ORIGINAL RESEARCH

Application of P-TOPALS for Smoothing Input Data for Population Projections 'At the Edge'

Sigurd Dyrting[1] · Andrew Taylor[1] · Tom Wilson[2]

Abstract

Sparsely populated areas of developed countries are regions of great demographic diversity and dynamism. While they remain strategically and economically important, trends in urbanization and technology have increased their relative sparsity and isolation making centralized government, service delivery and planning a challenge. Populations of their sub-jurisdictions are small and often exhibit significant heterogeneity in key demographic characteristics, not least between their Indigenous first residents and non-Indigenous citizens. Development of projection models for these areas is challenged by significant input data paucity, biases and structural issues related to the data collection and estimation architectures in place to gather input data across diverse and small populations. While this is the case, the demand for and importance of projections is no less for sparsely populated areas than elsewhere. Variants of the cohort component model are important tools for population projections for SPAs, with their grounding in the demographic accounting equation and modest input requirements. Nevertheless, to attain fit-for-purpose input data requires demographers to consider and select from a growing number of methods for smoothing issues with input data for projections for these regions. In this article we analyze the contributions of recent advances in methods for estimating fertility, mortality, and migration rates for small and diverse populations such as those in SPAs, focusing on the very sparsely populated jurisdiction of the Northern Territory of Australia. In addition to the contributions of our method itself, results at the detailed level demonstrate how abnormal and challenging 'doing' projections for sparsely populated areas can be.

Keywords Population projections · P-TOPALS · Northern territory · Smoothing · Expanding · Sparsely populated areas

---

Extended author information available on the last page of the article

Introduction

Deriving accurate input data for sub-national population projections is critically important to their accuracy and shelf-life (see Simpson et al., 2019). At sub-national levels, demographers face a range of impediments in transitioning raw fertility, mortality and migration data into smoothed and fit-for-purpose input data for projections. In examining mortality data, for example, Wilson (2018) identified a paucity of time series data on consistent geographical boundaries; small numbers of occurrences (births and deaths) across some areas and cohorts; and costs, complexities and time demands for statistically modeling in preparing data for inputting to projections and, indeed, the time and skills demands needed to program and develop the model itself (Wilson, 2018, p. 2). These challenges extend to fertility data (Dyrting, 2018; Saha et al., 2023) and migration data (Dyrting, 2020; Raymer et al., 2020) at the sub-national level, even for nations such as Australia which has a comparably robust national statistical system underpinned by a central statistical agency.

One cause of these issues is that often data is provided in abridged age groups which must be expanded into single year ages to provide a suitable fertility or mortality schedule for projections which are for single-year of age based outputs (Dyrting, 2018; Dyrting & Taylor, 2023). This is the case, for example, in Australia for sub-national jurisdictions with age specific death rates and overseas migration (see ABS, 2022a, 2022c). The choice of method for the expansion of abridged data into single year ages is important to the results of the projections as different methods will produce differential outcomes over time. At the sub-national level, choice of method can have a big influence on comparative projection results.

While an abundance of methods for estimating complete schedules from abridged data are available (for example, Baili et al., 2005), demographers continue to work to augment existing techniques or develop new ones. The main aim is, of course, to improve accuracy, however important secondary conditions are at play including ensuring flexibility in the application of the method and developing tools with relatively easy computational parameters and requirements (Schmertmann, 2014). The science of smoothing and expanding data is, therefore, an ongoing continuous improvement process for demographers.

For sparsely populated sub-national jurisdictions with comparably small populations, reliable population projections for planning, policy making, investment decisions and evaluative work are no less important than for more populous states. While not all sparsely populated areas (herein SPAs) have small populations (northern Russia, for example, has significantly sized major cities), in general their populations, and thus population data, are hallmarked by small numbers across all demographic components. This is reflective of settlement patterns for SPAs in nations such as Canada, the USA, Australia and the Scandinavian North. By and large, these feature one city which is relatively large in population size in comparison to the entire SPA, but which is also very small in comparison to the major cities within the same nation. The remaining population outside of such cities is spread very thinly over a very large land mass (Taylor, 2014). The Northern Territory (herein NT or Territory), Australia's least populated State or Territory with a resident population of just 250,600, as the focus for this study, capaciously meets this description.

Literature on projection models for small areas highlights there are additional and substantial challenges for the demographer (Baffour & Raymer, 2019), or as Wilson and colleagues (2022) put it, demographers have a limited toolbox from which to project for small populations. Input data are generally of poorer quality, trends in such data may be highly variable over successive time periods, and small differences in accuracy can produce large errors soon into the projection horizon (Taylor, 2011; Wilson et al., 2022). This means the shelf-life for projections or the point at which levels of uncertainty relative to population size become too high, is inversely related to size of the population being projected (Simpson et al., 2019; Wilson & Shalley, 2019).

In spite of these challenges and the likelihood that uncertainty will creep into the projections in a relatively short time interval from the base, Taylor (2011) identified there are secondary benefits to developing population projections for SPAs. Importantly, the coalescing of key stakeholders like governments, academics and projections users around developing the model, its assumptions and the input data provides the opportunity for input data quality to be collectively reviewed and for techniques to be applied or variations of methods to be developed to improve input data quality and, thus the accuracy of projections.

To highlight some of the challenges inherent in developing population projections for SPAs and to demonstrate the importance of new or enhanced methods to meet such challenges, in this paper we utilize a cohort component model to produce projection outputs for the NT by applying a smoothing and expansion method known as P-TOPALS, which has been developed by demographers to help minimize impacts from statistical noise on projections and general demographic modeling and analysis. We initially outline the core demographic attributes of the NT and provide context around some population projections work already undertaken. We then give an overview of the P-TOPALS method and highlight some of its salient properties. We then describe the methods associated with data preparation using P-TOPALS as well as the projections model and assumptions used to produce outputs with and without smoothing which can be compared to highlight the benefits and discuss these in relation to the amount and complexity of work required to assess and comment on the value of such work in the context of SPAs.

Population Projections and the Northern Territory of Australia

With a population of just 250,600 (ABS, 2022 g), the NT is the least populous and most sparsely populated of Australia's eight States and Territories. While it accounts for 17.5% of the national landmass, its population is just 1% (ABS, 2022 g; Geoscience Australia, 2004). The NT's demographic composition is markedly different to the other seven jurisdictions. First, it has a one-third population share of Indigenous Australians, way above the national share at 3.8% (ABS, 2022d). With comparably high fertility rates, especially for young mothers, and higher mortality rates at all ages, that population is far younger than the non-Indigenous population. Partly because of this, but also derived from the age profile for internal migration exchanges, the NT has a young population with a median age of 33.4 for men and 33.7 for women (ABS,

2023a), compared to the national median of 37.7 and 39.4. A further distinguishing feature is the population's male bias – with 103.3 men per 100 women compared to 98.5 nationally (ABS, 2023a).

The non-Indigenous population of the NT have high rates of internal migration exchanges and it generally loses population to elsewhere in Australia on a net basis (Dyrting et al., 2020). Nevertheless, a more stable population of long-term residents is evident and the age profile of these, combined with aging in the Indigenous population are underpinning a notable structural aging process (Taylor & Payer, 2017; Temple et al., 2020; Zeng et al., 2015). Still, the NTs annual rate of population turnover, or 'churn' is comparably high at around 20%. Contributing to this, and complicating notions of who is resident for projection purposes, several types of 'special populations' have ongoing representation. Defense force personnel (navy, air force and army) are one, including contingents of overseas forces such as the US Marines who maintain a permanent rotation of more than 2,000 troops in the north of the NT. Recent defense reviews and announcements indicate that the size of the defense workforce will grow significantly in the coming decade (for example, Australian Government, 2023). Other special populations include workers engaged on large and major projects, especially in the construction phase. For example, a recently constructed liquid natural gas processing facility in Darwin engaged more than 7,000 workers during construction, equating to 3% of the population at the time. Most of these lived elsewhere and it is not clear how many were captured in the 2016 Census which was undertaken at the same time as the peak of construction. Finally, there are always Fly-in-fly-out (FIFO) and drive-in-drive-out (DIDO) populations present across the NT, including in the resources and government or service sectors which are prominent.

In aggregate, the demographic characteristics and presence of heterogeneous populations outlined above make population estimation and projections for the NT a comparatively difficult task. As well as the issue of past swings in fertility, mortality and migration data from sparsity of numbers, biases in input data are prominent. Included are pre-existing biases in Census data, from which estimates are derived and subsequently used for jump-off populations (Wilson & Shalley, 2019) and introduced bias from the methods used to transform Census data into Estimated Resident Population data. While these biases exist for all States and Territories, the difficulties inherent in conducting a Census with a young, geographically dispersed and sparse population with a high Indigenous composition and male bias makes input data biases an order of magnitude larger for the NT (Taylor, 2011, 2014). This means the choice of techniques for overcoming biases and noise evident in input data through smoothing techniques is important.

Population projections for the Northern Territory are produced by the Australian Bureau of Statistics every five years, with the jump-off year based on rebased population estimates derived from the five-yearly Australian Census of Population and Housing (see for example, ABS, 2017). However, these do not disaggregate by Indigenous status; a critical issue for the Northern Territory given the one third Indigenous population and interactions between the Indigenous population and others (i.e., partnering and childbirth). Rather, the ABS produces a separate set of 'estimates and projections' for Indigenous Australians each five years (see ABS, 2019).

The absence of a model which concurrently projects by Indigenous status led to a partnership between local academics and the Northern Territory Government in 2009 which aimed to develop a locally suited cohort component projections model in which Indigenous population is projected separately and interactions between the two populations are facilitated (see Wilson, 2009). Over time, the model was continuously improved and has been used to produce three iterations of bi-regional (NT and Australia) and sub-NT regional projections by Indigenous status (NT Department of Treasury and Finance, 2019).

While the NT has locally developed model which is tailored to the needs of government and other users, the size of the task in updating the projections and the need to continually work on improving their accuracy means it is necessary to assess the relative strengths and accuracy of smoothing techniques for the input data sets.

P-TOPALS

In this section we give an overview of the P-TOPALS method and those of its properties useful to preparing inputs for a population projection. The method estimates a complete schedule of age-specific quantities r_x for single year of ages from $x = 0$ to a maximum age $x = \omega$ (equal to 60 for fertility and 110 for other rates) from g observed rates $_n\widetilde{r}_i$ for $i = 1, \ldots, g$. The observed quantity can be subject to noise (due to finite sample size or disclosure avoidance processes), and age abridgment ($n > 1$).

P-TOPALS (Penalised Tool for Population Analysis with Linear Splines) is a non-parametric relational framework that expresses the quantity of interest r_x at age x as a product of a standard curve $\widehat{r}_x$ and the exponential of a weighted sum of spline functions B_x

$$r_x = \widehat{r}_x \exp\left(B_x \cdot \theta\right), \quad (1)$$

where the weights θ are chosen to fit the observed rates by maximizing an objective function that balances fitting errors and smoothness of that age profile (Dyrting, 2018; Dyrting et al., 2020, 2022)

$$L\left(\theta\right) = L_r\left(\theta\right) - \frac{\lambda}{2}\theta' \cdot D' \cdot D \cdot \theta, \quad (2)$$

where D is the first-order difference matrix (Eilers & Marx, 1996), and $L_r\left(\theta\right)$ is the log likelihood function for a Poisson (for mortality, fertility, internal migration, and emigration rates) or normal (for immigration numbers) process

$$L_r\left(\theta\right) = \begin{cases} _nN' \cdot \left(_n\widetilde{r}\log_n r - _n r\right), & \text{Poisson} \\ -\iota' \cdot \frac{1}{2\sigma^2}\left(_n\widetilde{r} - _n r\right)^2, & \text{Normal} \end{cases} \quad (3)$$

Here $_nr$ is the abridged rate calculated from the smoothed curve, ι is a vector of ones, and $_nN$ is the observed sample size (population exposed). After specifying the smoothness penalty λ, the optimal set of spline weights θ can be found by iterated linear regression (see Dyrting et al. (2020) for details). There are several criteria for choosing the optimal value for λ (Eilers & Marx, 1996). Two popular methods are to minimize the Akaike information criterion (Akaike, 1974) or the Bayesian information criterion (Schwarz, 1978). These methods seek to balance the smaller deviance of an improved fit against the increase in effective dimension needed to achieve it, giving a small penalty when uncertainty is small ($_nN$ is large or σ^2 small) and a large penalty when it is large ($_nN$ small or σ^2 large) (Dyrting et al., 2020).

P-TOPALS can be viewed as both a non-parametric method, through the spline component $B \cdot \theta$, and a relational method, where the schedule is specified as modified version of an exogenous standard $\widehat{r}$. When viewed as a non-parametric method, the role of the B-spline is to improve the fit of the standard to the observed rates (Dyrting et al., 2020). This allows any method (e.g. parametric) to be included in the estimation problem via the standard. As a relational method, the role of the standard profile $\widehat{r}$ is bounded by the two extremes of a small and large penalty (Dyrting et al., 2020). For a small penalty its use is to model those parts of the age-specific profile that are non-polynomial. Examples of such features can be found in most inputs, and include for vital processes the rapid drop in death rates after birth and the rapid rise in fertility rates in young adults (Dyrting, 2018), for internal migration the student peak (Dyrting et al., 2020), and for the population age distribution stationary features from special populations and non-stationary features such as the change in numbers from the pre to post baby boom cohorts (Dyrting et al., 2022). For a large penalty, the standard determines the entire age profile up to a constant factor. This latter role of the standard also occurs when extrapolating rates into the open age interval because a constant set of weights over this age interval will have both a zero smoothness penalty and be sufficient to fit the rate referencing this interval (if any).

Data and Methods

The principal inputs to a population projection model are the jump-off population by age and sex and the age-specific rates for the four components of change, fertility, mortality, internal migration and overseas migration, all of which we estimate from data available on the ABS website (ABS, n.d.). For the Northern Territory the relatively small numbers in either the numerator or denominator means that estimates of rates as ratios of events and exposed populations will display significant random variations with age due to sample noise. Additional uncertainty in these quantities can be the result of disclosure-avoidance processes used by national statistical agencies when preparing data for publication. Examples of these processes include noise injection, abridgment, age top-coding, and rounding (ABS, 2021; USCB, 2022). Methods designed to protect the anonymity of small populations will have a larger impact on the estimation of rates for sparsely populated areas as well as numbers in advanced ages than for more populous jurisdictions and younger ages.

To estimate the components of changes from noisy data we use primarily P-TOPALS. In addition to P-TOPALS we use the calibrated splines estimator (CS) to estimate a standard for fertility and mortality data. CS is a non-parametric method for smoothing and expansion that produces demographically plausible profiles by penalizing fits that deviated significantly from a reference set of shapes derived from historical fertility and mortality datasets (Dyrting & Taylor, 2023; Schmertmann, 2014).

After estimating the starting values for the components of change, it is necessary to forecast their future values. Three years on from 2020, net overseas migration has increased to approximately two-thirds of the 2019 (pre COVID-19) level, However, Australian households continue to experience the impacts of the COVID pandemic on family formation and living costs (ABS, 2022f, 2023b), so that it is not yet clear what the post-pandemic trends in the components of change are or will be in the near future. For this reason, we selected future fertility, mortality, migration and immigration numbers based on 2021 values. Holding rates for input data constant also facilitates direct comparisons between outputs based on smoothed data and outputs from un-smoothed input data.

Once the input data is estimated, the population by sex and single year of age is projected using SASPOPP (State and Substate Population Projection Program), a bi-regional cohort-component model (Wilson, 2017). In SASPOPP, populations for the Northern Territory and the rest of Australia are projected forward in time at one year increments accounting for births, deaths, interstate migration, and overseas migration. Births are calculated by multiplying age-specific fertility rates by the female population from age 15 to 49. Deaths, interstate out-migration and overseas out-migration are calculated by multiplying the corresponding rates and population from ages 0 to 105+, and overseas in-migration by age is given by age-specific immigration flows. The population in the jump-off year was given by the 2021 estimated resident population by single year of year (ABS, 2022 g) with the 100+age group decomposed into single year of age groups and 105+ using a combination of the Extinct Cohort and modified Survivor Ratio methods (Wilson & Terblanche, 2018).

Results

Fertility

State and Territory births and population estimates for 1975 to 2021 by single year of age were obtained from ABS (2022b). Births and population for the rest of Australia were calculated by aggregating all states and the ACT. Single-year age-specific fertility rates for ages 15 to 49 were calculated by dividing births by population. The single-year rates were smoothed using fertility P-TOPALS (Dyrting, 2018) using a standard obtained by fitting the rates with fertility calibrated splines (Grigorieva et al., 2020; Schmertmann, 2014).

Figure 1 shows 2021 raw and smoothed fertility rates for the NT. The most salient feature of the NT's fertility schedule is that it is bimodal, with a peak near age 20 that is primarily driven by high fertility rates for young Indigenous women and a non-Indigenous peak near age 30. The latter is similar to the mode for the remainder of
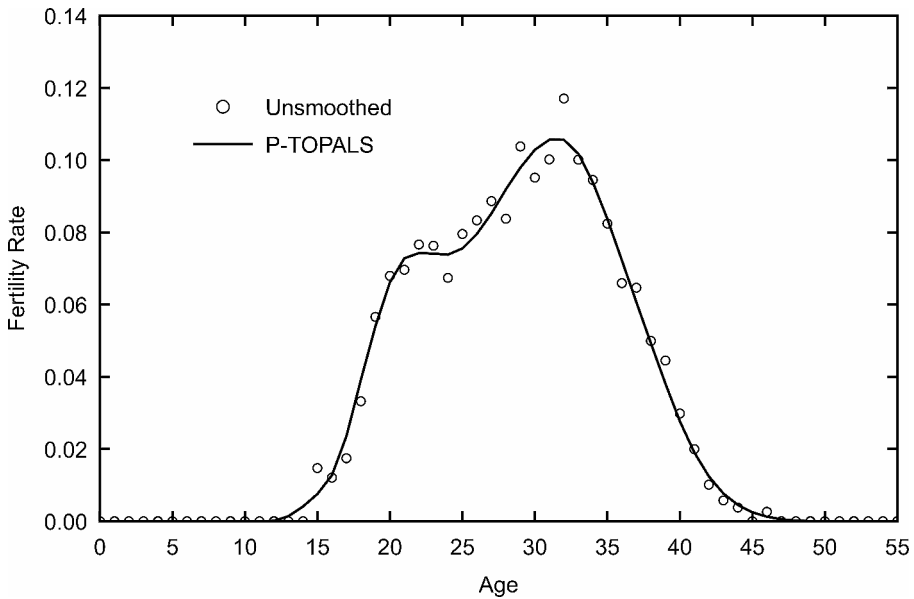
Fig. 1 Fertility rates, Northern Territory, 2021. Based on ABS data

Australia (Fig. 2). For both regions, fertility rates were trending downwards prior to 2020 and then increased in 2021.

Mortality

State and Territory deaths and population estimates for 1971 to 2021 by single year of age were obtained from ABS (2022a). Deaths and population for Australia (ex NT) were calculated by aggregating all states and the ACT. Single-year age-specific death rates for ages 0 to 99 and the open interval 100+were calculated by dividing deaths by population. Death rates were first smoothed using mortality calibrated splines (Dyrting & Taylor, 2023) and the resultant curve then used as a standard for mortality P-TOPALS (Dyrting, 2018), but the latter did not lead to a significant improvement in fit and so we report here the calibrated spline fit only.

Figure 3 shows 2021 raw and smoothed death rates for the NT. Figure 4 shows 2021 raw and smoothed death rates for Australian (ex NT). Differences in the size of respective populations combined with low absolute rate levels account for the much greater levels of noise seen in the NT death rates compared to the rest of Australia, with the NT showing some ages with zero death rates. For both regions life expectancy (Table 1) has been increasing over the period 2016 to 2021 driven by reductions in adult death rates.

Internal Migration

Northern Territory and rest-of-Australia one-year origin-destination tables for census years 2006, 2011, 2016, and 2021 by single year of age were obtained from ABS
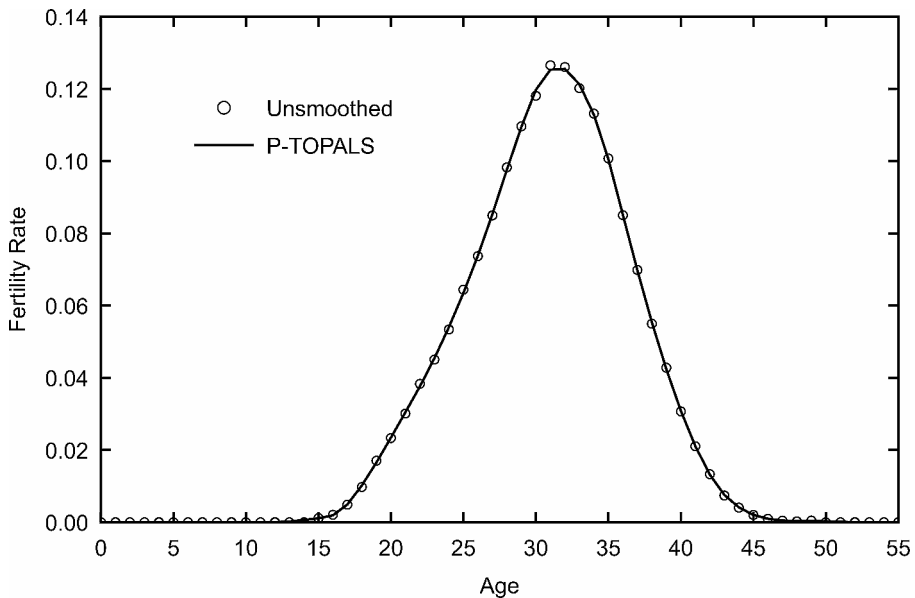
Fig. 2 Fertility rates, Australia (ex NT), 2021. Based on ABS data

(2022i). Single-year age-specific out-migration probabilities for ages 0 to 114 were calculated by dividing movers by the origin population. Migration probabilities were smoothed using migration P-TOPALS (Dyrting, 2020) using a standard obtained by fitting probabilities with the Student Peak Model Migration Schedule (Wilson, 2010).

Figures 5 and 6 show 2021 raw and smoothed out- and in-migration probabilities for the Northern Territory respectively, illustrating considerable dispersion due to the small absolute number of internal migrants at each age over a one-year interval. The smoothed schedules highlight some qualitative differences in out-migrants and in-migrants, with the former showing prominent student and retirement peaks while for the latter the labor force peak is the most significant feature. Migration intensities have decreased over the period 2016–2021 with the exception of in-migration of females in their twenties. From 2006 to 2016 the age at peak migration intensity had been increasing but in 2021 it decreased for both female and male in-migration (Table 1). Figures 5 and 6 also illustrate the ability in P-TOPALS to use the standard to introduce non-polynomial features to the fit, in this case the student peak.

Overseas Migration

State and Territory overseas arrivals and departures for 2004 to 2021 by five-year age group and rounded to the nearest multiple of 10 were obtained from ABS (2022c). Arrivals and departures for Australia (ex NT) were calculated by aggregating all states and the ACT. Abridged age-specific emigration rates for five-year age intervals 0 to 60 and open interval 65+ were calculated by dividing departures by population estimates. Single-year age-specific immigration numbers and emigration rates were calculated by expanding abridged numbers and rates using two methods: splines and
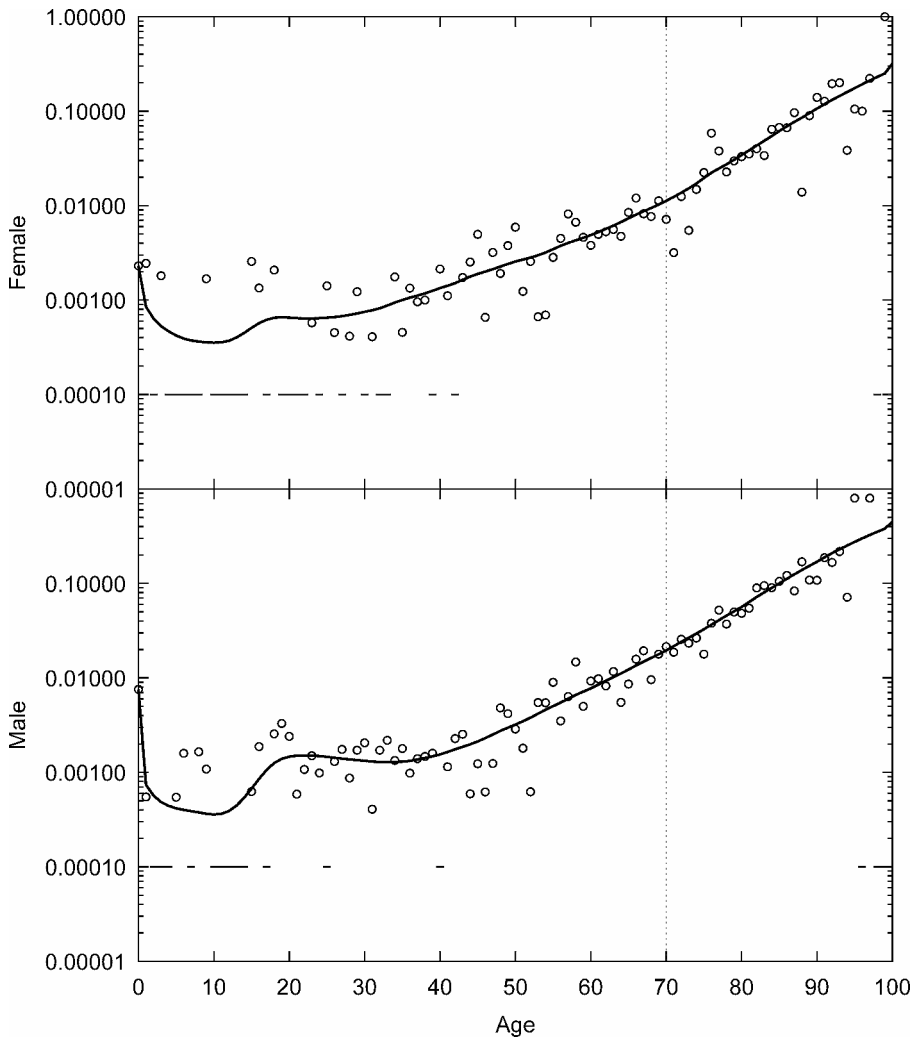
Fig. 3 Death rates, Northern Territory, 2021. The vertical line at age 70 indicates the 1951 birth cohort. Horizontal dashes indicate ages where the observed death rate is zero because there were no recorded deaths. Based on ABS data

P-TOPALS. The spline method models the age-specific schedules by a polynomial function in each age group. For each closed age group we use a cubic polynomial. For emigration rates, we use a constant polynomial to extrapolate over the open interval 65+, consistent with observed regularities in migration intensities beyond retirement ages (Rogers et al., 1978). For immigration numbers, we extrapolate with an exponential decay, that is, the logarithm of the number is linear in age. The second method used population P-TOPALS (Dyrting et al., 2022) for immigration numbers (with $\sigma^2 = 25/3$ in Eq. (3) to model uncertainty due to ABS rounding) and mortality P-TOPALS for emigration rates (Dyrting, 2018). The standards used in the
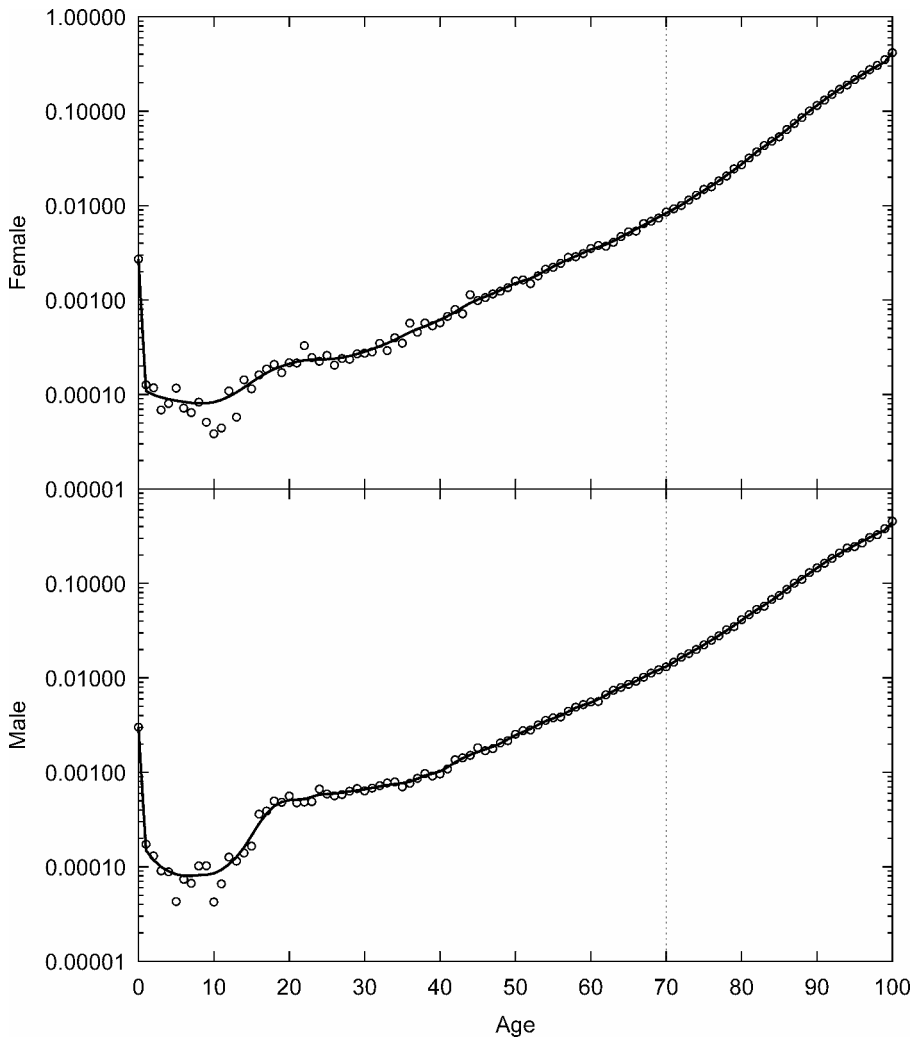
Fig. 4 Death rates, Australia (ex NT), 2021. The vertical line at age 70 indicates the 1951 birth cohort. Based on ABS data

P-TOPALS expansions were derived from single-year age-specific national overseas arrival and departure data from ABS (2022 h).

Figures 7 and 8 show 2021 abridged and expanded emigration rates for the Northern Territory and the rest of Australia respectively. The most noticeable feature of the NT's emigration is the large difference between female and male rates, in particular the peak in out-migration of males aged 20 to 25. In contrast, the rest of Australia shows only a small difference in female and male emigration, with a sex ratio of 105 compared to the NT's emigration sex ratio of 184 (Table 1). The disparity between female and male overseas departures has been a feature of emigration from the NT in recent decades, notwithstanding the decreased level of rates nationally for both sexes

Table 1 Population summary measures, Northern Territory and the rest of Australia, 2021. TFR, total fertility rate; MAC, mean age at childbearing; e_0^F, female life expectancy at birth; e_0^M, male life expectancy at birth; API^F, female age at peak migration intensity; API^M, male age at peak migration intensity; SR^E, emigration sex ratio; SR^I, immigration sex ratio

Component	Measure	Northern Territory	Australia (ex NT)
Fertility	TFR	1.82	1.70
	MAC	29.4	31.3
Mortality	e_0^F	83.3	86.0
	e_0^M	77.9	82.3
Internal migration	API^F	26.3	23.6
	API^M	26.9	24.4
Overseas migration	SR^E	184	105
	SR^I	132	106

in 2021 compared to the trend over the prior five years. Figures 7 and 8 also illustrate the ability in P-TOPALS to extrapolate schedules over open age intervals by 'borrowing' the profile of the standard.

Figures 9 and 10 show 2021 abridged and expanded immigration numbers for the Northern Territory and the rest of Australia respectively. Similar to emigration, there were more male overseas arrivals to the NT than female, although the disparity is not as great as for departures, with an immigration sex ratio equal to 132 for the NT and 106 for the rest of Australia (Table 1). The gendered nature of the NT's overseas migration has been a persistent feature since at least 2004, although the size of the disparity has been decreasing over time.

Projected Population

Figures 11 and 12 show the projected population in 2051 by age and sex for the Northern Territory and the rest of Australia. The most noticeable difference between the population projections with smoothed and unsmoothed inputs is a tendency for latter to be larger, particularly for ages greater than 70, regardless of sex or region. We also notice a greater age-on-age variation of the Northern Territory's population pyramid with unsmoothed inputs. This becomes more noticeable when we remove the age variation of the jump-off population by focusing on the projection for a single cohort. Figures 13 and 14 show the projected population of the cohort born in 1951 for both the Northern Territory and the rest of Australia. This cohort was 70 in 2021 (see vertical reference line in Figs. 3, 4, 5, 6, 7, 8, 9 and 10) and those remaining will be 100 in 2051. We see from Figs. 3 and 5, and Fig. 7 that for 70 year old Territorians, the three sources of decrement - death, interstate out-migration, and emigration - have comparable rates (0.01–0.02 per year) but the latter two show increasing levels of difference between unsmoothed and smoothed estimates because of either sample noise (out-migration) or abridgement (emigration).

For the rest of Australia, we see from Figs. 4 and 6, and Fig. 8 that for the 1951 cohort the highest sources for decrement are death and emigration, with the latter showing the greatest sensitivity to rate estimation method. Extrapolating past age 65 with a flat emigration rate means that, relative to the P-TOPALS profile, rates are lower below age 80 and higher afterwards which results in a 'squarer' shape to the otherwise smooth projected cohort profiles seen in Fig. 14.
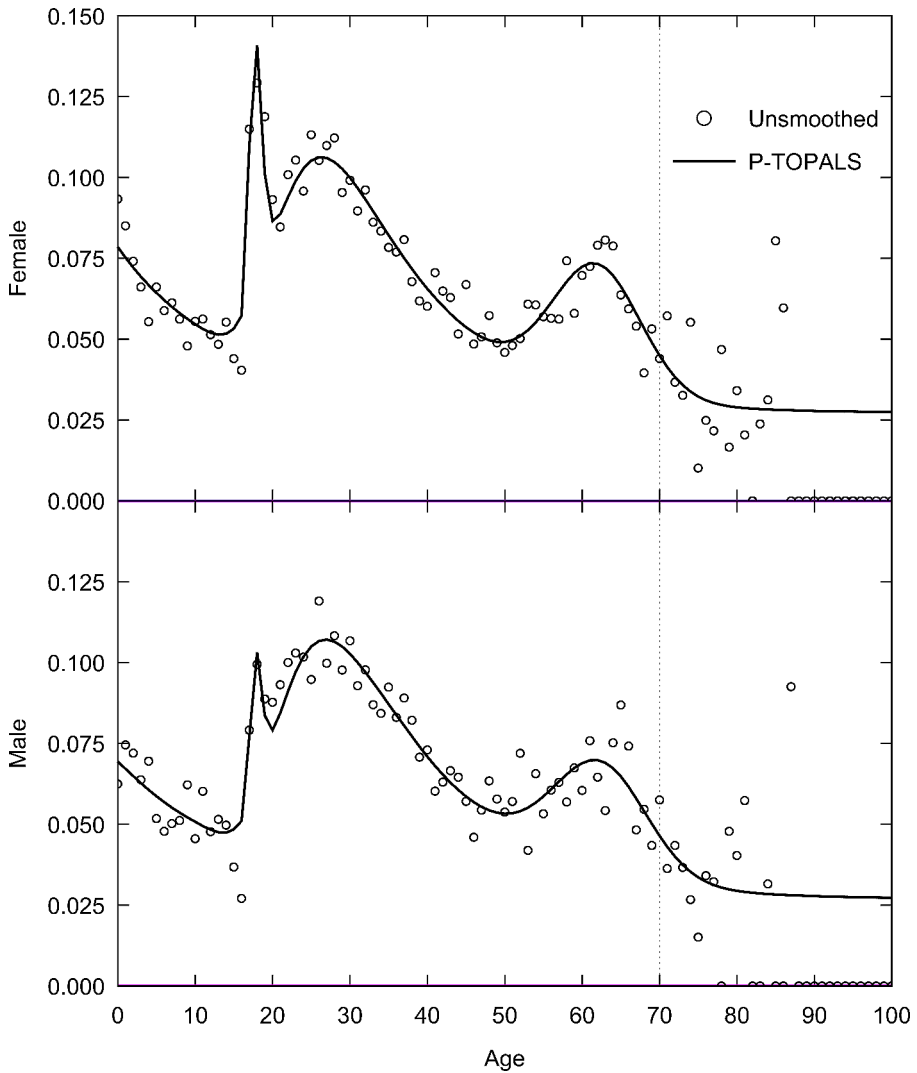
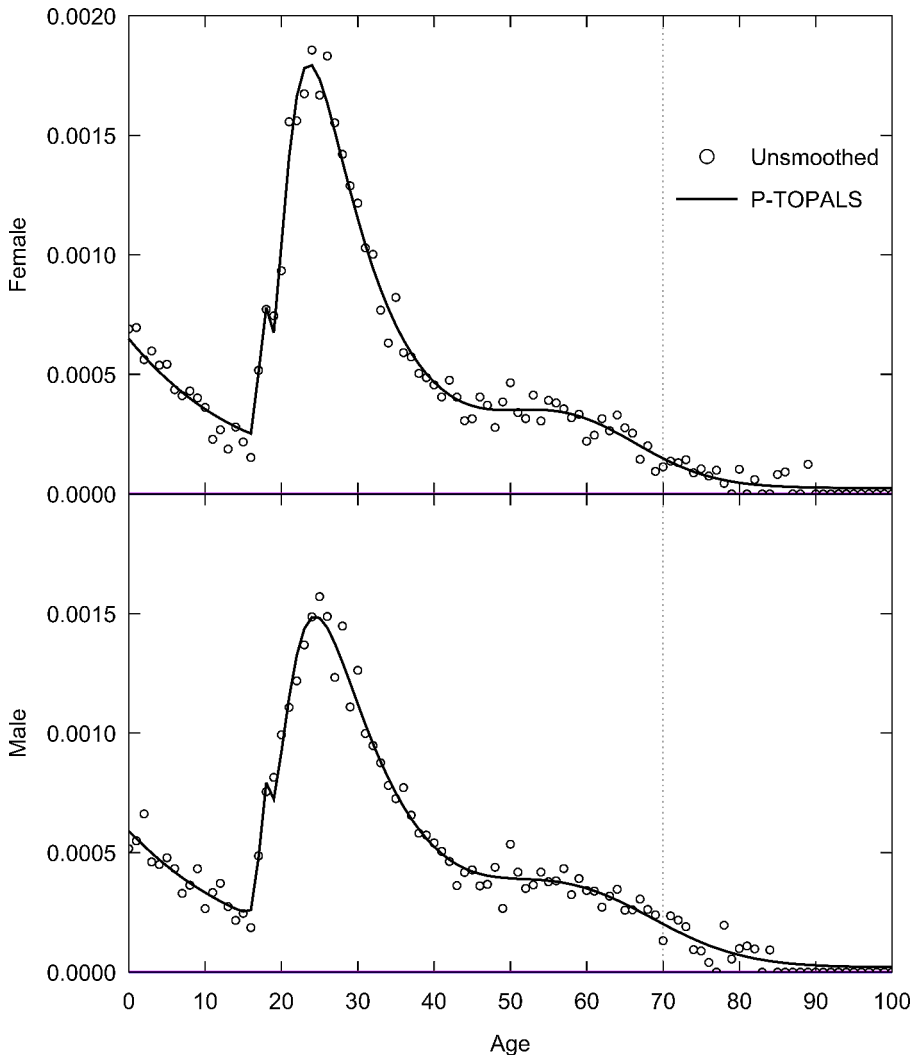**Fig. 5** Out-Migration probabilities, Northern Territory, 2021. The vertical line at age 70 indicates the 1951 birth cohort. Based on ABS data

Discussion

Our results highlight that the Northern Territory is a region of great demographic complexity where paradigms can be stretched to exaggerated forms. For example, as total fertility rate decreases and the mean age at childbearing increases it has been observed that some counties develop signs of a bimodal distribution in the age-specific fertility rates (Chandola et al., 1999, 2002; Lima et al., 2018; Peristera & Kostaki, 2007), but few have the two modes as prominent as they are for the Northern Territory (see Fig. 1). And, while for Australia (ex NT) the sex ratio of overseas

Fig. 6 In-Migration probabilities, Northern Territory, 2021. The vertical line at age 70 indicates the 1951 birth cohort. Based on ABS data

migration is close to parity, for the NT there are large differences in gender for both the level and the age profile. These demonstrate the need for careful consideration of smoothing and abridgement methods, as well as the role the demographer must play in carefully explaining projection results using their knowledge on the specificities of the input data. While this is not necessarily a function of the undertaking of projections for SPAs alone, these examples and the study here highlight the potential unconformity of input data for SPAs, even after smoothing is applied.

For the Northern Territory it is the post-retirement ages where smoothing methods have the greatest impact on forecasting, producing cohort projections that evolve smoothly with time as the population decreases. Realistic projections of this age
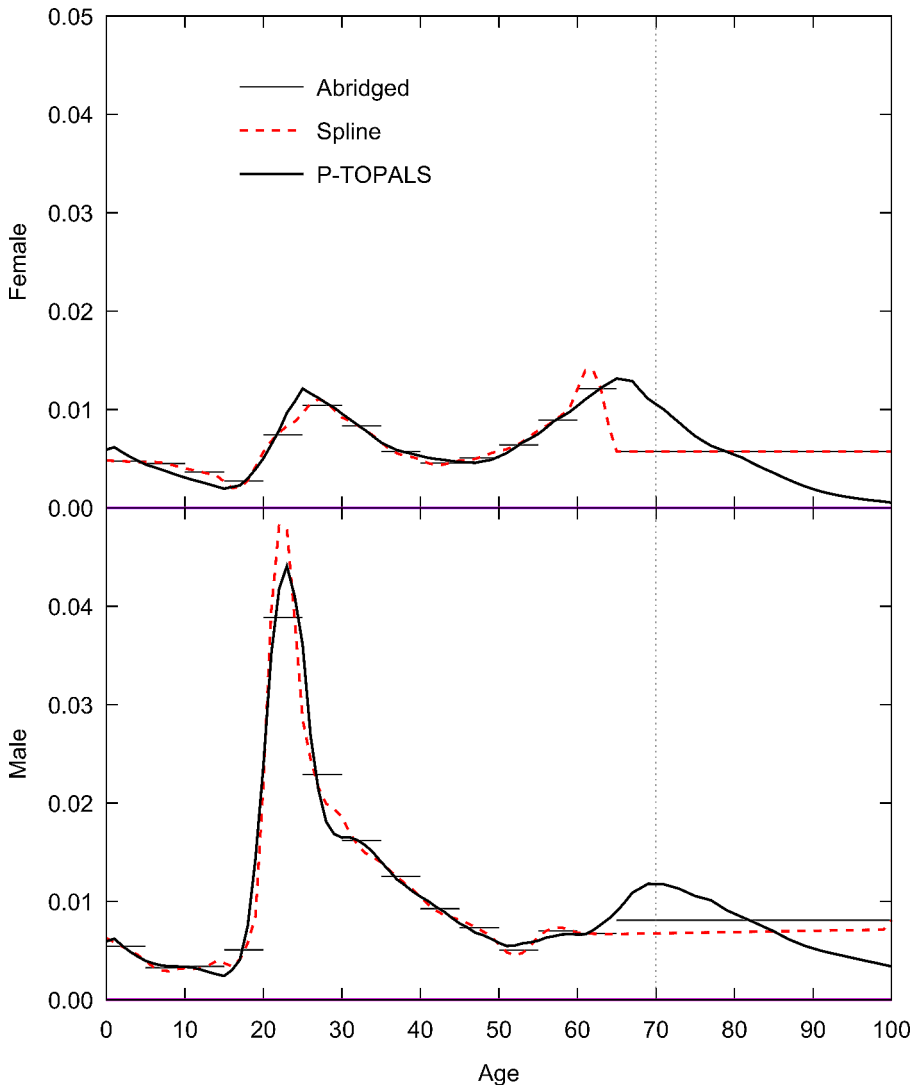
Fig. 7 Emigration rates, Northern Territory, 2021. Light solid line, abridged emigration rates calculated from grouped data; Dashed line, expanded schedule of rates using spline interpolation; Heavy solid line, expanded schedule of rates using P-TOPALS. The vertical line at age 70 indicates the 1951 birth cohort. Based on ABS data

group are important because they are the fastest growing section of the population, representing both a challenge for service provision (Zeng et al., 2015) and an important focus of population retention strategies in the Northern Territory (Dyrting et al., 2020) and in northern Australia more broadly where the same trend of structural aging is underway (Taylor et al., 2022).

Calibrating the inputs to a subnational population projection with a jump-off date one year after the onset in Australia of the COVID pandemic highlights the impor-
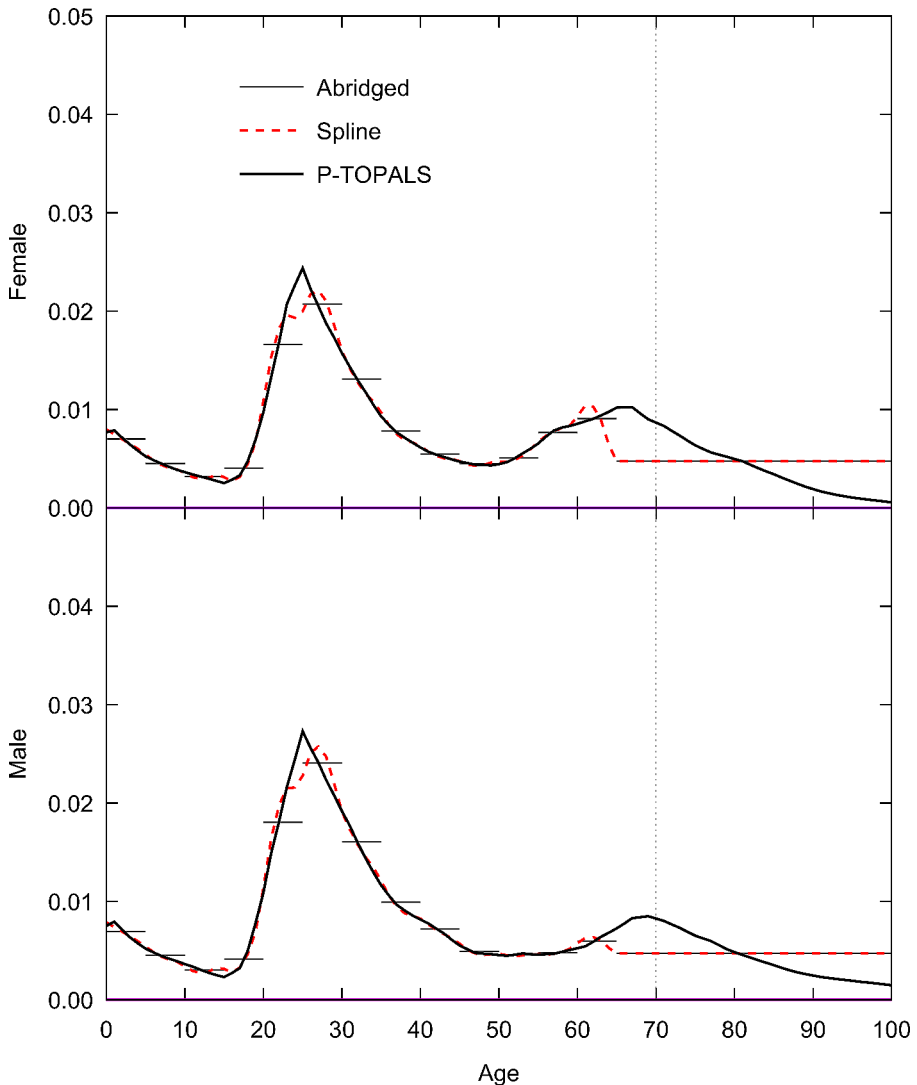
Fig. 8 Emigration rates, Australia (ex NT), 2021. Light solid line, abridged emigration rates calculated from grouped data; Dashed line, expanded schedule of rates using spline interpolation; Heavy solid line, expanded schedule of rates using P-TOPALS. The vertical line at age 70 indicates the 1951 birth cohort. Based on ABS data

tance of smoothing methods that are robust in the presence of significant uncertainties in the underlying rates. For example, to reduce the effect of sample noise the ABS publishes complete mortality schedules for Australia's States and Territories using data aggregated over three years (ABS, 2022e). However, it would be unwise to apply such averaging to fertility and overseas migration for a period where rates for these were an order of magnitude away from the linear trends observed prior to the COVID-19 pandemic. Meanwhile, for internal migration reliable data for the
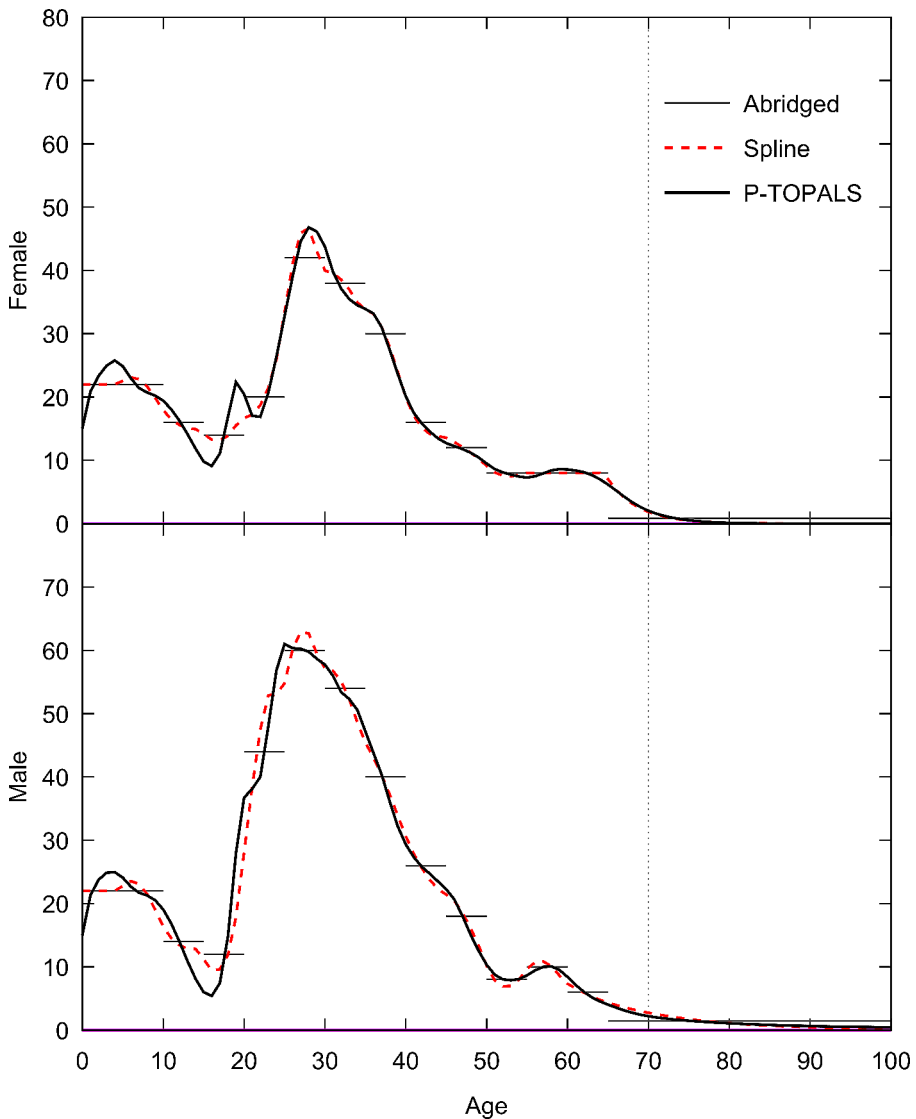
Fig. 9 Immigration numbers by single year of age, Northern Territory, 2021. Light solid line, abridged immigration numbers calculated from grouped data; Dashed line, expanded schedule of numbers using spline interpolation; Heavy solid line, expanded schedule of numbers using P-TOPALS. The vertical line at age 70 indicates the 1951 birth cohort. Based on ABS data

Northern Territory is only available from census data collected every five years. But as shown in Figs. 1, 3 and 5, and Fig. 6 data for a single year is likely to be very noisy.

Migration between the NT and the rest of Australia plays an important role in shaping the composition of the NT's population, with net arrivals of people in their twenties and departures in all other ages producing a bulge in the population pyramid beginning at age twenty (Fig. 11). Since, from a population perspective, the rest of
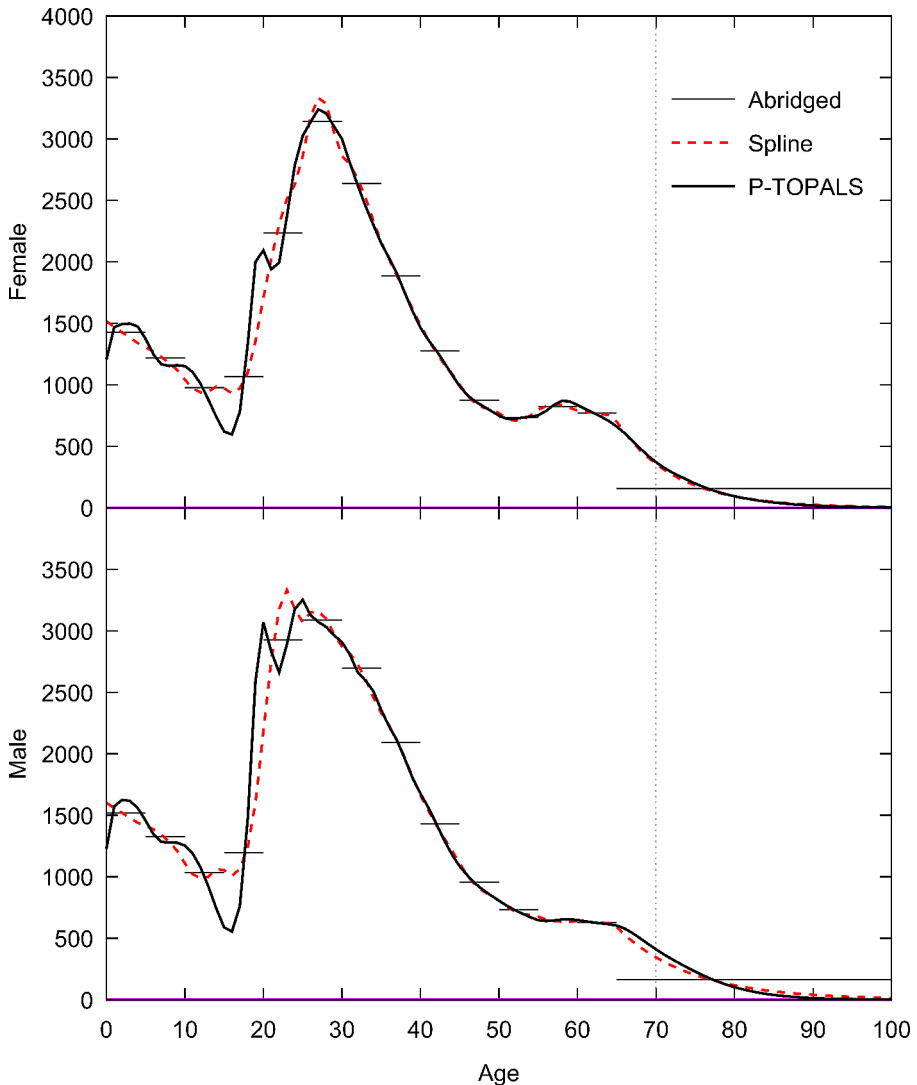
Fig. 10 Immigration numbers by single year of age, Australia (ex NT), 2021. Light solid line, abridged immigration numbers calculated from grouped data; Dashed line, expanded schedule of numbers using spline interpolation; Heavy solid line, expanded schedule of numbers using P-TOPALS. The vertical line at age 70 indicates the 1951 birth cohort. Based on ABS data

Australia is practically all of Australia, if we compare Figs. 3 and 4, we see that methods for producing projection inputs for SPAs must not only be robust in the presence of small sample noise but must also be capable of the great level accuracy which is required for national populations.

One of the strengths of the P-TOPALS method is that it is able to automatically adjust from a non-parametric to a relational method as the exposed population decreases by, in terms of the Eq. (2), trading goodness-of-fit for smoothness. This is
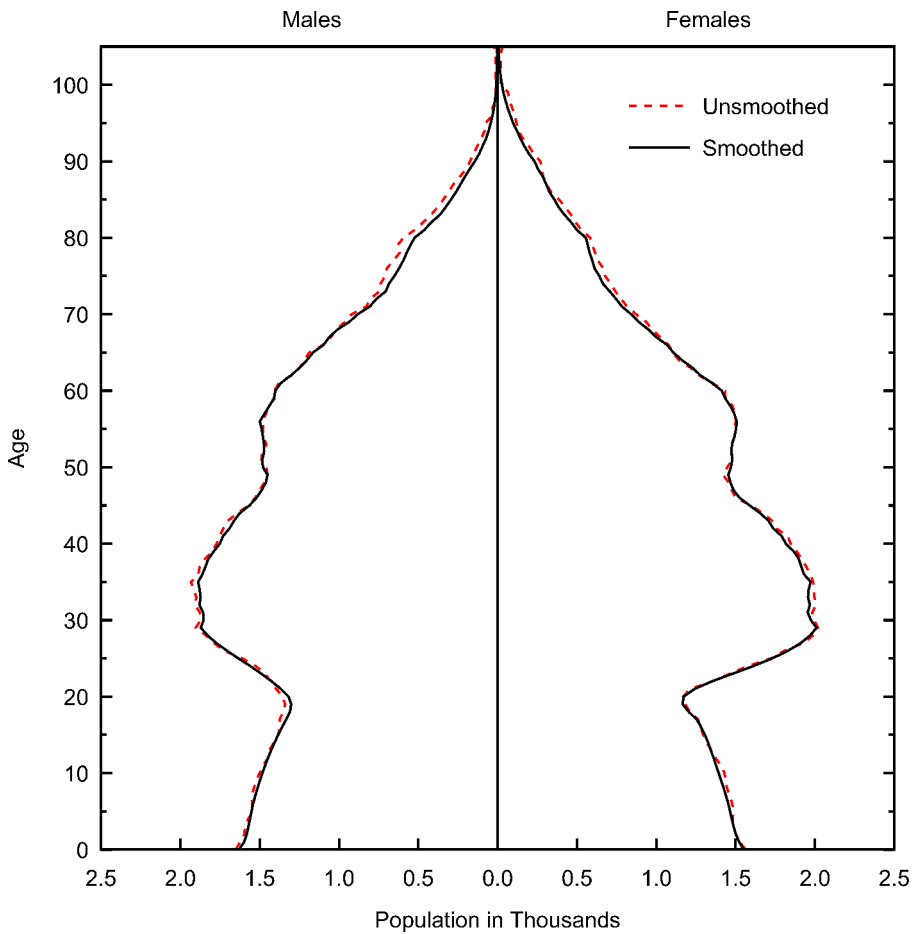
Fig. 11 Projected population for 2051, Northern Territory. Dashed line, projection from unsmoothed rates; Solid line, projection from smoothed rates. Based on ABS data

illustrated in Table 2, which compares the performance of the input smoothing for the NT and the rest of Australia using the sum of squared standardized errors.

$$\text{gof} = \frac{1}{g} \sum_{i=1}^{g} \frac{\left(_n\tilde{r}_i - _nr_i \right)^2}{\sigma_i^2}, \quad (4)$$

as a measure of goodness-of-fit. Here we use the factor $1/g$ to adjust for the different number of observed rates, and σ_i^2 is the variance of the i th observed rate, equal to $_nr_i/_nN_i$ for the Poisson variables and σ^2 for normal variables. We see that in 7 of the 9 cases gof decreases when going from the smaller to the larger population.

Methods for estimating inputs to population projection models for SPAs need to be both flexible enough to capture the demographic diversity of SPAs, and robust in the presence of noisy and incomplete data. They must be able generate plausible
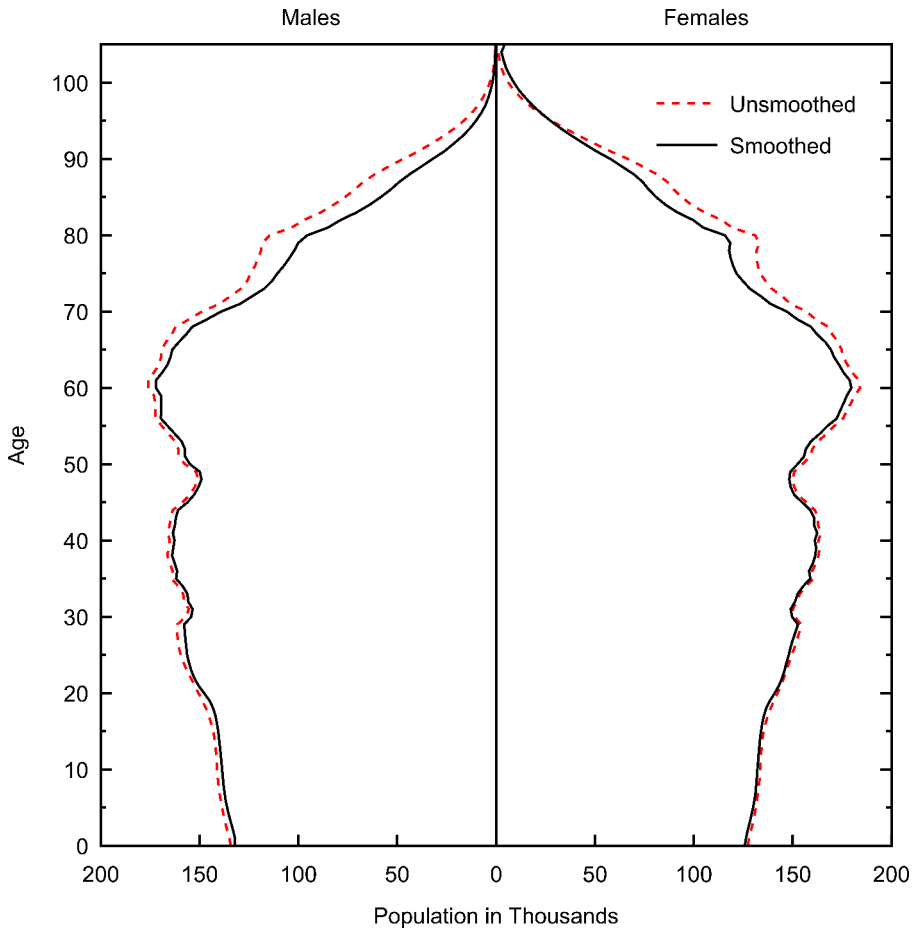
Fig. 12 Projected population for 2051, Australia (ex NT). Dashed line, projection from unsmoothed rates; Solid line, projection from smoothed rates. Based on ABS data

rate schedules across the full range of ages within the limits imposed by sample noise and anonymization. Recently developed methods such as P-TOPALS and CS are well suited to balancing these competing requirements because they are highly parameterized (through splines) as well as constrained in the profiles they generate when accuracy is low.

In developed countries, national migration trends are still advancing the share of population living in major cities, and this continues to be the case in Australia. By contrast, a lower proportion of internal migrants who move across State or Territory boundaries are selecting SPAs as their residential destination (Taylor & Carson, 2017). Since the end of the national mining 'boom', which took place in Australia during 2002 to around 2012, the whole of the north of the country has stagnated population-wise, and in many areas, population has declined (Taylor et al., 2022). Technology developments in the resources (notably automation) and other sectors, along with the widespread adoption and acceptance of remote working as a legacy
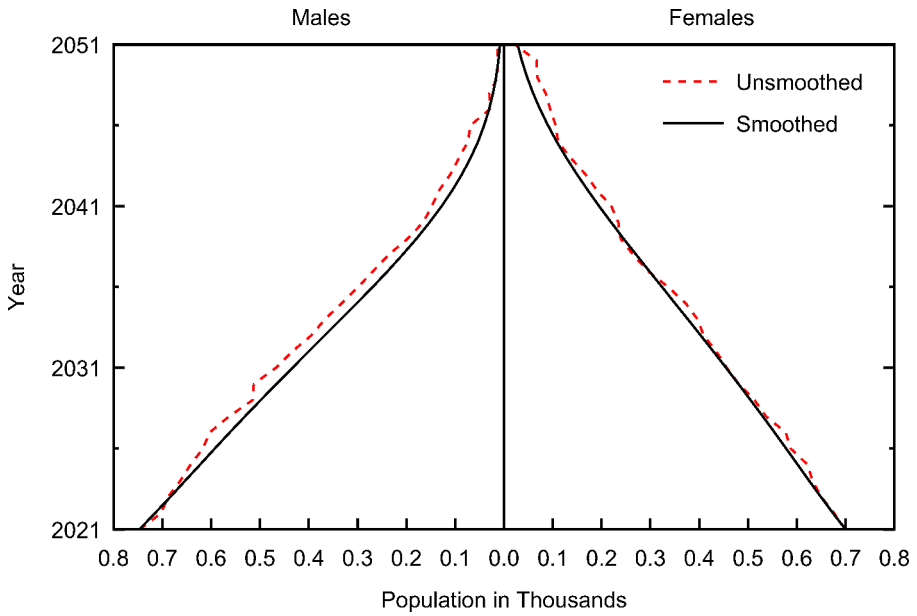
Fig. 13 Projected population for 1951 cohort, Northern Territory. Dashed line, projection from unsmoothed rates; Solid line, projection from smoothed rates. Based on ABS data
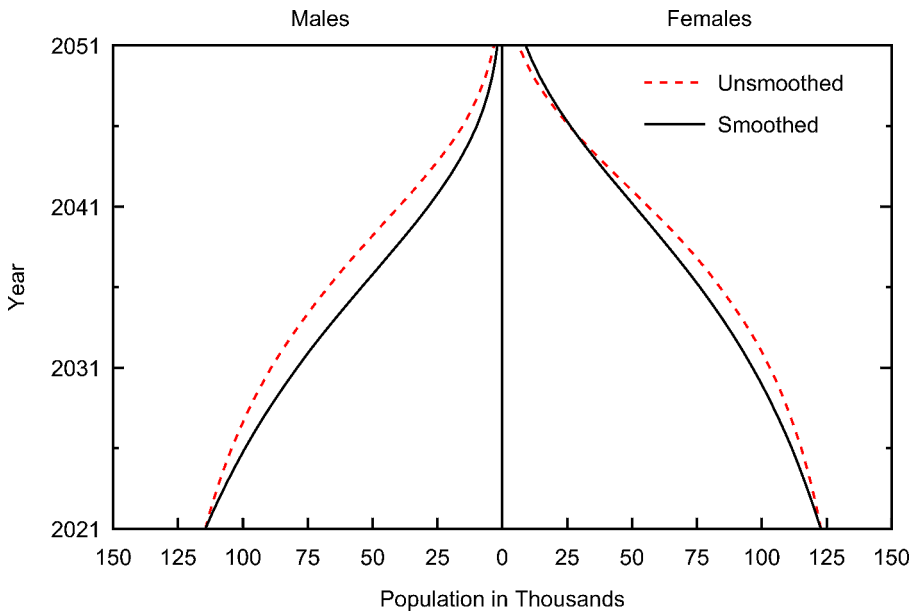

Fig. 14 Projected population for 1951 cohort, Australia (ex NT). Dashed line, projection from unsmoothed rates; Solid line, projection from smoothed rates. Based on ABS data

Table 2 Goodness of fit measure Eq. (4) for smoothed inputs, Northern Territory and the rest of Australia, 2021

Component	Sex	Northern Territory	Australia (ex NT)
Fertility	Female	1.10	1.35
Mortality	Female	1.62	0.78
	Male	1.22	0.55
Internal migration	Female	1.00	1.09
	Male	1.36	0.96
Overseas out-migration	Female	0.51	0.37
	Male	0.40	0.28
Overseas in-migration	Female	0.09	0.001
	Male	0.19	0.002

of the COVID-19 pandemic, mean that numbers associated with the demographic components for SPAs will continue to be small and noisy in the foreseeable future.

In addition to small numbers, levels of vital processes and internal migration are decreasing and there is an increasing focus by national statistical agencies such as the Australian Bureau of Statistics on maintaining the privacy of individuals in the face of advances in methods for database reconstruction. This means that demographic data for subnational areas like the Northern Territory will likely become more noisy not less. Methods for enhancing the estimation of rates through smoothing will, therefore, be increasingly important to the production of realistic population projections for these regions. Not least, given the amount of work demographers face in updating and continually improving projections, ensuring projections have a viable shelf-life (where projection errors are kept to a minimum for as far as possible into the horizon) for the benefit of users becomes particularly important. For SPAs and other subnational regions, critical investment and policy decisions are founded on projection outputs, making tools such as the smoothing methods deployed here a valuable investment in the process of 'projection making' in their own right.

Demographers preparing inputs for a cohort-component model have an number of methods they can choose from within the four broad classes of parametric (Bermúdez et al., 2012; de Beer & Janssen, 2016; Wilson, 2020), polynomial (Campbell, 1996; Grigoriev & Jdanov, 2015; Hsieh, 1991; McNeil et al., 1977), relational (Booth, 1984; Brass, 1971; Smith et al., 2013), and non-parametric (Bernard & Bell, 2015; Kostaki et al., 2009, 2011; Rizzi et al., 2015) methods. We have seen that for projecting the population of an SPA, a method must have a number of desirable features: producing plausible profiles in the presence of large uncertainty in observed rates (large noise), producing accurate profiles for the rest of the country (small noise), the ability to include non-polynomial features into the profile such as jumps or spikes in a rate (non-polynomial), flexibility in controlling the profile over an open age interval (extrapolate), the ability to include views on what a profile should look like (expert view), and the ability to handle deviations from orthodox profiles (flexibility).

Table 3 is a survey of the four main classes of methods for estimating population projection inputs together with P-TOPALS, summarizing the inputs they are used on (the first four rows) and whether the properties desirable for preparing inputs to a population projection for an SPA are normally present (Y), absent (N), or contingent

Table 3 Properties of a smoothing method for preparing inputs to a population projection model for sparsely populated areas, four classes and P-TOPALS.

Property	Parametric	Polynomial	Relational	Non-Parametric	P-TOPALS
Mortality	Y	Y	Y	Y	Y
Fertility	Y	Y	Y	Y	Y
Migration	Y	Y	-	Y	Y
Counts	N	Y	Y	Y	Y
Large noise	-	N	Y	-	Y
Small noise	-	Y	N	Y	Y
Non-polynomial	Y	N	Y	N	Y
Extrapolate	Y	N	Y	-	Y
Expert View	-	N	Y	N	Y
Flexible	N	Y	N	Y	Y

(-). Parametric methods are available for all types of inputs except population counts. In terms of uncertainty in observed rates they do not automatically adjust to the level of noise, and can be over-parameterized when noise is large (relational models being favored in this case) and under-parameterized when noise is small (where polynomial and non-parametric methods are used). Parametric methods do not readily allow views on the form of the schedules, and are the least flexible of methods in that they make strong assumptions on shape of the profiles. Polynomial methods are not stable when there is large uncertainty in the observed rates because they interpolate rather than smooth. They, by design, cannot model non-polynomial features in the profile, are poor at extrapolating into open intervals because make they strong assumptions on the profile after the end of the last closed interval (often constant or linear), and are closed to expert views because estimated profiles are fully determined by the observed rates.

The main weakness of relational methods is their lack of flexibility, in that all age-specific structure is taken from a standard curve and their consequent difficulty in fitting observed rates when uncertainty is small. Non-parametric methods generally perform well when uncertainty is small or moderate, although there are exceptions (Dyrting & Taylor, 2023; Grigorieva et al., 2020), and they are similar to polynomial methods in not being easily adjustable to expert views or auxiliary sources of information. However, if we consider relational and non-parametric method together in Table 3 we see that where one is negative or indeterminate the other is positive, and it is this complementarity of properties of the two methods that, when combined in P-TOPALS, make the latter well suited to estimating inputs for a SPA population projection model.

Conclusion

Providing accurate data for the input to population projection models is critical to both the accuracy and shelf-life for projections outputs. Users must be confident that output data on which significant decisions might be made is as accurate as it can be. This challenges the demographer to be diligent and detailed in the application of smoothing and adjustment methods which are applied to input data across the components, regardless of the type of model used. For sparsely populated areas, a range of biases mean there are additional challenges for the demographer in preparing fit-for-purpose input data. Not least, small exposure and event counts make the data particularly susceptible to uncertainty from sample noise and anonymization. Second, the presence of heterogeneity in the population can cause the age schedules to take on exaggerated forms, significantly different from the national age structure. And third, strong interaction with the remainder of the country through internal migration means that methods must also be accurate for national level data. These issues certainly apply to population data for Australia's most sparsely populated State or Territory; the Northern Territory, which was the focus of this study. We have found that, through projecting using the cohort component model SASPOPP, advances in the estimation of demographic rates through smoothing and expansion are effective in mitigating and meeting these challenges and can be used to improve population projections for sparsely populated areas. Benefits from the methods applied here are particularly notable for the post-retirement ages because, in these ages death, internal out-migration, and emigration rates are comparable.

In comparing smoothed and unsmoothed input and output data for the Northern Territory, the primary contribution of this article is to prosecute the case that modern quantitative methods are useful for improving population projections for subnational regions which might be described as 'edges'. Conversely, the need for key users such as planners, investment decision makers and policy makers for high quality and accurate projections can serve as inspiration to applied demographers for future innovations. For example, one of the limitations of the approach taken in this article is that population is not disaggregated by Indigenous status. Estimating inputs to an accurate projection of the Northern Territory's Aboriginal population at single year of age would need to confront additional issues such identification uncertainty in vital registrations, census data quality issues related to remoteness, notable age heaping, the age structure for internal migration, age misstatement, and non-western paradigms of mobility. The application of P-TOPALS can certainly assist but is by no means a simple remedy for the spectrum of data issues that demographers developing projections for SPAs face. For example, demographers have highlighted the parallels between sparsely populated areas in Arctic and non-Artic countries in relation to input data biases and the challenges in developing projections and population estimates (Karacsonyi & Taylor, 2023; Peters et al., 2016), emphasizing there are opportunities to apply the P-TOPALS method to comparative work for populations in the Arctic and beyond. A process of continual improvement and leaning on techniques and innovations applied in other SPAs, and in the demographic scholarly community more broadly, therefore remains the preferred approach.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

ABS (2017). *Population projections, Australia*. Retrieved February 20, 2023, from https://www.abs.gov.au/statistics/people/population/population-projections-australia/latest-release#northern-territory.

ABS (2019). *Estimates and projections, Aboriginal and Torres Strait Islander Australians*. Retrieved February 20, 2023, from https://www.abs.gov.au/statistics/people/aboriginal-and-torres-strait-islander-peoples/estimates-and-projections-aboriginal-and-torres-strait-islander-australians/latest-release.

ABS (2023b). *Selected living cost indexes, Australia*. Retrieved April 19, 2023, from https://www.abs.gov.au/statistics/economy/price-indexes-and-inflation/selected-living-cost-indexes-australia/dec-2022.

ABS (2023a). *National, state and territory population, Sep 2022*. Retrieved March 20, 2023, from https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/sep-2022/31010do002_202209.xlsx.

ABS (2022i). *Table Builder – 2021 Census, internal migration* [Data set]. Retrieved March 8, 2023, from https://www.abs.gov.au/statistics/microdata-tablebuilder/tablebuilder.

ABS (2022 h). *Overseas migration*. https://www.abs.gov.au/statistics/people/population/overseas-migration/2021-22-financial-year.

ABS (2022 g). *National, state and territory population*. https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/sep-2022.

ABS (2022f). *Marriages and divorces, Australia*. https://www.abs.gov.au/statistics/people/people-and-communities/marriages-and-divorces-australia/2021.

ABS (2022d). *Estimates of Aboriginal and Torres Strait Islander Australians*. https://www.abs.gov.au/statistics/people/aboriginal-and-torres-strait-islander-peoples/estimates-aboriginal-and-torres-strait-islander-australians/jun-2021.

ABS (2022e). *Life tables*. https://www.abs.gov.au/statistics/people/population/life-tables/2019-2021.

ABS (2022c). *Data Explorer - net overseas migration, arrivals, departures and net, state/territory, age and sex - calendar years, 2004 onwards* [Data set]. Retrieved March 10, 2023, from https://explore.data.abs.gov.au.

ABS (2022b). *Data Explorer - fertility, by age, by state* [Data set]. Retrieved February 22, 2023, from https://explore.data.abs.gov.au.

ABS (2022a). *Data Explorer - deaths, year of registration, age at death, age-specific death rates, sex, states, territories and Australia* [Data set]. Retrieved February 22, 2023, from https://explore.data.abs.gov.au.

ABS (2021). *Data confidentiality guide*. Retrieved April 18, 2023, from https://www.abs.gov.au/about/data-services/data-confidentiality-guide.

ABS. (n.d.). *Welcome to the Australian Bureau of Statistics*. Retrieved April 19 (2023). from https://www.abs.gov.au/.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Australian Government (2023). *National defence: Defence strategic review 2023*. Retrieved April 26, 2023, from https://www.defence.gov.au/about/reviews-inquiries/defence-strategic-review.

Baffour, B., & Raymer, J. (2019). Estimating multiregional survivorship probabilities for sparse data: An application to immigrant populations in Australia, 1981–2011. *Demographic Research*, *40*(18), 463–502. https://doi.org/10.4054/DemRes.2019.40.18.

Baili, P., Micheli, A., Montanari, A., & Capocaccia, R. (2005). Comparison of four methods for estimating complete life tables from abridged life tables using mortality data supplied to EUROCARE-3. *Mathematical Population Studies*, *12*, 183–198. https://doi.org/10.1080/08898480500301751.

Bermúdez, S., Blanquero, R., Hernández, J. A., & Planelles, J. (2012). A new parametric model for fitting fertility curves. *Population Studies*, *66*(3), 297–310. https://doi.org/10.1080/00324728.2012.695802.

Bernard, A., & Bell, M. (2015). Smoothing internal migration age profiles for comparative research. *Demographic Research*, *32*(33), 915–948. https://doi.org/10.4054/DemRes.2015.32.33.

Booth, H. (1984). Transforming Gompertz's function for fertility analysis: The development of a standard for the relational Gompertz function. *Population Studies*, *38*(3), 495–506. https://doi.org/10.1080/00324728.1984.10410306.

Brass, W. (1971). On the scale of mortality. In W. Brass (Ed.), *Biological aspects of demography* (Vol. 10, pp. 69–110). Taylor & Francis.

Campbell, P. R. (1996). *Population projections for states by age, sex, race, and Hispanic origin: 1995 to 2025* (PPL-47). U.S. Bureau of the Census, Population Division.

Chandola, T., Coleman, D. A., & Hiorns, R. W. (1999). Recent European fertility patterns: Fitting curves to distorted distributions. *Population Studies*, *53*, 317–329. https://doi.org/10.1080/00324720308089.

Chandola, T., Coleman, D. A., & Hiorns, R. W. (2002). Distinctive features of age-specific fertility profiles in the English-speaking world: Common patterns in Australia, Canada, New Zealand and the United States, 1970-98. *Population Studies*, *56*(2), 181–200. https://doi.org/10.1080/00324720215929.

de Beer, J., & Janssen, F. (2016). A new parametric model to assess delay and compression of mortality. *Population Health Metrics*, *14*, https://doi.org/10.1186/s12963-016-0113-1.

Dyrting, S. (2018). *Penalised TOPALS as a tool for expanding abridged fertility and mortality schedules for small populations*. https://doi.org/10.13140/RG.2.2.35835.41764.

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. (2023). Estimating age-specific mortality using calibrated splines. *Population Studies*, 1–18. https://doi.org/10.1080/00324728.2023.2228297.

Dyrting, S., Taylor, A., & Shalley, F. (2020). A life-stage approach for understanding population retention in sparsely populated areas. *Journal of Rural Studies*, *80*, 439–451. https://doi.org/10.1016/j.jrurstud.2020.10.021.

Dyrting, S., Flaxman, A., & Sharygin, E. (2022). Reconstruction of age distributions from differentially private census data. *Population Research and Policy Review*, *41*(6), 2311–2329. https://doi.org/10.1007/s11113-022-09734-2.

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, *11*(2), 89–121. https://doi.org/10.1214/ss/1038425655.

Geoscience Australia (2004). *Area of Australia - states and territories*. https://www.ga.gov.au/scientific-topics/national-location-information/dimensions/area-of-australia-states-and-territories viewed 30 March 2023.

Grigoriev, P., & Jdanov, D. (2015). *Splitting abridged fertility data using different interpolation methods: Is there the optimal solution?* Paper presented at Human Fertility Database: Expanding research opportunities, Member initiated meeting, PAA 2015 Annual Meeting, San Diego, 29 April 2015.

Grigorieva, O., Jasilioniene, A., Jdanov, D. A., Grigoriev, P., Sobotka, T., Zeman, K., & Shkolnikov, V. M. (2020). *Methods protocol for the Human Fertility Collection*. Human Fertility Collection.

Hsieh, J. J. (1991). Construction of expanded continuous life tables: A generalization of abridged and complete life tables. *Mathematical Biosciences*, *103*, 287–302. https://doi.org/10.1016/0025-5564(91)90057-p.

Karacsonyi, D., & Taylor, A. (2023). Understanding demographic and economic patterns in sparsely populated areas – a global typology approach. *Geografiska Annaler: Series B Human Geography*, *105*(3), 228–247. https://doi.org/10.1080/04353684.2022.2103445.

Kostaki, A., Moguerza, J., Olivares, A., & Psarakis, S. (2009). Graduating the age-specific fertility pattern using support Vector machines. *Demographic Research*, *20*(25), 599–622. https://doi.org/10.4054/DemRes.2009.20.25.

Kostaki, A., Moguerza, J. M., Olivares, A., & Psarakis, S. (2011). Support Vector machines as tools for mortality graduation. *Canadian Studies in Population*, *38*, 37–58. https://doi.org/10.25336/P6VS46.

Lima, E. E. C., Zeman, K., Sobotka, T., Nathan, M., & Castro, R. (2018). The emergence of bimodal fertility profiles in Latin America. *Population and Development Review*, *44*(4), 723–743. https://doi.org/10.1111/padr.12157.

McNeil, D. R., Trussell, T. J., & Turner, J. C. (1977). Spline interpolation of demographic data. *Demography*, *14*(2), 245–252. https://doi.org/10.2307/2060581.

NT Department of Treasury and Finance (2019). *Northern Territory population projections (2019 release)*. Retrieved May 8, 2023, from https://treasury.nt.gov.au/dtf/economic-group/population-projections.

Peristera, P., & Kostaki, A. (2007). Modeling fertility in modern populations. *Demographic Research*, *16*, 141–194. https://doi.org/10.4054/DemRes.2007.16.6.

Peters, P., Taylor, A., Carson, D. B., & Brokensha, H. (2016). Sources of data for settlement level analyses in sparsely populated areas. In A. Taylor, D. B. Carson, P. C. Ensign, L. Huskey, R. O. Rasmussen, & G. Saxinger (Eds.), *Settlements at the edge: Remote human settlements in developed nations* (pp. 153–176). Edward Elgar Publishing Limited.

Raymer, J., Bai, X., Liu, N., & Wilson, T. (2020). Estimating a consistent and detailed time series of immigration and emigration for sub-state regions of Australia. *Applied Spatial Analysis and Policy*, *13*(2), 411–439. https://doi.org/10.1007/s12061-019-09310-w.

Rizzi, S., Gampe, J., & Eilers, P. H. C. (2015). Efficient estimation of smooth distributions from coarsely grouped data. *American Journal of Epidemiology*, *182*(2), 138–147. https://doi.org/10.1093/aje/kwv020.

Rogers, A., Raquillet, R., & Castro, L. J. (1978). Model migration schedules and their applications. *Environment and Planning A*, *10*(5), 475–502. https://doi.org/10.1068/a100475.

Saha, U. R., Das, S., Baffour, B., & Chandra, H. (2023). Small area estimation of age-specific and total fertility rates in Bangladesh. *Spatial Demography*, *11*(1), 1–24. https://doi.org/10.1007/s40980-022-00113-1.

Schmertmann, C. P. (2014). Calibrated spline estimation of detailed fertility schedules from abridged data. *Revista Brasileira De Estudos De População*, *31*(2), 291–307. https://doi.org/10.1590/S0102-30982014000200004.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136.

Simpson, L., Wilson, T., & Shalley, F. (2019). The shelf life of official sub-national population forecasts in England. *Applied Spatial Analysis and Policy*, *13*, 715–737. https://doi.org/10.1007/s12061-019-09325-3.

Smith, S. K., Tayman, J., & Swanson, D. A. (2013). *A practitioner's guide to state and local population projections* (Vol. 37). Springer. https://doi.org/10.1007/978-94-007-7551-0.

Taylor, A. (2011). The forecasting of remote area populations: Numbers aren't everything. In D. Carson, R. O. Rasmussen, P. Ensign, L. Huskey, & A. Taylor (Eds.), *Demography at the edge: Remote human populations in developed nations* (pp. 39–52). Ashgate.

Taylor, A. (2014). Population projections for sparsely populated areas: Reconciling error and context. *International Journal of Population Research*, *2014*, https://doi.org/10.1155/2014/658157.

Taylor, A., & Carson, D. (2017). *Synthesising Northern Territory population research: A report to the Northern Territory Department of the Chief Minister*. Northern Institute, Charles Darwin University.

Taylor, A., & Payer, H. (2017). Population ageing in northern Australia: Seniors' voices on ageing in place. *Journal of Population Ageing*, *10*(2), 181–196. https://doi.org/10.1007/s12062-016-9156-0.

Taylor, A., Thurmer, J., & Karácsonyi, D. (2022). Regional demographic and economic challenges for sustaining growth in Northern Australia. *Regional Studies Regional Science*, *9*(1), 425–445. https://doi.org/10.1080/21681376.2022.2082316.

Temple, J. B., Wilson, T., Taylor, A., Kelaher, M., & Eades, S. (2020). Ageing of the Aboriginal and Torres Strait Islander population: Numerical, structural, timing and spatial aspects. *Australian and New Zealand Journal of Public Health*, *44*(4), 271–278. https://doi.org/10.1111/1753-6405.13001.

USCB (2022). *Data protection and privacy policy*. Retrieved April 18, 2023, from https://www.census.gov/about/policies/privacy.html.

Wilson, T. (2009). A multistate model for projecting regional populations by Indigenous status: An application to the Northern Territory, Australia. *Environment and Planning A: Economy and Space*, *41*(1), 230–249. https://doi.org/10.1068/a402.

Wilson, T. (2010). Model migration schedules incorporating student migration peaks. *Demographic Research*, *23*(8), 191–222. https://doi.org/10.4054/DemRes.2010.23.8.

Wilson, T. (2017). *SASPOPP: State and substate population projection program user guide*. Northern Institute, Charles Darwin University.

Wilson, T. (2018). Evaluation of simple methods for regional mortality forecasts. *Genus*, *74*, https://doi.org/10.1186/s41118-018-0040-z.

Wilson, T. (2020). Modelling age patterns of internal migration at the highest ages. *Spatial Demography*, *8*, 175–192. https://doi.org/10.1007/s40980-020-00062-7.

Wilson, T., & Shalley, F. (2019). Subnational population forecasts: Do users want to know about uncertainty? *Demographic Research*, *41*(13), 367–392. https://doi.org/10.4054/DemRes.2019.41.13.

Wilson, T., & Terblanche, W. (2018). New estimates of Australia's centenarian population. *International Journal of Population Data Science*, *3*(1), 1–10. https://doi.org/10.23889/ijpds.v3i1.447.

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*(3), 865–898. https://doi.org/10.1007/s11113-021-09671-6.

Zeng, B., Brokensha, H., & Taylor, A. (2015). *NOW YOU SEE US! A report on the policy and economic impacts from rapid growth in the number of senior Territorians*. Northern Institute, Charles Darwin University. https://digitalcollections.cdu.edu.au/nodes/view/4661?keywords=&type=all&lsk=317109e096a532720b98cdcf41d264bc.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Sigurd Dyrting[1] · **Andrew Taylor[1]** · **Tom Wilson[2]**

✉   Sigurd Dyrting
    sigurd.dyrting@cdu.edu.au

    Andrew Taylor
    andrew.taylor@cdu.edu.au

    Tom Wilson
    wilson.t1@unimelb.edu.au

[1]   Northern Institute, Charles Darwin University, Northern Territory, Darwin, Australia

[2]   Demography and Ageing Unit, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia