MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://monash.edu/business/ebs/research/publications

Forecasting the age structure of the scientific workforce in Australia

Rob J Hyndman, Kelly Vanh Nguyen

August 2025

Working Paper no/yr

AACSB ACCREDITED EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

Forecasting the age structure of the scientific workforce in Australia

Rob J Hyndman
Department of Econometrics & Business Statistics
Monash University
Clayton 3800
Australia
Email: Rob.Hyndman@monash.edu
Corresponding author

Kelly Vanh Nguyen
Department of Econometrics & Business Statistics
Monash University
Clayton VIC 3800
Australia

21 August 2025

JEL classification: C10,C14,C22

Forecasting the age structure of the scientific workforce in Australia

Abstract

A brief summary of our ideas

Keywords: blah; blah.

1 Introduction

In planning for the future labour market, it is necessary to forecast the age structure of the workforce, in order to enable informed decision-making on policies, especially concerning universities and immigration. We propose a statistical modelling approach to this problem, illustrated using various scientific disciplines in Australia, forecasting future workforce age structures over the next decade. The forecasts described have been used by the Australian Academy of Science as part of *Australian Science, Australia's Future: Science 2035*, an initiative assessing the capability of the national science system and its role in achieving Australia's ambitions (Australian Academy of Science 2025).

The economic implications of workforce age structure shifts are well-documented (e.g., Bloom et al. 2007), affecting productivity, pensions and superannuation, and skill shortages (Productivity Commission 2013; OECD 2019a,b; Hyndman, Zeng & Shang 2021). The social implications are also significant, with an aging workforce leading to changes in workplace dynamics, potential problems with intergenerational knowledge transfer, and the need for policies that support older workers. Yet this problem does not appear to have been previously addressed from a statistical modeling perspective.

Our approach builds on functional data models, introduced to demographic modelling by Hyndman & Ullah (2007). They combined nonparametric smoothing and functional principal components for age-specific demographic rates. These models were then used by Hyndman & Booth (2008) for mortality, fertility, and migration rates, providing stochastic data generating processes for the components of demographic balance equations. These separate component models were then simulated to form future sample paths, leading to age- and sex-specific stochastic population forecasts. The modelling framework was later extended by Hyndman, Booth & Yasmeen (2013) to ensure coherence of forecasts between sexes or other demographic groups.

We propose to adapt this framework to workforce dynamics by redefining the demographic components in two ways. First, we replace fertility with workforce entry, which functions more like a migration process than a birth process because graduates can enter the workforce at any age. Second, we allow workers to leave the workforce through two processes: retirement and death. Of course, people may also leave the workforce for other reasons, such as a career change or family commitments, but since we do not have data on these processes, we model them implicitly via a remainder term.

We describe the methodology in Section 2. By way of illustration, we apply the methodology to major scientific disciplines in Australia, focusing on the Natural and Physical Sciences. We describe the data sources in Section 3, with the results provided in Section 4. The aim of this analysis is to inform future workforce planning and policy decisions to support the growth of Australia's scientific community. Finally, we provide some discussion and conclusions in Section 5.

2 Methodology

Suppose our workforce is divided into I groups, indexed by $i = 1, \dots, I$. In our application, these are scientific disciplines, but in principle they could refer to any subdivision of workers. Let $P_{i,x,t}$ denote the number of equivalent full-time workers in group i who are aged x at the start of year t, where $x = 15, 16, \dots$. The starting age of 15 is chosen because it is the minimum age at which individuals are counted as part of the labour force in the Australian Census (Australian Bureau of Statistics 2021b). Forecasts are required for $P_{i,x,T+h}$ across all ages and groups, for some forecast horizon $h > 0$.

People can leave the workforce of a group through death, retirement, career change, or emigration. They can enter the workforce through graduation, immigration, or career change. Unfortunately, we typically do not have data on many of these processes, so we will combine career change and migration into a remainder term, which we denote as $E_{i,x,t}$. We usually do have data on numbers of deaths ($D_{i,x,t}$), numbers of retirements ($R_{i,x,t}$), and numbers of graduates ($G_{i,x,t}$); the numbers in each case are for people aged x at the start of year t.

The population can be forecasted using a model similar to the stochastic population model of Hyndman & Booth (2008), excluding the birth process:

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + G_{i,x,t} + E_{i,x,t}, \quad x = 15, 16, \dots \tag{1}$$

That is, the population each year is equal to the population from the previous year having aged 1 year, minus the deaths or retirements that occurred during the previous year, plus the new graduates, plus any other changes due to migration or career change (which may be negative). We assume that D, R and G are all non-negative, $D_{i,x,t} + R_{i,x,t} \leq P_{i,x,t}$, and $E_{i,x,t} = G_{i,x,t} = 0$ above some age threshold (say $x = 100$). Once $P_{i,x,t} = 0$ when x is above that threshold, all future populations $P_{i,x+k,t+k} = 0$, $k = 1, 2, \dots$. That is, when the cohort aged x in year t has all retired or died, they will not be replaced through new graduates or migrants of the same age.

As a first approximation, the components D, R, E and G can be assumed to behave independently for each combination of i, x and t. In reality, there may be a negative correlation between G and E as insufficient graduates would probably lead to employers finding people from overseas, while too many graduates would lead to scientists seeking work elsewhere.

To use Equation 1, data on P, D, R, and G is required, and the remainder component E will be calculated using Equation 1. It is unlikely that we have available separate death and retirement counts for each group, so we will instead use probabilities of death and retirement across all groups, giving $D_{i,x,t} = P_{i,x,t}q_{x,t}$ and $R_{i,x,t} = P_{i,x,t}r_{x,t}$ where $q_{x,t}$ is the probability of death at each x in year t, and $r_{x,t}$ is the probability of retirement at each x in year t. Similarly, graduation numbers are rarely available by discipline and age, so we will approximate $G_{i,x,t} = g_x G_t$ where $G_{i,t}$ is the total number of graduates in year t and g_x is the proportion of graduates by age across all disciplines. So our more realistic model becomes

$$P_{i,x+1,t+1} = P_{i,x,t}(1 - q_{x,t} - r_{x,t}) + g_x G_{i,t} + E_{i,x,t}, \quad x = 15, 16, \dots \tag{2}$$

To obtain forecasts of P, forecasts of q, r, G, and E must then be computed. Each of these can be estimated using a functional data model (Hyndman & Ullah 2007).

3 Data

3.1 Data sources

This analysis focuses on the Natural and Physical Sciences as defined in the Australian Standard Classification of Education (ASCED) by the Australian Bureau of Statistics (2001). We refer to ASCED's Narrow Fields as “disciplines”; these comprise Physics and Astronomy, Mathematical Sciences, Chemical Sciences, Earth Sciences, Biological Sciences, Other Natural and Physical Sciences, and Natural and Physical Sciences not further defined (n.f.d.). We define the population of workers in a discipline as those who are active in the labour market and hold a bachelor's degree or higher in that discipline.

3.2 Labour force participation

Data on $P_{i,x,t}$ are sourced from the *Census of Population and Housing* (2006, 2011, 2016, and 2021) (Australian Bureau of Statistics 2023). This dataset encompasses labour force participation status, age group, the highest level of completed non-school qualification level (QALLP), the corresponding field of study [QALFP; Australian Bureau of Statistics (2021c)], and the industries in which individuals work.

However, labour force participation data is available only for 2016 and 2021 (Australian Bureau of Statistics 2021a). To estimate worker numbers for 2006 and 2011, the average participation rates from 2016 and 2021 were applied, assuming overall patterns remain consistent. While participation rates were generally stable between 2016 and 2021, smaller groups experienced more volatility. Younger and older age groups within some disciplines were particularly affected by this volatility, especially when sample sizes were small. For example, in a group of five 15–19-year-olds, a single individual's employment status change could significantly shift the participation rate. Similarly, fluctuations in smaller cohorts of older scientists impacted participation rates. Averaging the rates from 2016 and 2021 helps mitigate this volatility and provides a more stable estimate for 2006 and 2011. Any discrepancies in participation rates are minimal due to the small cohort sizes, making this a more reliable approach than relying on a single year.

The census data on scientists active in the labour market, provided by the ABS in one-year age groups, is shown in Figure 1 as the thick lines. As data is only available in census years (2006, 2011, 2016, 2021), cohort interpolation is used to estimate values for the intervening years. Cohort interpolation methods (Stupp 1988) apply linear interpolation within each age cohort between census years.

3.3 Smoothed Disaggregation

Retirement intentions data is categorised by the industry of an individual's main job and is provided in broad age groups, as shown in Figure 2. While the ABS dataset covers 19 industries, only the key industries relevant to this study are displayed in the figure.

To estimate single-year age distributions, a monotonic cubic spline (Smith, Hyndman & Wood 2004) is applied directly to the cumulative values of these age groups, ensuring a smooth transition between them. This method preserves the overall distribution while generating continuous single-year estimates, as shown in Figure 3.

3.4 Retirement Distribution by Age

After smoothing the retirement intentions data, the appropriate industry-level retirement distribution or combination of distributions must be selected. Since scientists work across multiple industries,

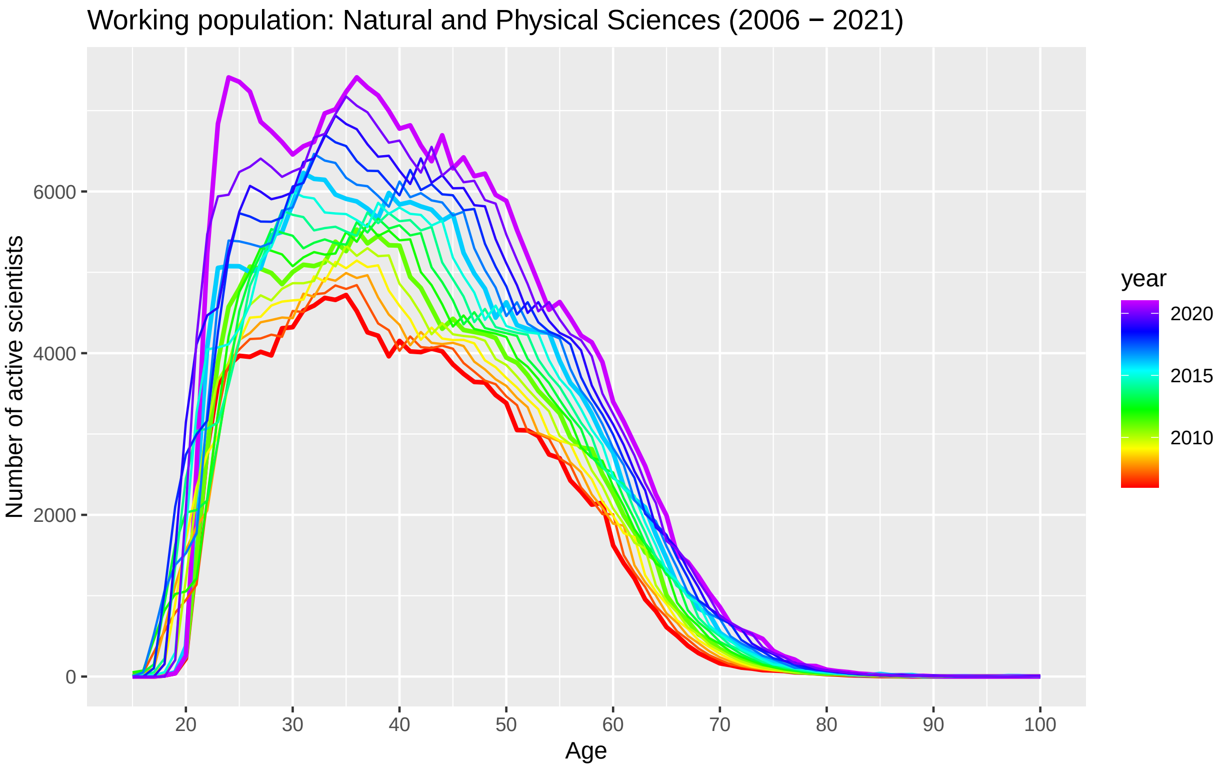

Figure 1: *Interpolated number of scientists in Australia, 2006–2021, using linear cohort interpolation between census years. Thicker lines are used to denote census years.*

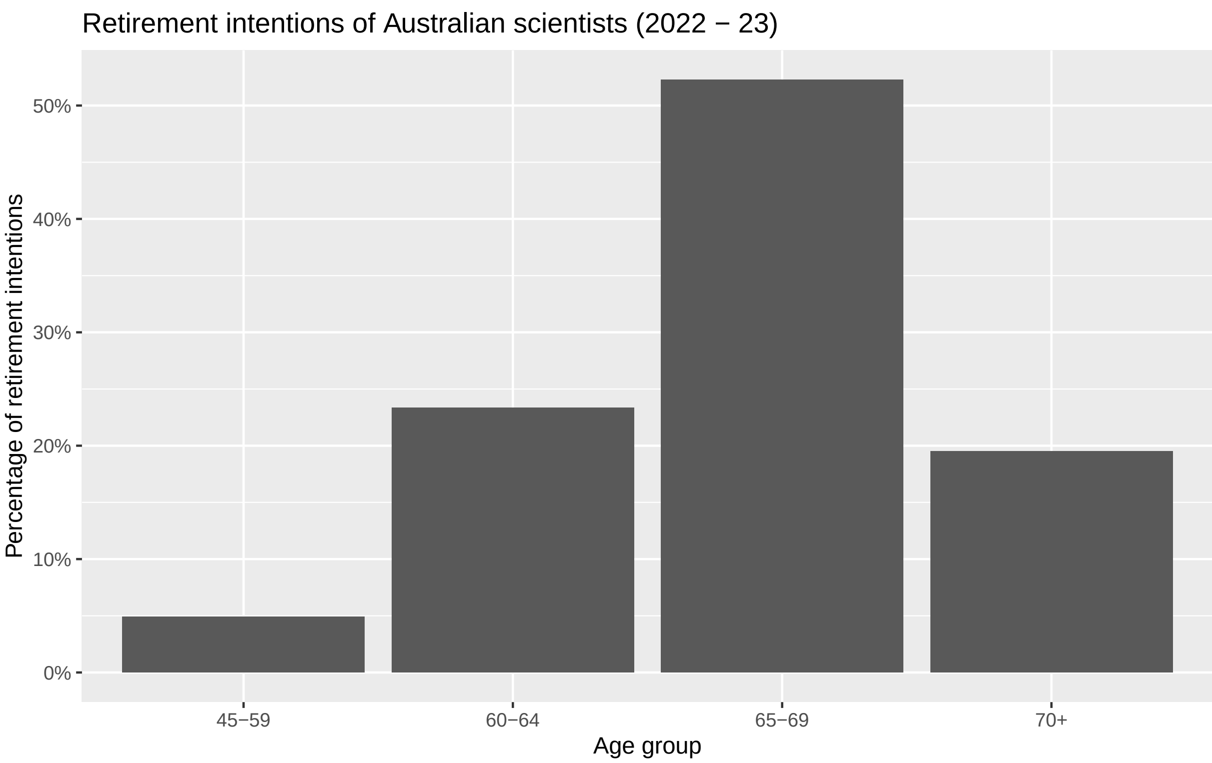

Figure 2: *Retirement intentions for key industries before smoothed disaggregation.*

industry-level data must be adapted to reflect discipline-specific trends. While distributions may vary, retirement trends generally follow similar patterns across sectors. By focusing on the top industries, which account for the majority of workers in each discipline, a stable foundation for analysis is ensured, minimising unnecessary variability arising from changing worker proportions in different industries over time.

For the Natural and Physical Sciences, the three primary industries in which the workforce is employed are Education and Training, Professional, Scientific and Technical Services, and Health Care and Social Assistance, representing 15.81%, 15.48%, and 14.65%, respectively. The proportions in other industries decrease significantly beyond this point. To estimate retirement intentions, a weighted average was calculated using these top three industries, with proportions rescaled to sum to 1.

The data show that a significant portion of workers plan to retire between the ages of 60 and 70. This distribution of retirement ages exhibits positive kurtosis, with the highest peak at age 67, where 13.16% of workers indicate their intention to retire, as shown in Figure 3.

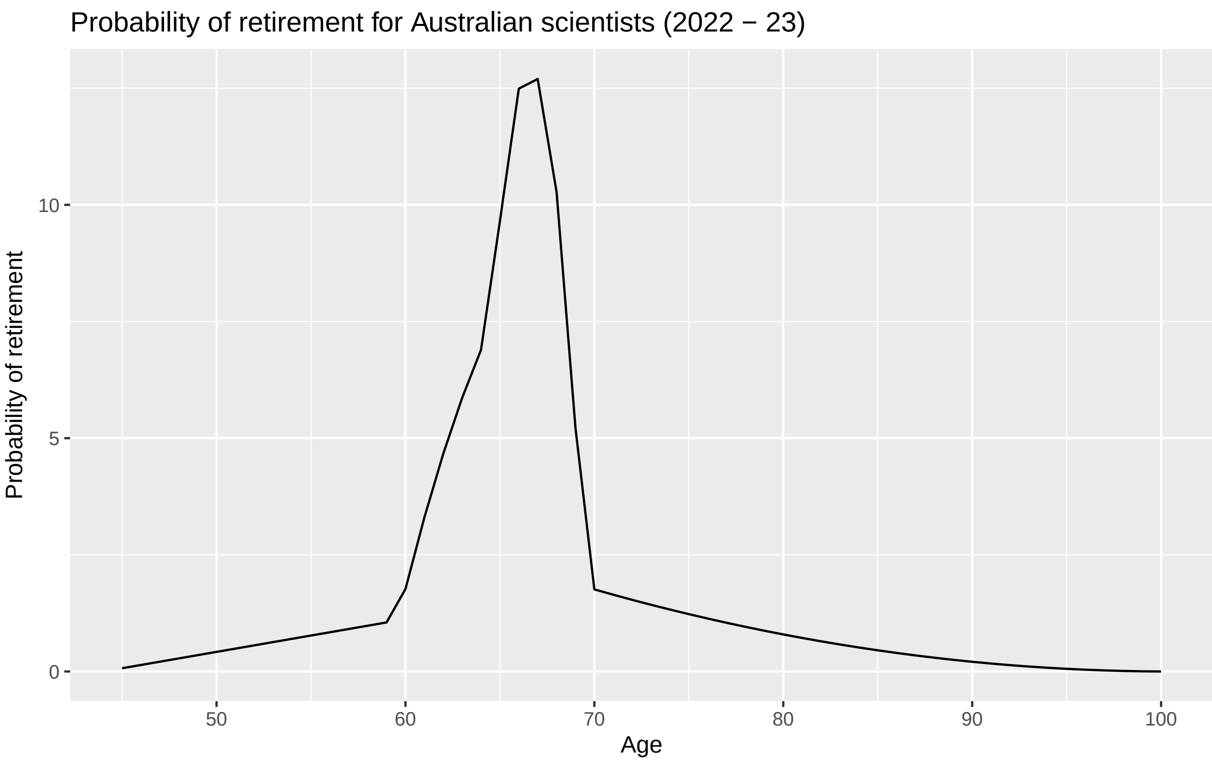

Figure 3: *Age distribution of retirement intentions for the 2022–2023 Australian financial year.*

From this distribution, the empirical probability of retirement at each age can be calculated using standard life table methods (Macdonald, Richards & Currie 2018). These probabilities are shown in Figure 4. These probabilities, which exhibit a significant increase after age 59, are used to estimate $R_{i,x,t}$ from $P_{i,x,t}$, and are treated as fixed for the simulation, assuming that retirement patterns will remain stable over time.

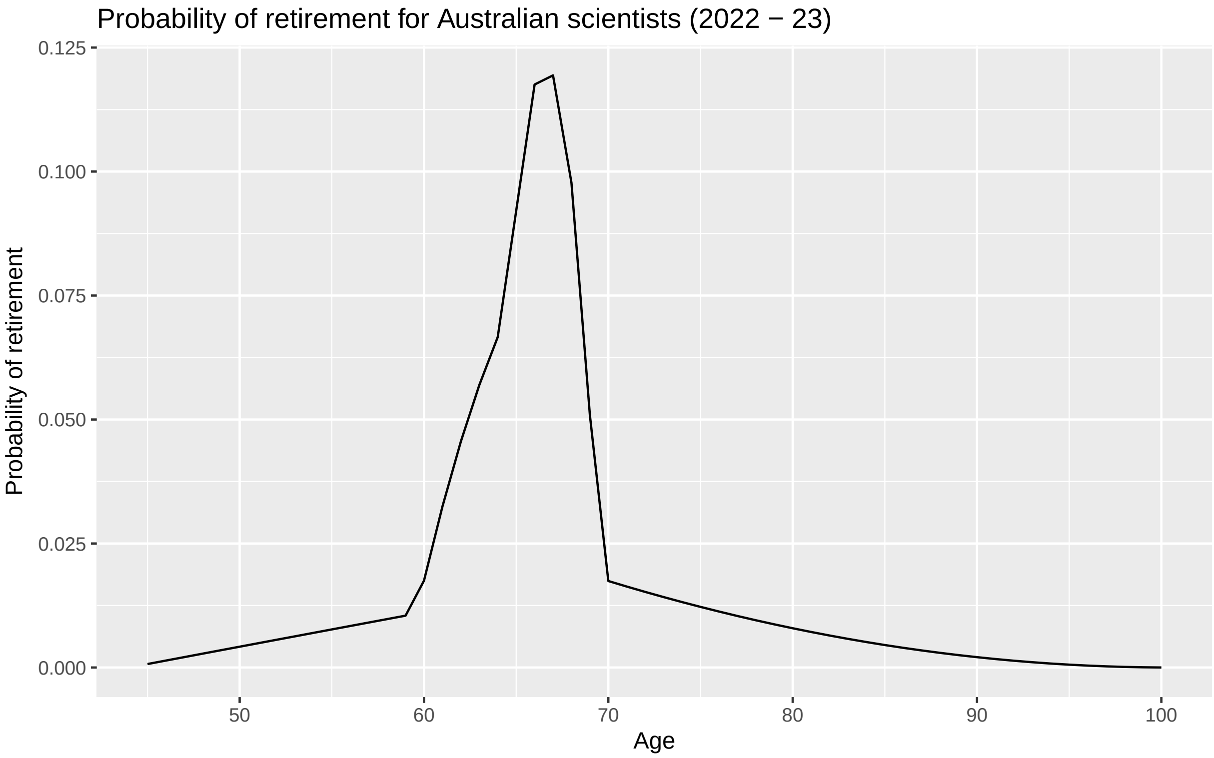

Figure 4: *Age-specific probability of retirement based on the 2022–2023 Australian financial year.*

In this approach, retirement is treated as a final state, meaning that once a scientist retires, they are not reintroduced into the workforce. This assumption considers workforce exit to be final. Mortality data, discussed in the next section, is incorporated into the calculation of the total probability of leaving the workforce, adjusting for the possibility of death before retirement. This ensures that the combined probability of exiting the workforce (through retirement or death) does not exceed 100%.

Projections extend over a 20-year period to 2041, with particular focus on workforce composition in 2035. These estimates account for key demographic factors, including mortality, retirement, graduation, and migration.

This analysis utilises Australian data sources to generate accurate forecasts. These include the ABS, the Human Mortality Database (HMD), and the Department of Education.

Australian age-specific death counts and population exposures from 1971 to 2021 were sourced from the Human Mortality Database (2024). This dataset includes death counts and population exposures categorised by age.

Retirement data is sourced from the *Retirement and Retirement Intentions* dataset (Catalogue 6238) for the 2022–2023 financial year (Australian Bureau of Statistics 2022-23). It provides insights into expected retirement ages for workers in age groups ranging from 45 to 70 across various industries.

Graduate completion statistics from 2006 to 2023, categorised by age group and including individuals with a bachelor's degree or higher, are derived from the *Award Course Completions* dataset (Department of Education 2024b). Additionally, the Department of Education provides data on the number of graduates with a bachelor's degree or higher, categorised by discipline and year, for the same period (Department of Education 2024a). These datasets include both domestic and international students.

Data limitations include the availability of labour force participation data only for 2016 and 2021, the absence of full-time worker data outside census years, and the aggregated nature of retirement intentions data. These gaps are addressed using specific handling of labour force participation data, cohort interpolation methods, and smoothed disaggregation, as detailed in Sections 3.2, **?@sec-cohort**, and 3.3.

Another limitation is that retirement intentions data is unavailable by educational discipline, with industry-level data serving as the closest proxy. Since scientists work across multiple industries, this does not capture discipline-specific retirement patterns but is used in this context through methods outlined in Section 3.4.

3.5 Death Probabilities by Age

Age-specific mortality rates from 1971 to 2021 are obtained from the Human Mortality Database (2024). Using standard life table methods, these rates are converted into age-specific probabilities of death, as shown in Figure 5. Over time, mortality probabilities have generally declined across all age groups, reflecting improvements in Australian life expectancy. Consequently, fewer scientists are expected to exit the workforce due to death, illness, or related factors. Similar to retirement, death is treated as a final state; once a scientist dies, they are no longer part of the workforce.

Although the final workforce forecasts focus on individuals aged 15 and older, mortality probability estimates are derived from data covering ages 0 to 100. These probabilities serve as inputs for estimating $D_{i,x,t}$, the number of deaths among full-time equivalent workers in scientific discipline i at age x in year t, based on $P_{i,x,t}$, the corresponding population of such workers. Death estimates are calculated by applying the forecasted mortality probabilities to the workforce population.

3.6 Graduate Completions by Age

Data on graduate completions by age and year is not available for individual disciplines. However, it is available in aggregate across all disciplines, allowing for an analysis of how the age distribution of graduates has changed over time.

Figure 6 shows the percentage distribution of graduate completions by age for each year from 2006 to 2023. Some missing values result in gaps in certain lines, but the overall pattern remains highly

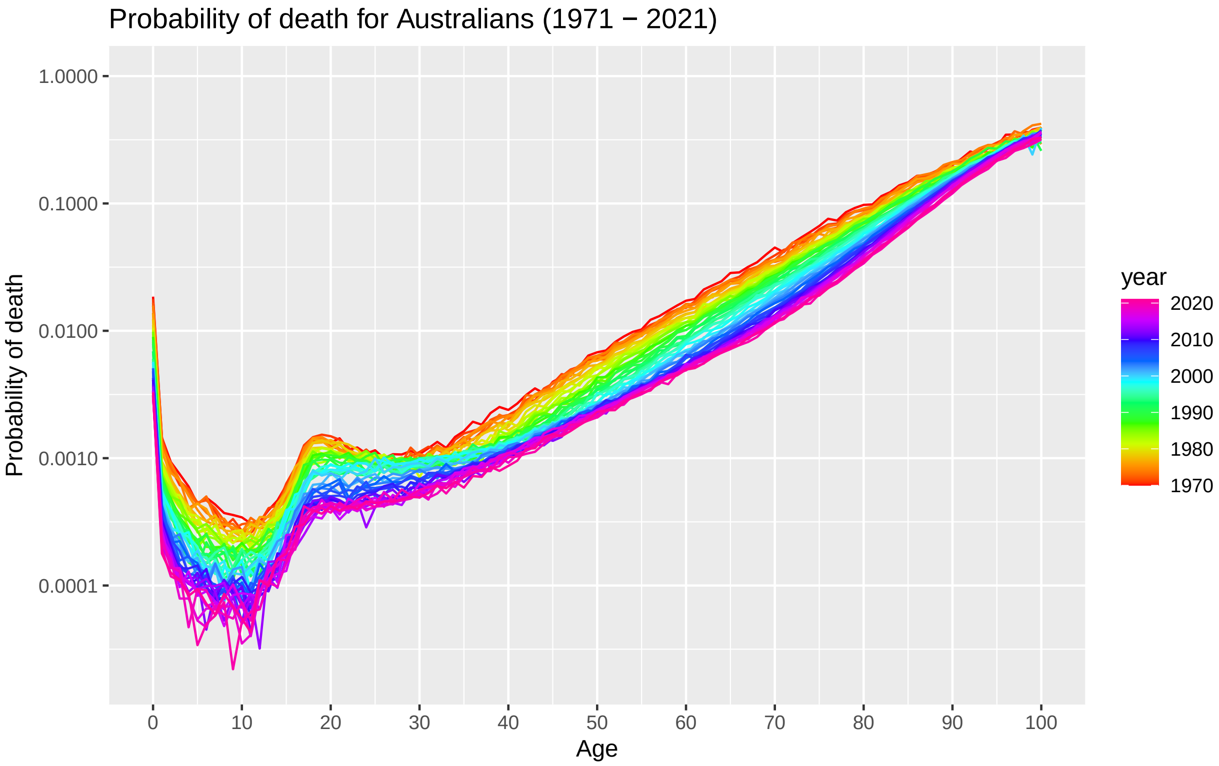

Figure 5: *Age-specific probabilities of death (on a logarithmic scale) for each year from 1971 to 2021.*

consistent across years. Given this consistency, the data is averaged across all available years to construct a representative age distribution of graduates at the bachelor's level and higher.

The resulting averaged distribution, shown in Figure 7, smooths out year-to-year fluctuations and provides a stable estimate of graduate completions by age. This transformation assumes that the overall age distribution of graduates remains relatively stable over time, allowing for a more reliable estimate when applied to specific disciplines.

Table 1 highlights the percentage distribution of graduates aged 20 to 25, demonstrating a pronounced peak in this range, with the highest percentage observed at age 21. This reflects the typical age at which students complete their tertiary education.

Table 1: *Percentage distribution of graduate completions for ages 20-25 for the years 2006 to 2023.*

Percentage of Graduates Ages 20-25	
Age	Percentage of Graduates
20	8.64%
21	12.78%
22	12.19%
23	10.48%

Table 1: *Percentage distribution of graduate completions for ages 20-25 for the years 2006 to 2023.*

Percentage of Graduates Ages 20-25	
Age	Percentage of Graduates
24	8.31%
25	6.13%

3.7 Net Migration

The demographic growth-balance equation (Equation 1), when rearranged, provides an estimate of net migration:

$$G_{i,x,t} = P_{i,x+1,t+1} - P_{i,x,t} + D_{i,x,t} + R_{i,x,t} - N_{i,x,t}, \quad x = 15, 16, \dots \tag{3}$$

In this equation, G represents the difference required to balance population changes over time after accounting for deaths (D), retirements (R), new graduates (N), and the population at time t. It serves as the residual value, reflecting the components influencing the workforce's age structure.

4 Results

Table 2 lists the scientific disciplines (as per Australian Bureau of Statistics (2001) Narrow Fields) included in this analysis, along with their corresponding Detailed Fields. This table provides essential context for understanding the specific areas classified under each discipline.

4.1 Working Population

The Australian Census provide working population data every five years. Figure 8 shows the age distribution of scientists active in the labour market as recorded in the Census for the years 2006, 2011, 2016, and 2021. There are notable differences over time, and between disciplines. The number of working scientists has increased over time, especially in Mathematical Sciences in the latest 2021 Census, as well as in Biological Sciences and Other Natural and Physical Sciences. The right tail has gradually grown heavier, reflecting an increasing proportion of older scientists, while the distribution has also become wider, indicating a broader age range in the workforce.

The working population for years between census are estimated using linear cohort interpolation (Stupp 1988), with the resulting population estimates shown in Figure 9. These interpolated values form the basis for forecasting the future working population, representing the values of $P_{i,x,t}$.

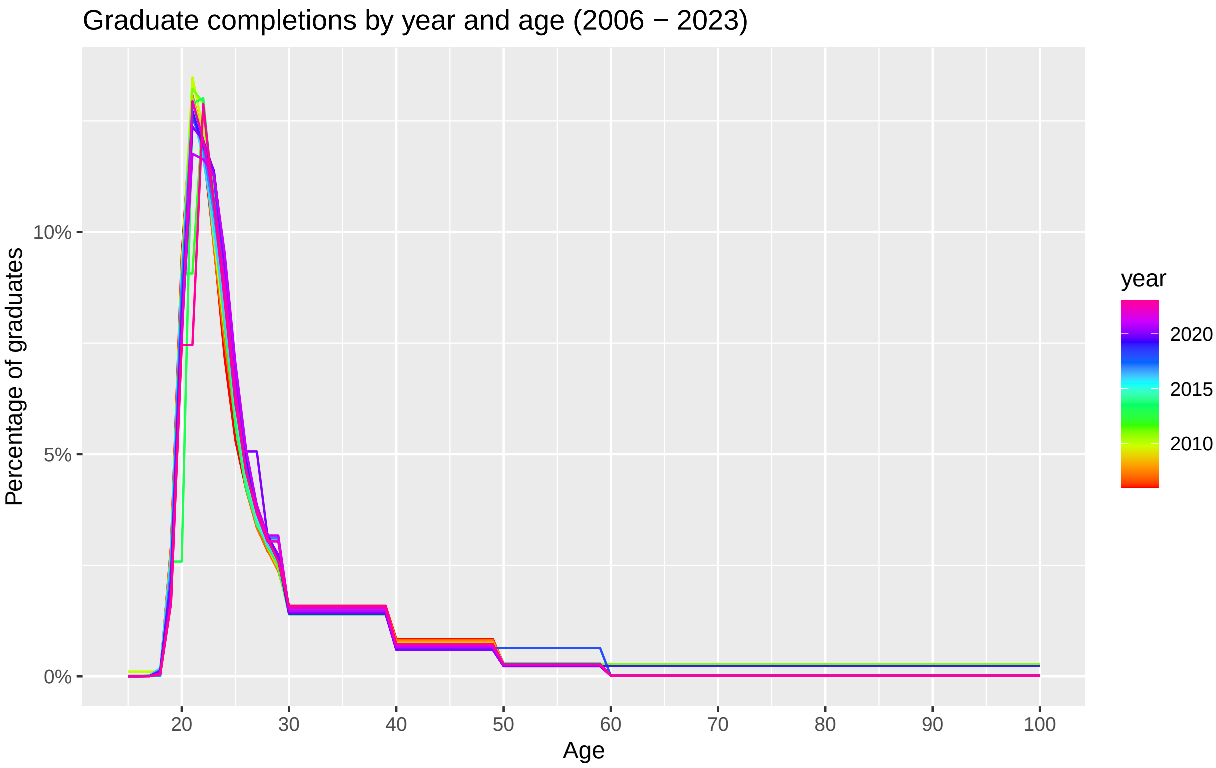

Figure 6: *Percentage distribution of graduate completions by age for each year from 2006 to 2023.*

Figure 7: *Percentage distribution of graduate completions by age, aggregated over the years 2006 to 2023.*

Table 2: *Classification of scientific disciplines, based on the ASCED Narrow Fields of Education within the Broad Field of Natural and Physical Sciences. The table lists their corresponding Detailed Fields. "n.e.c." stands for "Not Elsewhere Classified."*

Broad Field: Natural and Physical Sciences	
Narrow Fields	Detailed Fields
Physics and Astronomy	Physics, Astronomy.
Mathematical Sciences	Mathematics, Statistics, Mathematical Sciences, n.e.c.
Chemical Sciences	Organic Chemistry, Inorganic Chemistry, Chemical Sciences, n.e.c.
Earth Sciences	Atmospheric Sciences, Geology, Geophysics, Geochemistry, Soil Science, Hydrology, Oceanography, Earth Sciences, n.e.c.
Biological Sciences	Biochemistry and Cell Biology, Botany, Ecology and Evolution, Marine Science, Genetics, Microbiology, Human Biology, Zoology, Biological Sciences, n.e.c.
Other Natural and Physical Sciences	Medical Science, Forensic Science, Food Science and Biotechnology, Pharmacology, Laboratory Technology, Natural and Physical Sciences, n.e.c.

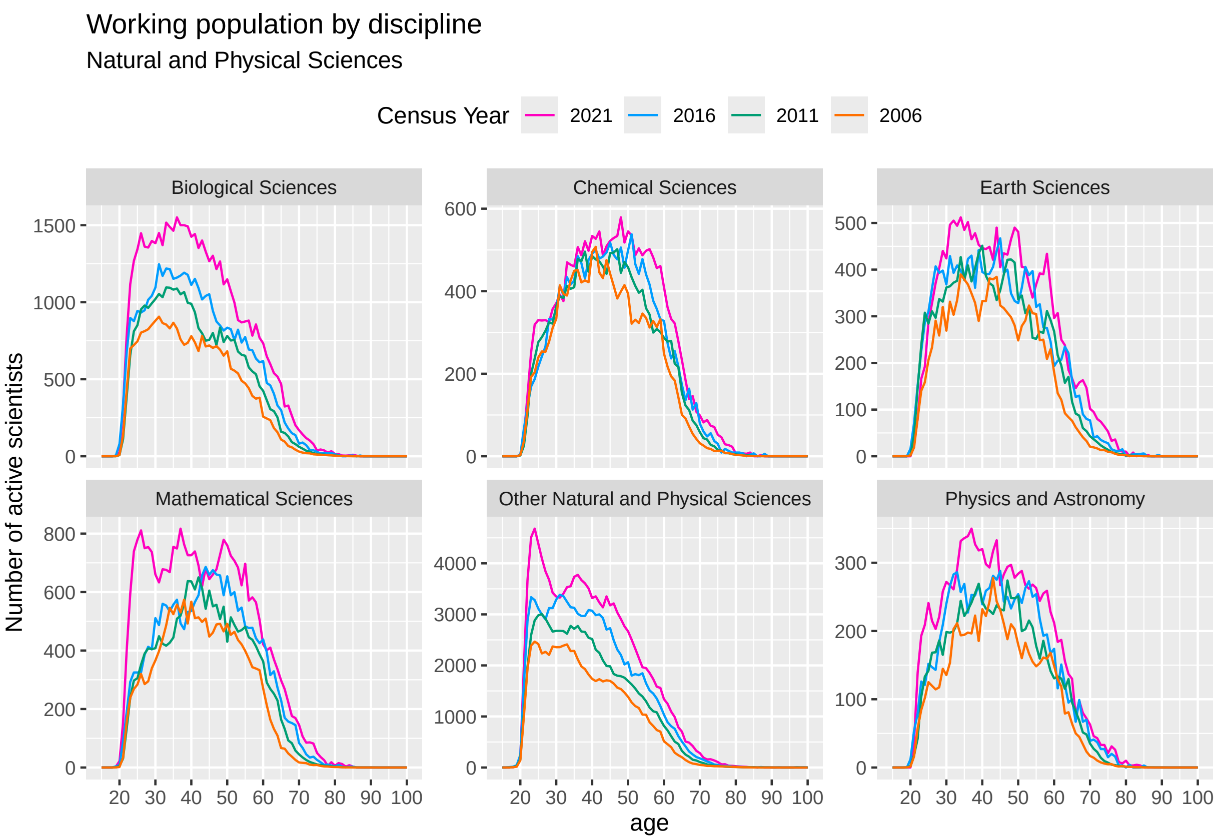

Figure 8: *Age distribution of the working population across scientific disciplines based on Census data from 2006, 2011, 2016, and 2021. Each panel represents a different discipline, with lines indicating the number of active scientists in the Australian labour market by age for each Census year.*

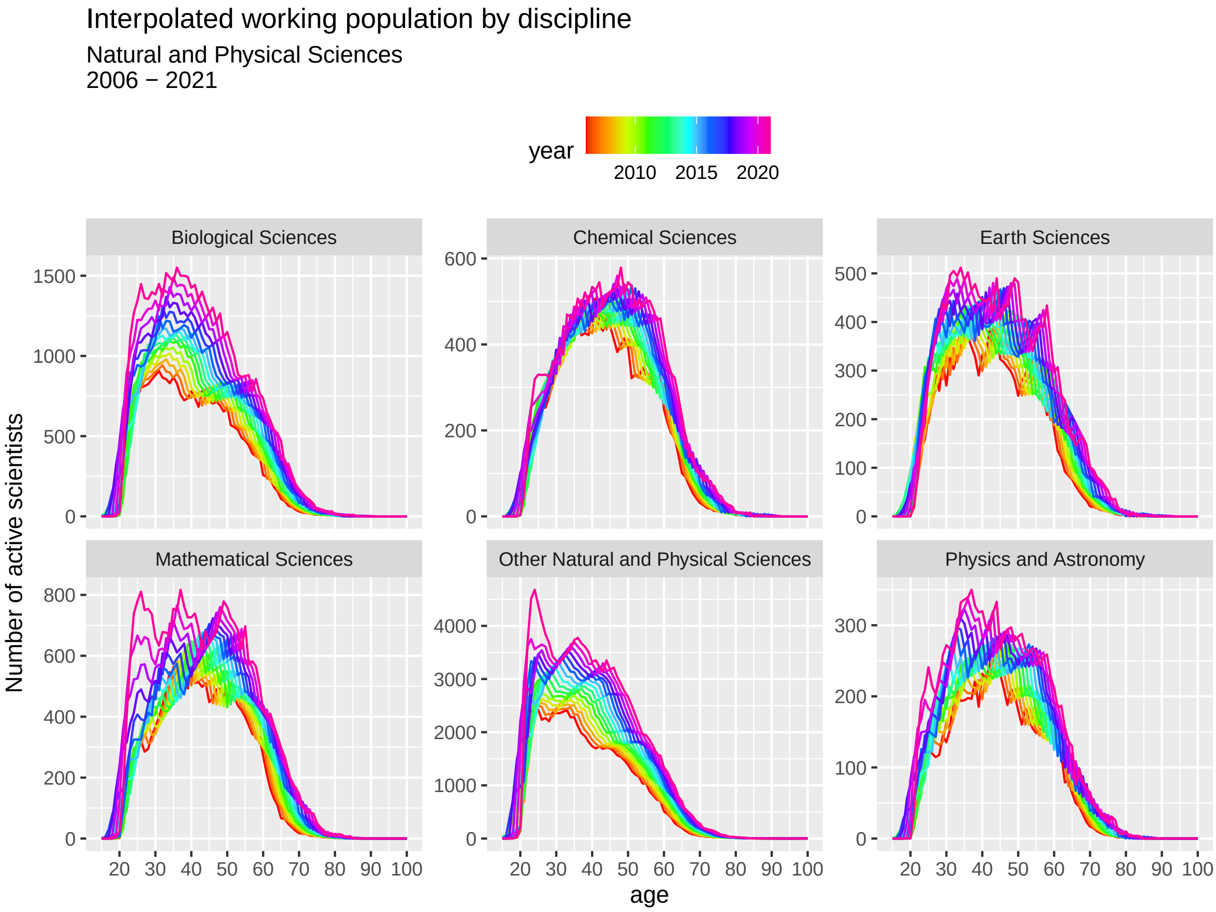

Figure 9: *Age distribution of the working population across scientific disciplines, interpolated from Census data using linear cohort interpolation. Each panel represents a different discipline, with lines indicating the number of active scientists in the Australian labour market by age and year.*

4.2 Graduate Completions

Figure 10 shows the number of total graduate completions from 2006 to 2023 ($\sum_x N_{i,x,t}$). A significant positive trend is observed in three disciplines: Mathematical Sciences, Other Natural and Physical Sciences, and Natural and Physical Sciences (n.f.d.). Note that the Y-axis scales vary across disciplines to reflect differences in the number of completions.

The increase in the working population observed in the 2021 Census for Mathematical Sciences, as mentioned earlier, can be partly attributed to the sharp rise in graduate numbers between 2016 and 2021. This surge in graduates during that period provides insight into the corresponding growth in the working population.

To forecast future graduate numbers, a global ARIMA model was employed, following the principles outlined by Hyndman & Montero-Manso (2021). ARIMA is a time series forecasting method that identifies underlying trends in past data, assuming these trends will continue. The global model captures overall trends across disciplines by scaling graduate data within each discipline, ensuring proportional contributions from all disciplines before fitting the global ARIMA model. This improves robustness by incorporating information across disciplines. The model is then used to generate discipline-specific forecasts, with simulations run 500 times to account for variability and generate 90% prediction intervals. The results, along with these intervals, are shown in Figure 10.

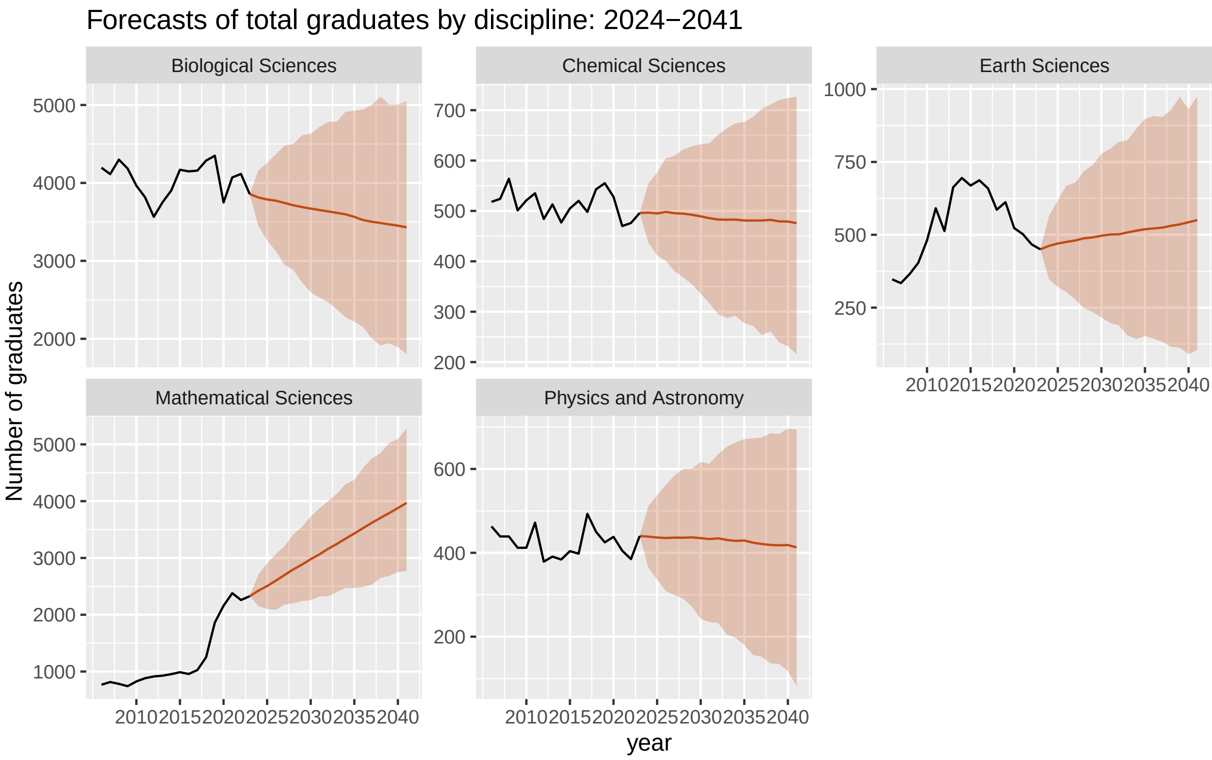

Figure 10: *Forecast number of graduates by discipline, 2024-2041, with 90% prediction intervals, based on historical data from 2006–2023.*

Since workforce projections require not only total graduate numbers but also their age distribution, the forecasted totals (depicted in Figure 10) are disaggregated by age using the percentage distribution of graduate completions by age across all disciplines (as shown in Figure 7). This enables an estimation of how many graduates will enter each age group in future years, ensuring consistency with observed patterns in the broader graduate population.

4.3 Net Migration

Net migration ($G_{i,x,t}$) is estimated using the demographic growth-balance equation, which is rearranged to calculate the net number of migrants each year (Equation 3). Figure 11 shows the age distribution of net migration among scientists between 2006 and 2020. Emigration during the early career stage is largely driven by international students graduating and returning home. This negative net migration in the early career stage is a significant factor across all disciplines.

Significant emigration is observed between the ages of 20 and 27, primarily driven by international students leaving Australia after completing their studies, with some domestic graduates also seeking employment overseas. As the graduate data includes both domestic and international students, this emigration is largely influenced by international students returning home.

Beyond this age range, net migration fluctuates between small levels of immigration and emigration. However, these variations are relatively minor, with annual net migration reaching no more than approximately 15 individuals per age group. Given the small magnitudes, these fluctuations may not reflect meaningful migration patterns but rather noise or systematic cohort effects due to input data limitations.

As outlined in the methodology, these estimates rely on forecasting the other components (D, N, and R). Therefore, only estimated historical values up to 2020 are shown here. While historical data on the working population extends to 2021, net migration values for 2021 can only be obtained by incorporating forecasted components due to the structure of the equation.

Other Natural and Physical Sciences and Natural and Physical Sciences (n.f.d.) show little to no immigration beyond this stage. Given their larger workforce sizes (as seen on the y-axis), these disciplines exhibit more stable net migration patterns.

In contrast, core scientific disciplines, which have smaller populations, show more fluctuation in migration levels after the early career stage. This variability is likely due to workforce mobility, the sensitivity of estimates for smaller populations, and cohort effects, rather than reflecting actual migration dynamics.

4.4 Simulating Future Populations

To forecast the future age structure of the scientific workforce in Australia, a simulation-based approach is employed. This approach uses a stochastic population model, adapted from the demographic growth-balance model described in Hyndman & Booth (2008) and excluding birth processes, to project future populations. This model incorporates uncertainties in demographic processes, generating a range of potential future scenarios for the scientific workforce. The following steps outline the process for projecting the future working population:

1. **Death Probabilities:** Fit a functional data model (Hyndman & Ullah 2007) to the death probabilities shown in Figure 5, with ARIMA models for the coefficients. This model is used to simulate values of age-specific death probabilities in future years.
2. **Net Migration:** Fit a separate functional data model to the age-specific migrant numbers shown in Figure 11, with ARIMA models for the coefficients. This model simulates values of age-specific net migrant numbers in future years.

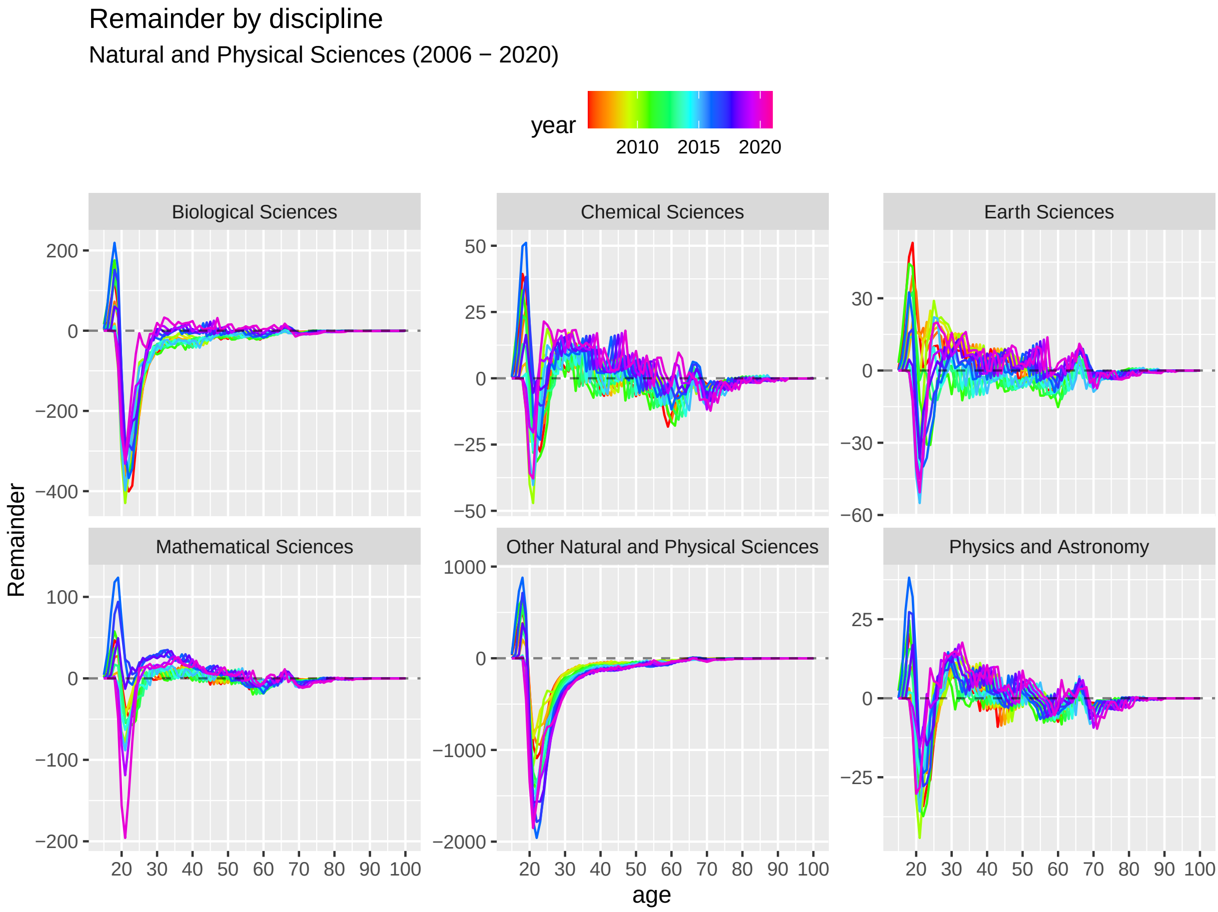

Figure 11: *Estimated net migration by age across scientific disciplines (2006–2020). Each panel represents a different discipline, showing net migration trends across age groups.*

3. **Graduate Numbers:** Fit an ARIMA model (Hyndman & Athanasopoulos 2021) to simulate total numbers of graduates in each future year from the numbers shown in Figure 10. These graduate numbers will then be disaggregated by age using the age distribution shown in Figure 7 to estimate age-specific graduate numbers in each year.
4. **Retirements:** Estimate retirements in each year by multiplying the number of people of working age in each age bracket (Figure 9) by the retirement probabilities shown in Figure 4. Fixed age-specific retirement probabilities are used.
5. **Deaths:** Estimate deaths in each year by multiplying the number of people of working age in each age bracket by the forecast death probabilities.
6. **Demographic Growth-Balance:** Use the demographic growth-balance equation Equation 1 to obtain the next year's age-specific population values. This step incorporates the graduate numbers, net migration, retirements, and deaths, to produce the forecasted population for the next year.

A total of 500 simulations are run to obtain a distribution of possible future population scenarios, allowing for quantification of forecast uncertainty. Figure 12 shows two example simulations. Historical data (2006–2021) are shown in gray, with simulated future populations (2022–2041) in colour. While both simulations exhibit similar trends, differences arise due to inherent uncertainties in model inputs, including variations in graduate numbers, migration patterns, and workforce exits through death and retirement. These sources of uncertainty are incorporated into the model, contributing to the range of possible future trajectories.

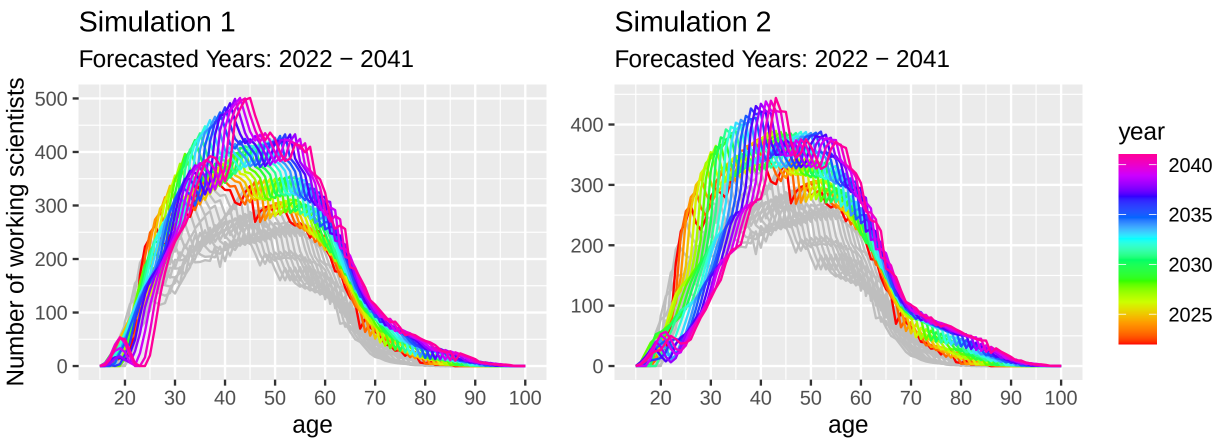

Figure 12: *Two example simulations from a set of 500, using historical data from 2006-2021 to forecast age-specific population distributions for 2022-2041.*

4.5 Forecast Results

The average of the 500 simulations provides the mean age-specific forecast, while quantiles estimate forecast uncertainty. **?@fig-forecast-working-population** presents the mean and 90% prediction intervals for 2025 and 2035.

In 2035, forecast variability is highest at younger ages, starting from 15 and widening before gradually narrowing as the workforce ages. The prediction interval becomes especially narrow during the retirement phase, where the workforce dynamics become more predictable, indicating greater certainty in this portion of the forecast.

Prediction intervals are also shown for 2025, as it represents a forecast rather than observed workforce data. This comparison helps illustrate how the age distribution is expected to evolve over the next decade.

Between 2025 and 2035, the forecasts indicate an aging workforce, with increases in the number of scientists in their late 30s, and from ages 45 to 80. This trend is particularly evident in the wider gap at ages 45-60 between the forecasted distributions over the ten-year period.

Additionally, patterns in the age distribution persist over time. A dip or trough observed at a given age in 2025 reappears 10 years later as the same cohort progresses through the workforce. While these shifts may be harder to distinguish due to the structure of single-year age cohort data, they reveal how workforce cohorts evolve and how observed patterns carry through in future projections.

Summing over age groups allows for estimating the total number of working scientists in each future year, as shown in Figure 13. This highlights whether the workforce is growing, stabilising, or declining. On average, projections indicate continued growth, but at a gradually slower pace. The lower bound remains nearly flat, suggesting workforce stagnation in a conservative scenario. Even in an optimistic scenario, growth only slightly exceeds the current pace.

After 500 simulations of the future population for each discipline, the projections ten years apart (2025 and 2035) shown in **?@fig-forecast-working-population** provide a basis for comparing estimates of the current and future workforce.

The projections indicate wider prediction intervals for younger age groups due to uncertainty in future graduate numbers, while mid-to-late career estimates primarily reflect the aging of existing cohorts. Across disciplines, the typical working population spans from the mid-to-late 20s (10th percentile) to the mid-50s and early 60s (90th percentile). Since retirements increase after the late 50s, workforce participation beyond 60 serves as a benchmark for identifying trends in delayed retirement and extended career duration.

The key findings from **?@fig-forecast-working-population** are summarised below:

- **Physics and Astronomy** *(Aging Workforce):* The 35–45 age group shows a higher average than the previous 2025 cohort. There is a growing number of scientists aged 45 to 60, and more individuals aged 60+ are remaining in the workforce longer. These patterns indicate a gradual transition toward an older age structure, reflecting an aging workforce.
- **Mathematical Sciences** *(Expansive Workforce):* Workforce growth is evident across a broad range of ages, particularly among individuals aged 25 to 70. The most significant increase occurs before 60, though there is also a clear rise in older workers (60+) remaining in the workforce rather than retiring. Overall, workforce growth is strong and evident across age groups.
- **Chemical Sciences** *(Stationary Workforce):* The average workforce size remains stable, with a relatively narrow prediction interval, indicating a predictable pattern of stability over time. While there is a minor increase in the number of workers aged 70, this change is not substantial enough to suggest a broader shift in workforce dynamics.
- **Earth Sciences** *(Aging Workforce):* Early-career workforce projections are more uncertain relative to its size. A decline in the working population is observed in the mid-thirties, followed

by an increase throughout the forties, though both remain modest. While the overall workforce remains relatively stable, these patterns indicate a slow shift toward an older age distribution.

- **Biological Sciences** *(Aging Workforce):* The number of workers aged 50 to 65 is increasing, indicating growth in later-career participation.
- **Other Natural and Physical Sciences** *(Expansive Workforce):* Workforce growth is strong across ages 30 to 70, characteristic of an expanding workforce.
- **Natural and Physical Sciences (n.f.d.)** *(Aging Workforce):* Growth is primarily concentrated in the 55–70 age range, indicating delayed retirement and an aging workforce.

These trends are further validated by the total workforce projections shown in Figure 13. This figure presents the total number of working scientists by discipline at each point in time, illustrating changes in workforce size over the forecast period. While all disciplines exhibit widening prediction intervals resembling an open funnel, this primarily reflects increasing uncertainty over time as forecasts extend further into the future. The extent of this uncertainty varies by discipline, with some exhibiting more stable trends while others show greater variability in potential workforce outcomes.

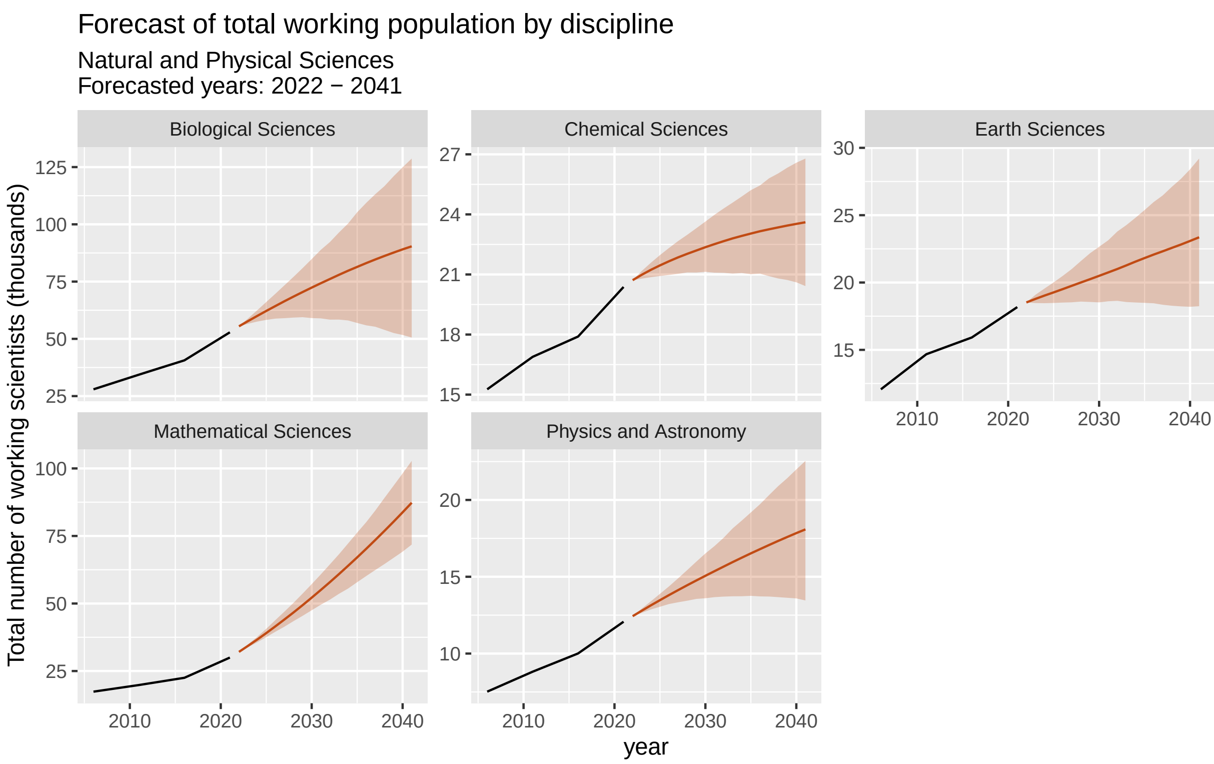

Figure 13: *Forecasted total number of working scientists across scientific disciplines from 2022 to 2041. The shaded region represents the 90% prediction interval, the coloured line indicates the mean estimate, and the black line represents historical data.*

From the plot, Mathematical Sciences, Other Natural and Physical Sciences, and Natural and Physical Sciences (n.f.d.) are projected to grow at an accelerating pace. Chemical Sciences, on the other hand, is more likely to decline, while the remaining disciplines are expected to experience more stable growth. However, it is important to consider the prediction intervals when interpreting these trends. These patterns are largely consistent with workforce structure classifications, with expansive disciplines experiencing faster growth, aging disciplines maintaining more stable trends, and constrictive disciplines showing signs of decline.

5 Conclusion

This report provides data-driven forecasts of Australia's science workforce age structures, outlining expected changes across disciplines by 2035. The Natural and Physical Sciences, along with Engineering, are projected to experience workforce aging, with growth concentrated in older age groups. Information Technology is expected to expand, though some disciplines remain uncertain due to high graduate numbers not translating into workforce participation. Smaller disciplines within Agriculture and Environmental Studies are projected to decline.

When interpreting these projections, policymakers should consider data limitations, particularly in discipline classification and reporting. Variations in self-reporting and ASCED categorisation may not fully capture industry or education changes, impacting accuracy, especially in smaller or evolving disciplines.

Additionally, while these projections provide a solid foundation for workforce planning, it is important to note that they are entirely driven by historical trends and do not account for possible new developments, such as the impact of AI and other emerging technologies on the labour market in different scientific disciplines. Other factors, such as policy changes or global economic shifts, may also influence workforce trends and should be taken into account.

While alternative demographic methodologies, such as the Lee-Carter model, could be considered for estimating demographic components, the methods employed in this report, including the functional data model, ARIMA model, and demographic growth-balance equation, are better suited to the specific characteristics of the data and the goals of this analysis.

Other quantitative studies, such as job market forecasting, may offer complementary insights into broader trends in the scientific workforce. Qualitative studies can also uncover latent factors that go beyond historical data. While this report focuses on forecasting based on past trends, these alternative approaches can provide additional perspectives on the scientific workforce in Australia.

Future work could enhance the accessibility of these projections through interactive visualisations, improve accuracy with new Census data, and broaden the analysis to additional ASCED fields. Refining estimation methods would further strengthen projection reliability and robustness.

This analysis contributes to the Australian Academy of Science's initiative *Australian Science, Australia's Future: Science 2035* by identifying shifts in the workforce, helping anticipate future demand for science capabilities, and highlighting potential workforce gaps as aging and retirement reshape workforce dynamics.

6 Acknowledgements

We thank Alexandra Lucchetti from the Australian Academy of Science for sourcing the data required for this project, and for helpful feedback on earlier versions of this report.

References

Australian Academy of Science (2025). *Australian Science, Australia's Future: Science 2035*. Australian Academy of Science. https://www.science.org.au/supporting-science/australian-science-australias-future-science-2035.

Australian Bureau of Statistics (2001). *Broad, Narrow and Detailed Fields*. ABS. https://www.abs.gov.au/statistics/classifications/australian-standard-classification-education-asced/2001/field-education-structure-and-definitions/structure/broad-narrow-and-detailed-fields.

Australian Bureau of Statistics (2021a). *Labour force participation flag (LFFP)*. ABS. https://www.abs.gov.au/census/guide-census-data/census-dictionary/2021/variables-topic/national-reporting-indicators/labour-force-participation-flag-lffp.

Australian Bureau of Statistics (2021b). *Labour force status (LFSP)*. ABS. https://www.abs.gov.au/census/guide-census-data/census-dictionary/2021/variables-topic/income-and-work/labour-force-status-lfsp.

Australian Bureau of Statistics (2021c). *Non-school qualification: field of study (QALFP)*. ABS. https://www.abs.gov.au/census/guide-census-data/census-dictionary/2021/variables-topic/education-and-training/non-school-qualification-field-study-qalfp.

Australian Bureau of Statistics (2023). *Microdata and TableBuilder: Census of Population and Housing*. (2006, 2011, 2016, 2021). ABS. https://www.abs.gov.au/statistics/microdata-tablebuilder/available-microdata-tablebuilder/census-population-and-housing (visited 25 Feb. 2025).

Australian Bureau of Statistics (2022-23). *Retirement and Retirement Intentions, Australia*. Table 9.2. ABS. https://www.abs.gov.au/statistics/labour/employment-and-unemployment/r etirement-and-retirement-intentions-australia/latest-release (visited 16 Dec. 2024).

Bloom, DE, D Canning, G Fink & JE Finlay (2007). Does age structure forecast economic growth? *International Journal of Forecasting* **23** (4), 569–585.

Department of Education (2024a). *Award Course Completions*. Higher Education Data Request (2006-2023). Department of Education. https://www.education.gov.au/higher-education-statistics/higher-education-statistics-data (visited 10 Feb. 2025).

Department of Education (2024b). *Award Course Completions for All Students by Age Group and Broad Level of Course*. Table 5 (2006-2016), Table 14.5 (2017-2023). Department of Education. https://www.education.gov.au/higher-education-statistics/student-data (visited 19 Dec. 2024).

Human Mortality Database (2024). *Human Mortality Database*. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). https://www.mortality.org (visited 18 Dec. 2024).

Hyndman, RJ & G Athanasopoulos (2021). *Forecasting: principles and practice*. 3rd ed. Melbourne, Australia: OTexts. https://www.OTexts.org/fpp3.

Hyndman, RJ & H Booth (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International J Forecasting* **24**(3), 323–342.

Hyndman, RJ, H Booth & F Yasmeen (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography* **50**(1), 261–283.

Hyndman, RJ & P Montero-Manso (2021). Principles and algorithms for forecasting groups of time series: Locality and globality. *International J Forecasting* **37**(4), 1632–1653.

Hyndman, RJ & S Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis* **51**(10), 4942–4956.

Hyndman, RJ, Y Zeng & HL Shang (2021). Forecasting the old-age dependency ratio to determine a sustainable pension age. *Australian & New Zealand J Statistics* **63**(2), 241–256.

Macdonald, AS, SJ Richards & ID Currie (2018). *Modelling Mortality with Actuarial Applications*. Cambridge, UK: Cambridge University Press.

OECD (2019a). *OECD Employment Outlook 2019: The future of work*. Paris. https://www.oecd.org/en/publications/oecd-employment-outlook-2019_9ee00155-en.html (visited 16 Apr. 2025).

OECD (2019b). *Working Better with Age*. Paris. https://www.oecd.org/en/publications/working-better-with-age_c4d4f66a-en.html (visited 16 Apr. 2025).

Productivity Commission (2013). *An ageing Australia: preparing for the future*. Canberra. https://www.pc.gov.au/research/completed/ageing-australia/ageing-australia-overview.pdf (visited 16 Apr. 2025).

Smith, L, RJ Hyndman & SN Wood (2004). Spline interpolation for demographic variables: the monotonicity problem. *J Population Research* **21**(1), 95–98.

Stupp, PW (1988). Estimating intercensal age schedules by intracohort interpolation. *Population Index* **54**(2), 209–224.