International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:12, No:6, 2018

Comparison of Automated Zone Design Census Output Areas with Existing Output Areas in South Africa

T. Mokhele, O. Mutanga, F. Ahmed

***Abstract*—**South Africa is one of the few countries that have stopped using the same Enumeration Areas (EAs) for census enumeration and dissemination. The advantage of this change is that confidentiality issue could be addressed for census dissemination as the design of geographic unit for collection is mainly to ensure that this unit is covered by one enumerator. The objective of this paper was to evaluate the performance of automated zone design output areas against non-zone design developed geographies using the 2001 census data, and 2011 census to some extent, as the main input. The comparison of the Automated Zone-design Tool (AZTool) census output areas with the Small Area Layers (SALs) and SubPlaces based on confidentiality limit, population distribution, and degree of homogeneity, as well as shape compactness, was undertaken. Further, SPSS was employed for validation of the AZTool output results. The results showed that AZTool developed output areas out-perform the existing official SAL and SubPlaces with regard to minimum population threshold, population distribution and to some extent to homogeneity. Therefore, it was concluded that AZTool program provides a new alternative to the creation of optimised census output areas for dissemination of population census data in South Africa.

***Keywords*—**AZTool, enumeration areas, small areal layers, South Africa.

I. Introduction

CENCUS data are a powerful tool for development and poverty reduction. It is a foundation for a wide range of research and analyses required to improve the standard of living of people in any country. Population projections are one of the most important analytical outputs based on census information [1], [2]. The characteristics of all individuals within a given area are recorded simultaneously in the census data collection. These data are utilised to inform government policy making, planning and administration. They are also used for demographics, social research and research to inform business, industry, labour and the public [2]-[8]. In addition, census data provide a sampling framework for surveys that provide further insights into demographic and socio-economic trends that could be used to assess, monitor, and evaluate the implementation of government policies and programs [2]-[4].

Many countries conduct censuses at regular intervals of five or ten years. In South Africa, the Statistics Act No. 6 of 1999 mandates Stats SA to carry out a census in a five-year cycle, but a decision was taken by Cabinet in 2004 that censuses would be undertaken in every ten years [4]. South Africa is one of the countries that have moved from using the same geographic unit for census enumeration and dissemination. For the 1991 and 1996 censuses, the same EAs were used for both census enumeration and dissemination. For the 2001 census, it was decided that census data must be released on an area larger than an EA due to confidentiality [4], [9], [10]. Stats SA then attached two names each EA, and a spatial layer was created from the name attributes (SubPlaces and MainPlaces). Most users of the census data believed that these areas were too large. This resulted in the creation of the SAL using a non-zone design approach with the aim of meeting South African census user needs. A similar non-zone design approach was also employed in the creation of SAL for the 2011 census data. As indicated earlier, the main objective of the SAL was to have a spatial area layer that corresponded as much as possible to the EA layer, but remained within the confidentiality limit of 500 people [10]. For instance, for the creation of SAL in 2005, the following criteria were set and adhered to as far as possible: firstly, EAs could only be merged if they are within the same SubPlace; secondly, EAs could only be merged if they have the same EA geography type; thirdly, an EA could only be merged if its population is less than 500; and lastly, the resulting small area polygons must have a population total of 500 and more [10].

In South Africa, it has not been established whether automated zone design generated census output areas could perform better than the existing official census dissemination areas with respect to certain design criteria or not. Automated zone design procedures tend to offer more efficient, systematic, and objective methodologies for designing optimised zoning systems than non-zone design methods. However, their success is dependent on the extent to which it is possible to model real-world phenomena and whether it is feasible to parameterise the required design criteria [11]. Applications of the automated zone design, especially the AZTool program, are well described in previous studies [11]-[16]. Automated zone design methods offer more efficient, systematic, and objective methodologies for designing optimised zoning systems than manual methods [11]. In the United Kingdom, [17] compared automated zone design program ‘‘A2Z’’ zones, developed by [18], with areal units identified subjectively by local government officers as communities in the city of Bristol. Their findings showed that the first automated zone design was much more successful in identifying homogenous deprivation areas than the subjective community (cf. the ICC values of 0.82 and 0.61), and was

Tholang Mokhele is with the Human Sciences Research Council, South Africa (e-mail: TAMokhele@hsrc.ac.za).

equally successful in identifying homogeneous areas of a particular housing type (0.51 and 0.51) even though zone design was much less compact in shape than the subjective areas. Their results further highlighted that automated zone design was close to replicating the subjective communities when the balance of objectives and boundary constraints was adjusted. In New Zealand, [14] compared the AZTool new geographies with existing official geographies. They found that the new geographies substantially out-performed the existing geographies across almost all of their optimisation criteria. In France, [16] compared the AZTool new zones with existing IRIS census areas to explore relationships between asthma and deprivation in Strasbourg. Their results indicated that the newly produced synthetic neighbourhood solution performed better than the then existing IRIS census areas, measured by improved statistical relationships between asthma and deprivation. Therefore, the objective of this article was to compare the newly AZTool developed census output areas with existing official census output geographies such as SALs and SubPlaces in South Africa with the aim of evaluating the AZTool application in South Africa.

II. Methods

The study areas were Free State and Gauteng provinces of South Africa, which were representative of rural and urban settings, respectively [19], [20]. The EAs from the 2001 census estimates data [21] were used as building blocks for the development of new census output areas using automated zone design procedure. The 2001 SubPlaces data and the 2001 SALs data, from Stats SA were used for comparisons with the newly created census output areas. The 2001 SAL data did not have dwelling type variable. For geotype, it had three categories; namely, urban, rural, and mixed instead of four categories that were in the 2001 census EAs data; namely, formal urban, informal urban, informal rural (tribal Areas); and formal rural (farms). Hence, the comparison on social homogeneity could not yield fruitful results. The 2001 SubPlaces data had similar variables with the EAs data which were used as building blocks for the newly developed output areas, therefore comparisons were made with all design criteria. The 2011 SALs data from Stats SA were also explored.

In order to create optimised census output areas in South Africa, the AZTool version 1.0.3 [11] was used. ESRI's ArcGIS 10.2 and Microsoft Excel were employed for data preparation to be used by the AZTool software and for displaying AZTool output results. As indicated in Table I, the design criteria were that all output areas must not breach a minimum population threshold of 500, must be as homogeneous with regard to dwelling type and geotype and be as compact with regard to shape as possible. The population mean target was also set in order to control the population distribution. In this study, like in other studies such as [13], [14], [22], [23], output areas have been developed by taking existing areas (the 2001 census EAs) and using them as building blocks to create larger areas that are optimised based on the required design criteria.

TABLE I
Design criteria for newly developed census output areas

Criteria	Description	Weighting
Minimum threshold population size[1]	500	N/A
Mean target population	1000	100
Homogeneity[2]	IAC score for dwelling type and geotype	100
Shape compactness[3]	Perimeter squared per area (P2A)	100

[1] Minimum population threshold used by Statistics South Africa in creation of SAL [10]. [2] Intra-Area Correlation (IAC) [13], [24]-[27]. [3] Shape compactness [17], [22]

In order to statistically validate the results from the AZTool program, further quantitative statistical analyses were undertaken using SPSS. These included one-way Analysis of Variance (ANOVA), Kolmogorov sminov test and a paired t-test.

III. Results

Table II shows the statistical characteristics of the newly developed output areas, the SALs and SubPlaces data for the 2001 census at all spatial levels in the Free State province. This table indicates that the confidentiality limit of 500 was respected at all spatial levels for the newly created output areas whereas for both the SALs and SubPlaces, this threshold was breached at all levels. Setting the mean target in output areas also made output areas to have much narrower and tighter population distribution than that of SALs. For instance, population distribution of the newly created output areas was compared with that of the SALs for Maluti-a-Phofung Municipality in Free State province. It is important to mention that maximum populations for output areas are a bit larger than those of the SALs at all levels. As indicated earlier, in many instances, the SubPlaces were too large for most census data users. This is illustrated in Table II as maximum population for a SubPlace could go as high as 93290 persons in the Free State province. With regard to social homogeneity, only the newly created output areas and SubPlaces could be compared as SAL social homogeneity could not yield fruitful IAC results due to lack of homogeneity variables. The IAC scores for the newly AZTool created output areas were slightly lower than those of SubPlaces at most levels except for provincial level where they both recorded the same IAC score of 0.59. When comparing compactness of the shapes, the output areas had slightly higher P2A mean values with lower standard deviations than the SALs at all spatial levels. This means that the newly created output areas were less compact in shape than the SALs in all regions. In general, there was statistically significant ($p < 0.05$) difference in P2A means between the output areas, the SALs and SubPlaces based on one-way ANOVA results. The LSD post-hoc test revealed that difference between the P2A means for the output areas and the SALs was not statistically significant ($p > 0.05$). The SubPlaces had higher P2A mean values with higher standard deviations than the output areas and the SALs. In addition, the P2A means difference between the SupPlaces and the output areas was statistically significant ($p < 0.05$) as well as between the SubPlaces and the SALs ($p < 0.05$). This shows that the

SubPlaces were less compact in shape compared to both the output areas and the SALs.

Similar analyses were undertaken in Gauteng province at all spatial levels in order to get an understanding of comparisons at urban settings (Table III). It is satisfying to note that AZTool created output areas adhered to the minimum population at all levels as it was in rural areas, which is very reassuring from a confidentiality perspective. As it was for rural areas, the SALs and SubPlaces breached the confidentiality threshold with minimum population of zero being record at all levels. Table 3 further shows that output areas recorded higher IAC scores than SubPlaces at all levels. In contrary to rural settings, this shows that the newly created output areas were more homogeneous than the SubPlaces based on dwelling type and geotype homogeneity variables. In terms of shapes, the AZTool optimised output areas had higher P2A mean values and standard deviations than the SALs showing that output areas were less compact than their counter-parts at all spatial levels. The difference between three P2A means was statistically significant ($p < 0.05$). On the contrary to the rural areas, the P2A mean difference between the output areas and the SALs was statistically significant ($p < 0.05$). Similar to rural areas, the SubPlaces had statistically significantly ($p < 0.05$) higher P2A mean values and standard deviations than the output areas and the SALs. It is interesting to see that all levels recorded the same P2A mean value of 29 even though their standard deviations tend to increase with spatial level, thus Pretoria Mainplace had standard deviation of 12, then 15 and 17 for City of Tshwane and Gauteng province, respectively.

TABLE II
STATISTICAL CHARACTERISTICS OF NEWLY DEVELOPED OUTPUT AREAS, THE 2001 SALS AND SUBPLACES FOR FREE STATE

			Population			Shape		
	Zones	Min	Max	Mean	SD	Mean	SD	IAC
Output Areas								
Phuthaditjhaba	48	649	2704	1113	346	28	9	0.21
Maluti-a-Phofung	349	610	2704	1027	232	32	13	0.5
Thabo Mofutsanyane	667	581	5292	1087	403	33	13	0.56
Free State	2440	547	9269	1101	489	31	12	0.59
Small Area Layers								
Phuthaditjhaba	68	408	1144	782	169	26	13	N/A
Maluti-a-Phofung	474	0	2071	761	248	30	14	N/A
Thabo Mofutsanyane	901	0	6701	806	359	30	13	N/A
Free State	3463	0	6701	782	318	29	13	N/A
SubPlaces								
Phuthaditjhaba	13	410	10507	4091	3565	43	29	0.29
Maluti-a-Phofung	110	0	22496	3280	4250	38	22	0.54
Thabo Mofutsanyane	223	0	25500	3255	3977	36	19	0.57
Free State	791	0	93290	3422	5974	34	22	0.59

TABLE III
STATISTICAL CHARACTERISTICS OF THE NEWLY DEVELOPED OUTPUT AREAS, THE 2001 SALS AND SUBPLACES FOR GAUTENG

			Population			Shape		
	Zones	Min	Max	Mean	SD	Mean	SD	IAC
Output Areas								
Pretoria	500	621	5026	1056	320	28	11	0.1
Tshwane	1276	502	8802	1203	514	27	10	0.5
Gauteng	7253	501	9627	1214	520	27	9	0.5
Small Area Layers								
Pretoria	662	0	4227	794	301	26	10	N/A
Tshwane	1723	0	8092	886	442	25	11	N/A
Gauteng	10177	0	8092	868	389	25	10	N/A
SubPlaces								
Pretoria	157	0	26773	3346	4599	29	12	0.1
Tshwane	315	0	82002	4848	8764	29	15	0.5
Gauteng	2222	0	131662	3977	7403	29	17	0.4

When comparing population distribution for AZTool newly created output areas and the SALs in Maluti-a-Phofung municipality, results showed that AZTool successfully respected the confidentiality rule by having more than 500 people in all the areas. Kolmogorov sminov test results showed that both the output areas and the SAL population distributions were not normal ($p < 0.05$ in both cases). Furthermore, the AZTool newly created output areas population distribution follows a normal curve more than the SALs. This shows that the newly created output areas, with population target set to 1000, had a much narrower and tighter population distribution than the SALs. This makes the newly created output areas more ideal from user's perspective as the individual areas could be comparable to each other with regard to population size.

In general, the percentages of areas breaching the population thresholds for the SALs were 6.3% and 4.7%, for Free State and Gauteng provinces, respectively. In the SubPlaces data, 24.7% of areas fell below the 500 population confidentiality limit in the Free State province and 21.2% in Gauteng province. None of the areas breached the population confidentiality for the newly created output areas in both rural and urban areas. The rural areas seem to be more homogeneous than the urban areas at all regions for both newly created output areas and SubPlaces.

In terms of shape compactness, a further examination was done at provincial levels for both rural and urban settings for the newly created output areas, SALs, and SupPlaces. Fig. 1 indicates that output areas had higher P2A mean values and lower standard deviations than the SALs in both rural and urban provinces. The SubPlaces recorded higher shape mean values for both provinces but their standard deviations were also too high. The fact that standard deviations overlapped indicate that there were not significant differences between the P2A means. This was proven by performing ANOVA, which further revealed that the P2A mean difference between these three groups was not statistically significant ($p > 0.05$). When comparing the two provinces for all three areas, the rural province had slightly less compact shapes than the urban province as P2A means were higher than those of urban province. However, the differences in P2A means were also not significant as the standard deviations were overlapping.

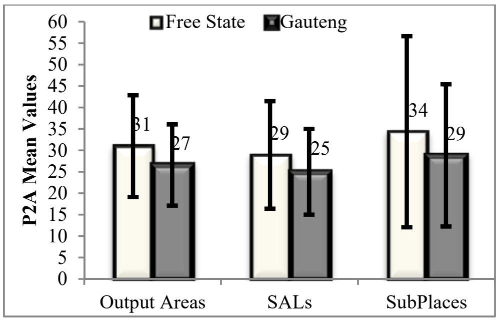

Fig. 1 Shape means and standard deviations of the output areas, the SALs and the SubPlaces for the Free State and Gauteng provinces

In an effort to compare newly developed output areas with SALs with all design criteria, the 2011 SAL data for the Free State was explored. Only Free State was used as indicated earlier that this province experienced low population growth and its provincial boundaries did not change from 2001. It is worth noting that, for the 2011 SAL, only populated areas were captured hence all zero-populated areas were not included in the data. Table IV indicates that minimum population for the 2011 SALs, the 2001 SALs and newly created output areas were 9, 0 and 547, respectively. Additionally, the newly created output areas were more compact in shape than the 2011 SALs and 2001 SALs. The 2001 SALs were less compact than the 2011 SALs. In terms of social homogeneity, the output areas were slightly less homogeneous with IAC score of 0.59 compared to 0.62 of the 2011 SALs. However, the geotype homogeneity variables for the 2011 SALs had only three categories (Urban, Rural and Farms), while the AZTool output areas had four categories which were Formal Urban, Informal Urban, Formal Rural (Tribal Areas) and Informal Rural (Farms).

TABLE IV
STATISTICAL CHARACTERISTICS OF THE OUTPUT AREAS, 2001 SALS, AND 2011 SALS FOR FREE STATE PROVINCE

		Population				Shape		
Free State	**Zones**	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**	**IAC**
Output Areas	2440	547	9269	1101	489	31	12	0.59
Small Area Layers 2001	3463	0	6701	782	318	29	13	N/A
Small Area Layers 2011	5129	9	5586	535	228	25	9	0.62

Populations for both the aggregated SubPlaces data and the original SubPlaces data from Stats SA for Phuthaditjhaba mainplace were displayed in Fig. 2. This shows that populations for the aggregated SubPlaces, derived from the 2001 EAs estimates data, were slightly higher than those of the original SubPlaces data in each individual areas. A paired t-test was performed to see if the means from these two datasets were the same. The results ($t = 3.944$, $p = 0.002$) showed that difference in mean populations from the aggregated data and the original data was statistically significant. The mean difference between the two datasets was 18. 77 with the 95% confidence interval ranging from 8.401 to 29.137. This indicates that, although the difference in means was statistically significant, it was actually relatively small. In order for these results to be valid, the differences between the paired values should be approximately normally distributed. Therefore, a simple Kolmogorov sminov test revealed that indeed the distribution of differences was normal ($p > 0.05$).

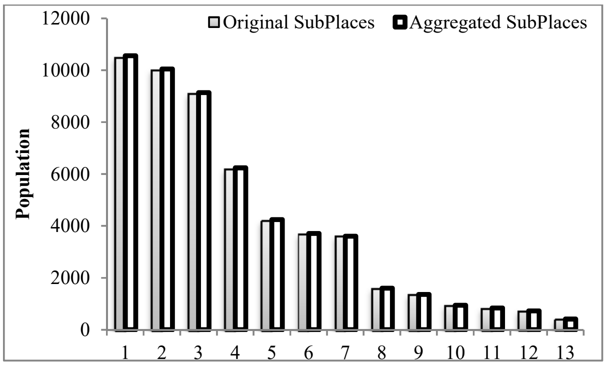

Fig. 2 Comparison of the original SubPlaces with the aggregated SubPlaces population data for Phuthaditjhaba mainplace

A further comparison of the aggregated SubPlaces data with the original SubPlaces data was performed using the AZTool results outputs for all levels in both rural and urban settings. Table V shows that statistical qualities of these areas were mostly the same. It is important to note that when comparing IAC scores at each spatial level for both aggregated SubPlaces and original SupPlaces, IAC scores were exactly the same for rural areas. The urban areas showed a slight difference in these IAC scores as the ones for aggregated SubPlaces were slightly higher than those of the original SubPlaces. These comparisons provide some confidence with regard to the use of the 2001 EAs estimates as building blocks in the development of the AZTool output areas.

TABLE V
STATISTICAL CHARACTERISTICS OF THE ORIGINAL SUBPLACES AND THE AGGREGATED SUBPLACES

		Population				Shape		
	Zones	Min	Max	Mean	SD	Mean	SD	IAC
Original SubPlaces								
Phuthaditjhaba	13	410	10507	4091	3565	43	29	0.29
Maluti-a-Phofung	110	0	22496	3280	4250	38	22	0.54
Thabo Mofutsanyane	223	0	25500	3255	3977	36	19	0.57
Free State	791	0	93290	3422	5974	34	22	0.59
Aggregated SubPlaces								
Phuthaditjhaba	13	412	10554	4109	3581	43	29	0.29
Maluti-a-Phofung	110	0	22594	3260	4281	38	22	0.54
Thabo Mofutsanyane	223	0	25612	3253	4001	36	19	0.57
Free State	791	0	93701	3397	5969	34	22	0.59
Original SubPlaces								
Pretoria	157	0	26773	3346	4599	29	12	0.07
City of Tshwane	315	0	82002	4848	8764	29	15	0.45
Gauteng	2222	0	131662	3977	7403	29	17	0.44
Aggregated SubPlaces								
Pretoria	157	0	26915	3363	4625	29	12	0.08
City of Tshwane	315	0	82440	4872	8815	29	15	0.46
Gauteng	2222	0	132363	3962	7415	29	17	0.45

IV. DISCUSSION

Results from this study show that the newly developed output areas using the AZTool are very much an improvement over the SALs and the SubPlaces. This was proven by the fact that newly developed output areas effectively satisfied minimum and target population thresholds, while the population distributions were much narrower in range than those of the existing SALs and SubPlaces. The confidentiality limit of 500 people was respected at all spatial levels in both rural and urban settings for the newly created output areas, whereas for both SALs and SubPlaces the confidentiality limit of 500 persons was breached at all levels. The fact that the AZTool generated output areas did not breach minimum population throughout all study areas is very reassuring from a confidentiality perspective. Similarly, [14] found that AZTool successfully constrained all tracts to be of at least the required minimum size.

The population target criterion also yielded positive results as the AZTool census output areas had a much narrower and tighter population distribution than that of the SALs. The summary of rules set for the creation of the SALs did not have population target, which would have made them to have a better distribution than the current one. The importance of tighter and narrower population distribution is that it makes the newly created output areas more ideal from a census data user's point of view as the individual areas could be easily compared in terms of their population size distribution. This supports previous arguments by [10] that, in many instances, the SubPlaces were too large for most census data users, hence the initiative was taken to develop the SALs in 2005. It is worth mentioning though that some of the AZTool output areas had very large population sizes. This is due to the fact that the 2001 EAs were used as building blocks for the creation of these output areas. The availability of data at the lower level than EAs, household level, would allow the optimisation algorithm to have more options in generating output areas that meet target population sizes as much as possible. Other studies such as [13], [14], [17], [22], [23] also identified similar challenges as they used existing areas as building blocks, hence the flaws of such areas were inherited into the generated output areas [28].

With regard to homogeneity, only the SubPlaces were comparable to the newly generated census output areas as the SALs did not produce IAC score due to insufficient homogeneity variables. The newly AZTool created output areas were slightly less homogeneous than the SubPlaces at most levels in rural areas. The provincial level was an exception as the output areas and SubPlaces shared the same degree of homogeneity in terms of dwelling type and geotype variables, with both having IAC score of 0.59. In contrary, the urban settings showed that the newly created output areas were more homogeneous than the SubPlaces based on dwelling type and geotype homogeneity variables at all levels [17].

A further attempt was undertaken to use the 2011 SALs which had both dwelling type and geotype variables in order to be able compare the AZTool output areas with the SALs. The census output areas generated from the AZTool program were slightly less homogeneous with IAC score of 0.59 compared to 0.62 of the 2011 SALs for Free State province. This might be due to the fact that the geotype homogeneity variable for the 2011 SALs had only three categories, while the AZTool output areas had four categories. Although these results are from two different censuses, it is believed that they are good indication of how homogeneity variable would perform in the comparisons as this province did change at all in terms provincial boundaries and did not change much in terms of population growth. It is important though to note that, due to infrastructure development the dwelling type variable, might have been affected from 2001 to 2011.

Findings from both rural and urban areas showed that the AZTool newly created output areas were less compact in shape compared with the SALs at all regions. This is in line with the previous findings by [14], [17], where automated zone design output areas were slightly less compact than original existing geographies. The SubPlaces had less compact shapes than both the output areas and the SALs.

Comparing the 2001 EAs estimates data, which were used as building blocks for the output areas, with the original 2001 SubPlaces data brought some confidence in the AZTool newly created output areas as these have to be close to reality as much as possible. This does not rule out the fact that the original EAs data from Stats SA would have been preferable had it been available. There are some positives to be drawn from this study as the comparison of automated zone design census output areas with existing official census output areas had not been reported before in South Africa. Therefore, findings from this study provide a new alternative to the creation of optimised census out areas for population census disseminations.

V. Conclusion

In general, the census output areas that were generated using the AZTool out-performed the existing official SALs and SubPlaces, non-zone design developed geographies. The AZTool generated output areas effectively satisfied minimum and target population thresholds criteria. In addition, the population distributions were much narrower in range than those of the existing official output area, SALs and SubPlaces. However, the AZTool census output areas were less compact in shape compared with the SALs at all spatial levels in both rural and urban areas. A comparison of automated zone design census output areas with existing official census output areas has not been reported before in South Africa. Therefore, it was concluded that findings from this paper provide a new alternative to the creation of optimised census output areas for future census disseminations in South Africa.

Acknowledgment

The authors thank Prof. David Martin and Prof. Samantha Cockings for their permission to use the AZTool software, which is copyrighted by David Martin, Samantha Cockings and University of Southampton. Stats SA and HSRC are also acknowledged for datasets. This paper is part of the doctoral thesis work that was submitted to the University of KwaZulu-Natal in 2015.

References

[1] C. Schwabe, *The South African census user's handbook: Analysing data from 1996 census.* Cape Town, RSA: Human Sciences Research Council Publishers, 2003.

[2] Stats SA, *Census 2011 methodology and highlights of key results.* Pretoria, RSA: Statistics South Africa, Report no. 03-01-42, 29 pp., ISBN 978-0-621-41389-2, 2003.

[3] H. Margeot and S. Ramjith, "The South African census 2001 spatial information system data capture problems", *International Conference on Spatial Information for Sustainable Development,* 2 – 5 October 2001, Kenya.

[4] Stats SA and HSRC, *Using the 2001 census: Approaches to analysing data.* RSA: Statistics South Africa and Human Sciences Research Council, 244 p, 2007.

[5] P.O. Owiti, "Embedded GIS for census mapping", Nairobi, Kenya, 2008. Available from: http://www.map-gis-rs.blogspot.com/2008/03/gis-for-census.html (Accessed 8 January 2012).

[6] UN, *Handbook on Geographic Information Systems and digital mapping. Studies in Methods, Series F No. 79.* New York: United Nations Publication, 2000.

[7] UN, "Integration of GPS, digital imagery and GIS with census mapping," *United Nations Expert Group Meeting to Review Critical Issue Relevant to the Planning of the 2010 Round of Population and housing Censuses*, 15 – 17 September 2004, New York.

[8] UN, *Handbook on geospatial infrastructure in support of census activities. Studies in Methods, Series F No. 103.* New York: United Nations Publication, 2009.

[9] Stats SA, *Census 2001. How the count was done.* RSA: Statistics South Africa, 2003.

[10] H. Verhoef and N. Grobbelaar, "The development of a Small Area Layer for South Africa for census data dissemination", Statistics South Africa, 2005. Available from: http://www.cartesia.org/geodoc/icc2005/pdf/poster/TEMA26/HELENE%20VERHOEF.pdf (Accessed 19 November 2009).

[11] S. Cockings, A. Harfoot, D. Martin and D. Hornby, "Maintaining existing zoning systems using automated zone-design techniques: methods for creating the 2011 census output geographies for England and Wales," *Environment and Planning A*, vol. 43, pp. 2399 – 2418, 2011.

[12] S. Cockings, A. Harfoot, D. Martin and D. Hornby, "Getting the foundations right: spatial building blocks for official population statistics," *Environment and Planning A*, vol. 45, pp. 1403 – 1420, 2013.

[13] R. Flowerdew, D.J. Manley and C.F. Sabel, "Neighbourhood effects on health: Does it matter where you draw the boundaries?" *Social Science and Medicine*, vol. 66, pp. 1241 – 1255, 2008.

[14] M. Ralphs and L. Ang, *Optimized geographies for data reporting: Zone design tools for census output geographies*, Wellington: Statistics New Zealand, Statistics New Zealand Working Paper No 09–01, 2009.

[15] D. Martin, S. Cockings and A. Harfoot, "Development of a geographical framework for census workplace data," Journal of Royal Statistical Society, vol. 176 (2), pp. 1 – 18, 2013.

[16] C. E. Sabel, W. Kihal, D. Bard and C. Weber, "Creation of synthetic homogeneous neighbourhoods using zone design algorithms to explore relationships between asthma and deprivation in Strasbourg, France," *Social Science and Medicine*, vol. 91, pp. 110 – 121, 2013.

[17] R. Haynes, K. Daras, R. Reading and A. Jones, "Modifiable neighbourhood units, zone design and residents' perceptions," *Health and Place*, vol. 13, pp. 812 – 825, 2007.

[18] K. Daras, "An information statistics approach to zone design in the geography of health outcomes and provision", unpublished PhD Thesis. University of Newcastle, England, 2006.

[19] T. Mokhele, O. Mutanga and F. Ahmed, "Development of census output areas with AZTool in South Africa," *South African Journal of Science*, vol. 112 (7/8), pp. 1 – 7, 2016.

[20] T. Mokhele, O. Mutanga and F. Ahmed, "Effects of different building blocks designs on the statistical characteristics of Automated Zone-design Tool output areas," *South African Journal of Geomatics*, vol. 6(2), pp. 155 – 171, 2017.

[21] HSRC, "2001 census EA estimates. Human Sciences Research Council in collaboration with Prof DJ Stoker", unpublished. Pretoria, South

Africa, 2005.
[22] S. Cockings and D. Martin, "Zone design for environment and health studies using pre-aggregated data," *Social Science and Medicine*, vol. 60, pp. 2729 – 2742, 2005.
[23] R. Haynes, A. Jones, R. Reading, K. Daras and A. Emond, "Neighbourhood variations in child accidents and related child and maternal characteristics: does area definition make a difference?" *Health and Place*, vol. 14, pp. 693 – 701, 2008.
[24] M. Tranmer and D. Steel, "Using census data to investigate the causes of the ecological fallacy," Environment and Planning A, vol. 30, pp. 817 – 831, 1998.
[25] M. Tranmer and D. Steel, "Using local census data to investigate scale effects," in *Modelling scale in geographical information science*. N.J. Tate and P.M. Atkinson, Eds. Chichester: John Wiley and Sons, 2001, pp. 105 – 122.
[26] D. Martin, A. Nolan and M. Tranmer, "The application of zone-design methodology in the 2001 UK Census," *Environment and Planning A,* vol. 33, pp. 1949 – 1962, 2001.
[27] R. Flowerdew, "How serious is the Modifiable Areal Unit Problem for analysis of English census data?" Population Trends, vol. 145, pp. 106 – 118, 2011.
[28] A. Drackley, K. B. Newbold and C. Taylor, "Defining socially-based spatial boundaries in the Region of Peel, Ontario, Canada," International Journal of Health Geographics, vol. 10 (38), pp. 1 – 12, 2011.