



Charles Darwin University

Estimating internal migration between Northern Territory SA3 regions

Dyrting, Sigurd; Taylor, Andrew

Published in:
Northern Institute Research Brief Series

Published: 01/02/2025

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):
Dyrting, S., & Taylor, A. (2025). Estimating internal migration between Northern Territory SA3 regions. *Northern Institute Research Brief Series*.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Estimating internal migration between Northern Territory SA3 regions

February 2025

AUTHORS

Sigurd Dyrting and Andrew Taylor
Northern Institute



Summary

Projecting migration between small areas such as SA3s in the Northern Territory, and other areas with relatively small populations, is difficult given the small number of migration 'events' on which to estimate probabilities. The aim of this paper is to explore the feasibility of estimating internal migration probabilities for a multi-regional population projection model at 1-year age and time intervals. To do so, we examine the feasibility of applying a polynomial smoothing technique called PTOALS to estimate migration probabilities between the Northern Territory's nine SA3 regions at one-year intervals by age, Indigenous status, and sex.

We find that it is feasible to calculate demographically plausible estimates for internal migration between NT SA3 regions. Much of the process can be automated in R scripts, but some steps are manual (extraction of data from TableBuilder, review of the calibration reports) and subjective (assessments of plausibility and adjustment of method parameters).

With minor modifications the method we use here could be adapted to SA3-specific interstate migration. It could also be adapted to estimate inter-SA2 migration, but migration ratios at the SA2 level will likely display very high levels of noise. To overcome this, it might be possible to get reasonable estimates by pooling data from multiple Censuses. Otherwise, the SA3-level estimates given here could be used to assign reasonable values to SA2-level ratios with some additional assumptions.

Introduction

Internal migration probabilities are important for projecting changes in population compositions and to understand the composition and size of flows to and from geographic regions with relatively small populations. The current Northern Territory Government population projection model uses SA3 net migration rates for 5-year age groups and time intervals (Department of Treasury and Finance, 2024). Net migration rates are useful for combining internal and interstate migration into a single variable which is easier to estimate than its constituent rates but can sometimes imply implausible demographic futures (e.g., negative populations). The aim of this paper is to explore the feasibility of estimating internal migration probabilities for a multi-regional population projection model at 1-year age and time intervals. In the next section we describe the data sources and the method we applied before outlining results in which we explore the implications of different choices of age and time intervals, as well as investigating the appropriate parameter settings for handling sample and perturbation noise in the input data. In the closing section we summarise our recommended approach and discuss how it can be adapted for estimating SA3 interstate migration and SA2 internal migration probabilities for the NT and elsewhere.

Data and Methods

For each SA3 we constructed a two-region geographic framework consisting of the SA3 itself (for example Barkly) and for the second region we aggregated the remaining SA3s into the remainder of the NT (i.e., not-Barkly). We then used TableBuilder (Australian Bureau of Statistics, 2012, 2017, 2022) to extract counts of movers from the SA3 to the remainder, by Indigenous status (Indigenous, Non-Indigenous), sex (Female, Male), age, and region of usual residence (e.g. Barkly, not-Barkly), filtered on usually residing in that region one or five years previous. This was done for the three age groupings given in TableBuilder (1-year, 5-year, and 10-year), giving 6 tables for each region (two time intervals by three age groupings). These tables were used to calculate sample aggregate internal migration probabilities (the number of people who moved to another SA3 divided by the number of people who remained in the NT) and sample out-migration probabilities for each region (the number of people who moved out of the region to another region of the NT divided by the number of people previously in that region and who remained in the NT). For each region we also we extracted counts of movers by Indigenous status, sex, age, region of usual residence, filtered on usually residing in that region one or five years previous. These tables were used to calculate the 8 sample migration ratios for each region (the number of people who moved to a given destination divided by the number of people who moved out of the region and who remained in the NT).

Figure 1 shows the workflow we used to estimate migration probabilities from TableBuilder data. In the first step, we converted region-specific spreadsheet files exported from TableBuilder into dataframes suitable for manipulation in R (Posit team, 2023; R Core Team, 2023). In the second step, we created objects used by the estimation library (Dyrting, 2018). In the third step, we estimated the probabilities from the objects using the P-TOPALS/P-spline method (Dyrting & Taylor, 2024). In this method, the probability $m(O, D, x)$ of migrating from origin O to destination D at age x is equal to the product of the out-migration probability $m(O, x)$ and the migration ratio $c(O, D, x)$

$$m(O, D, x) = m(O, x) \times c(O, D, x).$$

Estimation is in three stages: first aggregate migration is smoothed, then out-migration is smoothed by adjusting the aggregate migration to best fit the sample probabilities, then migration ratios are smoothed by adjusting a constant ratio to best fit the sample values. At each stage we specify the type and amount of smoothing through a set of smoothing parameters. In the fourth and fifth steps we create diagnostic plots and compile them into separate calibration reports for each population. In the sixth step we review the calibration reports. If any of the estimates do not appear plausible, we adjust

the relevant estimation parameters and return to the estimation step. This continues until we consider all estimates plausible. All steps in the workflow are automated by R scripts (see Table 1 in the Appendix) except for the review of the calibration reports and adjustment of the estimation parameters, which are conducted manually. We conducted this workflow for Census years 2011, 2016, and 2021.

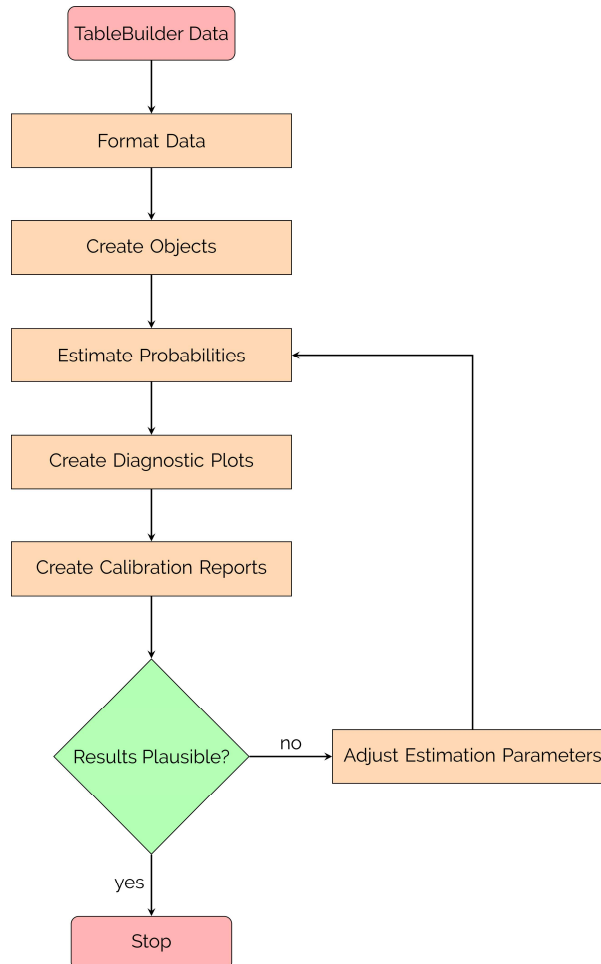


Figure 1: Workflow for estimating internal migration probabilities from Census data

Results

Aggregate Migration

Aggregate migration measures the probability that a person who remains in the NT migrates to another SA3. Of the three probabilities we will estimate, sample aggregate migration probabilities display the least dispersion because they are calculated for the largest exposed population (the population of NT that did not move interstate). Furthermore, since it measures the average inter-SA3 migration probability, it can be used as the standard for a relational method that estimates origin-specific out-migration, or as a proxy for a region's out-migration.

We estimated aggregate migration using polynomial smoothing. In the P-TOPALS method, this is achieved by setting the standard migration curve to equal 1 at all ages. We accounted for TableBuilder's perturbation noise using the method given in Dyrting (2020). This method approximates the effect of perturbations by adjusting the exposed population downwards by an amount proportional to the variance of the perturbations which we estimated from the difference between the total population

and the sum of stayers and movers. Nonetheless, sometimes we did see evidence that perturbations were affecting the estimates at older ages and, when this was the case, we adjusted the default estimation parameters. When we reviewed the calibration reports, we found it useful to compare estimates based on different age groups, because an unusually high or low sample probability that was present in one age-grouping but not in another was likely to be the result of perturbation.

The most important parameter for improving estimates with P-TOPALS is the penalty, which controls the amount of smoothing that is applied. The greater the penalty, the smoother the estimate, but this increased smoothness is at the cost of an increase in fitting error. Our default parameters (see Table 3 in the Appendix) specify that the penalty should be determined automatically using the Bayesian Information Criterion (BIC, Schwarz, 1978). Sometimes the default penalty would under-smooth, in which case we would increase it to one of the alternate values given in Table 3. Much of the age structure of aggregate migration will flow through to any schedule that is determined from it using a relational method. For this reason, it is important to ensure aggregate migration estimates are free from implausible features.

Figures 2, 3, and 4 show estimates of 2021 aggregate migration over a 1-year time interval and 1-year, 5-year, and 10-year age groups respectively. We see that Indigenous internal migration probabilities are lower than non-Indigenous, and the age structure of Indigenous migration is far less pronounced, only becoming visually apparent for 5-year and 10-year age groups.

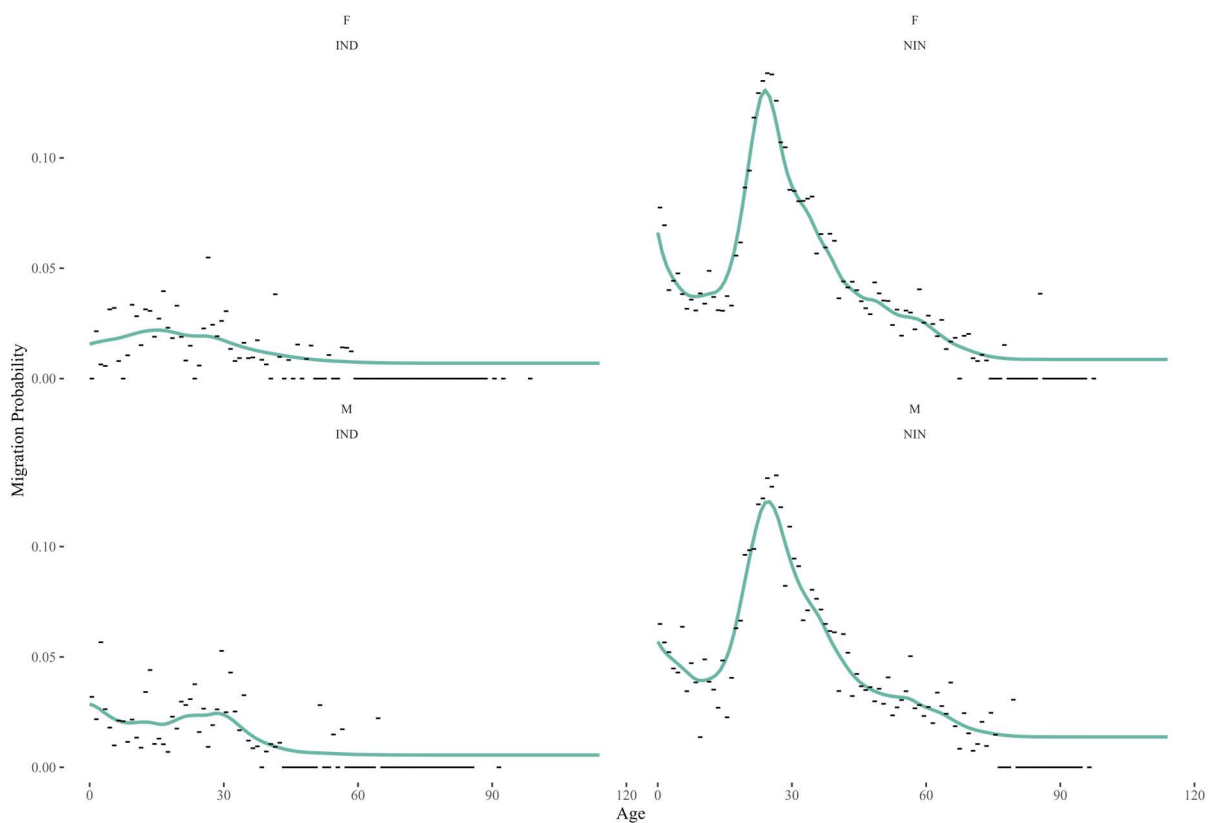


Figure 2: Aggregate migration probability by sex, Indigenous status, and age: Census 2021, 1-year interval, 1-year age groups. Points, sample probabilities for each age; Curves, estimated probabilities. See Table 2 in the Appendix for a glossary of codes.

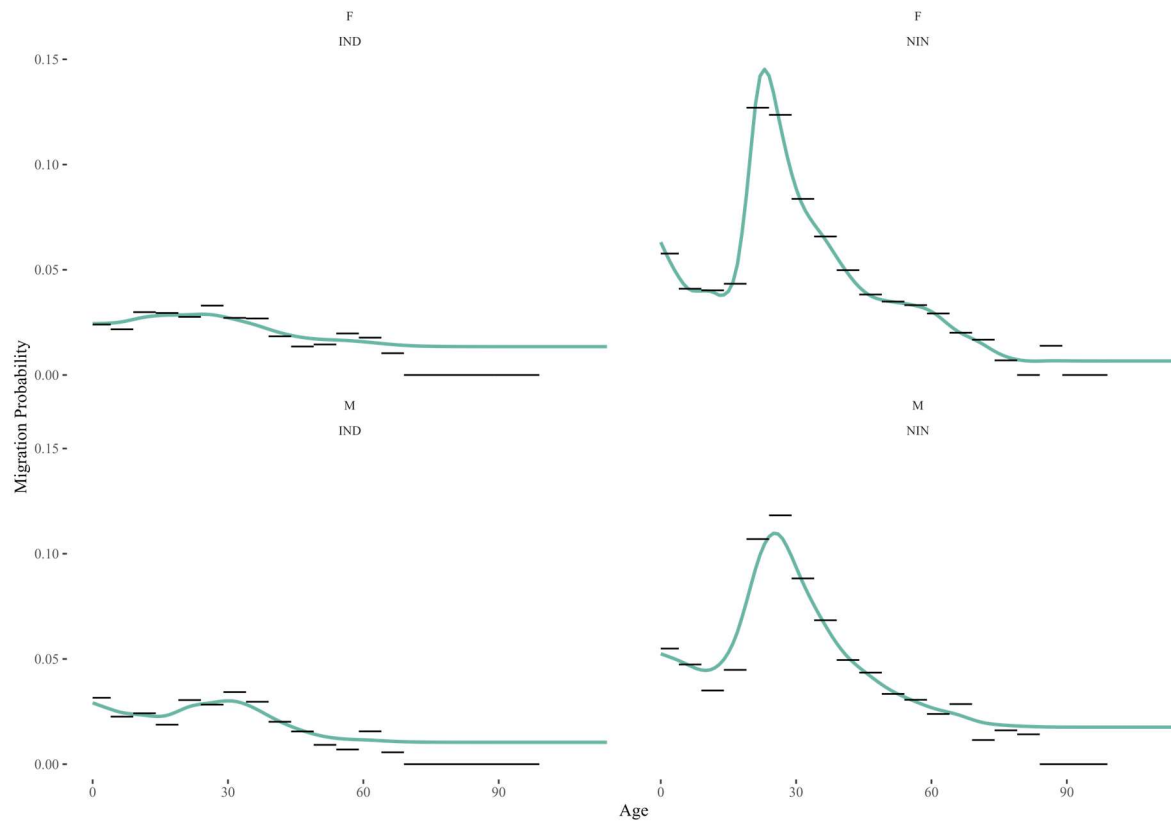


Figure 3: Aggregate migration probability by sex, Indigenous status, and age: Census 2021, 1-year interval, 5-year age groups. Horizontal lines, sample probabilities for each age group; Curves, estimated probabilities. See Table 2 in the Appendix for a glossary of codes.

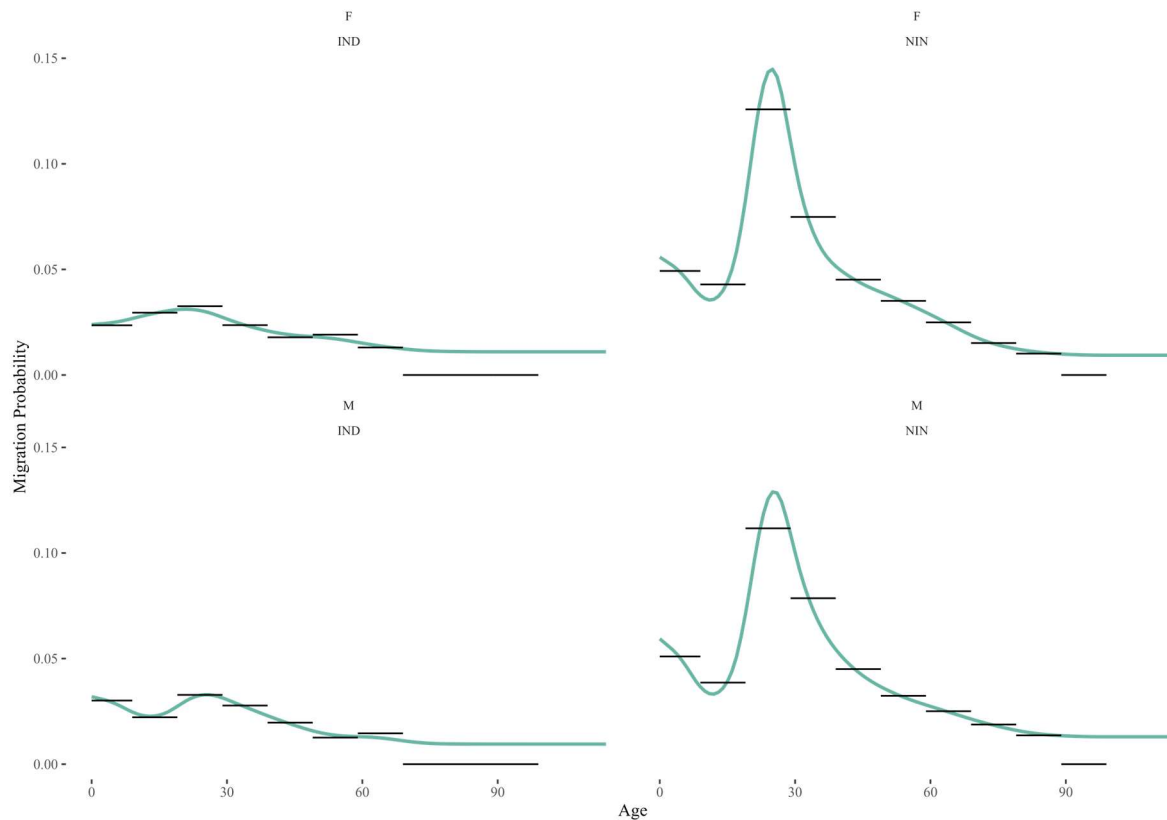


Figure 4: Aggregate migration probability by sex, Indigenous status, and age: Census 2021, 1-year interval, 10-year age groups. Horizontal lines, sample probabilities for each age group; Curves, estimated probabilities. See Table 2 in the Appendix for a glossary of codes.

Out-Migration

Out-migration measures the probability that a person from a specific region within the NT who remains living in the NT migrates to another region within the NT. Out-migration sample probabilities display more dispersion than for aggregate migration because the exposed population is smaller, being the number of people in the region at the beginning of the interval who did not leave the NT, compared to the overall number of people who did not leave the NT for aggregate migration.

We estimated out-migration by the P-TOPALS method, using aggregate migration as the standard and the default penalty given in Table 3. For Indigenous females and males there is significant dispersion in the sample probabilities even when using 5-year or 10-year grouped ages (see Figure 5). Consequently, we needed to increase the penalty to one of the alternate values to obtain plausible schedules from both these age groups. For non-Indigenous females and males, sample out-migration probabilities from SA3s outside Greater Darwin, other than for Alice Springs, showed high levels of dispersion (e.g. Daly-Tiwi-West Arnhem (DTW-SA3) in Figure 6). For these regions we sometimes needed to increase the penalty to one of the alternate values.

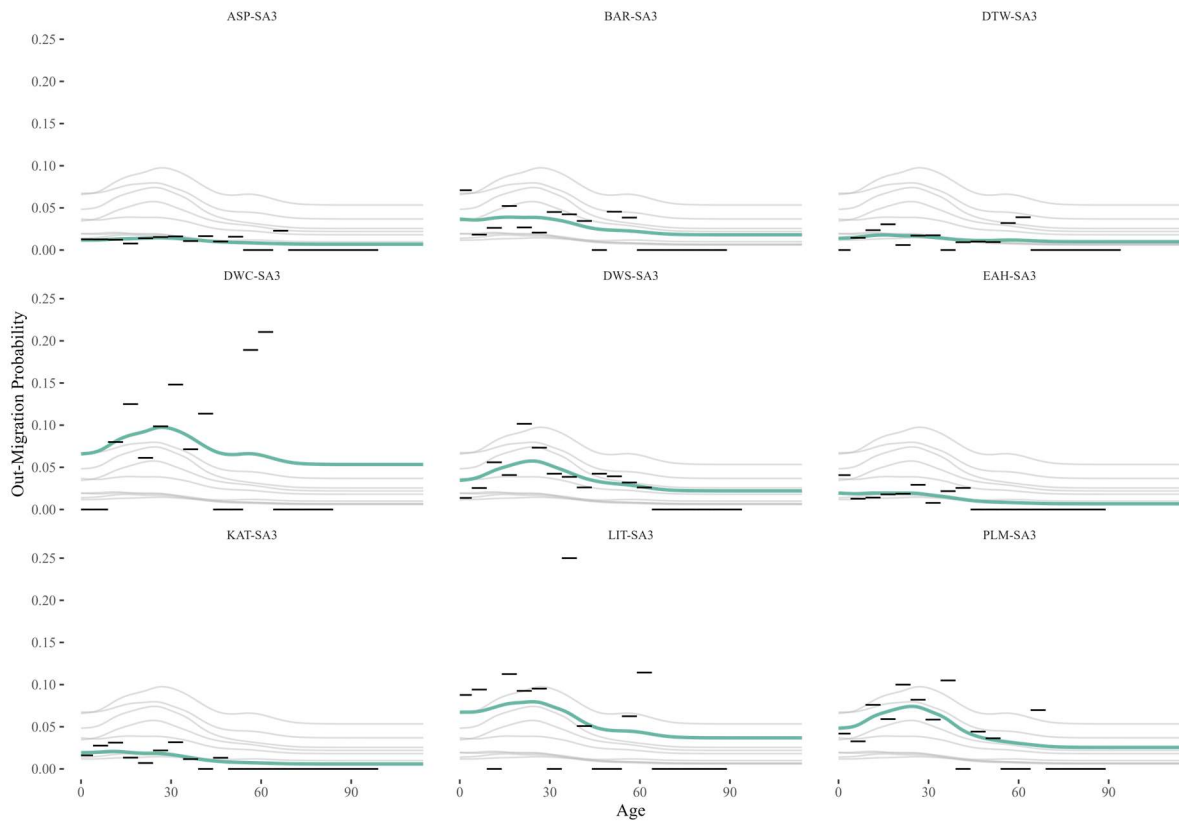


Figure 5: Indigenous female out-migration probability by region and age: Census 2021, 1-year interval, 5-year age group. Horizontal lines, sample probabilities for each age group; Curves, estimated probabilities. See Table 2 in the Appendix for a glossary of codes.

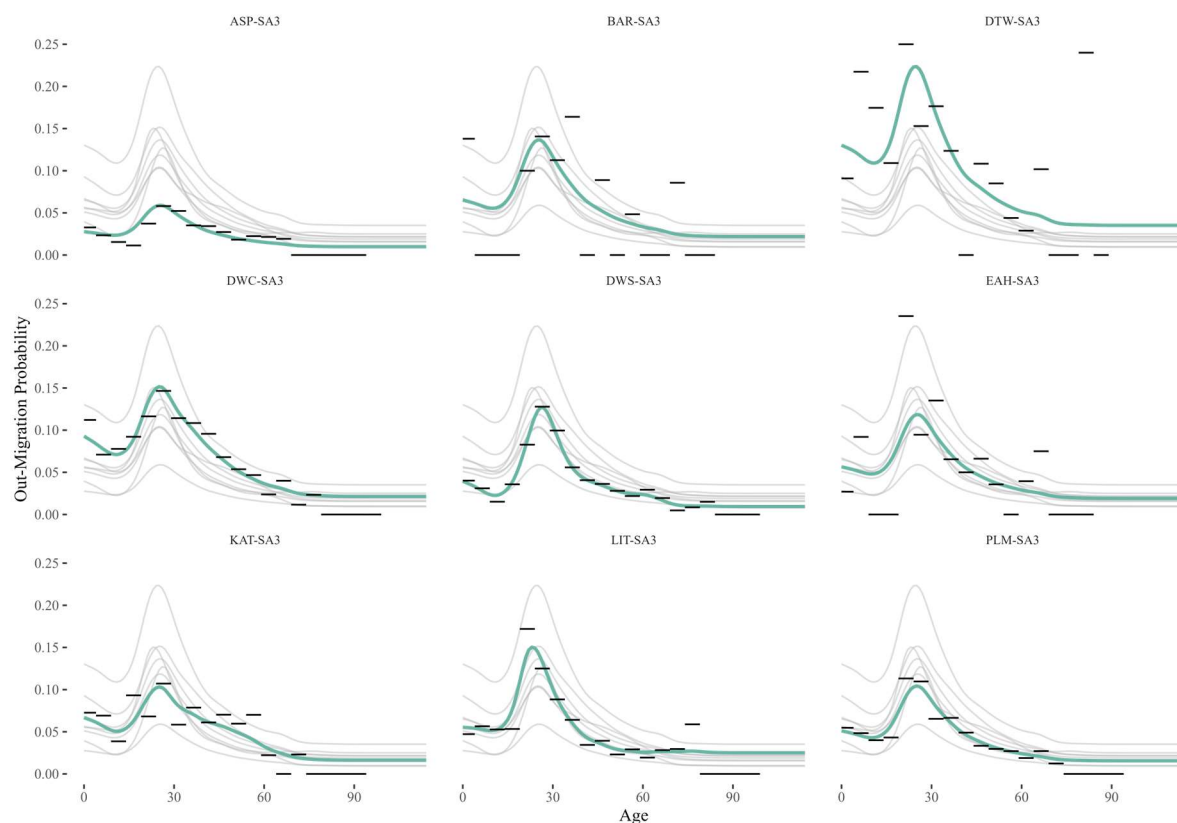


Figure 6: Non-Indigenous male out-migration probability by region and age: Census 2021, 1-year interval, 5-year age group. Horizontal lines, sample probabilities for each age group; Curves, estimated probabilities. See Table 2 in the Appendix for a glossary of codes.

Migration Ratios

Migration ratios measure the probability that a person moves to a specific destination region conditional on them leaving a given origin region. Of the three probabilities we estimate, migration ratios display the greatest dispersion because the exposed population, the number of people who leave a region, is the smallest. For example, Figure 7 shows Census 2011 sample migration ratios by destination region for non-Indigenous males migrating from Darwin Suburbs. We see that, even for this relatively mobile population and with 10-year age groups, the age-structure is unclear. Therefore, we felt that we could only use the data to estimate the average over all age groups, which we did by using a very high penalty in the P-spline estimation method (see Table 3). Even so, in 6 cases no migration ratios could be estimated because the TableBuilder tables of movers by destination gave no movers for any age (either because there were none or because the actual number had been perturbed to zero). All such tables were for a 1-year migration interval and a 1-year (5 cases) or 5-year (1 case) age group.

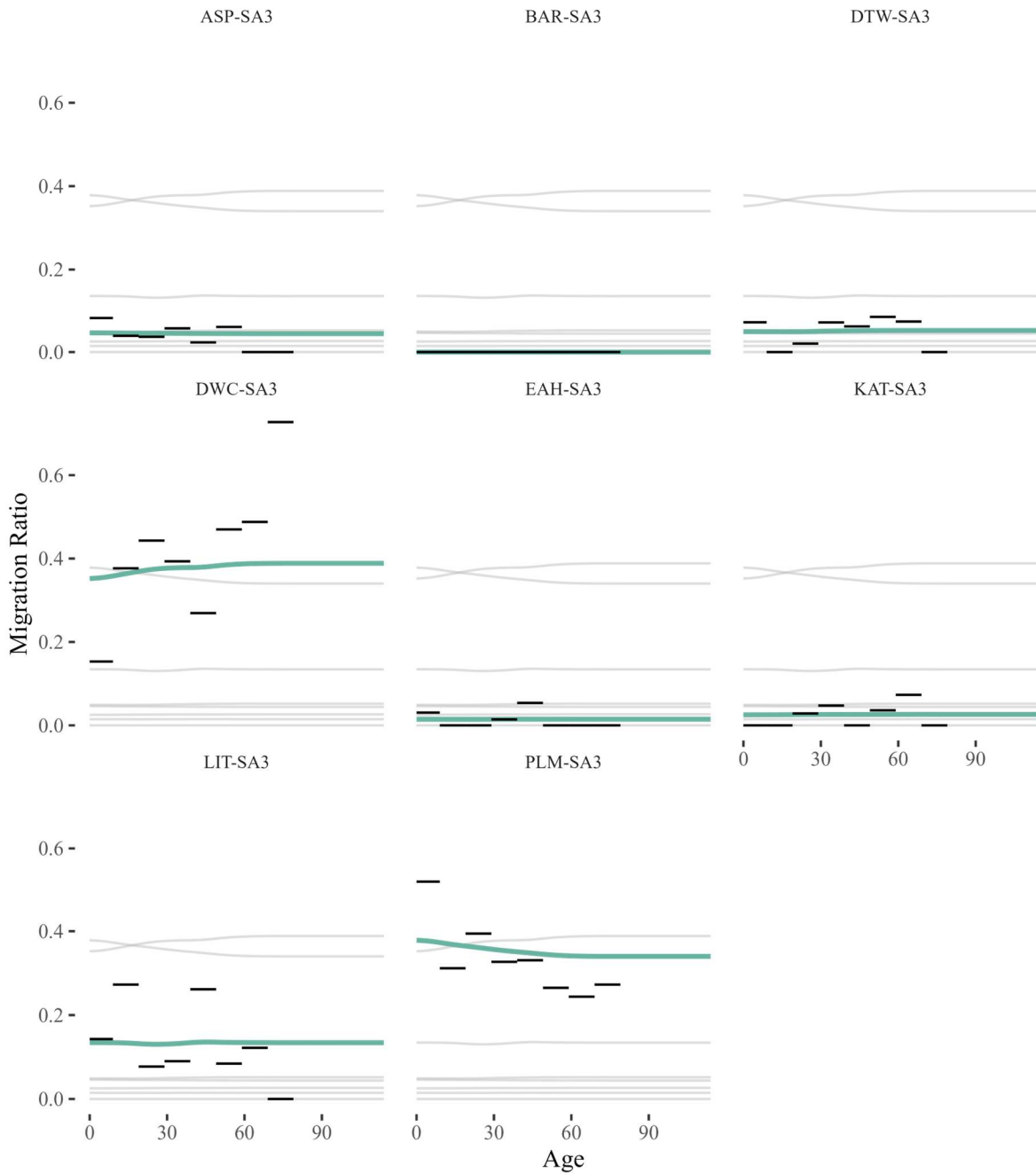


Figure 7: Darwin Suburbs non-Indigenous male migration ratios by destination region and age: Census 2011, 1-year interval, 10-year age group. Horizontal lines, sample ratios for each age group; Curves, estimated ratios. See Table 2 in the Appendix for a glossary of codes.

Treating migration ratios as almost constant across ages can be justified by previous findings showing they do not display as strong a dependence on age as out-migration probabilities (Dyrting & Taylor, 2021; Rogers et al., 2002).

Discussion

The three probabilities (aggregate migration, out-migration, and migration ratios) have successively smaller exposed populations (people that have remained in the NT, people previously in a region, and people previously in a region who have out-migrated to another region). This is reflected in the

increasing amount of dispersion of these probabilities due to sample and perturbation noise and consequently makes resolving their age structure increasingly difficult. We have found that it is feasible to calculate demographically plausible estimates for internal migration between NT SA3 regions. Much of the process can be automated in R scripts, but some steps are manual (extraction of data from TableBuilder, review of the calibration reports) and subjective (assessments of plausibility and adjustment of method parameters), although, within the estimation approach we employed, subjective choice was confined to a single variable: the smoothing penalty.

We found there were persistent features across the three census years. For example, Indigenous out-migration probabilities could be grouped into high and low clusters corresponding to Greater Darwin and Rest of NT respectively (Figure 5). For non-Indigenous out-migration, Alice Springs always had the lowest probabilities (Figure 6). We sometimes found it useful to compare estimates from different age groups to get a feeling for which features of the data were ‘real’ and which might be due to perturbation noise. But if only a single age-grouping was desired for each probability then we suggest:

- Giving preference to 5-year age groups to estimate aggregate and origin-specific out-migration, and
- Giving preference to 10-year age groups to estimate migration ratios, or alternatively, calculate ratios from data with no age breakdown.

Given the large dispersion of sample migration ratios, as further improvements to the approach used here, we also suggest:

- Investigate pooling data from multiple Censuses, and
- Keep ratios constant in time when projecting future probabilities.

With minor modifications the method we use here can be adapted to SA3-specific interstate migration. For out-migration, the origin labels remain the same and destinations D label the destination states. For in-migration, destination labels remain the same and the origins O label the state of origin. The method can also be adapted to estimate inter-SA2 migration. For example, first estimate aggregate SA2 migration. Then use aggregate SA2 migration as a standard for estimating SA2 migration aggregated at each SA3 level $m_2(O_3, x)$. This is the probability that a person in SA3 O_3 moves out of their SA2. The quantities estimated in these first two steps should not be noisier than SA3 aggregate and out-migration because the denominators (exposed population at NT and SA3 levels) are the same and the numerator (SA2 moves) includes all inter-SA3 moves. Aggregate SA2 migration at the SA3 level can then be used either as a standard for estimating out-migration for the constituent SA2s or as a proxy for out-migration from those SA2s.

The last step is to estimate inter-SA2 migration ratios. Migration ratios at the SA2 level will likely display very high levels of noise. It might be possible to get reasonable estimates by pooling data from multiple Censuses. Otherwise, the SA3-level estimates given here could be used to assign reasonable values to SA2-level ratios with some additional assumptions. For example, let O_2 and D_2 be SA2 origin and destination and O_3 and D_3 be the corresponding SA3s, and assume we have estimated SA3 out-migration $m(O_3, x)$, SA3 ratios $c(O_3, D_3, x)$, and SA3-aggregated inter-SA2 probability $m_2(O_3, x)$. Figure 8 illustrates the two possibilities for a move from O_2 to D_2 . The destination can be in the same SA3 O_3 (an intra-SA3 move) or it can be in a different SA3 D_3 (an inter-SA3 move). Since an inter-SA3 move is also an inter-SA2 move we must have

$$m_2(O_3, x) \geq m(O_3, x).$$

For an intra-SA3 move we could assign the migration ratio

$$c(O_2, D_2, x) = \left[1 - \frac{m(O_3, x)}{m_2(O_3, x)} \right] \times \frac{1}{|O_3| - 1}.$$

The first term on the right is the probability of remaining in the origin SA3 conditional on out-migrating from the origin SA2. The second term is the probability of migrating to D_2 , assuming all possible destinations within O_3 are equally likely¹, where $|O_3|$ is the number of SA2s in O_3 . For an inter-SA3 move we could assign the migration ratio

$$c(O_2, D_2, x) = \frac{m(O_3, x)}{m_2(O_3, x)} \times c(O_3, D_3, x) \times \frac{1}{|D_3|}.$$

The first term on the right is the probability of moving out of the origin SA3 conditional on out-migrating from the origin SA2. The second term is the probability of migrating to D_3 conditional on out-migrating from O_3 . The third term is the probability of migrating to D_2 , assuming all possible destinations within D_3 are equally likely², where $|D_3|$ is the number of SA2s in D_3 .

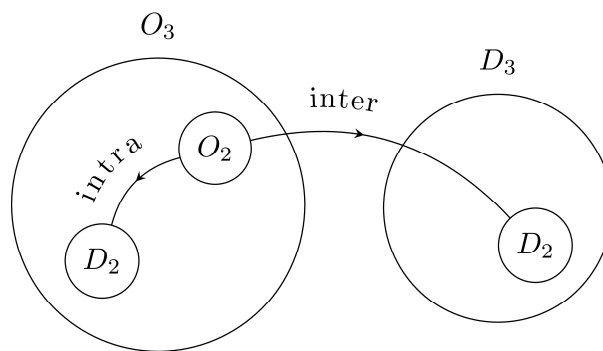


Figure 8: Migration between SA2s: inter-SA3 and intra-SA3 moves

¹ Alternate assumptions would be easy to implement. For example, destinations could be weighted by their population.

² Ibid.

References

- Australian Bureau of Statistics. (2012). *Table Builder - 2011 Census, internal migration [Internet]* [Dataset]. Canberra: ABS. <https://www.abs.gov.au/statistics/microdata-tablebuilder/tablebuilder>
- Australian Bureau of Statistics. (2017). *Table Builder - 2016 Census, internal migration [Internet]* [Dataset]. Canberra: ABS. <https://www.abs.gov.au/statistics/microdata-tablebuilder/tablebuilder>
- Australian Bureau of Statistics. (2022). *Table Builder - 2021 Census, internal migration [Internet]* [Dataset]. Canberra: ABS. <https://www.abs.gov.au/statistics/microdata-tablebuilder/tablebuilder>
- Department of Treasury and Finance. (2024). *Northern territory population projections 2024 release*. Northern Territory Government. https://treasury.nt.gov.au/__data/assets/word_doc/0003/1402797/NT-Population-Projections-2024-Release.docx
- Dyrting, S. (2018). *Osier: A library for demographic calculations*. <https://doi.org/10.13140/RG.2.2.13400.39682>
- Dyrting, S. (2020). Smoothing migration intensities with P-TOPALS. *Demographic Research*, 43(55), 1607–1650. <https://doi.org/10.4054/DemRes.2020.43.55>
- Dyrting, S., & Taylor, A. (2021). Smoothing destination-specific migration flows. *Annals of Regional Science*, 67, 359–383. <https://doi.org/10.1007/s00168-021-01051-4>
- Dyrting, S., & Taylor, A. (2024). Estimating complete migration probabilities from grouped data: A methods protocol for developing a global human internal migration database. *PLoS ONE*, 19(1), e0315389. <https://doi.org/10.1371/journal.pone.0315389>
- Posit team. (2023). *RStudio: Integrated development environment for R*. Posit Software, PBC. <http://www.posit.co/>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rogers, A., Raymer, J., & Willekens, F. (2002). Capturing the age and spatial structures of migration. *Environment and Planning A*, 34, 341–359.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>

Appendix

Table 1: Scripts and configuration files used in each stage of the workflow.

Stage	Scripts
Format Data	nt_internalmig_sa3_make_migdata.R
Create Objects	nt_internalmig_sa3_make_migobjs.R
Estimate Probabilities	nt_internalmig_sa3_expand_migdata.R, ControlExpandMigdata.xlsx
Create Diagnostic Plots	nt_internalmig_sa3_make_plots.R
Create Calibration Reports	nt_internalmig_sa3_make_reports.R, nt_internalmig_sa3_report.Rmd

Table 2: Glossary of codes

Code	Description
F	Female
M	Male
IND	Indigenous
NIN	Non-Indigenous
DWC-SA3	Darwin City SA3
DWS-SA3	Darwin Suburbs SA3
LIT-SA3	Litchfield SA3
PLM-SA3	Palmerston SA3
ASP-SA3	Alice Springs SA3
BAR-SA3	Barkly SA3
DTW-SA3	Daly-Tiwi-West Arnhem SA3
EAH-SA3	East Arnhem SA3
KAT-SA3	Katherine SA3

Table 3: Parameters used to estimate internal migration probabilities

Migration Type	Method	Default Penalty	Alternate Penalties
Aggregate Migration	PTOPALS	BIC	100, 10*BIC
Out-Migration	PTOPALS	BIC	100, 10*BIC, 100*BIC
Migration Ratios	PSPLINE	1000	