Abstract: Exploring approaches to improve small area population projections in Namibia.

Introduction

Access to accurate subnational population figures is essential for local and regional planning, resource allocation, and service delivery, especially in sectors like health and education. National statistics offices (NSOs) produce various statistic products, including intercensal population projections. These population projections are used to estimate changes between censuses, typically employing the cohort component method (CCM), as preferred by demographers and NSOs [1](Wilson et al., 2022). However, CCM demands substantial data and assumes demographic components remain constant throughout the projection period, posing challenges in small areas where change happens more frequently. As a result, cannot adequately capture significant demographic shifts within a decade, making these traditional methods less suitable for smaller-area projection (Wilson et al., 2022). This study aims to contribute to ongoing research on alternative data and methods targeting to enhance the production of small area population projections. The process is done by applying datasets from population and housing census, household surveys, and geospatial and mobile phone data to attempt to improve subnational population projections in Namibia. The results are assessed against existing 2020 census mapping data. After that, we will validate the results using the most recent Namibia Census 2023 population data, which is expected to become available in March 2024.

Data and methods

In the initial approach, the CCM is implemented using the 2011 census data and then incorporating survey data to bring new data into the method; these are summarised in Table 1. In step one the national and regional population projections are generated by simulating the NSA CCM, using the 2011 census as the base population, and maintaining constant component rates. Then, step two will incorporate survey data from the mid-term 2016 Namibia Intercensal Demographic Survey (NIDS). Here, age-specific fertility and mortality rates will be generated to run the projections from the base year (2011), update at the survey year, 2016 and forward to target year, 2020. Step three will follow the same pipeline as the previous step; however, it will update the rates with the 2013 Demographic Health Survey

[1] Wilson, T., Grossman, I., Alexander, M., Rees, P., Temple, J., 2022. Methods for Small Area Population Forecasts: State-of-the-Art and Research Needs. Population Research and Policy Review 41, 865–898. https://doi.org/10.1007/s11113-021-09671-6

(DHS). Both rates generated from the surveys data are applied in the projection on the fourth run, resulting in two-step component updates for this run. Within this context, we will analyse and compare the various sets of projection adjustments (refer to Table 1) against the 2023 census results to assess the performance of these approaches.

Table 1: Input data and approaches to test for subnational population projections.

Input data	**Approach**	**Projection horizon time**	**Data sources**
2011 census total population Age-specific mortality and Age-specific fertility rates Survival rates (migration)	Cohort component method (NSA version)	2020 - 2020	NSA 2011 Census Population and Housing
Compute updated age-specific fertility and mortality age-specific schedules from survey data 2013DHS	Updated cohort component with Survey data	2011-2013-2020	2013 Namibia Demographic Health Survey
Compute age-specific fertility and mortality age-specific schedules from survey data 2016NIDS	Updated cohort component	2011-2016-2020	2016 Namibia Intercensal Demographic Survey (NIDS-2016)
Both NIDS2016 and DHS 2013 are used to update fertility and mortality components	Updated cohort component	2011 – 2013 - 2016 -2020	2013DHS and 2016NIDS
Geospatial covariates including landcover for example, network of roads, waterways, protected areas, public facility locations, land use, lights at night, derived slope estimates	Random forest methods	2011 - 2020	Author generated projected population data; Stevens et al. (2015)
Migration component – mobile phone (call detail records), survey population data, census 2023 population data	Gravity-type based models	2011 - 2020	Lai et al. (2019)

The second approach explores the use of geospatial datasets in top-down methods to disaggregate the regional estimates into constituency-level projections, as illustrated in earlier research work [2](Stevens, 2015). During this stage, we will integrate migration rates derived from aggregated mobile phones data[3](Lai et al., 2019) to update the migration component of the projections. Geospatial data, mainly information on human settlements, has proven invaluable in identifying small settlements and urban expansions, contributing to more accurate population estimation [4](Wardrop et al., 2018). These details can be challenging

[2] Stevens, 2015. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLOS ONE. https://doi.org/10.1371/journal.pone.0107042

[3] Li, T., Dooley, C.A., Tatem, A.J., 2019. Exploring the use of mobile phone data for national migration statistics. Palgrave Communications 5, 34. https://doi.org/10.1057/s41599-019-0242-9

4 Wardrop, N.A., Jochem, W.C., Bird, T.J., Chamberlain, H.R., Clarke, D., Kerr, D., Bengtsson, L., Juran, S., Seaman, V., Tatem, A.J., 2018. Spatially disaggregated population estimates in the absence of national population and housing census data. Proceedings of the National Academy of Sciences 115, 3529–3537. https://doi.org/10.1073/pnas.1715305115

to capture precisely in census data and may not be adequately represented in population projections. Similarly, migration as one of the key components for population projections at small areas, is shadowed by outdated and non-existing data at these scales. Nonetheless, research analyses done in Namibia showed that mobile data hold promising opportunities to complement existing official statistics to update key indicators on local population mobility (Lai et al., 2019). These alternative data sources can capture factors related to the geographically uneven processes driving population change, which are often difficult to predict and account for using standard projection models at subnational levels.

Preliminary findings

Compared to the recent 2020 census mapping dataset, the preliminary results from the traditional cohort component method revealed potential discrepancies with areas experiencing significant changes over the decennial period. To identify the differences, the absolute percent error (APE) measure is computed by subtracting the observed census counts per region from the projected population and dividing the results by the observed census counts, to get the rates is multiplied by a hundred. In Figure 1, the maximum difference seen was 35.9 percent for Kunene region, while the average was 14.76 percent. These findings show significant variations in the results obtained through the CCM across the regions, with a standard deviation of 9.9 percent. Regions with fifty percent of rural area such as Otjozondjupa and Hardap has minimal APE under ten percent, showed in shades of purple in Figure 1.C and displayed in green in Figure 2.

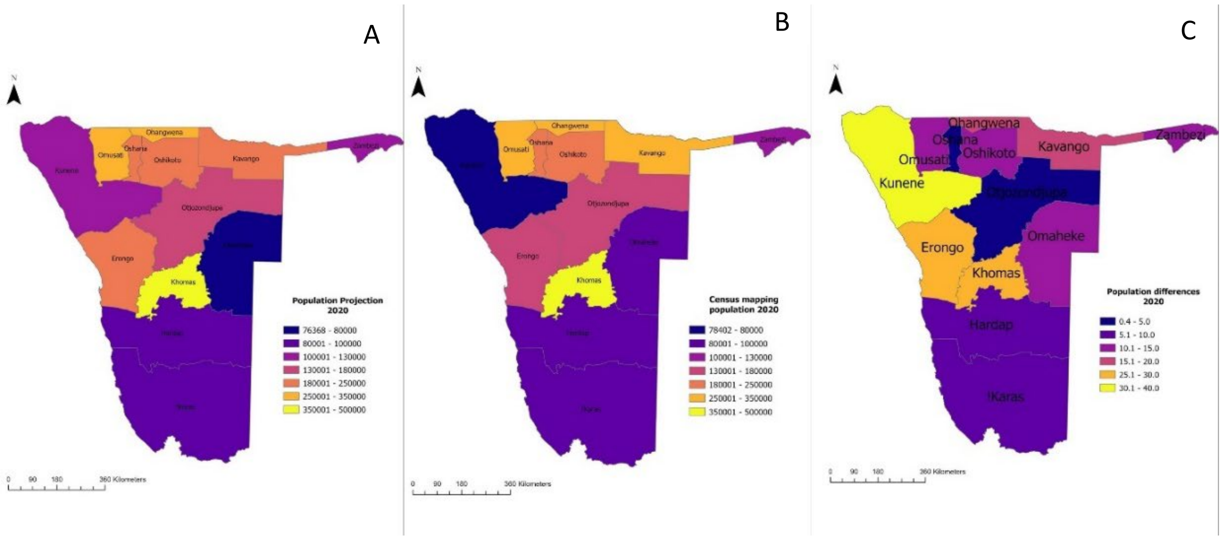

Figure 1 Map display the Absolute percent difference between projections and census mapping population by region for the target year 2020. (A) shows the population projected for 2020, (B) shows the census mapping population counts 2020 and (C) shows the absolute percent differences between the two population totals by regions

In Figure 2, we explore the correlation between these variations and the relationship between base period growth rates and the proportion of urban population within a given region. The blue regions, which formed less than 50 percent of the urban population and experienced a low growth rate (less than 2 percent) during the base period, exhibit a substantial APE exceeding 10 percent, is primarily attributable to demographic fluctuations during the projection period. On the other hand, purple regions, predominantly urban in nature, not only maintained a high growth rate of over 2.5 percent during the base period but also sustained this growth during the horizon period, also saw a high APE. Lastly, the green regions, characterized by approximately half of their population residing in urban areas, experienced a growth rate ranging from 1 to 2 percent during the base period. These maintained this growth rate throughout the horizon period, hence seeing a low APE.

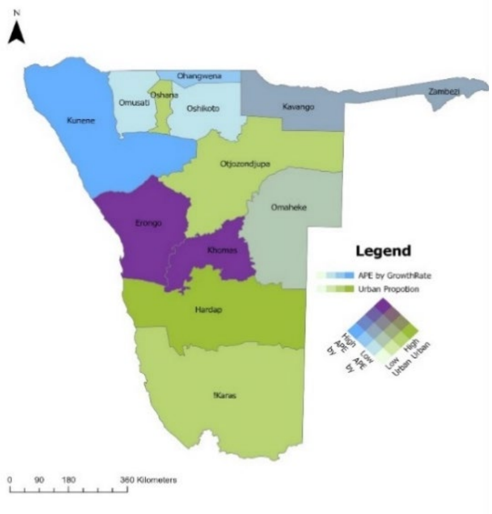

Figure 2 Map displaying the relationship between the percent differences by base growth rate and the proportion of urban for each region.

We aim to contribute to the ongoing research on enhancing subnational population projections by testing these alternative data sources. We seek to achieve this by leveraging updated, cost-effective datasets that can provide a more accurate and comprehensive understanding of population dynamics at lower geographical scales. The goals will be achieved by testing various approaches where survey data are introduced in the cohort component method to update the demographic component rates at mid-point and produce a series of projections, then test model-based approaches including top-down random forest methods and gravity-type based methods to produce migration statistics. These model outputs will be validated against the 2020 census mapping and 2023 census to examine their performances.