Statistical and Machine Learning Approaches for Municipal Population Forecasting in Gauteng, South Africa

Chukwuma OKOLIE, Mawande NGIDI, Zukisa SOGONI, Lesego TSHUWA, and Samukele NGEMA, South Africa

African Mayoral Leadership Initiative (AMALI) Data Programme - African Centre for Cities, School of Architecture, Planning and Geomatics, University of Cape Town, Cape Town, South Africa

Keywords: Geospatial data, population forecasting, population growth, statistics, machine learning.

1. SUMMARY

Population data and population forecasts are crucial for planning all around the world, especially in urban areas as cities are expected to ensure basic service provision to communities. This is especially important for African cities which are experiencing rapid population growth and significant resource constraints. Despite the importance of accurate population forecasts, there is no consensus on the best approach for its modelling. While statistical time-series models have been widely used, machine learning approaches have shown promise in recent studies. The relative performance of both techniques is still an ongoing debate in the literature. We address this research gap by comparing the accuracy of ten (10) statistical and machine learning models in five municipalities of Gauteng Province, South Africa (Johannesburg, Tshwane, Ekurhuleni, Sedibeng and West Rand). The models were trained and tested using authoritative mid-year population estimates, and then future forecasts were generated. The trends in historical population are reasonably well captured by both approaches, although statistical models dominated in the forecasts. Application of these modelling approaches has the potential to stir up the urban governance paradigm by giving wider access to data-driven decision making in local municipalities. Further research is needed to provide more conclusive evidence on the comparative performance of both approaches, particularly under different data conditions, spatial contexts, and forecasting horizons.

Statistical and Machine Learning Approaches for Municipal Population Forecasting in Gauteng, South Africa

Chukwuma OKOLIE, Mawande NGIDI, Zukisa SOGONI, Lesego TSHUWA, and Samukele NGEMA, South Africa

2. INTRODUCTION

Population is a fundamental geospatial dataset (Schwabe, 2019, UN-GGIM, 2018), and is crucial for sustainable development. It represents the spatial distribution of people within a geographical area, and is a core element required by cities to plan for various needs such as land management, housing, waste management, sanitation and hygiene, public health, local economic development, infrastructure provision, disaster management and job creation. The global urban population is projected to increase by several billions over the next couple of years. About 45% of the global population reside in cities, with projections indicating this number will increase to 68% by 2050 (United Nations, 2025; Copernicus, 2025; Destatis, 2026). Africa is at the centre of the global population growth – accounting for the largest relative growth; its population expanded from 283 million in 1960 to more than 1.5 billion in 2024, and is projected to touch 2.5 billion by 2050 (UNECA, 2024). Given the reality of the rising world population, population forecasts are very crucial to support urban planning and inform policy decisions.

Statistical time-series forecasting models are suitable for analysing and predicting population growth. These models can effectively capture the temporal dependence (Bandara et al., 2021), trends, and seasonality and patterns (Hewamalage et al., 2022) in population data. Recently, the field of machine learning has become very popular, setting the stage for contemporary applications in time-series forecasting (Fatima and Rahimi, 2024). Statistical time-series models are effective for capturing linear trends in time-series data, whereas machine learning models are reputed for their ability to capture complex, non-linear relationships and patterns (Chaudhuri, 2024). However, comparative assessments of the performance of machine learning to the classical statistical time-series approaches for population forecasting are scarce. This paper presents preliminary findings on the comparative performance of both approaches for population forecasting at municipal level.

3. STUDY AREA

The study area is Gauteng Province, one of the nine provinces of South Africa (Figure 1). The provinces of South Africa are divided into 8 metropolitan municipalities and 44 district municipalities. These districts are the second level of administrative division in South Africa, below the provinces (first level), and above the local municipalities (third level). The metropolitan municipalities are single-tier municipalities with significant powers, while the district municipalities are responsible for district-wide planning across the local municipalities.

Figure 1: Map of Gauteng showing the five municipalities

Gauteng is the smallest province (land size) but with the largest population and contributor to Gross Domestic Product (GDP) of South Africa. It is the most populated province with over 15.9 million people (25.3%), and between 2002 – 2024, its share of the total population of South Africa increased from 20 to 25.3 percentage points (Statistics South Africa, 2024a).

The province encompasses three metropolitan municipalities (City of Johannesburg, City of Tshwane, City of Ekurhuleni), two district municipalities (Sedibeng and West Rand), and six local municipalities located within Sedibeng and West Rand. The administrative capital of South Africa (Pretoria) is located within the City of Tshwane. Over the years, the province has experienced rapid urbanisation and a surge in demand for services and amenities. Population forecasts are essential to support planning in the province.

4. METHODOLOGY

The mid-year population estimates (MYPE 2024) for the five municipalities were accessed from the online portal of Statistics South Africa (Statistics South Africa, 2024b). This dataset contains yearly projected population data (2002 – 2024) grouped by age and sex. We summed up the projected population for the different demographic groups (age, sex) to obtain the total projected population for each municipality on a yearly basis. Ten statistical and machine learning models were deployed on the data. These include exponential smoothing models (Exponential Smoothing, ETS), linear models (Linear Regression, Elastic Net) and tree-based models (Decision Tree, Random Forest, Extra Trees, Gradient Boosting, Extreme Gradient Boosting, Categorical Boosting). A training/test ratio of approximately 80:20 was adopted. Training data was split into 3 folds for cross-validation using an Expanding Window Splitter. The forecast horizon was defined, and model performance was assessed using a scale-free metric, the mean absolute scaled error (MASE).

5. RESULTS AND DISCUSSION

The mid-year population estimates show that Johannesburg accounts for the largest share of Gauteng's population at 37%, followed by Tshwane and Ekurhuleni, each at 25% (Figure 2). The district municipalities contribute relatively small proportions, with Sedibeng at 7% and West Rand at 6%. Together, the three metropolitan municipalities accommodate 87% of the province's residents, while the two district municipalities account for only 13%. The population of the municipalities is generally increasing over time (upward trend).

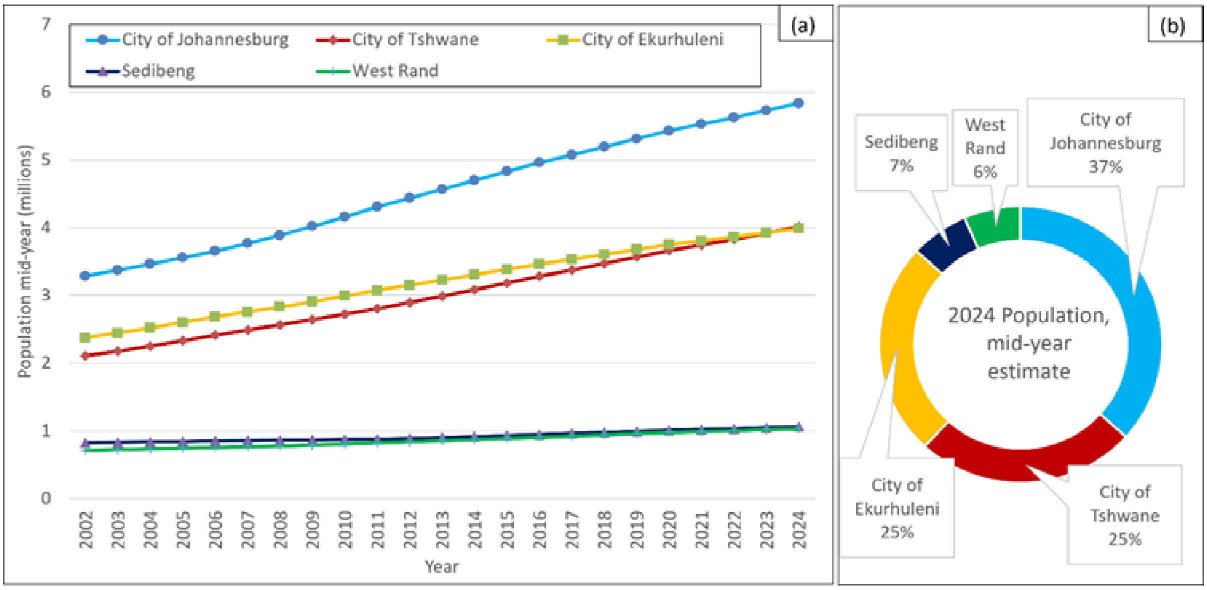

Figure 2: Illustration of the (a) trend in population growth, and (b) 2024 population distribution in Gauteng's municipalities, based on the 2024 mid-year population estimates

From 2002 to 2024, all municipalities exhibited steady population increase, though at different scales. Johannesburg shows the most pronounced growth, rising from just under 4 million in 2002 to close to 6 million by 2024. Tshwane and Ekurhuleni also display strong upward trajectories, with their populations converging at around 4 million in 2024. By contrast, Sedibeng and West Rand show slower and more modest growth over the entire period. Evaluation of the trained models based on average cross-validated scores is shown in Figure 3.

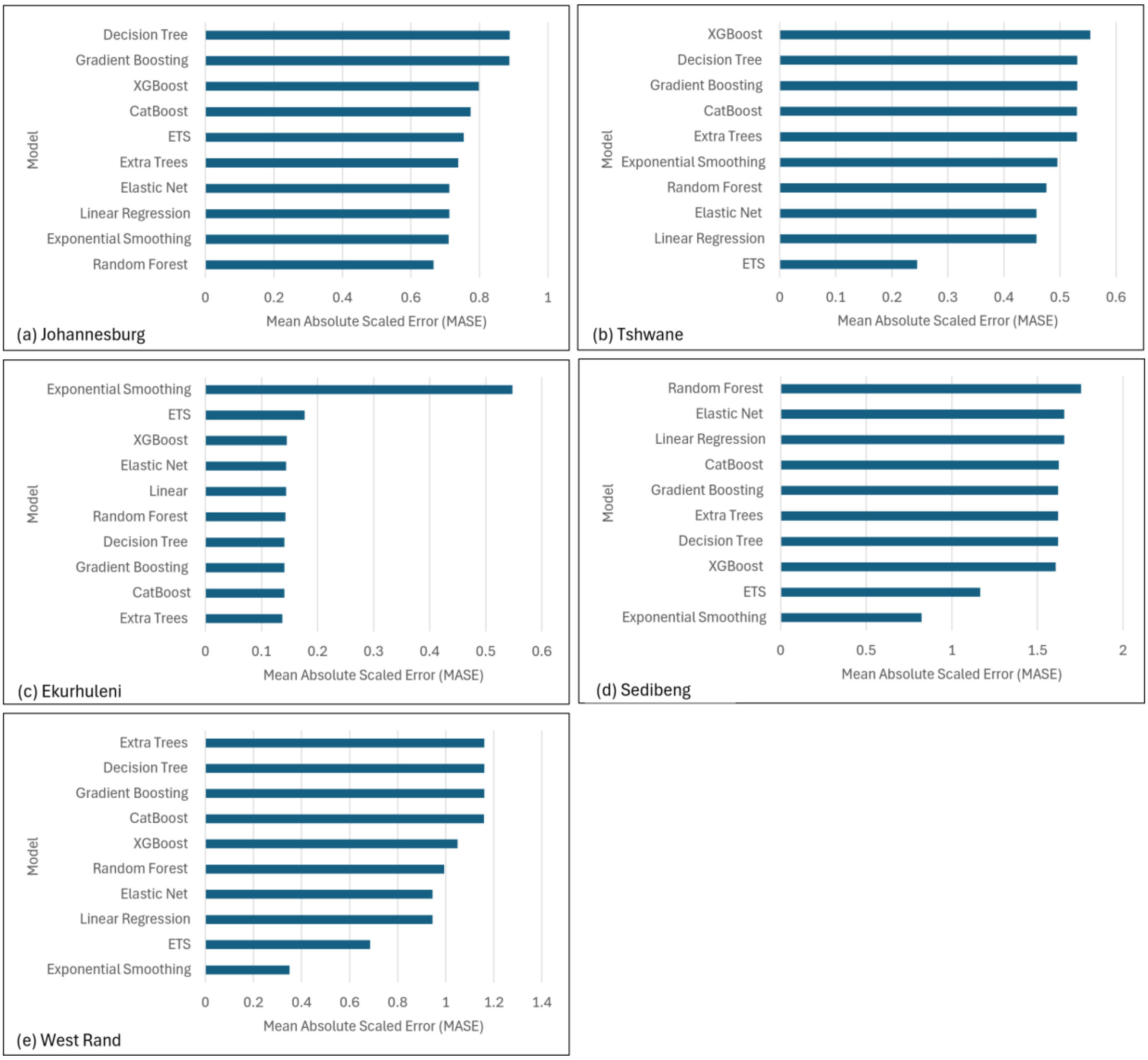

Figure 3: Comparison of the cross validation mean absolute scaled errors – (a) Johannesburg (b) Tshwane (c) Ekurhuleni (d) Sedibeng (e) West Rand

In Johannesburg, Random Forest is the best overall performer with an MASE of 0.666. In second position is Exponential Smoothing with an MASE of 0.710. Linear Regression and Elastic Net have nearly identical performance, and are tied in third place with an MASE of 0.712. In Tshwane, the ETS statistical model achieved the best accuracy with an MASE of 0.245. Linear Regression and Elastic Net followed in second place with an MASE of 0.458, followed by Random Forest with an MASE of 0.476. Exponential Smoothing is ranked in fourth place. In Ekurhuleni, Extra Trees is ranked in first place (MASE: 0.137), followed by CatBoost and Gradient Boosting (MASE: 0.141). In Sedibeng, Exponential Smoothing (MASE: 0.823), and ETS (MASE: 1.166) demonstrated exceptional accuracy.

XGBoost and Decision Tree are ranked third and fourth place respectively. In West Rand, Exponential Smoothing ranked highest achieving an MASE of 0.350. The final model performance scores after hyperparameter tuning are shown in Table 1. Although Random Forest delivered the lowest MASE for Johannesburg, it manifested a downward forecast trend which does not align with the observed population growth pattern in Gauteng. However, Exponential Smoothing with a slightly higher MASE had a better goodness-of-fit (R^2 = 0.866), and it generated an upward trend. Thus, Exponential Smoothing was adopted for the forecasts in Johannesburg.

Table 1: Final model performance scores, and forecast trend

Municipality	**Model**	**MASE**	**R^2**	**Forecast trend**
Johannesburg	Random Forest	0.417	0.803	Downward
	Exponential Smoothing	0.458	0.866	Upward
Tshwane	ETS	0.245	0.958	Upward
Ekurhuleni	Extra Trees	0.137	0.988	Upward
Sedibeng	Exponential Smoothing	0.823	0.691	Upward
West Rand	Exponential Smoothing	0.350	0.928	Upward

These top performing models were adopted for the population forecasts from 2025 to 2030 shown in Figure 4. Johannesburg, Tshwane and Ekurhuleni show strong upward trends throughout the forecast period. The 2025 forecast for Johannesburg is 6.10 million, and this increases to 6.84 million in 2030. In Tshwane, population estimates rise steadily from 4.14 million in 2025 to approximately 4.62 million in 2030. Ekurhuleni's population increases from 4.16 million in 2025 to about 4.53 million in 2030. Sedibeng exhibits a stable growth trajectory across the forecast horizon, with a population increase from 1.09 million in 2025 to 1.17 million in 2030. West Rand's population also shows a clear upward trend, rising from approximately 1.07 million to 1.16 million between 2025 and 2030. Generally, the future projections show continued upward trajectory for most Gauteng municipalities.

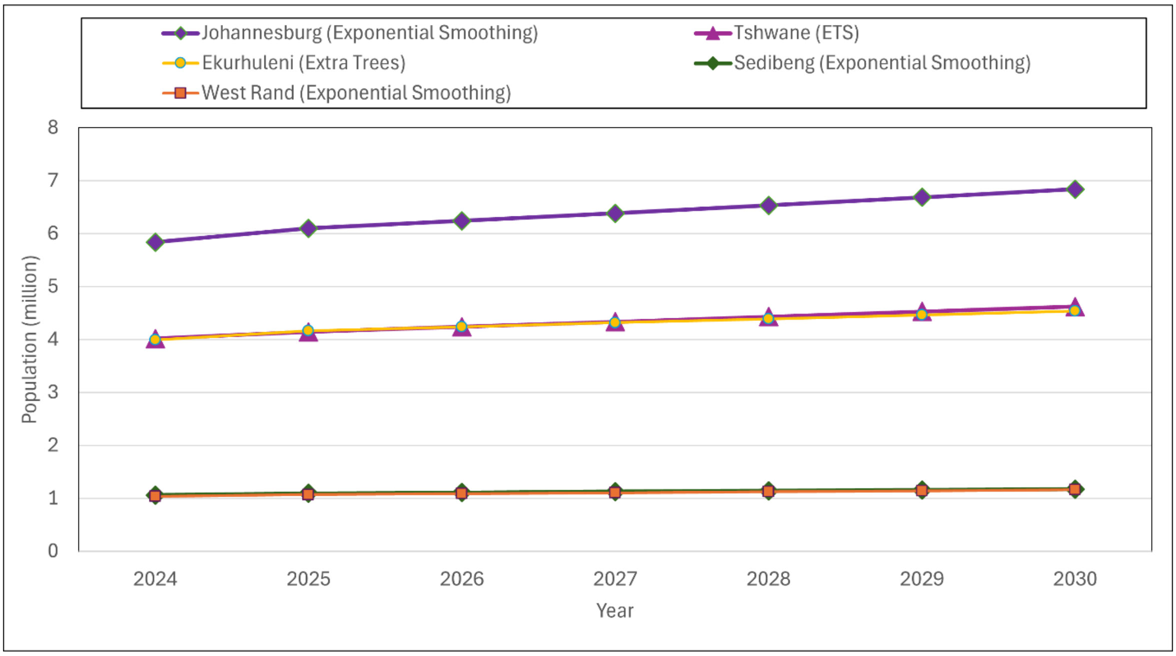

Figure 4: Population trend showing future forecast from 2025 – 2030

6. CONCLUSION

This study has proven the application of statistical and machine learning models for producing reliable population forecasts. Generally, there is a continued upward trajectory in future population growth for most Gauteng municipalities. Model performance was influenced by municipality population size, and this highlights the need for tailored approaches. These results are very informative for population forecasting, which is generally trend-based and seasonal. Statistical time-series models remain highly competitive in comparison to more advanced machine learning, and there is no single best model for all scenarios. The use of a tailored approach on a municipality-by-municipality case, speaks to how attentive municipal authorities need to be when considering which demographic modelling tools to adopt. This brings to light that population forecasts are not just numbers, but can be a strategic tool for empowering city authorities to transition from reactive governance to proactive planning and anticipating demographic shifts. This can create an environment for the design of smarter policies and lay the groundwork for sustainable urban futures. Future research should explore the combination of methods for improving the accuracy of population forecasts.

REFERENCES

Chaudhuri, S. (2024). Can Machine Learning Outperform Statistical Models for Time Series Forecasting? https://pub.towardsai.net/can-machine-learning-outperform-statistical-models-for-time-series-forecasting-73dc2045a9c5 (Date accessed: 15th September, 2025)

Copernicus (2025). Observer: Seven things we learned from the UN World Urbanization Prospects. https://www.copernicus.eu/en/news/news/observer-seven-things-we-learned-un-world-urbanization-prospects (Date accessed: 15th January, 2026).

Destatis (2026). International statistics - The largest cities worldwide in 2025. https://www.destatis.de/EN/Themes/Countries-Regions/International-Statistics/Data-Topic/Population-Labour-Social-Issues/DemographyMigration/UrbanPopulation.html (Date accessed: 15th January, 2026).

Fatima, S. S.W. & Rahimi, A. (2024). A Review of Time-Series Forecasting Algorithms for Industrial Manufacturing Systems. *Machines*, *12*(6), 380. https://doi.org/10.3390/machines12060380

Hewamalage, H., Bergmeir, C., & Bandara, K. (2022). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008

Bandara, K., Hewamalage, H., Liu, Y.H., Kang, Y., & Bergmeir, C. (2021). Improving the accuracy of global forecasting models using time series data augmentation. Pattern Recognition, 120, 2021, 108148. https://doi.org/10.1016/j.patcog.2021.108148

Schwabe, C. (2019). AfricaScope. South Africa's Population – A Fundamental Geospatial Dataset. https://africascope-sa.com/2019/03/20/south-africas-population-a-fundamental-geospatial-dataset/ (Date accessed: 12th September, 2025)

Statistics South Africa (2024a). Mid-year population Estimates, 2024. Presentation by the Statistician-General, Statistics South Africa. https://www.statssa.gov.za/publications/P0302/MYPE%20Presentation%202024.pdf (Date accessed: 13th September, 2025).

Statistics South Africa (2024b). P0302 - Mid-year population estimates, https://www.statssa.gov.za/?page_id=1854&PPN=P0302&SCH=73952 (Date accessed: 13th September, 2025).

UNECA (2024). As Africa's Population Crosses 1.5 Billion, The Demographic Window Is Opening; Getting The Dividend Requires More Time And Stronger Effort. https://www.uneca.org/stories/%28blog%29-as-africa's-population-crosses-1.5-billion%2C-the-demographic-window-is-opening-getting (Date accessed: 26th January, 2026).

UN GGIM (2018). The Global Fundamental Geospatial Data Themes. https://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/Fundamental_Data_Publication.pdf (Date accessed: 15th January 2026).

United Nations (2025). Press Release | Cities are home to 45 per cent of the global population, with megacities continuing to grow, UN report finds. https://www.un.org/sustainabledevelopment/blog/2025/11/press-release-wup2025/ (Date accessed: 15th January, 2026).

BIOGRAPHICAL NOTES

Dr Chukwuma Okolie is a geospatial professional, data scientist and licensed surveyor. He has a PhD in Geomatics from the University of Cape Town; research interests in data fusion, GIScience and remote sensing, Geospatial artificial intelligence, and urban and environmental modelling; and a proven track record of impactful teaching and research. He is a City Data Advisor with the AMALI Data Programme at the African Centre for Cities.

Mr Mawande Ngidi has close to 15 years of experience supporting public sector leaders with strategic and data-driven decision-support to improve government delivery. He is the AMALI Data Programme Lead, where he steers efforts to strengthen data systems and use across African cities. Mr Ngidi holds a Master of Philosophy in Urban and Regional Science from Stellenbosch University.

Mr Zukisa Sogoni is a skilled geospatial analyst, and town and regional planner with substantial experience in supporting South African cities through various predictive geospatial analytics. Zukisa plays a key role as a City Data Advisor in the AMALI Data Programme at the African Centre for Cities. Mr Sogoni-holds a Master of Science in Town and Regional Planning from the University of KwaZulu-Natal.

Ms Lesego Tshuwa is a built environment professional focused on working in and alongside governments to improve city operations. She is a City Data Advisor in the AMALI Data Programme, through which she contributes to supporting city leaders and their teams to improve practices in data management and use data to make government more efficient, accountable and equitable. Ms Tshuwa holds a Master of Science in Town and Regional Planning (Urban Studies) from the University of the Witwatersrand.

Dr Samukele Ngema is an urban and regional planner with experience in large infrastructure planning. He has experience advising on streamlining policies for implementation by government entities. As a City Data Advisor with the AMALI Data Programme, he coaches African mayors to effectively implement stated goals through data-driven decision making. He holds a PhD in Urban and Regional Planning from Stellenbosch University.

CONTACTS

Dr Chukwuma Okolie
Organisation: African Mayoral Leadership Initiative (AMALI) Data Programme
Address: African Centre for Cities, School of Architecture, Planning and Geomatics, University of Cape Town
City: Cape Town
COUNTRY: South Africa
Tel. +234 803 133 2527
Email: chuks.okolie@uct.ac.za
Web site: https://www.africancentreforcities.net/

Mr Mawande Ngidi
Organisation: African Mayoral Leadership Initiative (AMALI) Data Programme
Address: African Centre for Cities, School of Architecture, Planning and Geomatics, University of Cape Town
City: Cape Town
COUNTRY: South Africa
Tel. +27 66 302 1969
Email: mawande.ngidi@uct.ac.za
Web site: https://www.africancentreforcities.net/

Mr Zukisa Sogoni
Organisation: African Mayoral Leadership Initiative (AMALI) Data Programme
Address: African Centre for Cities, School of Architecture, Planning and Geomatics, University of Cape Town
City: Cape Town
COUNTRY: South Africa
Tel. +27 83 408 8945
Email: zukisa.sogoni@uct.ac.za
Web site: https://www.africancentreforcities.net/

Ms Lesego Tshuwa
Organisation: African Mayoral Leadership Initiative (AMALI) Data Programme
Address: African Centre for Cities, School of Architecture, Planning and Geomatics, University of Cape Town
City: Cape Town
COUNTRY: South Africa
Tel. +27 78 188 5293
Email: lesego.tshuwa@uct.ac.za
Web site: https://www.africancentreforcities.net/

Dr Samukele Ngema
Organisation: African Mayoral Leadership Initiative (AMALI) Data Programme
Address: African Centre for Cities, School of Architecture, Planning and Geomatics, University of Cape Town
City: Cape Town
COUNTRY: South Africa
Tel. +27 81 549 1660
Email: sam.ngema@uct.ac.za
Web site: https://www.africancentreforcities.net/