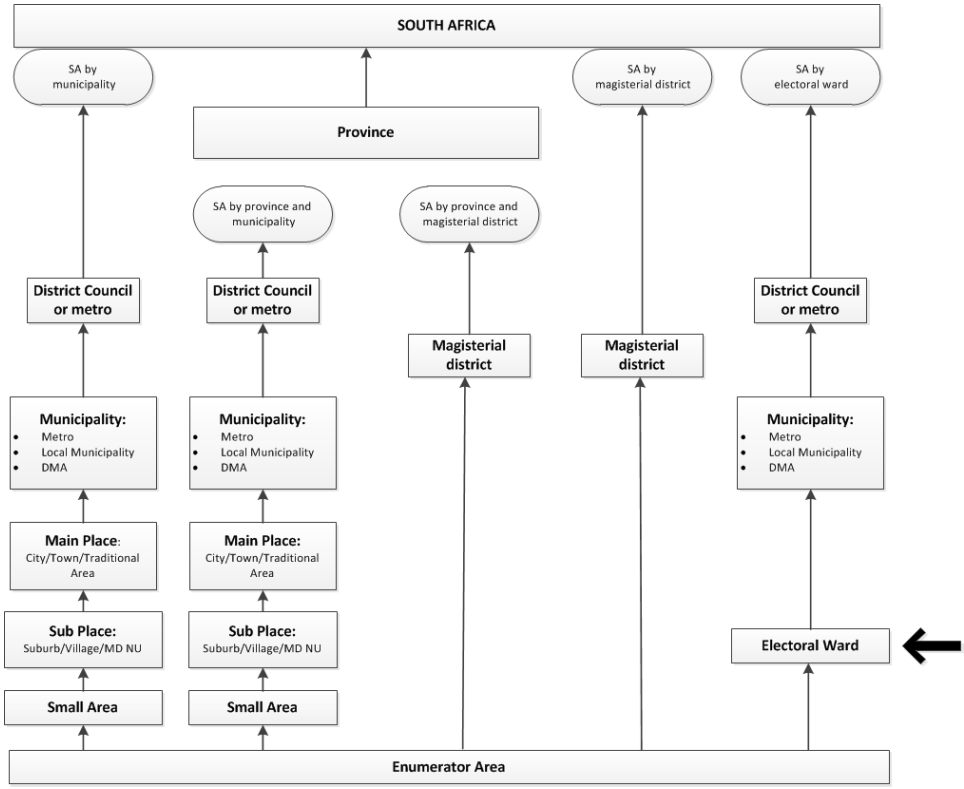

Figure 1: Spatial units used in SA from (Grobbelaar 2005: 2).

2.1. The Need for Migration Information

In South Africa the national census is now conducted only once every ten years. This is a long time period when considering that the time period for local municipal planning is much shorter – Integrated Development Plans[4] (IDP) are 5 year plans. Municipalities indicated a need for inter-census data given the extent of settlement change they observed within their areas within the last decade. During 2008, the Department of Science and Technology (DST) commissioned the Council for Scientific and Industrial Research (CSIR) and the Human Sciences Research Council (HSRC) to develop an information and modelling platform, now known as the Spatial and Temporal Evidence for Planning in South Africa platform (STEPSA), to support integrated planning, development and service delivery in South Africa (STEPSA.org 2013). For the component of the project that focused on developing *regional spatial profiles,* a number of living lab sessions was initiated with three district and one local municipality – the purpose was through engagement, to determine what information municipalities require support with their analysis and planning. Migration information was identified by all as a key data layer given that by then – 2008 - the 2001-based information was already very much outdated.

[4] Strategic plans undertaken by local and district municipalities in South Africa.

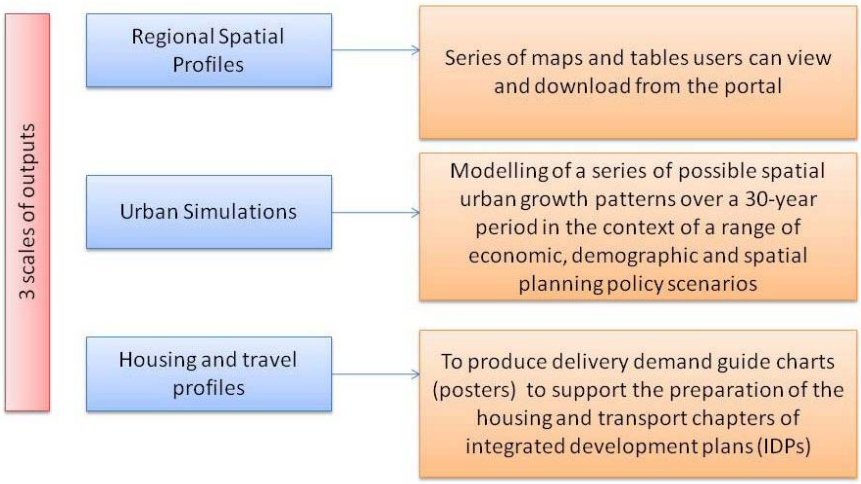

Figure 2: Components of the STEPSA project (STEPSA.org 2013).

In addition, the other components of the STEPSA project dealing with urban simulation and housing and travel profiles (*Figure 2*) also required disaggregated migration information. After an initial search for suitable alternative migration data sources it was decided to approach the Independent Electoral Commission (IEC) to explore voter registration data as a migration proxy data option.

2.2. Data from the Independent Electoral Commission (IEC)

The Independent Electoral Commission is South Africa's independent election management body and one of its obligations is to maintain a voters' roll. To be able to vote in South Africa, an eligible person must register his/her details in the voter's roll of the voting district where they reside. If a voter has not moved since the previous election his/her registration remains valid in that voting district. If however a voter has moved into a new voting district and wants to vote in the new voting district, he/she first has to re-register. Registration is also largely a measure to prevent voters from voting in multiple voting stations. Voting districts are geographical areas principally determined on the basis of geographical size and number of eligible voters. Urban voting districts generally contain approximately 3000 voters located within a radius of approximately 7.5 kilometres from the voting station. Rural voting districts accommodate approximately 1200 voters located within a radius of up to 10 kilometres of the voting station (IEC 2012). The importance given to elections since the end of Apartheid and the extent to which results are scrutinised, results in a voters' roll that is well maintained and accurate. As such it is viewed as a suitable and trustworthy dataset to track the movement of people (represented here by registered voters) over different time periods. The overall assumption though is that when people relocate that they would reregister at the voting district (VD) where they reside.

3. Methodology

The following section will briefly describe the process to obtain and process the IEC data to make it suitable for spatial- and time series analysis.

3.1. Sourcing IEC Voter Registration Data

Due to the sensitivity of voters' role information the IEC agreed to provide the unit record information in an anonomised fashion i.e. the identifiers of individuals namely the person's national identification number (ID) was replaced with an alternate number whilst retaining age and gender features in the data. The variables received from the IEC in 2010 were the following (Kok 2010):

- Variable 1: Anonymised person identifier, which is a unique number for every (unidentifiable) person in the data set
- Variable 2: Gender
- Variable 3: The two-digit year of birth obtained from the person's ID Number
- Variable 4: Four-digit birth year as obtained from the Department of Home Affairs (where available) to be used as a check
- Variable 5: The VD where the person was registered in 2009
- Variable 6: The VD where the person was registered in 2006 (if registered)
- Variable 7: The VD where the person was registered in 2004 (if registered)
- Variable 8: The VD where the person was registered in 2000 (if registered)
- Variable 9: The VD where the person was registered in 1999 (if registered)
- Variable 10: Whether the person voted in the local government elections in 2006 (if the person's participation information was received)
- Variable 11: Whether the person voted in the national election of 2009 (if the person's participation information was received)

Three data sets containing voter registration and behaviour data for the five elections between 1999 and 2009 was provided along with the number of registered voters in every VD for each of the five elections. For the 2009 national election 23 181 997 individual voters registered, which makes it a good dataset considering that the 2009 mid-year total population estimates for South Africa was 49 320 500 (Statistics South Africa, 2010). A user agreement stipulated that the registration data should not be provided to third parties without the consent of the IEC.

3.2. Processing the IEC information

The information was supplied for each of the preceding elections as indicated in *Table 2*. In addition to the tabular data, the voting districts were also supplied in GIS file format for each election.

Election year	Election type	Number of registered voters
1999	National/Provincial	18 168 072
2000	Municipal	18 476 516
2004	National/Provincial	20 674 926
2006	Municipal	21 054 957
2009	National/Provincial	23 181 997

Table 2: List of elections, type and number of registered voters (IEC 2012)

Within the IEC the Delimitation Directorate is responsible for delimiting the geographic area of South Africa into voting districts. Drawing the outer municipal boundaries is the responsibility of the Municipal Demarcation Board[5]. The IEC's voting districts do not have political significance (as wards do) but have been created for electoral efficiency and planning purposes. Before each election, the geospatial extent of voting districts in municipalities are inspected by municipal IEC representatives with a view to aligning the geography of voting districts in accordance with settlement, demographic and political changes that may have occurred since the previous election (EISA, 2002). This means however that voting districts between consecutive elections differ. Depending on the extent of settlement change the voting districts would change significantly or not at all.

In this process the last election period - 2009 - was selected as the base spatial unit and all prior election data would need to be related to it[6]. The 2009 voting district spatial layer became the target layer for all prior election periods. Using ArcMap GIS the proportional change needed to be calculated between the VD areas for the election periods 1999, 2000, 2004, and 2006 and related to the 2009 VD areas.

Figure 3 illustrates an example of the differences between the voting districts of two election periods – 1999 and 2009 VD areas. Crucially the former homeland territories[7] differed significantly from the rest of South Africa – only settlements were demarcated in 1999 as voting districts with large areas in-between excluded. These in-between spaces were also not free-standing but grouped into a single geospatial feature group.

[5] The Municipal Demarcation Board is an independent authority responsible for the determination of municipal boundaries (Municipal Demarcation Board 2013).

[6] More recently all information was to be related to the 2011 election – an identical process will be followed.

[7] "Former homelands" refer to self-governing territories for black African ethnic groups established under the Apartheid policy (Cahoon 2013).

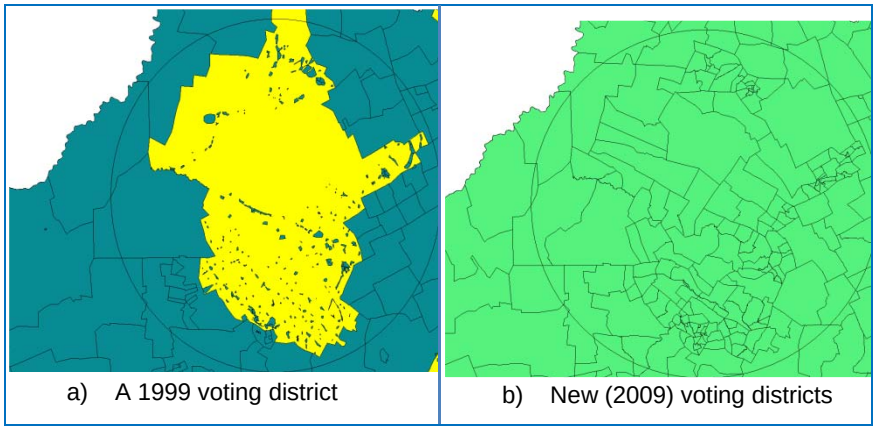

Figure 3: Comparison of two election period spatial units (Maritz 2012).

The second problem was that the boundaries of the older voting district datasets for 1999 and 2000 differed from the 2009 spatial data (*Figure 4*) resulting in overlaps and slivers. This can be attributed to less accurate, old boundary data for South Africa which had been captured at different scales from old topographic maps. Subsequently the South African coastline has been captured using more recent technologies such as high resolution satellite imagery and aerial photography.

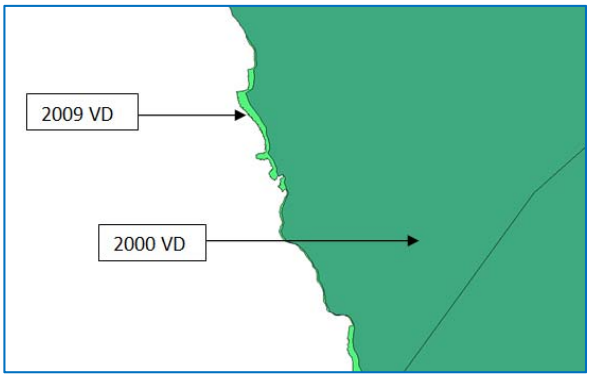

Figure 4: Comparison of spatial freatures from 2009 with 2000 (Maritz 2012).

All voting districts for elections prior to 2009 were combined individually with the 2009 voting districts by applying a union-analysis function to calculate a geometric intersection of all features. Using an area-based approach the proportional change of voting districts from the prior election (now part of the 2009 voting districts), could be calculated. The end results would indicate how much (percentage wise) of a prior voting district would 'shift' to a new voting districts, and thus also how many of the registered voters needed to be shifted as well. Assuming an even population spread this can then be translated to the allocation of registered voters between the two election periods. Using an area proportioning approach the 1999/2000/2004/2006 areas was apportioned to the 2009 VDs.

The next step was to determine the most likely 2009 VD for each person on the voters' roll at the time of the 1999, 2000, 2004 and 2006 elections.

This required a randomisation of the various eligible 2009 VDs. The spatial extent that each 2009 VD overlapped with the 1999/2000/-2004/2006 VD concerned was treated as a selection probability. The randomisation procedure that was adopted to determine the most appropriate 2009 VD for each person registered as a voter for the 1999 election. The randomised selection was based on the so-called tabled distribution. Where there was a one-to-one overlap no randomisation needed to take place. The same procedure was repeated for 2000, 2004 and 2006 to determine the most likely 2009 VD for the person concerned at the time of each election. In election years for which the person was not registered to vote, no allocation of 2009 VDs was made (Kok, 2010). These allocations provided the basis for the subsequent migration analyses.

3.3. Constrains and limitations

One of the obvious limitations of using voter registration data is that it only represents registered voters. It excludes those that are not eligible to vote such as foreigners and children (under 18 years of age). It also excludes those who simply do not vote and do not bother to register. The IEC migration data does not represent the entire population and as such it therefore does not replace other migration data such as those recorded though the censuses. It can be referred to as 'proxy' migration data because it does not necessarily reflect the actual migration situation in all cases. For example, when voters move their residence but fail to re-register in the voting district of their new place of residence they will be regarded as non-migrants. Although less likely, nothing also prevents voters from registering in another VD even if they do not reside there. When working at a more aggregate scale than voting district such issues are less important and the data becomes more reliable, since highly localised problems tend to be neutralised at higher spatial levels. Knowing the limitations of the IEC data is important if errors in analysis are to be avoided.

3.4. Validation of IEC migration data: A demographic analysis of migration levels and rates among registered voters in South Africa between 1999 and 2009

Once the IEC data was processed and spatialised the following analyses was undertaken to investigate the comparison of the IEC data with other data. The data provided by the IEC gives researchers a unique opportunity to study the migration levels[8] of South African registered voters at a spatially detailed level, the voting districts (VDs), and over two consecutive five-year periods (1999-2004 and 2004-2009) and one ten-year period (1999-2009). The tables below depict the migration levels (i.e. proportions of the total

[8] Migration level is also known as the "Crude Migration Intensity (CMI)" is calculated by expressing the total number of migrants (M) in a given time period as a percentage of the population at risk (P) such that CMI = 100M/P (UN Population Division 2013).

number of registered voters) in the various provinces who migrated between different municipalities (Table 3).

Province	1999-2004	2004-2009	1999-2009
Western Cape	8.7%	9.6%	18.6%
Eastern Cape	10.5%	9.5%	20.1%
Northern Cape	11.1%	8.8%	19.8%
Free State	7.5%	7.3%	14.9%
KwaZulu-Natal	9.9%	11.1%	21.1%
North West	12.0%	10.5%	22.5%
Gauteng	10.7%	11.2%	21.4%
Mpumalanga	9.7%	9.7%	19.6%
Limpopo	9.5%	7.3%	17.2%
SOUTH AFRICA	10.0%	9.9%	20.0%

Table 3: Inter-municipality migration levels over two consecutive five-year periods and the entire ten-year period for each of the nine provinces and South Africa in general, using the 2009 municipal boundaries as basis for the spatial analyses.

Table 3 confirms the observation, based on census data, that migration levels in South Africa have remained remarkably constant over time. Kok, O'Donovan, Bouare and Van Zyl (2003:54) found that "despite dramatic political, social and economic changes in the country (including the abolition of apartheid's migration-related measures such as influx control and group-area demarcations), the overall level of migration between the late 1970s and early 1990s did not change significantly". Kok and Collinson (2006:7) have since confirmed that the trends regarding migration levels which had been observed for the periods 1975-1980 and 1992-1996 by Kok et al (2003), mentioned above, were continued during the period 1996-2001. Looking at Table 3, it is clear that the general trend also remained very much the same for the five-year periods 1999-2004 and 2004-2009.

The proportion of migrants in the population at any particular age is called the migration rate for that age. Age-specific migration rates therefore reflect the age selectivity of migration. A migration analysis of the IEC data shows that the age-specific migration rates over the ten-year period 1999-2009 for the two sexes combined can be described reasonably well (with a goodness-of-fit R-squared value of 88%) by the following equation (for a fifth-order polynomial):

$$y = 8E\text{-}09x^5 - 2E\text{-}06x^4 + 0.0001x^3 - 0.0042x^2 + 0.0615x - 0.1546$$

However, from a demographic perspective it would be more appropriate to describe the observed migration rates in terms of the model migration schedules (MMSs) originally described by Rogers and Castro (1981) and first applied to South Africa by Hofmeyr (1988).

Migration studies in various countries (see, for example, Castro & Rogers 1983) have shown "a common age-dependent characteristic" of the type depicted in Figures 5-7, which indicate the "fundamental age pattern of migration with peaks occurring at infancy, young adulthood, and at retirement" (Hofmeyr, 1988:24). Castro and Rogers (1983) describe the use of model migration schedules (MMSs) to summarize any age profile of migration into a single equation, giving us the so-called Gross migraproduction rate (GMR). In Figure A the age profile of migration for the period 1992-1996, together with its associated MMS equation, is shown, where the absence of any significant retirement peak is noticeable, indicating an elongated retirement age, possibly due to a post-apartheid adjustment in labour participation patterns. The mathematical expression underlying the "full" or "basic" model migration schedule (containing 11 parameters) depicted in Figure 5 is given in Equation 1:

$$
\begin{aligned}
M(x) \quad = \quad & a_1 \exp(-\alpha_1 x) \\
& + a_2 \exp(-\alpha_2(x-\mu_2) - \exp(-\lambda_2(x-\mu_2))) \\
& + a_3 \exp(-\alpha_3(x-\mu_3) - \exp(-\lambda_3(x-\mu_3))) \\
& + c
\end{aligned}
\qquad [1]
$$

where x = 0, 1, 2, 3, ..., z (where z represents the upper open age interval, e.g. 85 years and older).

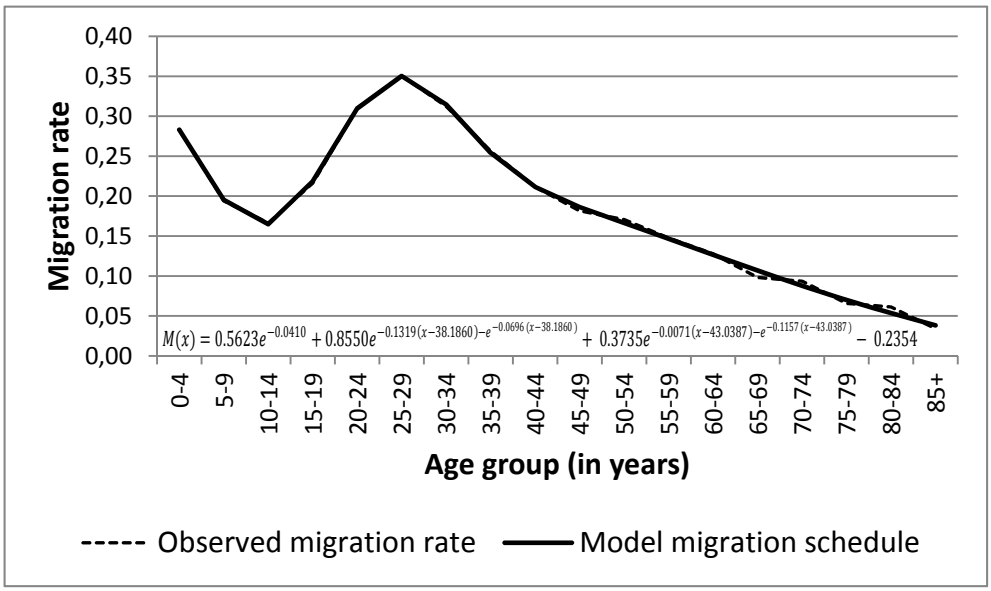

Figure 5: Age profile of inter-district migrants of both sexes combined during the period 1992-1996, with the equation of the relevant model migration schedule (MMS). [Source: derived from Statistics South Africa, Census '96 Migration Community Profile data).]

In Table 4 the findings from the various modelling exercises are shown for the parameters and variables discussed so far. The table also contains the findings in respect of the female and male populations represented in the IEC data and the 1996 census. The entries in Table 4 can be used to

construct model migration schedules for the observed migration patterns for the total, female and male populations during any of the three migration periods. In Table 4 the complete set of parameters and variables for the 1992-1996 period (from Stats SA census data) and the 1999-2004 and 2004-2009 periods (from the IEC data) is provided for reference purposes. The corresponding equations are as follows:

1992-1996: $M(x) = 0.5623 \exp(-0.0410\ x) + 0.8021 \exp(-0.1319\ (x-37.9186) - \exp(-0.0696\ (x-37.9186))) + 0.3853 \exp(-0.00911\ (x-43.0387) - \exp(-0.1157\ (x-43.0387))) - 0.2354$ [see Figure 5]

1999-2004: $M(x) = 0.1170 \exp(-0.0483\ (x-22.0979) - \exp(-0.5741\ (x-22.0979))) + 0.0413 \exp(-0.0301\ (x-52.8995) - \exp(-0.1542\ (x-52.8995))) + 0.0221$ [see Figure 6]

2004-2009: $M(x) = 0.1753 \exp(-0.0475\ (x-23.0069) - \exp(-0.3255\ (x-23.0069))) + 0.0540 \exp(-0.0386\ (x-55.0487) - \exp(-0.1616\ (x-55.0487))) + 0.00734$ [see Figure 7]

Parameters and variables	Migration period								
	1992-1996			1999-2004			2004-2006		
	Total	F	M*	Total	F	M	Total	F	M
GMR_f^*	3.28	3.16	3.47	--	--	--	--	--	--
GMR_r^{**}	--	--	--	4.41	4.16	4.83	4.58	4.22	5.09
E	0.29%	0.42%	0.53%	0.45%	0.51%	0.56%	0.46%	0.43%	0.48%
a_1	0.56	0.64	0.56	--	--	--	--	--	--
α_1	0.04	0.04	0.04	--	--	--	--	--	--
a_2	0.80	0.63	0.85	0.12	0.15	0.09	0.17	0.21	0.14
α_2	0.13	0.14	0.13	0.05	0.05	0.04	0.05	0.05	0.04
λ_2	0.07	0.06	0.07	0.57	0.43	0.87	0.33	0.27	0.42
μ_2	37.92	41.41	38.19	22.10	22.18	22.07	23.01	23.20	22.82
a_3	0.38	0.41	0.37	0.04	0.03	0.07	0.05	0.04	0.08
α_3	0.009	0.005	0.007	0.03	0.006	0.14	0.04	0.01	0.13
λ_3	0.12	0.13	0.12	0.15	0.16	0.12	0.16	0.18	0.18
μ_3	43.04	42.22	43.04	52.90	49.01	64.23	55.05	50.49	62.00
c	-0.23	0.32	-0.23	0.02	0.01	0.04	0.01	-0.005	0.02
δ_{12}	0.70	1.02	0.66	--	--	--	--	--	--
β_{12}	0.31	0.25	0.31	--	--	--	--	--	--
$x_{l(ow)}$	12	12	12	--	--	--	--	--	--
$x_{h(igh)}$	27	27	27	28	27	28	28	29	28
$x_{r(etirement)}$	--	--	--	58	54	64	61	61	62
x_{shift}	15	15	15	--	--	--	--	--	--
A	30	30	30	--	--	--	--	--	--
B	0.18	0.16	0.21	--	--	--	--	--	--

Table 4: Parameters and variables defining the observed migration schedules for the periods 1992-1996, 1999-2004 and 2004-2009: Total, female and male populations.

* Please note: The model for male migrants during the period 1992-1996 did not converge fully.

* GMR_f = Gross migraproduction rate for the "full" model

** GMR_r = Gross migraproduction rate for the "reduced" model (without ages 0, 1, 2, ..., 19 years)

When one looks at the columns in Table 4 that are based on the IEC data (for the migration periods 1999-2004 and 2004-2009) it is clear that the goodness-of-fit indices, E, are quite small (ranging from 0.43% to 0.56%), indicating that the IEC data provides a good model fit even though it does not cover the entire age range. A number of important conclusions can be drawn from Table 4. The first relates to the observation that the index of child dependency, δ_{12}, is notably higher for the female population (1.02) than for the total (0.70) and male (0.66) populations. These relatively high values, especially for women, show a pronounced child dependency. A second conclusion is that the low values for the index of parental-shift regularity, β_{12}, which range between 0.25 and 0.31 and are therefore quite far from unity as would normally be expected, probably indicate problems with the 1996 census data on child migration. Thirdly, the GMRs are relatively small and consistent (ranging from 3.16 to 3.47), which seem to confirm the viability of model migration schedules (MMSs) of this nature for South Africa.

In the rest of this section only the age bracket 20+ years will be used in the migration analyses. (Since the IEC data is relevant only for persons aged 18 years or older a first age group of 15-19 years would be incomplete – no persons aged 15-17 years – hence the age-profile descriptions given here are restricted to the ages 20+ years.) In Figure 6 the age profile of migrant voters for the period 1999-2004 and its associated MMS equation are shown, and to complete the picture the equivalent graph and equation for the subsequent period 2004-2009 are shown in Figure 7.

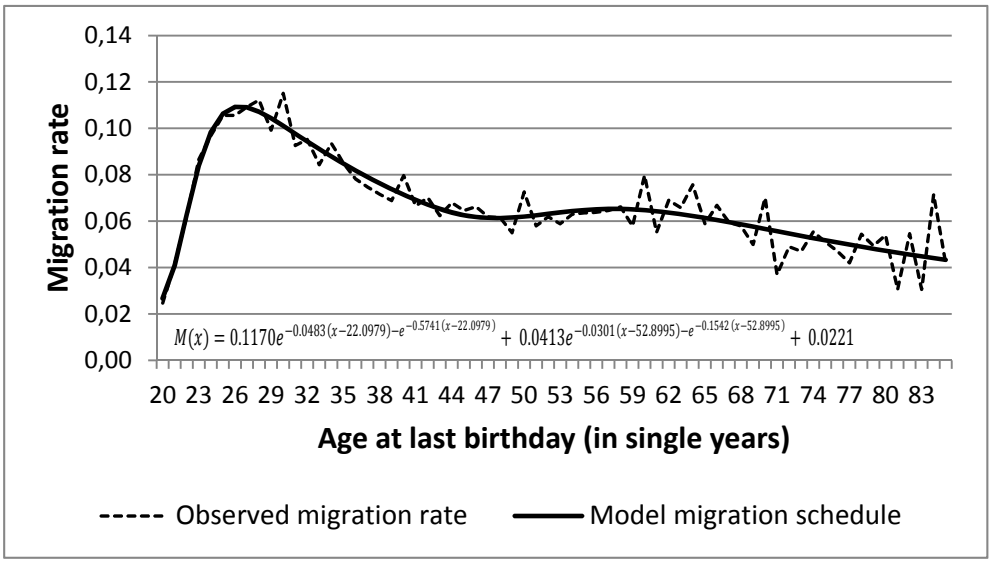

Figure 6: Age profile of inter-municipality migrant voters of both sexes during the period 1999-2004, with the equation of the relevant model migration schedule (MMS). [Source: derived from the IEC voters roll 2009.]

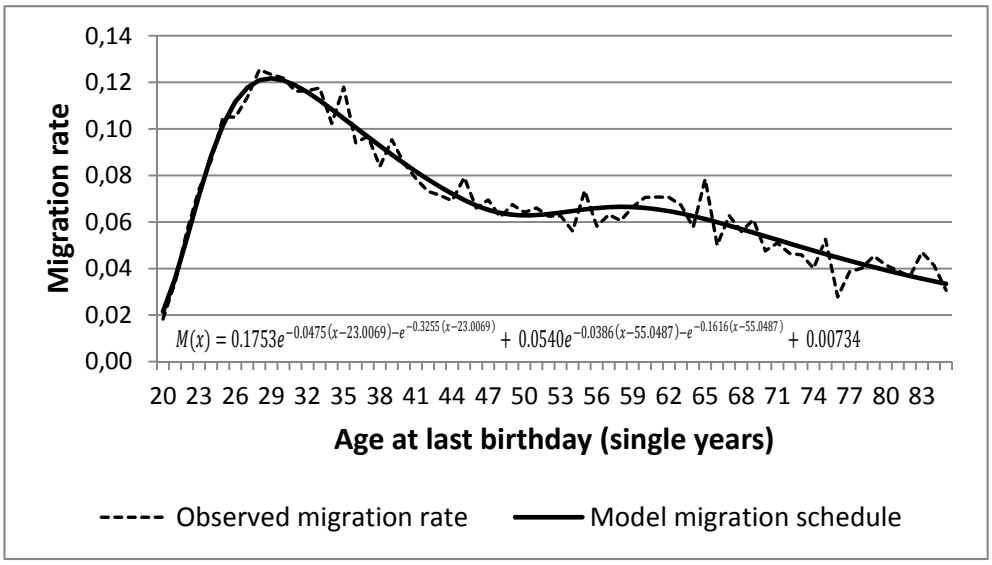

Figure 7: Age profile of inter-municipality migrant voters of both sexes during the period 2004-2009, with the equation of the relevant model migration schedule (MMS). [Source: derived from the IEC voters roll 2009.]

It is clear that the profiles for the two consecutive periods 1999-2004 and 2004-2009 are quite similar, showing that the age selectivity of migration in South Africa remains consistent over time: people in the early labour-force years, i.e., aged 24-35, are most migratory, and there is a slight retirement peak around age 60 years.

3.5. Visualisation of migration data

Once the voting district data was adjusted and randomised the resulting data tables for each election period was summarised as the data still reflected individual records. Summarising the information meant that flow between an origin and destination zone could be summed in ArcMap GIS in order to determine extent of FROM voting districts and the TO voting districts. This process had to be repeated to determine FROM and TO flows at a local municipal level. Once calculated the tables could be linked to the 2009 local municipality spatial file. Mapping migration from and to spatial zones was not sufficient as users wanted to spatially see where flows occurred thus being able to observe both the origin and destination of the flows. Upon investigating methods to map flows it was decided to create flow lines between origin and destination, much like creating a spider diagram.

Flow data model tools is a toolset developed by Alan Glennon (Department of Geography, University of California) and it consists of several Visual Basic for Applications (VBA) macros designed to work in ArcGIS. The toolset (*Figure 8*) was created to perform two functions: to integrate the functionality of *Waldo Tobler's FlowMapper into ArcGIS*, and to allow import and export of data into the Flow Data Model (Glennon 2008).

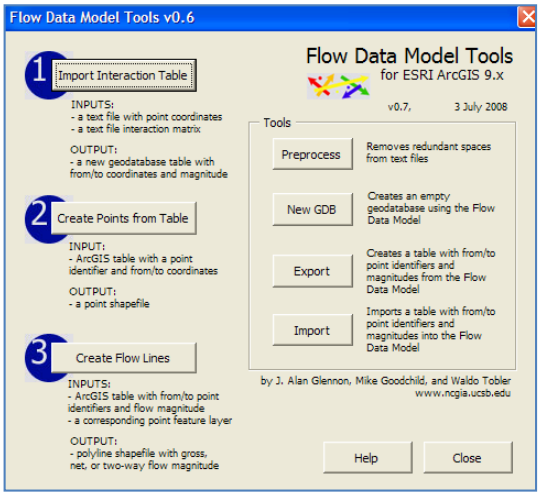

Figure 8: Flow data model tool (Glennon 2008).

To import the flow the data has to be in a specific format as illustrated in the following table namely:

FIELD	DESCRIPTION
InputnodeID	Refers to origin ID local municipality
X	Origin x-coordinate
Y	Origin y-coordinate
OutputnodeID	Refers to destination ID of local municipality
X2	Destination x-coordinate
Y2	Destination y-coordinate
Magnitude	Magnitude of flow (voters)

Table 5: Fields for importing flowdata (Glennon 2008).

Once the file was imported the tool can create flow lines between origin and destination flows. Three types of flows can be created namely two-way flows which create flow lines in both directions, net flows which cancel out flows of the same volume, and lastly lines that add all flows to create gross flows. The tool creates lines between all origin and destination pairs with a flow value. This results in a large number of lines all over the map creating an unreadable image. The user has to select the values to illustrate; for example showing only the highest flows.

4. Mapped results and initial application

Although the IEC data in non-spatial form also contributes significantly to migration analysis, for the authors it is especially valuable to apply the information spatially. Often in the South African context decision makers are overloaded with documents and reports which they might not even read. Mechanisms to display or summarise information visually are favoured as a means to easily and quickly inform. This was one of the considerations

when the STEPSA portal (http://stepsa.org) was developed. A number of relevant map themes including migration, are provided through an online map viewer to support local and regional planners on the website.

Apart from using the information at voting district unit level the data was also extracted at a more aggregate level - that of local municipality. Using the flow data model tool flow lines were created indicating all flows for all municipalities for the period 1999-2009. Using all flows however is messy and unreadable spatially, therefore only the main flows (flows larger than 800) were extracted to indicate where the more substantial migration occurred. *Figure 9* illustrates the result – inter-municipal migration flows for South Africa for the period 1999-2009.

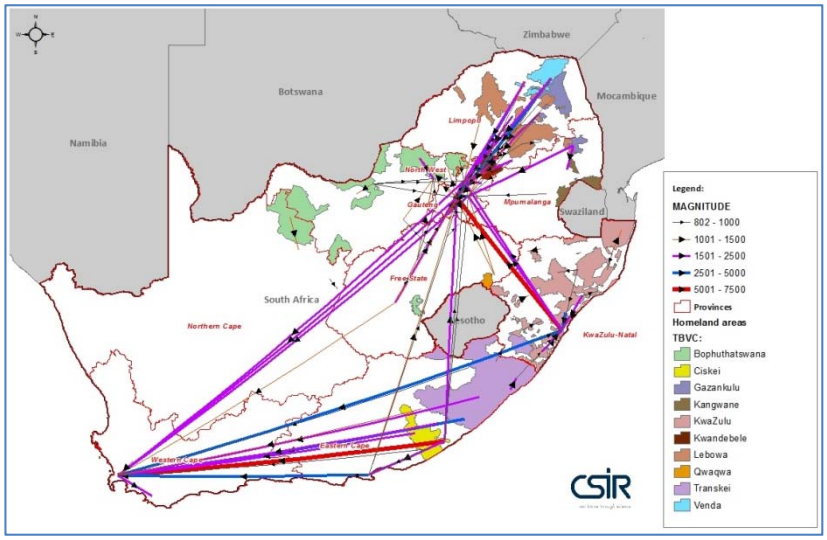

Figure 9: Main flows considering ex-homeland territories (Maritz 2012).

Using this information combined with the location of former homeland territories it is possible to observe the main migration trends that relate to these areas. Given the lack of development and high unemployment within these territories, there is an expectation that outmigration is occurring to larger centres. When analysing the visualised inter-municipal flows it is interesting to observe that there are definite net flows from areas in the Eastern Cape to the Western Cape – specifically Cape Town – as well as to Gauteng – especially southern Johannesburg. Flows from the northern Limpopo province (Vhembe district) have also taken place to Gauteng (Johannesburg and Ekurhuleni).

At the same time it is interesting to observe substantial flows occurring between South Africa's metropolitan areas, especially the Gauteng metros and Cape Town. Naturally this does not provide the full picture as smaller flows have not been indicated. The reasons for such movements can only be clarified when also considering other data. Reading the spatial information a conclusion can be drawn that, when considering migration, residents of the ex-homeland territories have shown strong trends of moving to larger ur-

ban centres. Follow-up studies will take the results and observation drawn from this data spatialisation exercise and explain the trends in more detail.

5. Conclusions

Although South Africa is in a better situation than many other developing countries it still requires more frequent and finer scale information especially for local planning and policy development. Migration is one of the key issues identified by users (local and district municipalities). Given the low frequency of migration data collected nationally the IEC's voting district-based data holds much potential as a migration proxy dataset. A key advantage is that the information is captured and related to voting districts which in themselves are relatively small spatial units. Comparisons of IEC data with other migration information have shown that it is a viable complementary data source. The IEC data does have limitations – largely due to human behaviour, however the effects of this potential problem can for the most part be overcome by the aggregation of data and by combining the analysis with other socio-demographic information.

There are empirical regularities that characterise observed migration schedules. As Rogers and Castro (1981) point out, these regularities are no less important than the corresponding and well-established regularities in observed fertility or mortality schedules. However, Rogers and Castro (1981:45) correctly ask, "Of what specific use is the model migration schedule that has been described in this study? What are some of its concrete practical applications?"

> "The model migration schedule may be used to graduate observed data, thereby smoothing out irregularities and ascribing to the data summary measures that can be used for comparative analysis. It may be used to interpolate to single years of age, [even when] observed migration schedules ... are reported for wider age intervals. Assessments of the *reliability* of empirical migration data and indications of appropriate strategies for their correction are aided by the availability of standard families of migration schedules. Finally, such schedules also may be used to help resolve problems caused by missing data" (Rogers & Castro, 1981:45 & 47).

It was shown above that it is indeed possible to use the parameters of an model migration schedule (MMS) to model the South African census and IEC migration data.

In applying the data to the issue of the former homeland territories the spatialised results gave a clear indication that the major migration streams indicated a move to the larger metropolitan centres, and regional centres such as Cape Town, Ekurhuleni, Johannesburg and Tshwane. At the time of this paper no analysis had been conducted on gender and age categories within the dataset. Such analyses will also allow for future analyses of population segments. The IEC migration data will in future be used more intensively to understand migration behaviour. This in turn will hopefully positively impact on national, regional and local planning and policy making.

References

Anon (2008) INDEPTH resource kit for demographic surveilance systems [Homepage of INDEPTH network], [Online]. Available: http://www.indepth-network.org Accessed on 1 April 2013.

Cahoon B (2013) Former black homelands (bantustans) [Homepage of World Statemen.org], [Online]. Available: http://www.worldstatesmen.org/South_African_homelands.html. Accessed on April 18, 2013.

Castro LJ, Rogers, A (1983) What the age composition of migrants can tell us. Population Bulletin of the United Nations, No. 15:63-79.

EISA (2002) South Africa: delimitation process and voting stations [Homepage of Electoral Institute for Sustainable Democracy in Africa], [Online]. Available: http://www.eisa.org.za. Accessed on 12April 2013

ESRI (2012) ArcGIS Resource Centre: Normal QQ plot and general QQ plot. 1 edn. Redlands, US: Environmental Systems Research Institute. Available from: <http://help.arcgis.com>.

Freund JE, Williams FJ and Perles BM (1988) Elementary business statistics - the modern approach. Fifth edn. New Yersey: Prentice-Hall, Inc.

Glennon JA (2008) Flow data model tools. [Homepage of Alan Glennon], [Online]. Available: http://www.alanglennon.com/flowtools/. Accessed on 1 June 2010

Grobbelaar N (2005) The development of a Small Area Layer to serve as the most detailed geographical entity for the dissemination of census 2001 data. *AfricaGIS 2005:* Beyond Talk: Geo-information working for Africa, 31 October to 4 November 2005

Hofmeyr BE (1988) Application of a mathematical model to South African migration data, 1975-1980. Southern African Journal of Demography, 2(1):24-28

IEC (2012) Independant Electoral Commission - about us. [Homepage of IEC], [Online]. Available: http://www.elections.org.za Accessed on 12 April 2013

Kok P (2010) Preparing the IEC data for migration analyses: approaches and methods adopted. Feedback to IEC edn. Pretoria: P. Kok.

Kok P and Collinson M (2006) Migration and urbanisation in South Africa. 03-04-02. Pretoria: Statistics South Africa

Kok P, O'Donovan M, Bouare O and Van Zyl J (2003) Post-apartheid patterns of internal migration in South Africa. Cape Town: HSRC Publishers

Maritz J (2012) Impact of migration trends (Part of Urban Growth in South Africa Parliamentary Grant project, 2012/13. CSIR: Internal document.

Municipal Demarcation Board (2013) About - history [Homepage of Municipal Demarcation Board], [Online]. Available: www.demarcation.org.za Accessed on 12 April 2013

Polzer T (2010) Population Movements in and to South Africa: Migration factsheet. Johannesburg: University of the Witwatersrand: Forced Migration Studies Programme

Presidency (2006) National Spatial Development Perspective 2006. Pretoria: The Precidency, South Africa

Rwelamira JK, Kirsten JF (2003) The impact of migration and remittances to rural migration-sending households: The case of the Limpopo Province South Africa. Agricultural Economic Association of South Africa (AEASA) annual conference on 2-3 October 2003, Pretoria, South Africa. AEASA

Skeldon R (2008) Migration and Development. UN/POP/EGM-MIG/2008/4. Bangkok, Thailand: United Nations Economic and Social Commission for Asia and the Pacific

Skeldon R (1990) Population Mobility in Developing Countries: A Reinterpretation. First edn. London: Belhaven Press

Standing G (1984) Conceptualising territorial mobility. In: R.E. BILSBORROW, A.S. OBERAI and G. STANDING, eds, Migration surveys in low income countries: guidelines for survey and questionnaire design. 1 edn. London: Croom Helm, pp. 31-59

Statistics South Africa (2010) Mid-year population estimates - 2009. P0302. Pretoria, South Africa: Statistics South Africa

Statistics South Africa (2001) Census 2001 - concepts and definitions. 03-02-26 (2001). Pretoria: Statistics South Africa

STEPSA.ORG (2013) About - stepsa.org [Homepage of CSIR], [Online]. Available: http://stepsa.org [April 11, 2013]

Steyn L (2013) Measuring the waves of migration. Mail and Guardian, Business, pp. Online

UN Population Division (2013) Cross-national comparisons of internal migration: An update on global patterns and trends. Unpublished Technical Paper No. 2013/1. New York: United Nations

Waugh D (2000) Geography, an integrated approach. Second edn. Cheltenham,United Kingdom: Nelson Thornes