NSW GOVERNMENT Transport for NSW

Performance and Analytics

2016 Travel Zone Projections (TZP2016 v1.51) for Population, Workforce & Employment in the Sydney Greater Metropolitan Area

Technical Guide

Version 1.51

Contents

Part 1: Model Overview and Results

Author:	Prepared based on a report by SGS Economics & Planning for TPA
Date:	2019
Version:	1.51
Reference:	TPA and SGS Economics & Planning
Division:	Customer Strategy and Technology
Review date:	2019

NSW GOVERNMENT | Transport for NSW

Performance and Analytics

Part 1: Model Overview and Results

1 Glossary and abbreviations

Term	Definition	Abbreviation
Average Annual Growth Rate (%)		AAGR
Australian and New Zealand Standard Industry Classifications	(ABS Cat. 1292.0)	ANZSIC (2006)
Australian Bureau of Statistics		ABS
Australian Standard Geographical Classification	The Australian Standard Geographical Classification was used from 1984 to 2011 by the ABS for the collection and dissemination of geographically classified statistics.	ASGC
Australian Statistical Geography Standard	The Australian Statistical Geography Standard is the ABS' new geographical framework and it is effective from July 2011, replacing the ASGC. The vast majority of ABS spatial data will be based on the ASGS by 2014.	ASGS
Bureau of Transport Statistics	(see Transport Performance and Analytics).	BTS
BTS Land Use Regions	Intermediate geographies which are an aggregation of travel zones and aligned to LGAs. These are a temporary requirement and will be redundant once DPE projections shift from ASGC to ASGS.	BTSLUR
Transport Performance and Analytics	Previously called Bureau of Transport Statistics (BTS).	TPA
Department of Planning and Environment		DPE
Employment	Employed persons by place of work	Emp
Estimated Resident Population	The total number of people that live within a defined area. This includes both people residing in private and non-private dwellings (i.e. college dormitories, jails, nursing homes).	ERP
Freight Movement Model		FMM

Term	Definition	Abbreviation
Greater Metropolitan Area	The Greater Metropolitan Area is the area used for TfNSW's Strategic Travel Model. The GMA includes the Sydney Greater Capital City Statistical Area (GCCSA), the Southern Highlands and Shoalhaven SA4, Illawarra SA4, Newcastle and Lake Macquarie SA4 and Lower Hunter, Port Stephens, Maitland SA3s as defined by the ABS.	GMA
Greater Sydney Commission	An independent organisation funded by the NSW Government to coordinate and align the planning that will shape the future of Greater Sydney.	GSC
Household size	The ratio of *persons in occupied private dwellings* to *occupied private dwellings* (e.g. currently the *household ratio* in Melbourne (C) is 1.91. This means on average there are 1.91 persons in each occupied private dwelling).	
Housing Supply Forecast Model	The 2017 Sydney Housing Supply Forecast Model provides estimates of future housing supply that will be built over the next 5 years. Estimates are prepared by the NSW Department of Planning & Environment to inform infrastructure planning and service delivery, as well as to inform decisions on future land use zoning. TPA has HSFM projections that go out to 2036	HSFM
Household Travel Survey	Is the largest and most comprehensive source of personal travel data for Sydney	HTS
Intergenerational Report	2015 report by the Australian Government which assesses long-term changes including Australia's population size, age profile, economic growth and workforce	IGR
Iterative Proportional Fitting	Statistical method which aligns known totals to an estimated distribution	IPF
Workforce	For any group, persons who are employed or are unemployed but are actively seeking work.	WF
Land Use		LU
Local Government Area		LGA

Term	Definition	Abbreviation
New South Wales		NSW
Not Private Dwelling	Communal accommodation provided by institutions such as hospitals or prisons and transitory accommodation such as hotels and motels.	NPD
Occupied Private Dwellings	A private dwelling that is occupied on Census night. Also represents households.	OPD
Occupancy Rate	The ratio of occupied to unoccupied private dwellings	OR
People in Occupied Private Dwellings	Estimated resident population who reside in private dwellings	POPD
People in Non-Private Dwellings	This includes persons in communal or transitory type accommodation (i.e. prisons, boarding school, hospital, defence establishments).	PNPD
Place of Institution	Refers to variables which are based on education locations (e.g. the number of people attending a tertiary institution within a particular zone)	PoI
Place of Usual Residence	Refers to variables which are based on the home location of population (e.g. the number of people who live in a particular zone)	PUR
Place of Work	Refers to variables which are based on employment locations (e.g. the number of 'Retail Trade' industry jobs within a particular zone)	PoW
Population Synthesiser	Model uses land use data to create 'agents' for input into STM	Pop Syth
SGS Economics and Planning		SGS
Strategic Transport Model	The STM combines travel behaviour with likely population, employment and transport networks to estimate future travel under different strategic land use and transport scenarios.	STM
Structural Private Dwelling	A privately owned building or structure that people live in. This may include a house, an apartment, or it may be a mobile dwelling such as a caravan.	SPD

Term	Definition	Abbreviation
Sydney Greater Metropolitan Area	Defined as the Sydney Statistical Division, Newcastle Statistical Subdivision and Illawarra Statistical Division.	Sydney GMA
Transport for New South Wales	Transport for NSW is the lead agency of the NSW Transport cluster. Its role is to lead the development of a safe, efficient, integrated transport system that keeps people and goods moving, connects communities and shapes the future of our cities, centres and regions.	TfNSW
Travel Zone	Travel Zones (TZs) are the smallest standard geography used for a number of transport datasets in NSW. They represent geographical areas that are used in origin-destination transport modelling. Latest version was created in 2011 and largely aligns with 2011 ABS Destination Zones. See https://opendata.transport.nsw.gov.au/dataset/travel-zones-2011 for additional information	TZ or TZ11
Trip Attractors	Variables relate to destinations. These destinations range from places of work and education to destinations such as shopping centres.	
Trip Generators	Variables relate to the origin location of travel (i.e. place of residence)	
Unemployment Rate	Number of unemployed persons expressed as a percentage of the workforce.	UR

2 Executive Summary

2.1 Background

Transport Performance and Analytics ('TPA') within Transport for NSW ('TfNSW') produces Travel Zone ('TZ') level projections (population, enrolments, workforce and employment) for Greater Sydney ('Sydney GMA' or 'GMA') as an input into the Strategic Travel Model ('STM'). The Travel Zone projections are also used for a range of other strategic and policy work across government and the private sector.

Projections are regularly updated through major and interim updates. Major updates realign to ABS Census data releases and geographies, while interim updates incorporate other updated datasets and also approach improvements.

This update (Travel Zone Projections 2016 Version 1.51 or TZP2016 v1.51) reflects an interim update. This interim update is based on a re-basing of the projections from the 2016 Census using the 2016 release from the Department of Planning and Environment ('DPE') population projections.

SGS Economics and Planning ('SGS') worked in partnership with the TPA Land Use team to refine the approach and update the projections for Travel Zone Projections 2016.

2.2 Report audience and package of outputs

This is a technical report which documents the overall approach, assumptions, data sources and summary results from Travel Zone Projections 2016. It is intended for a reader with a general understanding of the economic and projection techniques and should be read in conjunction with the model spreadsheets and results which provide additional details/results.

2.3 Cautions

It is important to remember the TZP estimates are developed to support a strategic view of Sydney and are calibrated with that city wide view in mind. Therefore, caution is advised when focusing solely on individual zones or variables as this is not the intention of the data. For a strategic transport model what is critical is that all people/workers/students are at least allocated down to the 'general area' to then create travel flows from one location to another. In addition, more data and information is available to support projections to 2036.

Also, when modelling the possible land use, it should be understood there is no one single future. Therefore, these projections seek to represent the most likely urban future based on current data, trends and an understanding of policy/structural changes. It does not reflect a policy aspiration or project specific scenario which would need to be further developed as a separate scenario.

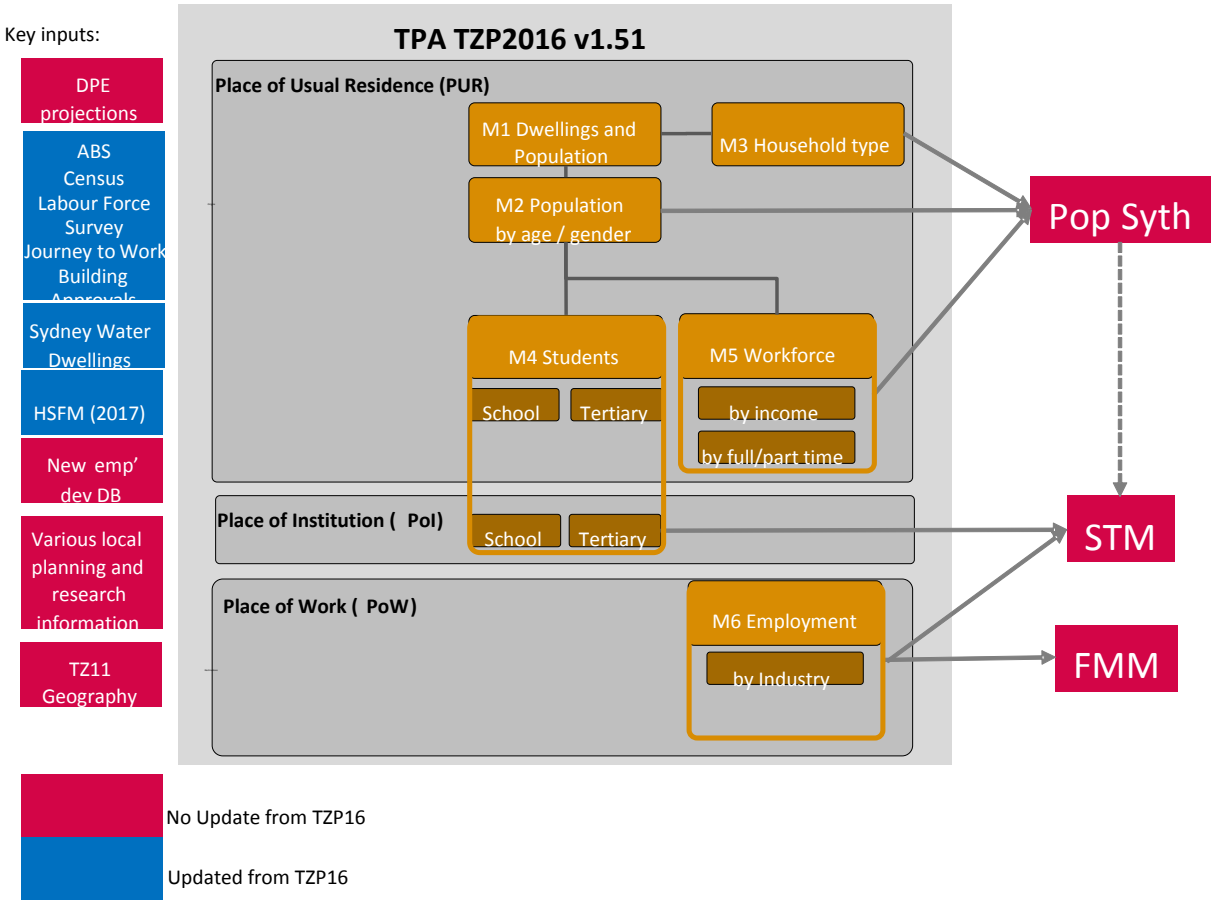
At a high level, the data covers three dimensions:

- **Time Period:** 5 yearly time periods from 2016 to 2056
- **Geography:** 2,949 travel zones across the Sydney GMA
- **Profiling Variables:** 60+ variables covering people, employment and students.

2.4 Travel Zone Projections 2016 Version 1.51 Approach Overview

The following diagram highlights the key data inputs and steps used to create the TZP2016 v1.51. It also highlights the key links to the Population Synthesiser ('Pop Syth') and STM/FMM models.

Figure 1 Travel Zone Projections 2016 v1.51 approach overview[1]



At a high level, official DPE Local Government Area population and household projections are disaggregated to travel zones using a number of small area input sources on current and future development trends. From this, population is further segmented by age/sex, household type, education and workforce status.

Given no official employment projections exist, workforce estimates are then combined with macro-economic trend analysis to project employment for the GMA. Employment by industry is then disaggregated to regions and then travel zones using a number of small area input sources on current and future employment.

The following section steps through the key aspects of the TPA TZP2016 process which are discussed in detail through this report.

2.4.1 Step 1: Structural Private Dwellings (SPD) by Travel Zone

SPD (i.e. occupied and unoccupied dwellings) are the first variable to be estimated. All Place of Usual Residence (PUR) data is then essentially a disaggregation of this travel zone level dwelling distribution. The following components are taken into account:

[1] HSFM refers to the Housing Supply Forecast Model provided by the Department of Planning &Environment

- Control totals. Household data (OPD) (i.e. occupied private dwellings) from DPE is sourced by LGA from 2011 to 2041. A number of small adjustments are then made before they are used as control totals.
- Base distribution (2016). The base SPD dwelling distribution is sourced from ABS 2016 Census data by Statistical Area 1 (SA1). This is then realigned to Travel Zones. ABS Buildings Approvals and Sydney Water Connections Data are then used to bring this base dwelling distribution to 2021. It is then benchmarked to the adjusted LGA control totals to align with DPE projections.
- Projection distribution. The Housing Supply Forecast Model (HSFM) 2017 version forms the primary input into the projected distribution of dwellings. A number of other localised greenfield and planned development information is also sourced. All dwelling development data is treated as 'possible capacity' with a 'preferred timing'. Adjusted LGA control totals are then distributed down to this 'possible capacity'.

2.4.2 Step 2: Population and segmentation

Dwellings are then systematically disaggregated to occupied private dwellings, population, age-sex, household types, student enrolments and workforce status. All population segments refer to persons living in private dwellings (i.e. excludes persons in non-private dwellings, nursing homes, jails, hotels, etc.).

Base 2016 disaggregation ratios (i.e. household size, age-sex splits, etc.) are largely sourced from the ABS 2016 Census. Data is then benchmarked back to control total data. Control total data is largely sourced from DPE by LGA and then adjusted similar to the SPD data.

An Iterative Proportional Fitting (IPF) approach is used to evolve the travel zone distribution over time to align to the Adjusted LGA controls totals, whilst still reflecting the variation at the travel zone level. Seed values for new residential locations are sourced from the respective local region.

2.4.3 Step 3: Workforce control totals (workers at their place of usual residences)

Participation rates by age and sex for the GMA are projected using the Productivity Commission's recommended approach and drawing on the latest ABS Labour Force Survey data. These are then applied to the DPE population by age and sex data to estimate the total workforce. A standard long term unemployment rate is also set at 5.0 per cent.

Workforce is then disaggregated to SA4s and then Travel Zones. It is also split into full/part time employment and five income bands. In this process SA4 and travel zone variances in participation rates, unemployment rates, full/part time and income bands are captured drawing on current trends from the ABS 2016 Census and the projected age-sex composition of the workforce.

2.4.4 Step 4: Employment (workers at their place of work or jobs in a location)

GMA employment by industry is projected using detailed trends analysis of employment by industry from the ABS Labour Force Survey, Census Journey to Work data, projected workforce and analysis of major-economic factors (i.e. regarding structural changes in the broader economy drawing from state and national publications). A number of indicator series are created to understand how the employment by industry projections align with both recent trends and key age segments (i.e. age cohorts).

GMA employment is first disaggregated to SA3s and then travel zones. Again, trend analysis and indicator series are used at the SA3 level to breakdown the employment

by industry projections. This ensures population serving employment (i.e. retail, education, etc.) is shifted to where population growth is projected to occur while the spatial distribution of other industries remains consistent with recent trends. A 'new developments database' forms the primary input into shifting employment projections from a base trend. Similar to the HSFM data for dwellings, this database is used as a 'possible capacity' and 'preferred timing' dataset, while there is a feedback loop to adjust broader trends where the new developments database is signalling a clear shift in policy and base trends.

2.5 Results overview

By 2056 there are projected to be more than 5 million jobs and more than 10 million people in the Sydney GMA (see Figures 2 and 3). The distribution of projected population and employment growth will have a marked impact on how the city functions.

In broad terms we will continue to see the urban footprint of Sydney expand as new growth areas establish to the North West and South West of Sydney. At the same time infill development through both major redevelopments and small scale incremental developments will provide for significant population growth within the existing urban footprint.

The economy will continue to undergo a fundamental restructure away from traditional manufacturing based sectors to one based on services – retail, health, professional and other. This transition has fundamental implications on spatial distribution of new jobs which will congregate around major nodes and population growth areas.

The following tables provide a summary of population and employment growth by district.

Figure 2 Population projections by Greater Sydney Commission (GSC) districts: 1996-2056

GSC District (name as at 2016)	ERP 1996	ERP 2016	ERP 2036	ERP 2056	AAGR 1996-2016	AAGR 2016-2036	AAGR 2036-2056
Central	751,700	1,030,300	1,337,700	1,707,000	1.85%	1.49%	1.38%
North	726,400	887,000	1,081,600	1,383,800	1.11%	1.10%	1.40%
South	620,800	741,200	944,900	1,220,100	0.97%	1.37%	1.46%
South West	525,300	710,200	1,089,400	1,408,700	1.76%	2.67%	1.47%
West	299,600	346,700	442,900	573,000	0.79%	1.39%	1.47%
West Central	664,100	972,900	1,525,300	1,968,400	2.32%	2.84%	1.45%
Total GSC Sydney	**3,587,900**	**4,688,300**	**6,421,800**	**8,261,000**	**1.53%**	**1.85%**	**1.43%**
xGSC Newcastle	762,600	944,500	1,149,800	1,323,400	1.19%	1.09%	0.75%
xGSC Wollongong	372,100	453,600	523,300	579,900	1.10%	0.77%	0.54%
Total GMA	**4,722,600**	**6,086,400**	**8,094,900**	**10,164,300**	**1.44%**	**1.65%**	**1.28%**

Note: ERP is the estimated resident population (no. persons)

AAGR is the Annual Average Growth Rate (%)

GMA is the Greater Metropolitan Area (see Section 4.4 Spatial Geographies for additional details)

Figure 3 Employment projections by GSC districts: 1996-2056

GSC District (name as at 2016)	EMP 1996	EMP 2016	EMP 2036	EMP 2056	AAGR 1996-2016	AAGR 2016-2036	AAGR 2036-2056
Central	630,100	938,800	1,244,800	1,543,500	2.45%	1.63%	1.20%
North	353,500	480,500	595,200	731,600	1.80%	1.19%	1.15%
South	199,800	243,000	303,900	379,600	1.08%	1.25%	1.25%
South West	158,000	249,700	373,900	511,100	2.90%	2.49%	1.83%
West	89,600	131,000	196,100	251,100	2.31%	2.48%	1.40%
West Central	291,800	448,900	675,700	874,200	2.69%	2.53%	1.47%
Total GSC Sydney	**1,722,700**	**2,491,700**	**3,389,600**	**4,291,100**	**2.23%**	**1.80%**	**1.33%**
xGSC Newcastle	275,000	407,400	492,700	591,600	2.41%	1.05%	1.00%
xGSC Wollongong	132,200	180,400	214,200	253,600	1.82%	0.94%	0.92%
Total GMA	**2,129,900**	**3,079,500**	**4,096,500**	**5,136,300**	**2.23%**	**1.65%**	**1.27%**

Note: Emp refers to the jobs available.
AAGR is the Annual Average Growth Rate (%)
GMA is the Greater Metropolitan Area (see Section 4.4 Spatial Geographies for additional details)

3 Introduction

3.1 Background

The TPA produces land use TZ projections (population, workforce and employment) as an input into the STM. The data produced is also available to download from the TfNSW Open Data Hub (https://opendata.transport.nsw.gov.au/) and is used as an input into a diverse range of local planning and research work by a wide range of practitioners.

These TZPs are developed to support a strategic view of Sydney and are calibrated with that city wide view in mind. When modelling the possible land use, it should be understood there is no one single future. Therefore, these projections seek to represent the most likely urban future based on current data, trends and an understanding of policy/structural changes that may impact the future. In addition, more data and information is available to support projections to 2036.

In 2015, SGS was engaged to undertake a review of the land use modelling undertaken by the TPA. This review was documented in the SGS report – Appraisal of BTS Land Use Projections (November 2015). In general, SGS found the approach was reasonable and rigorous based on available data and resources allocated. However, a number of ways to improve the approach were identified. In order to leverage internal resources, ensure consistency and manage risk it was recommended that the TPA land use modelling approach be further reviewed and improved in stages (or modules) with a clear plan developed for an ultimate overhauled structure leveraging best practice, data, approaches and software.

SGS was engaged to work within the TPA to further review and refine the land use Travel Zone Projection ('TZP') approach. This work was completed in stages to first focus on automation of existing processes; and then to include improvements to data, approaches and linkages between modules.

SGS then worked with the TPA team to feed the latest data into the improved model framework to create the final Travel Zone Projections 2016 presented in this report.

This interim update was developed by SGS in conjunction with TPA and completed in January 2019

3.2 Audience and objective of this report

This technical report is intended to provide a technical account of the entire approach, data and method used to create TZP2016. It is expected that the reader has some technical understanding of models and concepts and it is not intended for a general lay audience.

For readers who are familiar with the previous data set known as LU16, there is commentary that explains the differences in methodology and approach in this 2016 release.

It should be noted that further detail regarding the functionality of each module is also contained with the spreadsheets. Detailed datasheets also include the final TZP2016 v1.51 results along with a number of additional summary breakdowns by Districts, LGAs and Precincts.

3.3 Report structure

The remainder of the report is structured as follows:

Section 4	Model framework and data specification
Section 5	TZP2016 v1.51 results overview
Section 6	Module 1: Population and Households
Section 7	Module 2: Age - sex
Section 8	Module 3: Household Type
Section 9	Module 4: Enrolments
Section 10	Module 5: Workforce segmentation
Section 11	Module 6: Employment by Industry

4 Model framework and data specification

This chapter is a comprehensive summary of the data generated by the Travel Zone Projection modelling team and how it links to internal and to external requirements. Understanding the overall linkages between each module of the model is critical to understand the role and function of each.

4.1 Overview

TPA creates Travel Zone Projections of population, workforce and employment using a population based statistical model, which are used as inputs in the Pop Syth, STM and FMM models for transport forecasting.

The population synthesiser is used to create 'agents'. The resulting agents, along with additional TZP inputs, are then used by STM to create travel flows. Some of the TPA small area data is also published online and used for other purposes.

At a high level, the data covers three dimensions:

- **Time Period**: 5 yearly time periods from 2016 to 2056
- **Geography**: TZP2016 v1.51 is available in TZ11 geography with 2,949 travel zones across Sydney Greater Metropolitan Area (GMA)[2].
- **Profiling Variables**: 60 variables covering people, households, employment and students.

These various attributes are currently created using a series of interrelated MS Excel based modules.

The remainder of this section provides additional detail around the TZP framework, time periods, geographies and variables.

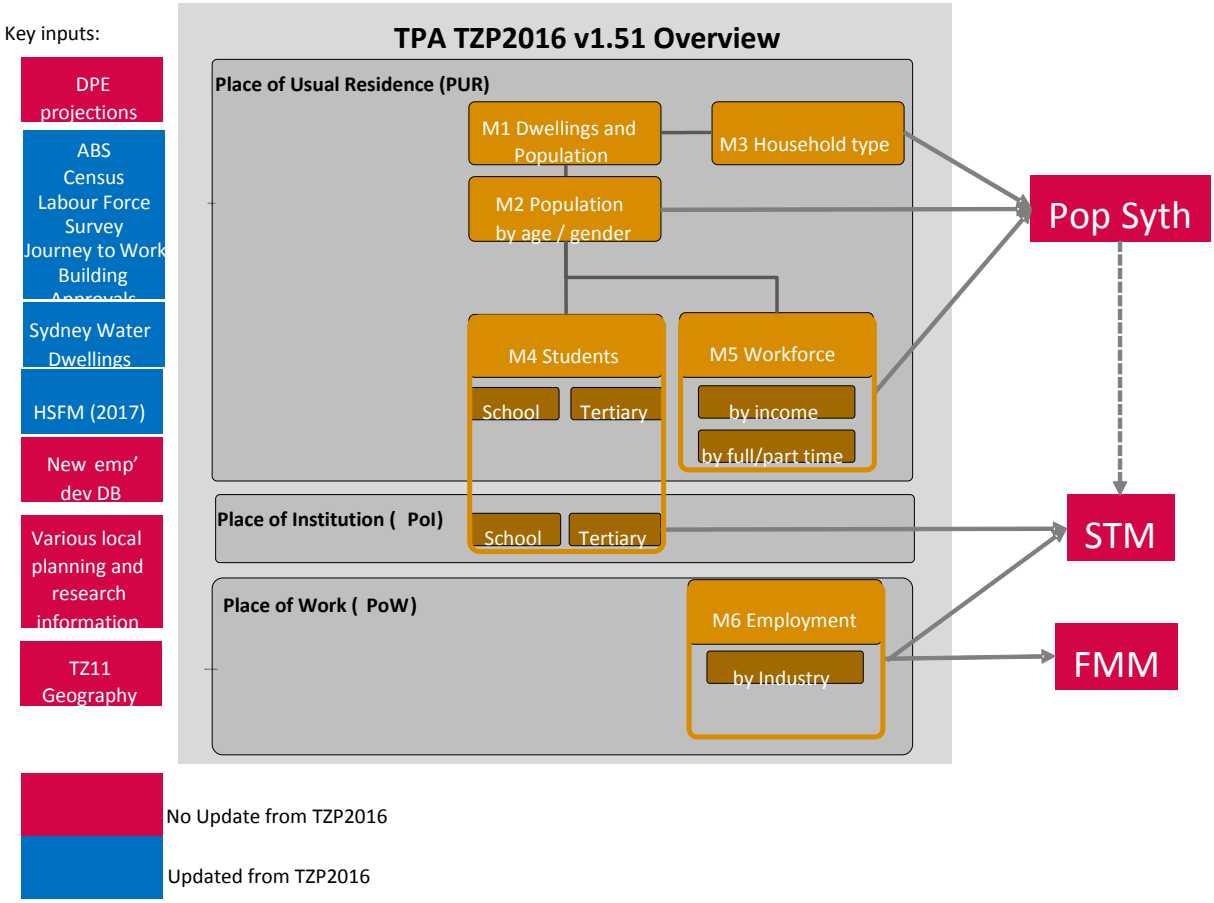
4.2 TZP framework

The following diagram highlights the key data inputs and steps used to create the Travel Zone Projections 2016 Version 1.51 data set. It also highlights the key links to the Pop Syth and STM.

Each module (presented visually as orange boxes) draws on a number of external datasets along with previous module outputs to create a specific set of data which is then used in other modules or for external purposes (presented visually as pink boxes).

[2] Note this definition of the Sydney GMA refers to the STM model area. It does not align with the old ABS SD and SSD definition.

Figure 4 TZP modelling relationships[3]



4.2.1 Modules 1/2/3: Dwellings, population and segmentation

SPD (i.e. occupied and unoccupied dwellings) are the first variable to be estimated. All Place of Usual Residence (PUR) data is then essentially a disaggregation of this travel zone level dwelling distribution.

- **Control totals**. Household data (OPD) (i.e. occupied dwellings) from DPE is sourced by LGA from 2011 to 2041. A number of small adjustments are then made before they are used as control totals.
    - The data is extended to 2056 using a trend analysis.
    - It is converted to 'Adjusted LGAs' a geography created by TPA which largely reflects 2011 LGAs with adjustments to fit Travel Zones and more closely align to Statistical Area 2 (SA2).
    - It is scaled up to an SPD estimate using a standard occupancy rate from the 2016 ABS Census

[3] Note: PUR – Place of Usual Residence. Refers to variables which are based on the home location of population
PoW - Place of Work. Refers to variables which are based on employment locations
PoI - Place of Institution. Refers to variables which are based on education locations

- **Base distribution (2016)**. The base SPD dwelling distribution is sourced from ABS 2016 Census data by Statistical Area 1 (SA1). This is then realigned to Travel Zones. ABS Buildings Approvals and Sydney Water Connections Data are then used to bring this base dwelling distribution to 2021. It is then benchmarked to the adjusted LGA control totals to align with DPE projections.
- **Projection distribution**. The HSFM forms the primary input into the forecast distribution of dwellings. A number of other localised greenfield and planned development information is also sourced. All dwelling development data is treated as 'possible capacity' with a 'preferred timing'. Adjusted LGA control totals are then distributed down to this 'possible capacity'.
    - If the control totals are too high then 'capacity' is brought forward to meet the DPE projection or eventually distributed across the dwelling stock. This effectively means demand is stronger than supply and so developments will come online sooner than planned.
    - Conversely if the control totals are too low, 'capacity' is pushed out to meet the DPE projection and in some circumstances may never be fully realised (in the modelling horizon). This effectively means demand is lower than planned supply, so developments are delayed or occur slower than anticipated.

Dwellings are then systematically disaggregated to occupied private dwellings, population, age-sex, household types, enrolments and workforce status. All population segments refer to persons living in private dwellings (i.e. exclude people in non-private dwellings, nursing homes, jails, etc.).

Base 2016 disaggregation ratios (i.e. household size, age-sex splits, etc.) are largely sourced from the ABS 2016 Census for travel zones. Data is then benchmarked back to control total data by Adjusted LGAs. Control total data is largely sourced from DPE by LGA and then adjusted as per the SPD steps above (i.e. extended to 2056 and converted to 'Adjusted LGAs')

An Iterative Proportional Fitting (IPF) approach is used to evolve the travel zone distribution over time to align to the Adjusted LGA controls totals while still reflecting variation at the travel zone level. The population profile of seed values for new residential locations (i.e. where no base distribution exists) are sourced from the respective local region.

The following primary inputs are used to project population and dwellings:

- DPE LGA forecasts (2011 – 2036)
- Sydney Water Dwellings data (2018)
- HSFM Data (2017 – 2036)
- Various local planning and research information
- ABS Census data (2016)

A process map of modules 1,2 and 3 are included in Appendix C: Travel Zone Projections 2016: Process Maps.

4.2.2 Module 4: Student enrolments

Student Enrolments by place of residences are first estimated then translated into a place of institution.

Population by age is translated into primary, secondary and tertiary enrolments using a similar process as for other population segmentation. A base distribution is sourced from 2016 ABS Census and 2011 ACARA enrolment data. Official enrolment

forecasts are then disaggregated to a SA4 and then Travel Zone level using the IPF approach.

A gravity model is then used to covert enrolments by place of residence to place of institution. High level enrolment caps for schools are then put in place to redirect students to their next closest school.

The following key inputs are used to project enrolments at the Primary, Secondary and Tertiary level:

- Current enrolment from ACARA 2011
- Australian University Enrolments
- Population projections 2011-2056 (M01a)

4.2.3 Module 5: Workforce and Income

Participation rates by age and sex for the GMA are projected consistent with the Productivity Commission's recommended approach and draw on the latest ABS Labour Force Survey data. This is then applied to the DPE population by age and sex data to estimate the total workforce. A standard long term unemployment rate is also set at 5.0 per cent which aligns with the 2015 Intergenerational Report long term unemployment range of 4.0 – 6.0%.

Workforce is then disaggregated to SA4s and then TZs. It is also split into full/part time employment and five income bands ($0 - $20,799, $20,800 - $31,199, $31,200 - $41,599, $41,600 - $67,599, >=$67,600) for use in the TPA Pop Syth. In this process SA4 and TZ variances in participation rates, unemployment rates, full/part time and income bands are captured drawing on current trends from the ABS 2011 Census and projected age-sex.

The following primary inputs are used to project workforce:

- ABS Labour Force Survey data
- ABS Census data
- Intergenerational Report NSW 2015

4.2.4 Module 6: Employment

GMA employment by industry is projected using detailed trends analysis of employment by industry from the ABS Labour Force Survey, Census Journey to Work, projected workforce and analysis of major-economic factors regarding structural changes in the broader economy drawing on state and national publications. A number of indicator series are created to understand how the employment by industry projections align with recent trends, and align with key age segments (i.e. age cohorts).

GMA employment is first disaggregated to SA3s and then travel zones. Again, trend analysis and indicator series are used at the SA3 level to breakdown the employment by industry projections. This ensures population serving employment (i.e. retail, education, etc.) is shifted to where population growth is projected while the spatial distributions of other industries remain consistent with recent trends. A 'new developments database' forms the primary input into shifting employment projections from a base trend. Similar to the HSFM data for dwellings, this database is used as a 'possible capacity' and 'preferred timing' dataset, while there is a feedback loop to adjust broader trends where the new developments database is signalling a clear shift in policy and base trends.

The following primary inputs are used to project employment:

- ABS Labour Force Survey data
- ABS Census data
- New Developments Database, which contains various assumed employment developments across Sydney. These employment assumptions were compiled by TPA and reviewed by DPE (see Appendix A: New development database).

4.3 Time periods

Results are generated for a base year followed by projections which run to 2056. The current base year is 2016. The following future periods are also produced:, 2021, 2026, 2031, 2036, 2041, 2046, 2051, and 2056.

4.4 Spatial geographies

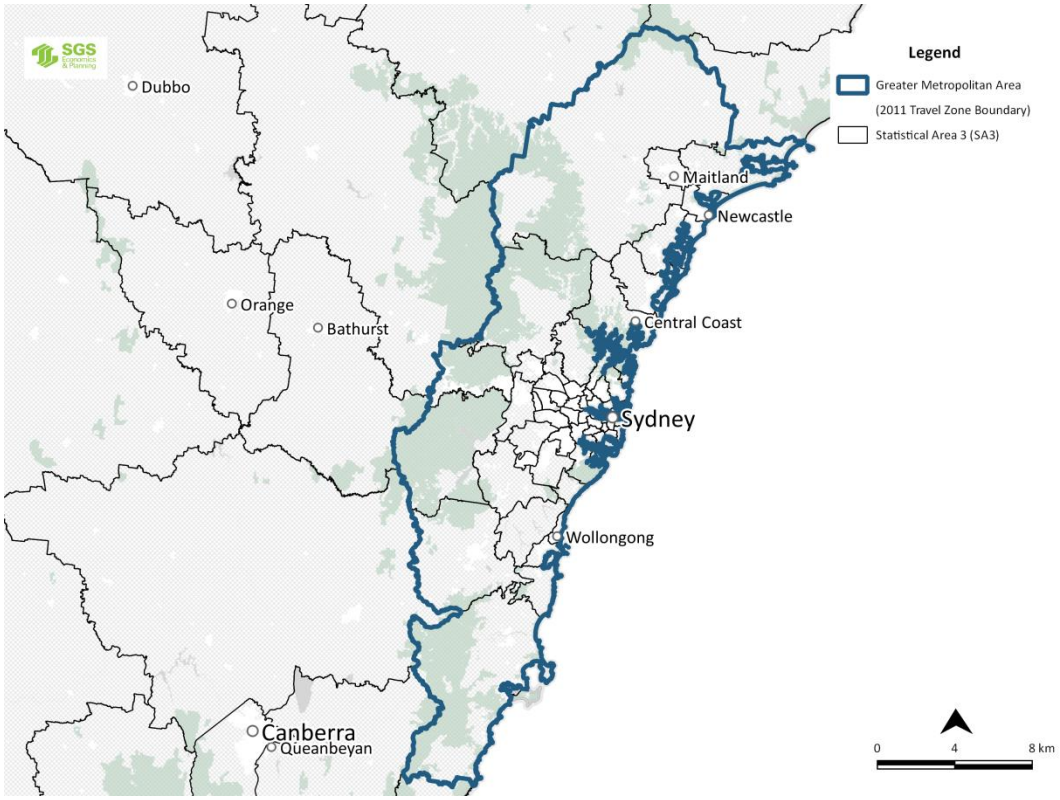
4.4.1 2011 Travel Zones

All land use variables are generated by **TPA 2011 Travel Zone** (TZ11) geography classification. The TZ11 geography aligns with the 2011 ABS Census Destination Zones and the associated ASGS (SA2, SA3, and SA4) structure. There are 3,514 TZ11s across all of NSW. However, projections are created for the 2,949 TZ11s which comprise the 'Sydney GMA' (see

Figure 5 Sydney GMA – showing SA3s. An interactive map is provided on the TfNSW website[4]. The GMA includes the Sydney Greater Capital City Statistical Area (GCCSA), the Southern Highlands and Shoalhaven SA4, Illawarra SA4, Newcastle and Lake Macquarie SA4 and Lower Hunter, Port Stephens, Maitland SA3s as defined by the ABS. Note this Sydney GMA definition does not exactly align with the old ABS SD and SSD definition of the Sydney GMA (see Table 1 Geographies used in TPA land use modelling).

[4] https://www.transport.nsw.gov.au/data-and-research/forecasts-and-projections/travel-zone-explorer

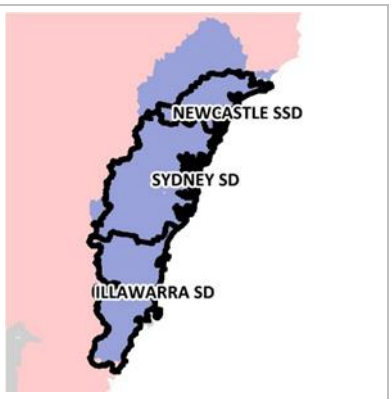
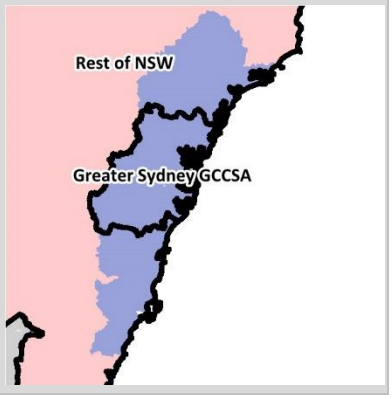
Figure 5 Sydney GMA – showing SA3s

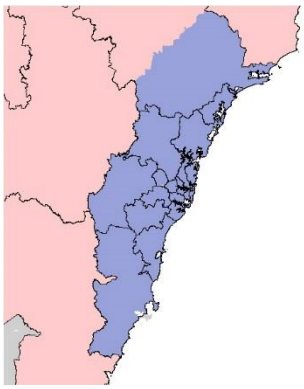
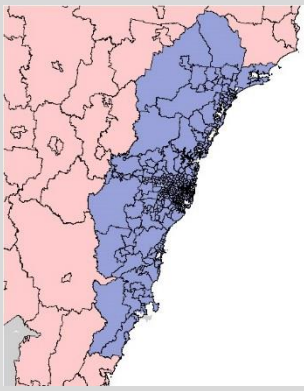
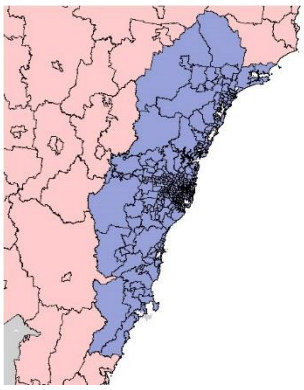
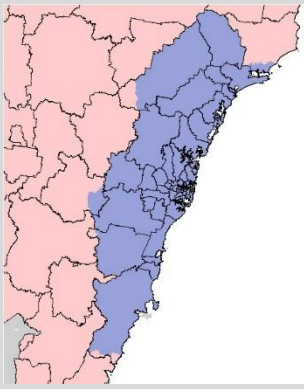


4.4.2 Other geographies

There are a number of other relevant geographies which are used in the development of the TZPs (see table below). As much as possible these are standardised and consistent across modules and with external data sources.

Table 1 Geographies used in TPA land use modelling

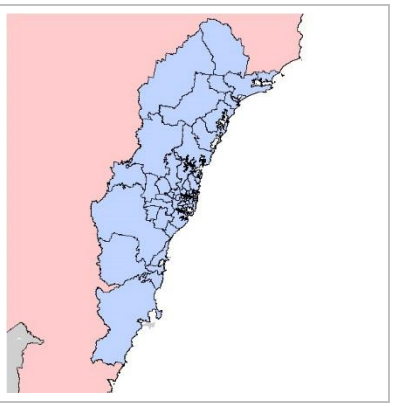
OLD Sydney GMA **Formally defined as Sydney Statistical Division, Newcastle Statistical Subdivision and Illawarra Statistical Division. Statistical (sub) divisions ('SSD') are an old geographic structure created by the ABS and were discontinued during the 2011 Census.** **This old definition of the Sydney GMA should not be confused with the STM model area which is also called the Sydney GMA.**	
Sydney GCCSA and Rest of NSW (RoNSW) Part of ABS ASGS (Cat. 1270.0.55.001) **The GCCSA is the ABS latest definition of 'Greater Sydney'. It excludes Newcastle and Illawarra which are included in 'Rest of NSW'.**	

Statistical Area 4 (SA4) **Part of ABS ASGS (Cat. 1270.0.55.001)** **There are 30 SA4's across NSW with 19 covering the 2,949 modelled travel zones. One SA4 is only partly included in this modelled area.** **They provide a sound sub-regional geographic scale for the TZP work to align with. Increasingly economic and demographic data is provided via this geography (i.e. Labour Force Survey, Census).**	
Statistical Area 2 **Part of ABS ASGS (Cat. 1270.0.55.001)** **This geography is approximately the same size of a suburb and can be useful for reporting and reviewing of results at a local neighbourhood level.**	
Precincts (including strategic centres) **This geography aligns to 'known places' and 'key strategic areas' based around TZs. These align with GSC defined centres while also including additional centres and other areas across the remainder of the model area (GMA).** **These precincts form a key role in the review and reporting process.**	
Local Government Areas (LGA) **These are political boundaries which may not always align with functional land use areas. Some LGAs have recently been redefined in the recent NSW Council amalgamations. As such, many statistical agencies have moved away from using them.** **This geography should only be used for summary reporting, if at all.**	

BTS Land Use Regions (BTSLUR)- or Adjusted LGAs

There are 56 regions which provide an interim geography between SA2 and LGAs.

These are being phased out with focus on shifting to the new ASGS geographic structure.



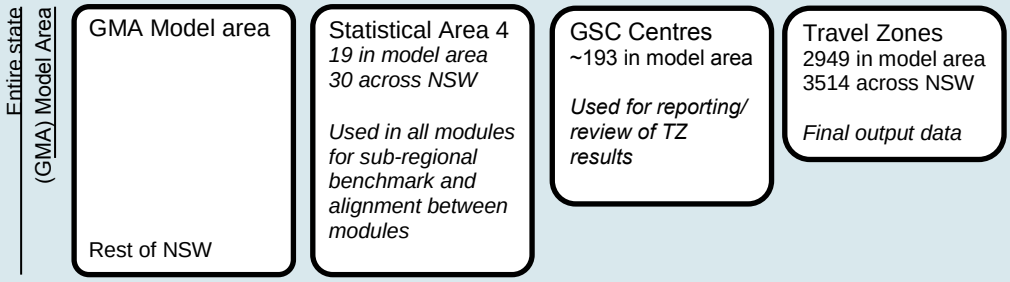
A note on TZ geographic framework going forward

Travel Zone data is the core output geography for the TZP2016 v1.51 and is created using a range of different geographies that largely 'input' into this process. These vary for different modules based on the input data source structure.

A consistent and stepped geographic framework which is used across all modules and used for both inputting data and outputting/reviewing results enables consistency, better cross module links over time and staged development.

The following presents a four stage framework for how modules should incrementally align to key geographic areas in the future. *Currently some modules still align to LGAs or Adjusted LGAs.*

Figure 6 **TPA's TZ Spatial Framework**



This framework provides a number of advantages:

- SA4 results provide clear regional level picture and enable easy linkages across modules (i.e. between workforce and jobs). It also allows for the modules created to be further developed to project the whole of NSW.
- Precincts: An intermediate geography that aligns to 'known places' and 'key strategic areas' that have been created. This geography can continue to evolve over time based on the government policy focus areas. As such these precincts are used for summary reporting and review. Some precincts may include just one or a small group of Travel Zones which aligns with that location.
- Centres are classified as either 'Strategic', 'District' or 'Other' to reflect there are of influence and priority.

4.5 Variable breakdowns

A number of variables are created in each module. These are used for both external requirements and as inputs into subsequent modules. Detailed breakdowns are required to create an accurate picture of trip attraction and generation. However, too many variable breakdowns may undermine the data quality and can make the process overly complicated. Understanding the exact requirements and tailoring module structure around this is critical.

4.5.1 Core Transport Model TZP variable requirements

The core purpose of the TPA TZP data is for input in the Pop Syth, STM and FMM.

The TZP variables required for these transport models fall into two primary categories.

- **Generator variables** that relate to the origin location of travel (i.e. place of usual residence (PUR))
- **Attractor variables** that relate to destinations. These destinations range from places of work (PoW) and educational institutions (PoI) to destinations such as shopping centres.

Table 2 summarises the 25 inputs required by the population synthesiser as well as additional 4 fields which need to be generated but not required.

Table 2 Population Synthesiser data input requirements

Place of count	Category	Sub category	Variable
PUR	POPD	POPD	POPD
PUR	POPD	Age-Sex	Males 0 - 19 Males 20 - 39 Males 40 - 59 Males ≥60 Females 0 - 19 Females 20 - 39 Females 40 - 59 Females ≥60
PUR	POPD (>15)	In Workforce (iWF) Not in Workforce (NiWF)	Full-time workers Part-time workers Unemployed[#] Not in Workforce[#]
PUR	POPD	POPD (<15) POPD(15+) (note includes people employed and not-employed)	Children (≤15 years old) $0 - $20,799 $20,800 - $31,199 $31,200 - $41,599 $41,600 - $67,599 ≥ $67,600
PUR	POPD	Enrolled students Not students	Primary School Student Secondary School Student Tertiary Institution Student Not a Student[#]
PUR	OPD	Household Types	Couples with Children Couples Only Single Parent Single Person Other Types

[#] *Generated variables not required by the Pop Syth.*

Table 3 lists the 19 attractor variables which are produced by the TPA Land Use Forecasting team for use by the STM as well as additional 8 variables that are generated but not required as STM inputs.

Table 3 STM data input requirements

Place of count	Category	Sub category	Variable
PUR	POPD	POPD	POPD
PoW	Total Employment	Total Employment	Total Employment
PoW	Total Employment	Employment income bands (currently 2011 dollars)	$0 - $20,799 $20,800 - $31,199 $31,200 - $41,599 $41,600 - $67,599 >=$67,600
PoW	Total Employment	Service industries	Accommodation and Food Services Financial and Insurance Services Rental, Hiring and Real Estate Services Professional, Scientific and Technical Services Administrative and Support Services Education and Training* Health Care and Social Assistance Arts and Recreation Services Other Services
		Other industries	Retail Trade Agriculture, Forestry and Fishing[#] Mining[#] Electricity, Gas, Water and Waste Services[#] Construction[#] Wholesale Trade[#] Transport, Postal and Warehousing[#] Information Media and Telecommunications[#] Manufacturing[#]
PoI	Enrolments	School enrolments	Primary School Enrolments Secondary School Enrolments
		Tertiary enrolments	Tertiary Institution Enrolments*

*Tertiary Institution Enrolments were proxies by employment in the tertiary education sector, which is a subset of the 'Education and Training' industry.

[#] *Generated variables not required by the STM*

In addition to inputs for the STM, the TPA also provides 34 attraction variables for use in the Freight Movement Model (FMM). The data segmentation required by the FMM is detailed in Table 4.

Table 4 FMM data input requirements

Place of count	Category	Sub category	Variable
PoW	Total Employment	Non-Manufacturing (1 digit ANZSIC)	Agriculture, Forestry and Fishing
			Mining
			Electricity, Gas, Water and Waste Services
			Construction
			Wholesale Trade
			Retail Trade
			Accommodation and Food Services
			Transport, Postal and Warehousing
			Information Media and Telecommunications
			Financial and Insurance Services
			Rental, Hiring and Real Estate Services
			Professional, Scientific and Technical Services
			Administrative and Support Services
			Education and Training
			Health Care and Social Assistance
			Arts and Recreation Services
			Other Services

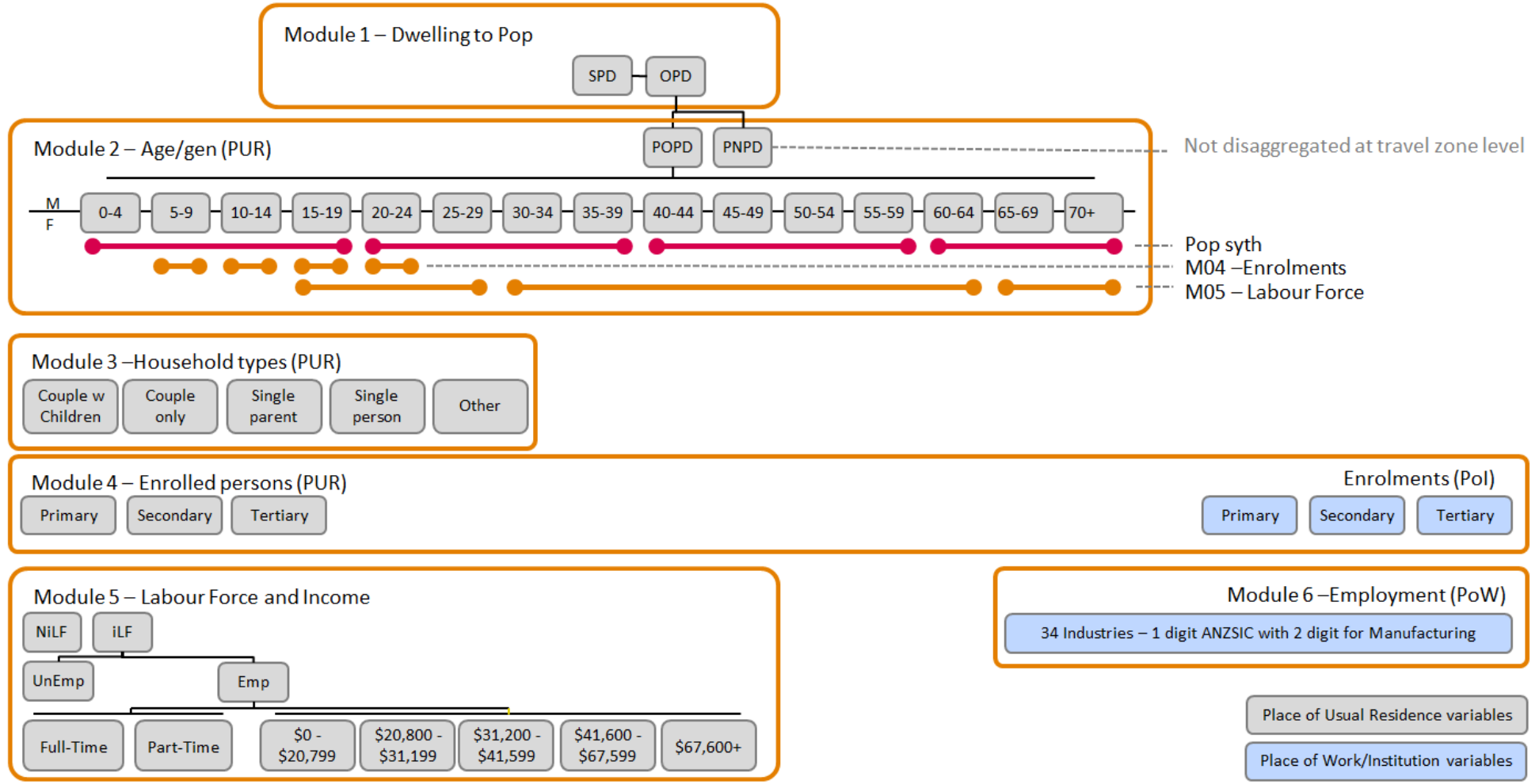
Place of count	Category	Sub category	Variable
PoW	Total Employment	Manufacturing (2 digit ANZSIC)	Food Product Manufacturing
			Beverage and Tobacco Product Manufacturing
			Textile, Leather, Clothing and Footwear Manufacturing
			Wood Product Manufacturing
			Pulp, Paper and Converted Paper Product Manufacturing
			Printing (including the Reproduction of Recorded Media)
			Petroleum and Coal Product Manufacturing
			Basic Chemical and Chemical Product Manufacturing
			Polymer Product and Rubber Product Manufacturing
			Non-Metallic Mineral Product Manufacturing
			Primary Metal and Metal Product Manufacturing
			Fabricated Metal Product Manufacturing
			Transport Equipment Manufacturing
			Machinery and Equipment Manufacturing
			Furniture and Other
			Manufacturing
			Food Product Manufacturing
			Beverage and Tobacco Product Manufacturing

4.5.2 Broader TZP variables requirements (by each module)

Beyond the core transport model requirements, the data is also available to download from the TfNSW Open Data Hub and used as an input into a diverse range of local planning and research work by a wide range of practitioners. As highlighted earlier in this chapter, there are also interrelationships between Modules and associated variable requirements. For example, detailed age/sex breakdowns are required for the development of Workforce variables.

Figure 7 seeks to define the entire scope of variables that the Land Use Forecasting team needs to produce. This covers requirements for Transport Modelling, website publishing and Module interdependencies.

Figure 7 Model variable generation and module dependencies



5 TZP2016 v1.51 results overview

This chapter provides a high level overview of the Travel Zone Projections 2016 Version 1.51 results. For an additional fine grain level view of results, data is presented on a data explorer visualisation with an interactive map, accessible on the TfNSW website.[5]

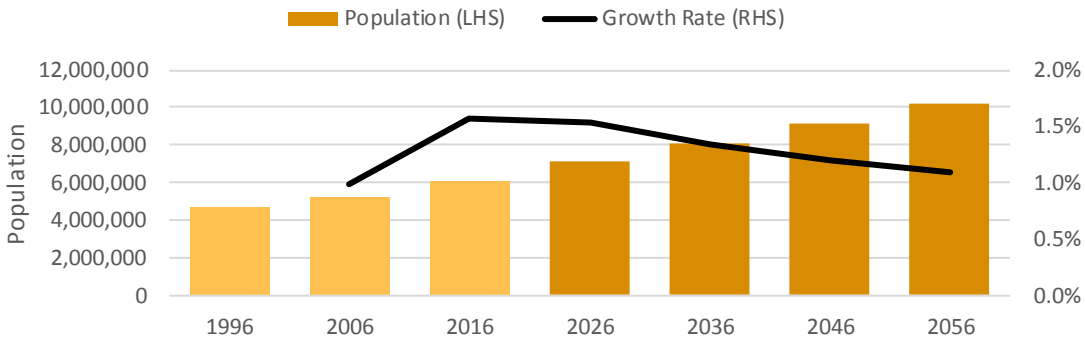
5.1 Greater Sydney Land Use Narrative

By 2056 there are projected to be more than 5 million jobs and more than 10 million people in the Sydney GMA. The distribution of projected population and employment growth will have a marked impact on how the city functions. The size and complexity of the Sydney urban system will be a significant challenge and the future distribution of growth will respond to the provision of future infrastructure, metropolitan strategic planning and evolving market and economic demands and preferences.

In broad terms the urban footprint of Sydney will continue to expand as new growth areas become established to the North West and South West of Sydney. At the same time, infill development through both major redevelopments and small scale incremental developments will provide for significant population growth within the existing urban footprint.

The economy will continue to undergo a fundamental restructure away from traditional manufacturing based sectors to one based on services – retail, health, professional and other. This transition has fundamental implications on the spatial distribution of employment with agglomeration around major nodes and population growth areas.

Figure 8 Greater Sydney employment and population growth



[5] https://www.transport.nsw.gov.au/data-and-research/forecasts-and-projections/travel-zone-explorer

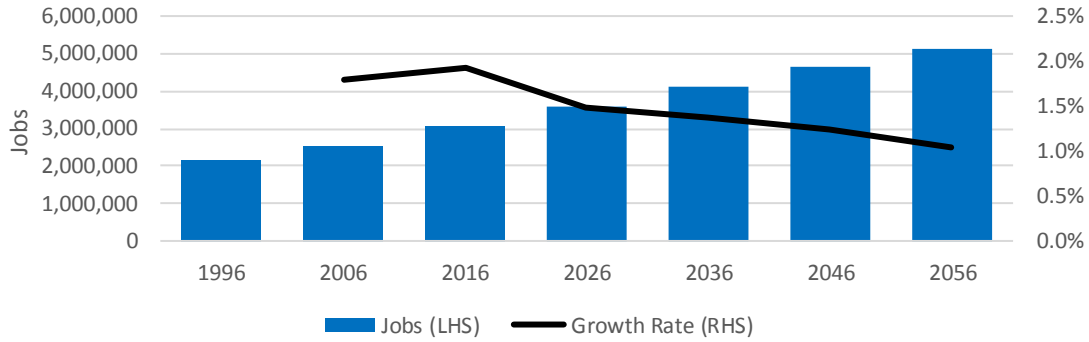
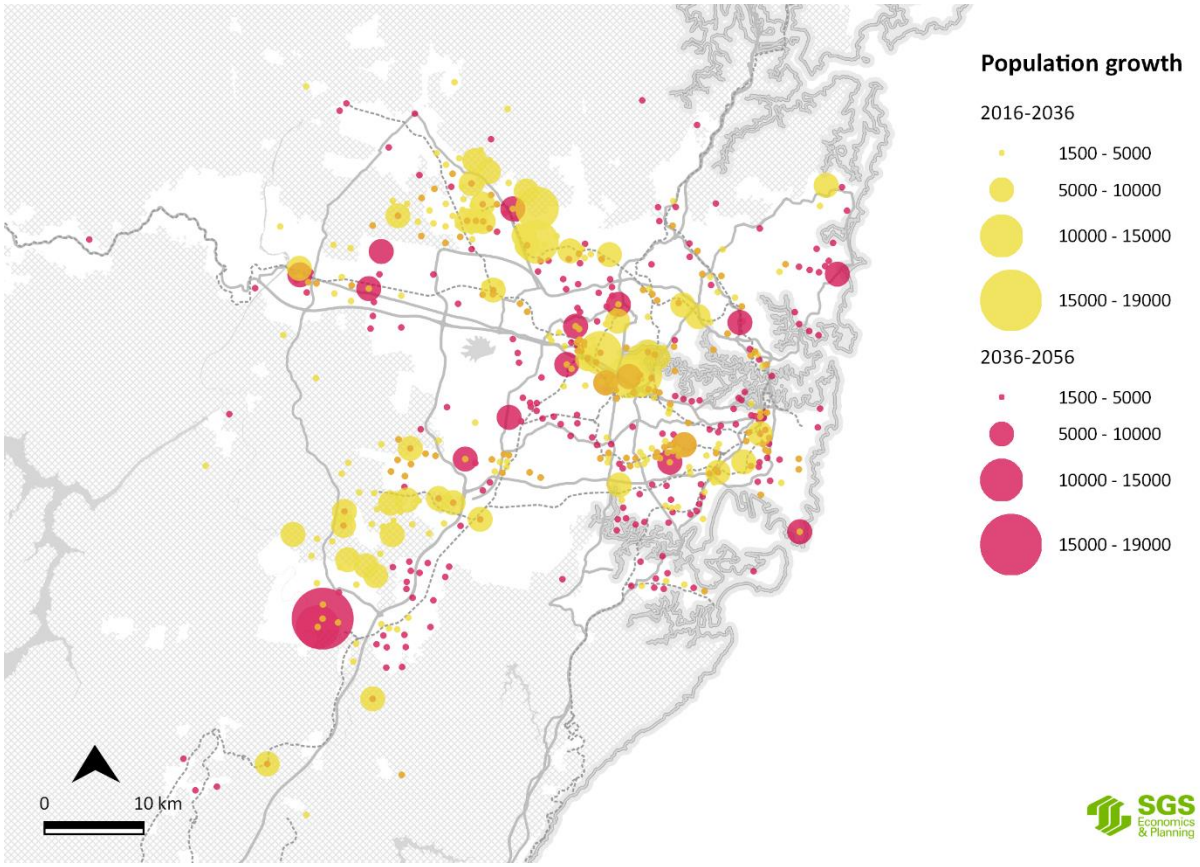
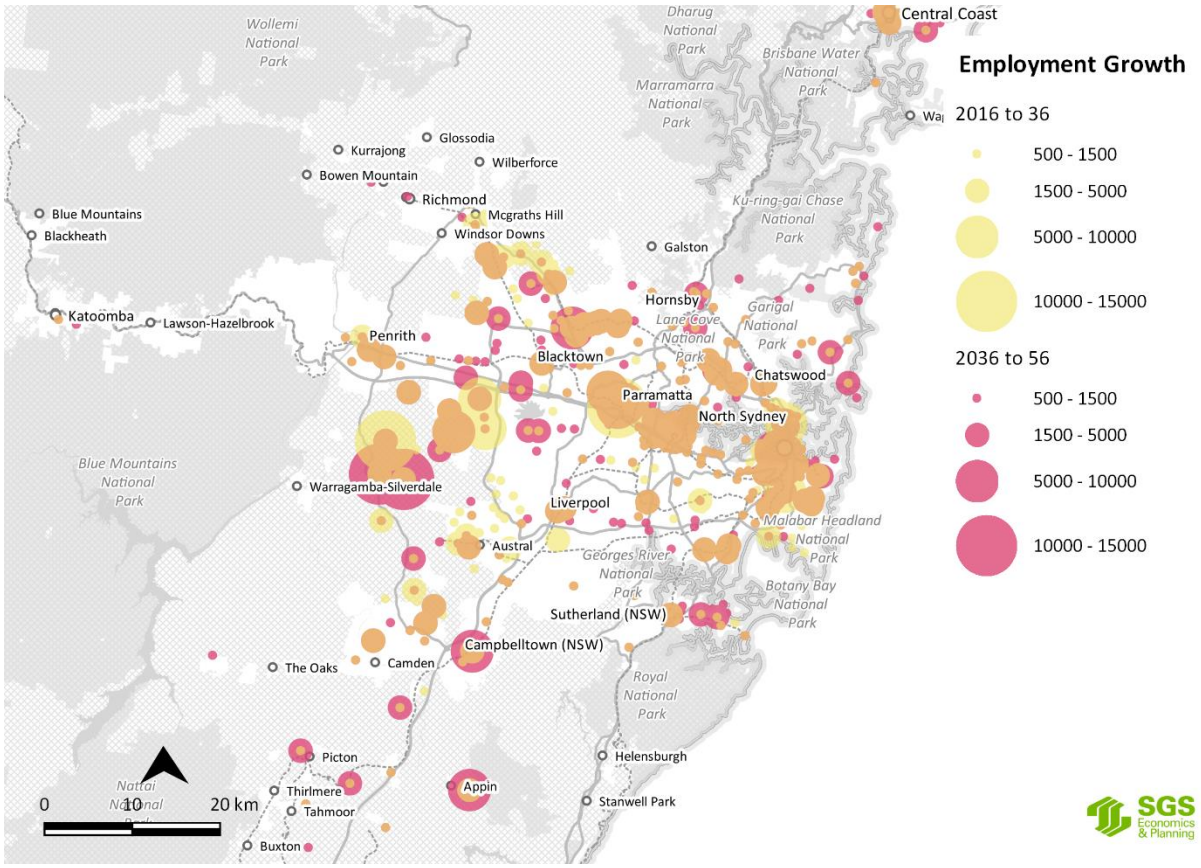


Figure 9 Distribution of Greater Sydney employment and population growth





5.2 GSC District Summary

The GSC Districts represent broad economic sub-regions across Greater Sydney and form the basis of significant sub-regional planning and coordination. Two additional 'Districts' for the broader Newcastle and Wollongong regions have been added to the official GSC Districts to cover the entire Sydney GMA.

5.2.1 Population and Dwellings

Table 5 and Note: *SPD refers to Structured Private Dwellings.*

AAGR is the Annual Average Growth Rate (%)

GMA is the Greater Metropolitan Area (see Section 4.4 Spatial Geographies for additional details)

Table 6 below show the structural private dwelling and population projections by district. The results indicate a rapid growth within the GMA, which will house an additional two million residents by 2036, representing an annual average growth rate of 1.4 percent.

Much of this growth is expected to occur within Greater Sydney, which is projected to reach 6.4 million residents by 2036 and 8.2 million by 2056. When considering the sub-regional distribution, the South West and West Central (which contains Parramatta) districts are expected to record the most significant growth, with annual average growth rates of 2.5 and 2.6 percent respectively between 2016 and 2056. This high growth rate means that, by 2036, the population within the Central district will be eclipsed by that of West Central.

Table 5 Structural private dwelling projection by districts: 1996-2056

GSC District (name as at 2016)	SPD 1996	SPD 2016	SPD 2036	SPD 2056	AAGR 1996-2016	AAGR 2016-2036	AAGR 2036-2056
Central	348,800	443,400	628,000	816,600	1.36%	2.08%	1.50%
North	298,900	355,200	455,100	589,200	0.94%	1.41%	1.47%
South	233,600	269,600	365,600	480,100	0.77%	1.78%	1.57%
South West	177,700	234,000	392,500	519,000	1.58%	3.39%	1.61%
West	113,200	132,500	180,100	235,300	0.85%	1.80%	1.53%
West Central	239,100	329,100	562,200	744,400	1.88%	3.54%	1.62%
Total GSC Sydney	**1,411,200**	**1,763,700**	**2,583,600**	**3,384,700**	**1.25%**	**2.32%**	**1.55%**
xGSC Newcastle	239,100	413,200	537,200	628,200	3.64%	1.50%	0.85%
xGSC Wollongong	166,000	203,400	249,800	281,000	1.13%	1.14%	0.62%
Total GMA	**1,915,600**	**2,380,300**	**3,370,600**	**4,293,800**	**1.21%**	**2.08%**	**1.37%**

Note: SPD refers to Structured Private Dwellings.

AAGR is the Annual Average Growth Rate (%)

GMA is the Greater Metropolitan Area (see Section 4.4 Spatial Geographies for additional details)

Table 6 Population projection by districts: 1996-2056

GSC District (name as at 2016)	ERP 1996	ERP 2016	ERP 2036	ERP 2056	AAGR 1996-2016	AAGR 2016-2036	AAGR 2036-2056
Central	751,700	1,030,300	1,337,700	1,707,000	1.85%	1.49%	1.38%
North	726,400	887,000	1,081,600	1,383,800	1.11%	1.10%	1.40%
South	620,800	741,200	944,900	1,220,100	0.97%	1.37%	1.46%
South West	525,300	710,200	1,089,400	1,408,700	1.76%	2.67%	1.47%
West	299,600	346,700	442,900	573,000	0.79%	1.39%	1.47%
West Central	664,100	972,900	1,525,300	1,968,400	2.32%	2.84%	1.45%
Total GSC Sydney	**3,587,900**	**4,688,300**	**6,421,800**	**8,261,000**	**1.53%**	**1.85%**	**1.43%**
xGSC Newcastle	762,600	944,500	1,149,800	1,323,400	1.19%	1.09%	0.75%
xGSC Wollongong	372,100	453,600	523,300	579,900	1.10%	0.77%	0.54%
Total GMA	**4,722,600**	**6,086,400**	**8,094,900**	**10,164,300**	**1.44%**	**1.65%**	**1.28%**

Note: ERP refers to Estimated Resident Population.
AAGR is the Annual Average Growth Rate (%)
GMA is the Greater Metropolitan Area (see Section 4.4 Spatial Geographies for additional details)

5.2.2 Workforce and Employment

A summary of the workforce projections by districts are shown below in Table 7. Overall, the workforce is expected to increase by nearly 2 million persons within the GMA by 2056. Workforce growth is highly correlated with population growth so is projected to have the fastest rate of workforce growth (2.98%) in the West Central district.

Table 7 Workforce projections by districts: 2016-2056

GSC District (name as at 2016)	Workforce 2016	Workforce 2036	Workforce 2056	AAGR 2016-2036	AAGR 2036-2056
Central	565,500	737,200	949,000	1.52%	1.44%
North	458,800	572,300	753,200	1.24%	1.58%
South	346,400	450,100	593,900	1.50%	1.60%
South West	321,700	493,700	626,300	2.67%	1.34%
West	174,000	218,200	286,000	1.27%	1.55%
West Central	474,700	757,500	956,500	2.98%	1.31%
Total GSC Sydney	**2,341,100**	**3,228,900**	**4,164,900**	**1.90%**	**1.45%**
xGSC Newcastle	444,100	540,800	627,100	1.09%	0.80%
xGSC Wollongong	190,900	221,100	245,800	0.79%	0.56%
Total GMA	**2,976,200**	**3,990,900**	**5,037,800**	**1.70%**	**1.31%**

Note:

AAGR is the Annual Average Growth Rate (%)

GMA is the Greater Metropolitan Area (see Section 4.4 Spatial Geographies for additional details)

The growth of workforce is also driven by increasing participation rates over the next 20 years due to a variety of factors such as the continuing trend of increasing female participation rates (

Figure 10) and people staying in the workforce for longer (Figure 11).

Figure 10 Participation rates: Female 25-29 years

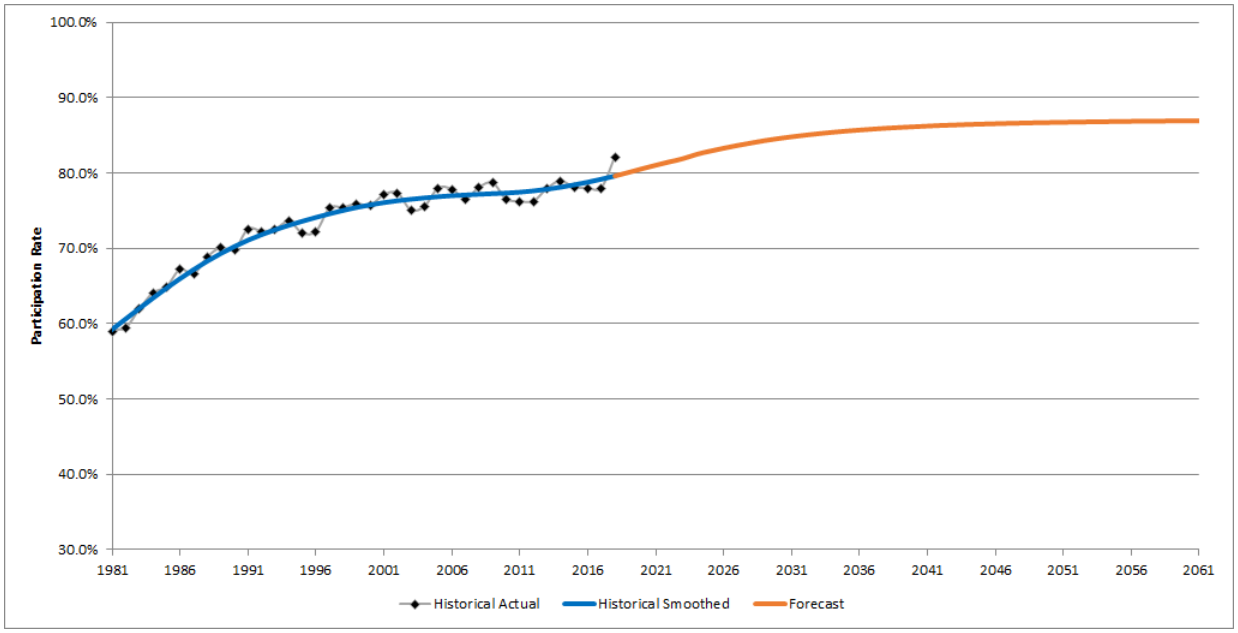
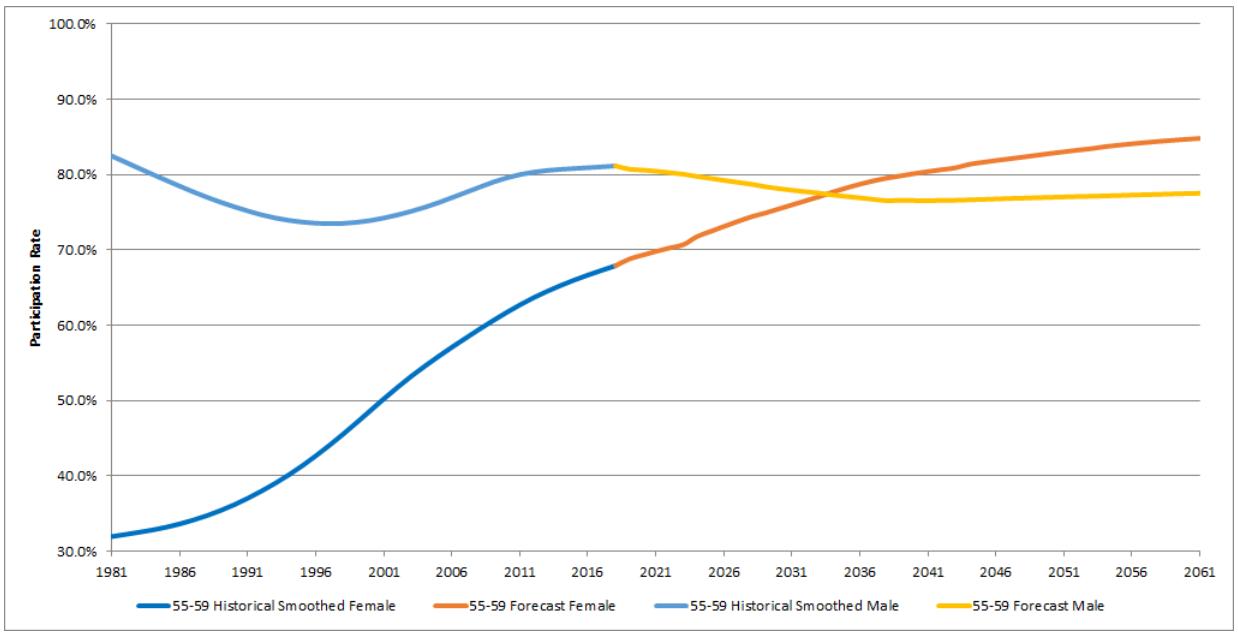


Figure 11 Participation rates: Male and female 55-59 years



Employment projections by district are shown below in Table 8. Employment is expected to continue to grow most strongly within the West Central, South West and West districts.

Table 8 Employment projections by districts: 1996-2056

GSC District (name s at 2016)	EMP 1996	EMP 2016	EMP 2036	EMP 2056	AAGR 1996-2016	AAGR 2016-2036	AAGR 2036-2056
Central	630,100	938,800	1,244,800	1,543,500	2.45%	1.63%	1.20%
North	353,500	480,500	595,200	731,600	1.80%	1.19%	1.15%
South	199,800	243,000	303,900	379,600	1.08%	1.25%	1.25%
South West	158,000	249,700	373,900	511,100	2.90%	2.49%	1.83%
West	89,600	131,000	196,100	251,100	2.31%	2.48%	1.40%
West Central	291,800	448,900	675,700	874,200	2.69%	2.53%	1.47%
Total GSC Sydney	**1,722,700**	**2,491,700**	**3,389,600**	**4,291,100**	**2.23%**	**1.80%**	**1.33%**
xGSC Newcastle	275,000	407,400	492,700	591,600	2.41%	1.05%	1.00%
xGSC Wollongong	132,200	180,400	214,200	253,600	1.82%	0.94%	0.92%
Total GMA	**2,129,900**	**3,079,500**	**4,096,500**	**5,136,300**	**2.23%**	**1.65%**	**1.27%**

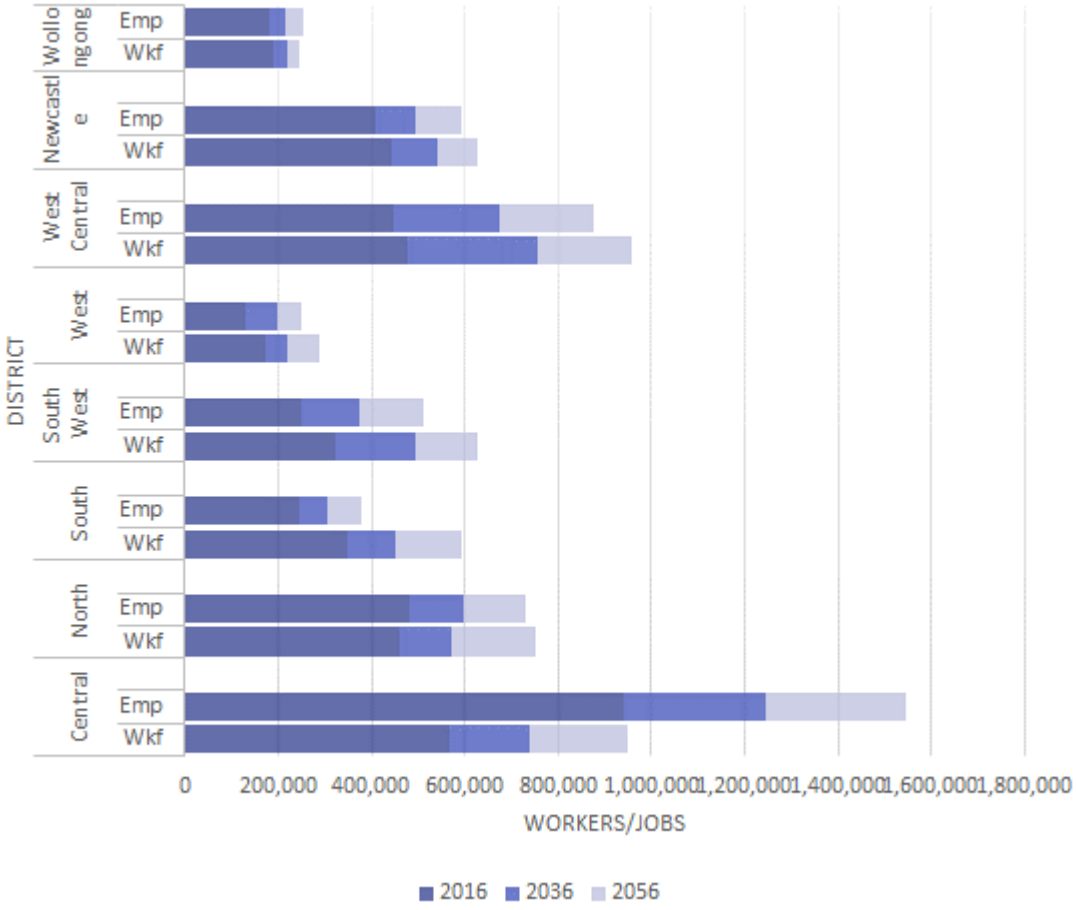
Note: Emp refers to the jobs available.
AAGR is the Annual Average Growth Rate (%)
GMA is the Greater Metropolitan Area (see Section 4.4 Spatial Geographies for additional details)

Comparing district-wide employment to workforce,

Figure 12 shows the persistence of distributional features.

The Central district will continue to have a large net inflow of workers, while districts such as South and South West have significant outflows.

Figure 12 Workforce and employment by District (name as at 2016)



5.3 Precinct (including GSC Centres) summary

193 Precincts were defined across the GMA. These are aggregations of Travel Zones around key centres or suburbs. The following section profiles a small selection of these precincts focusing on those that will contain most of the population and employment growth or that are the focus of policy work. The boundaries of these Centres are shown in the Appendix section of this report.

5.3.1 Central Sydney and Parramatta

Significant development and policy work is occurring around the two major hubs of Greater Sydney - Central Sydney and Parramatta. By 2056 they are projected to have 289,000 and 191,000 residents and 896,000 and 208,000 jobs respectively. These precincts have been further broken in to a number of sub-precincts to better understand the scale of growth within these areas.

Figure 13 presents population levels and gross density (based on travel zone area) within Central Sydney and Parramatta sub-precincts.

Figure 14 presents the same for employment levels and gross density.

Significant dwelling and employment growth is projected in these areas. While the majority of jobs growth will be focused in the Sydney CBD, population growth will be spread across the surrounding precincts. By 2056, Parramatta CBD followed by Broadway will have the highest gross population density. However, Sydney CBD gross employment density will remain significantly higher than any other precinct.

Figure 13 Population by Sub-Precinct, Central Sydney and Parramatta

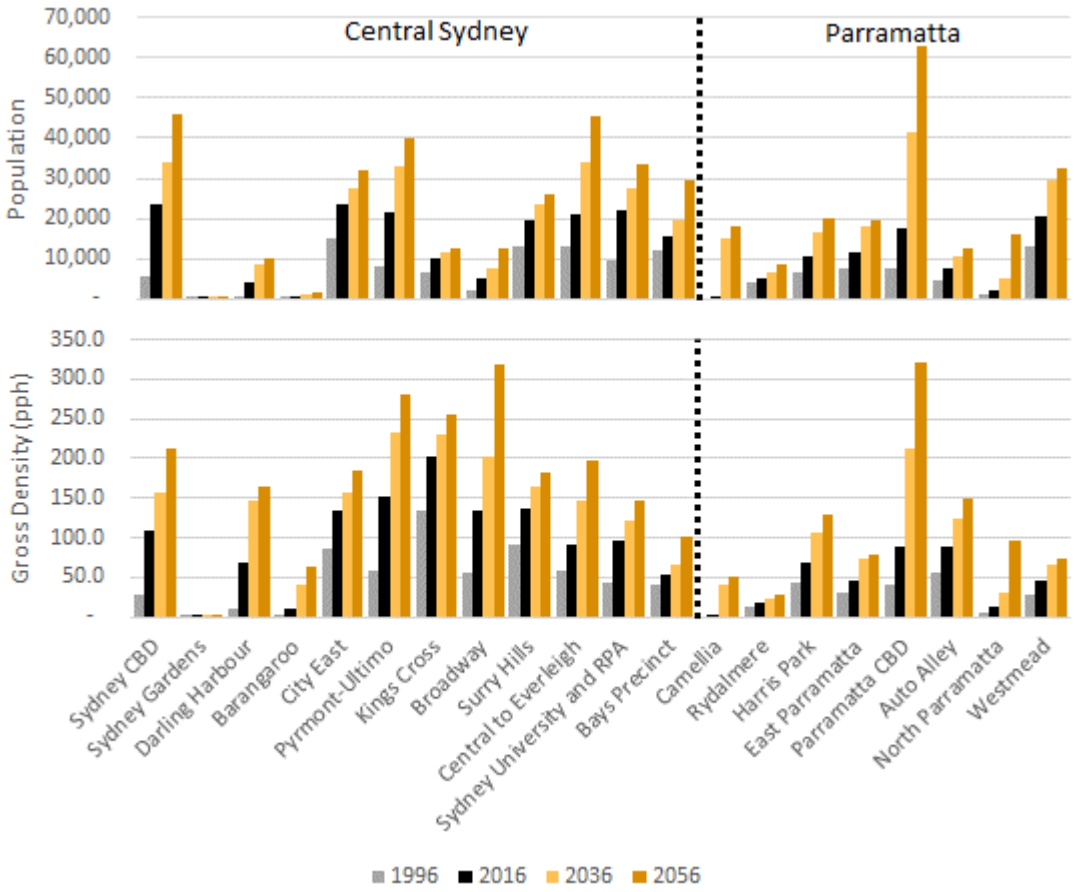
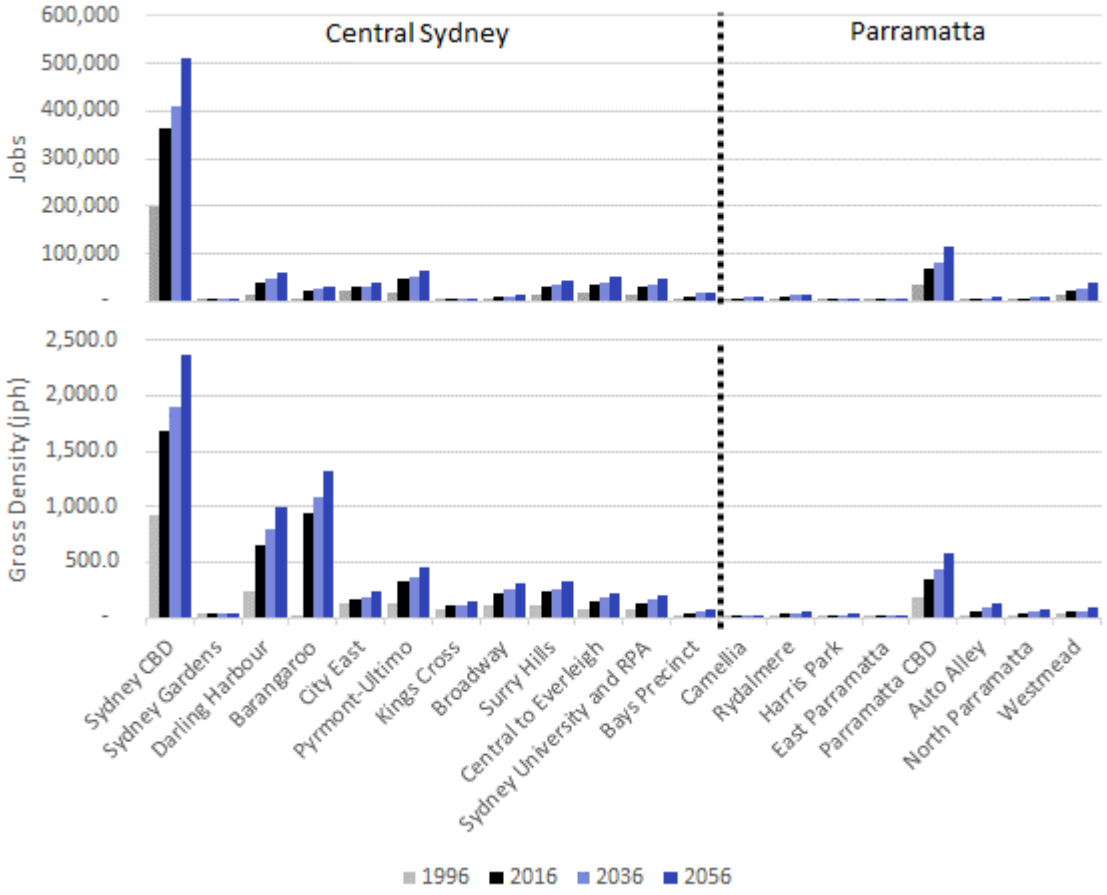


Figure 14 Employment by Sub-Precinct, Central Sydney and Parramatta



5.3.2 Strategic Centres

There are 20 Strategic Centres across Greater Sydney. Population and employment across these strategic centres (excluding Central Sydney) has been presented in

Figure 15 and Figure 16. They have been sorted by distance to Central Sydney.

Figure 15 highlights the significant population and employment growth projected for Greater Parramatta given significant policy and infrastructure focus around this area. It also illustrates that significant population growth is projected for Green Square and Penrith, along with a number of other key centres. In terms of employment, Western Sydney Airport will see significant growth as will Macquarie Park.

Figure 15 Population by Precinct, strategic centres (ex Central Sydney)

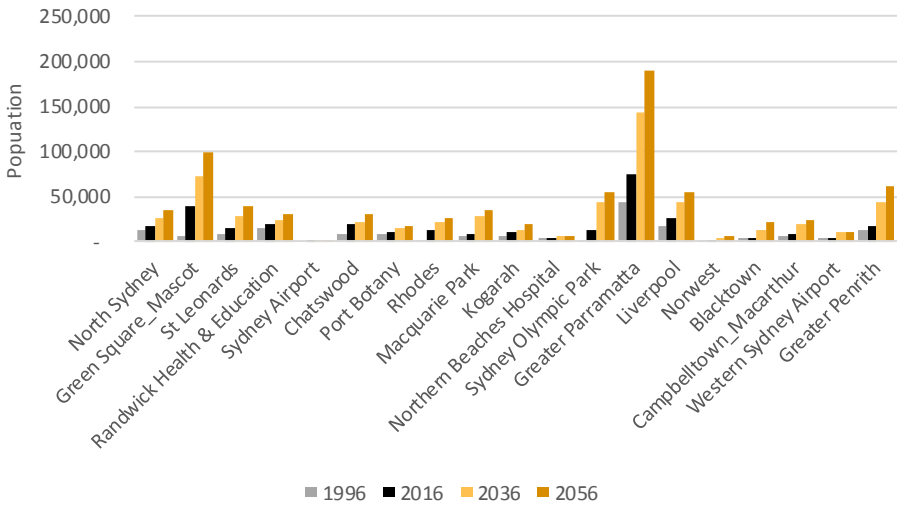
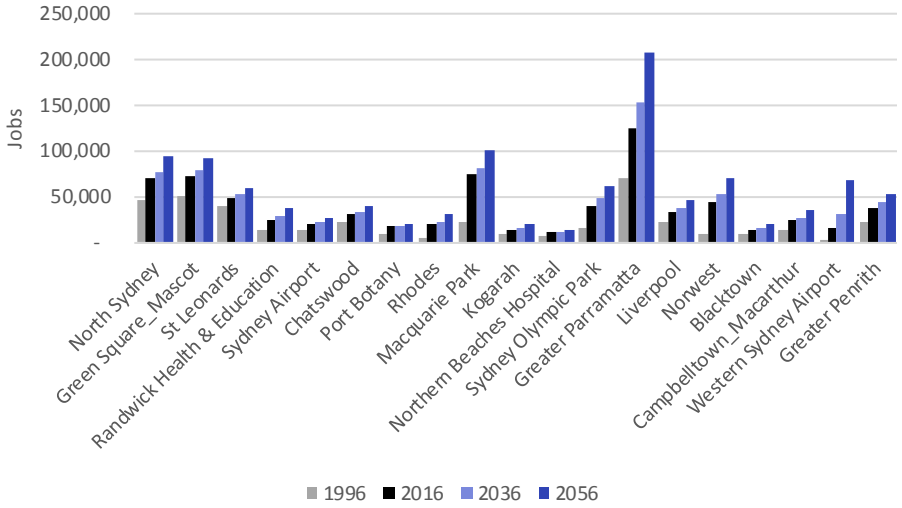


Figure 16 Employment by precinct, strategic centres (ex Central Sydney)



5.3.3 Summary of selected precincts

Figure 17 presents the 'top 15' precincts in terms of population and employment at 2016 and projected growth from 2016 to 2056. Greenfield release areas within the South West Growth Area are the drivers behind the high projections for population, in addition to urban renewal areas such as Parramatta and strategic sites in Central Sydney.

In terms of employment, the top three employment growth precincts are Central Sydney, Parramatta and Western Sydney Airport.

Figure 17 Population, selected top15 Precincts

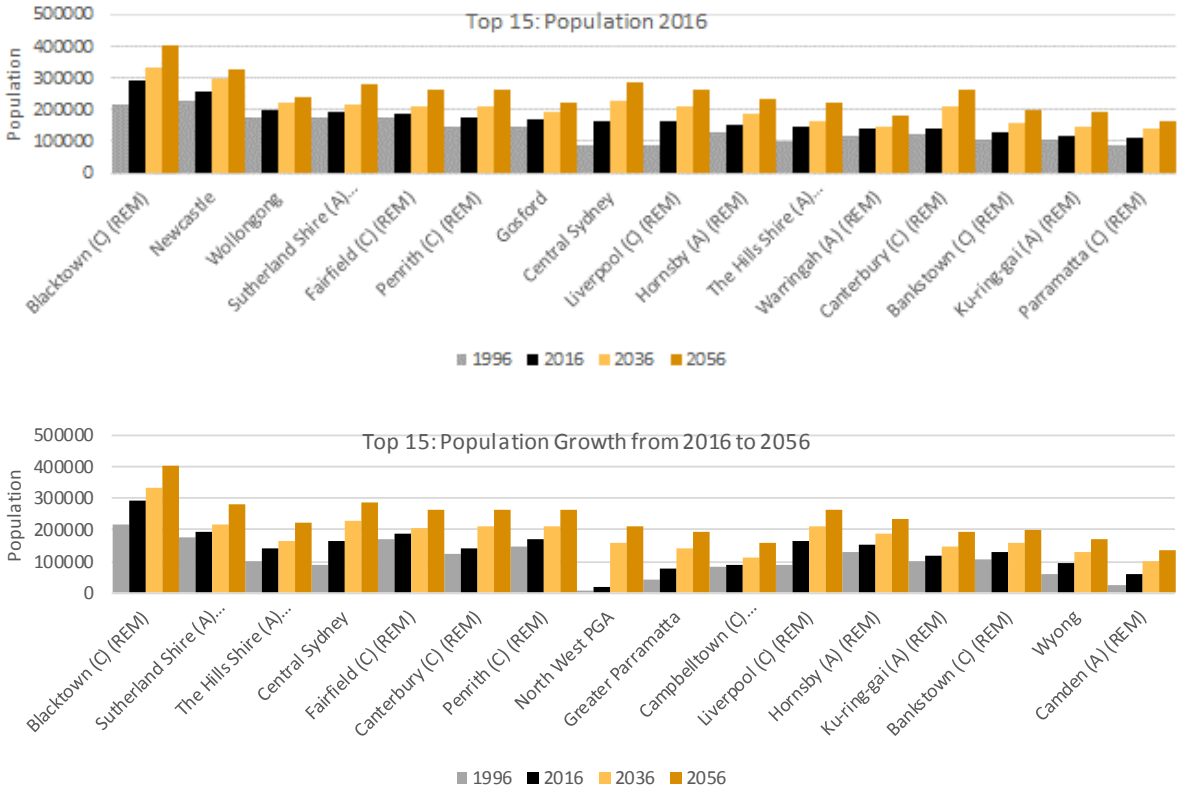
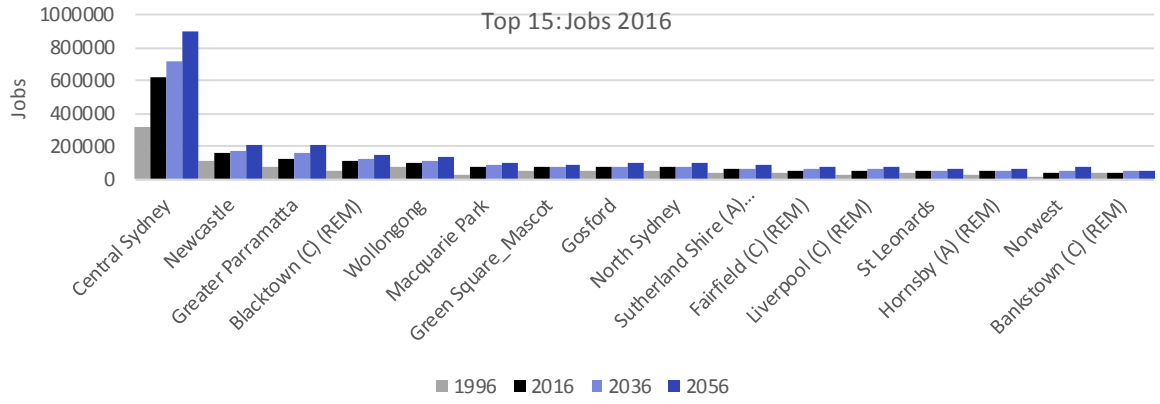
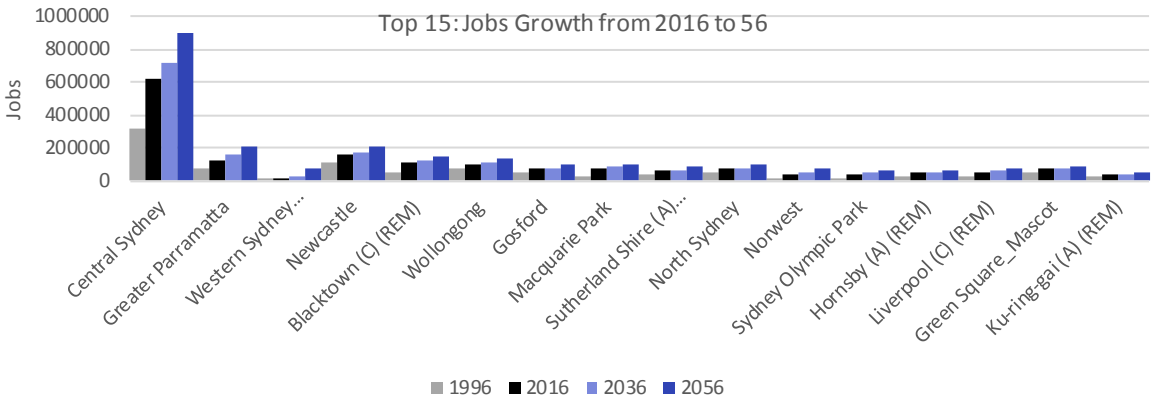


Figure 18 Employment, selected top15 Precincts





Additional TZP2016 v1.51 breakdowns and raw data.

The above section provides a snapshot of the TZPv1.51 2016 dataset to highlight a few of the key trends likely to occur within the GMA over the coming decades. Along with this report several spreadsheets containing the raw data, along with a number of summary tables by Districts, Local Government Areas and all precincts, have been created. This includes additional breakdowns such as dwelling, population, population by age, workforce status, employment by industry and enrolments by subsector. These spreadsheets also include summary growth levels, growth rates and density estimates.[6]

The remainder of this report further details the key assumptions, data sources and approach used in each module.

6

NSW GOVERNMENT | Transport for NSW

Performance and Analytics

Part 2: Detailed Model Methodology

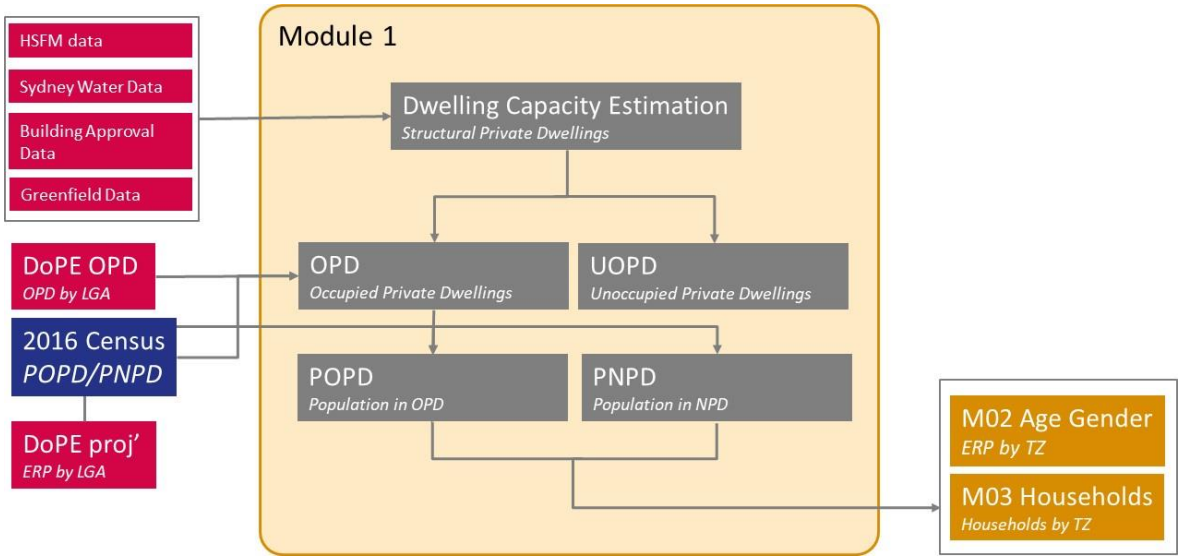
6 Module 1: Population and dwellings

This chapter provides a summary of the revised dwellings and population module of TZP2016 v1.51. This approach differs from the TZP2016 projection release in the treatment of private and non-private dwellings.

6.1 Module 1 overview

The diagram below presents an overview of the module and the relationship to the other modules.

Figure 19 Module 1 overview



This module is made up of four components:

- **Dwelling Capacity Estimation** (Structural Private Dwellings) – Incorporating HSFM forecasts for the Metropolitan area, Sydney Water data, Building Approvals and Greenfield data, and adjusted Land Use 2014 data to determine dwelling capacity for each travel zone.
- **Dwelling Projections (Occupied and Unoccupied Private Dwellings)** – Incorporating the capacity estimation, Occupied Private dwellings and Unoccupied Private Dwellings by travel zone are projected.
- **Population in Occupied Private Dwellings** – Combining the Private Dwelling projection and average household sizes by travel zone, the DPE population projections for persons in OPD are projected at travel zone level.
- **Population in Non Private Dwellings** – Combining the non-private dwellings projected and historical household sizes, DPE population projections for persons in NPD are projected at travel zone level.

6.2 Adjustments to Methodology in TZP2016 v1.51

TZP2016 v1.51 provides an update to the previous TZP2016.

The main difference is that the base year of the model has moved from 2011 to 2016 using 2016 ABS census data and a number of other updated inputs.

In TZP2016, TPA were provided LGA projections for Population in Occupied Private Dwellings (POPD) and Population in Non-Private Dwellings (PNPD).

In TZP2016 v1.51, 2016 Census data was used to re-base these numbers for the year 2016 to calculate LGA totals for POPD and PNPD. The 2016 split of POPD/PNPD was then applied to all subsequent forecast periods through to 2056.

Occupancy rates and household sizes were applied as per ABS Census 2016[7]

Further details on this methodology update is available in Appendix F.

6.3 Module 1 technical model overview.

The following sections provide a brief overview of how the model spreadsheets operate.

6.3.1 M01a – Capacity Estimation

This module determines dwelling capacity at a travel zone level drawing on a range of data sources. This first module has been built to incorporate the most up to date and available datasets available at the time. In future iterations of the model, this module will need to be restructured to incorporate updated and new data sources to inform future dwelling capacity.

This module requires inputs from the following sources:

Inputs

- HSFM forecasts for the Metropolitan area
- Sydney Water dwelling stock data aggregated at the Mesh block
- ABS Buildings Approval Data
- Previously sourced data by TPA (2014) – Adjusted Land Use 2014
    - Illawarra Greenfield and Target Data
    - Lower Hunter Target Data

Functions

- HSFM forecasts, Sydney Water and Target Data inform the additional priority capacity available in each travel zone.
- The secondary additional capacity is driven by the Adjusted Land Use 2014 remaining capacity identified.

6.3.2 M01b – Dwelling Projections

This module distributes the DPE Occupied Private Dwellings projections by LGA to travel zone based on the capacity estimates from M01a.

---

[7] 2016 Census - Counting Persons, Place of Enumeration (MB) SA1 by DWTD Dwelling Type; Counting: Persons Location on Census Night and 2016 Census - Counting Dwellings, Place of Enumeration (MB) SA1 by DWTD Dwelling Type; Counting: Dwellings Location on Census Night

Inputs

Key inputs:

- Occupied Private Dwellings (DPE Adjusted LGA Control Totals)
- Available Residential Land by travel zone
- Capacity Inputs from M01a
- Capacity Calibration

Functions

This module has two primary functions:

- The spreadsheet ‘Control Totals’ converts Occupied Private Dwellings to Structural Private Dwellings using a regional occupancy rate.
- The spreadsheet ‘Capacity Calibration’ allows the user to input manual calibrations to override the capacity inputs from M01a. For zones which require a manual override, enter a value of 1 in the "Capacity Override" column. Next, input the additional yearly capacity and define the fraction which should be allocated as priority capacity. Finally, entering a value of 1 in the "Lock to priority" columns will peg dwelling growth for that year to priority capacity calibration
- The spreadsheet ‘Calculations’ determines the final dwelling stock in each travel zone based on dwelling capacity and DPE control totals by LGA. For years 2016 to 2056, the growth of Structural Private Dwellings is distributed based firstly on the priority capacity. In the case there is a residual of dwelling growth (i.e. there is not enough capacity); the secondary capacity is used to allocate the remainder of the growth. If the both priority and secondary priority capacities are exhausted and a residual remains – future dwelling capacity is brought forward to meet the dwelling growth.

6.3.3 M01c – Population by Travel Zone

This module distributes the DPE population forecasts for Occupied Private Dwellings and Non-Private Dwellings.

Inputs

- Estimated Population in Occupied Private Dwelling and Population within Non-Private Dwellings
- Occupied Private Dwellings projected from M01b
- SA1 distribution of Population in Occupied Private and Non-Private Dwellings and the number of Occupied Private dwellings and Non-Private dwellings.

Functions

This module has several key functions:

- The spreadsheet ‘POPD Household Size’ calculates household sizes based on SA1 Census data. These household sizes are assumed to remain the same throughout the projection. The user can override the Census data using the ‘Override Indicator’. The user can change the indicator to ‘1’ and add in the household size for those travel zones.
- The spreadsheet ‘Final POPD Projection’ distributes the DPE population projections in two steps:
- First Cut: The growth in dwellings is multiplied by the assumed household size for each travel zone

- Final Cut: Population growth (DPE) by LGA is distributed based on the first cut projection.
- The spreadsheet ‘PNPD Distribution’ allows the user to input additional NPD Developments that are confirmed to be developed. The household size is required as it will override the current Census distribution.
- The spreadsheet ‘Final PNPD Projection’ distributes the LGA projected for PNPD (DPE forecast) by NPD household size for each travel zone as determined by Census distribution and any additional new development input.

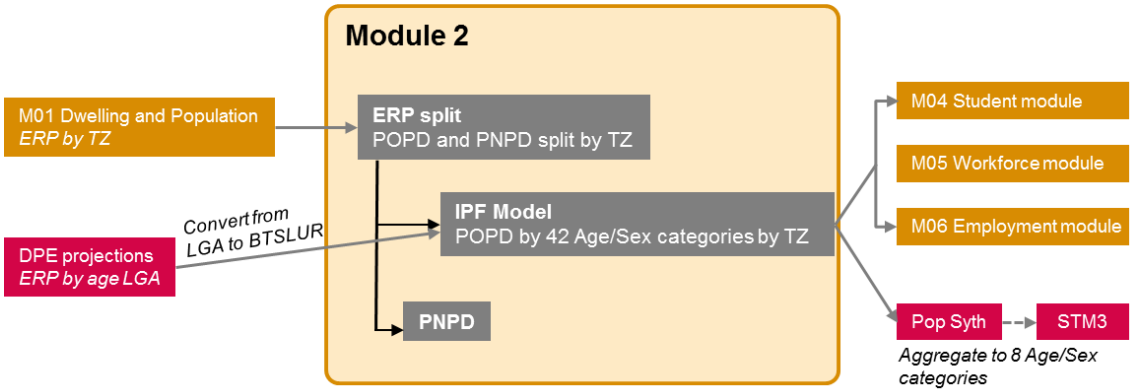
7 Module 2: Age-Sex

This chapter provides a summary of the revised age-sex module of TZP2016 v1.51. This approach is the same as used for the TZP2016 projection release.

7.1 Module 2 overview

The following diagram presents an overview of the population by age/sex module and how it interacts with the broader TZP framework. Understanding where inputs are sourced and what outputs are required are critical when operating the module.

Figure 20 Module 2: Age sex module approach overview



Estimated resident population (ERP) is comprised of two segments:

- People in occupied private dwellings (POPD) – Estimated resident population who reside in private dwellings. This represents approximately 98 per cent of the population.
- People in non-private dwellings (PNPD) – people who live in colleges, aged-care facilities, jails and other non-conventional dwelling forms. This segment of the population is not used for trip generation in the STM and is therefore not further disaggregated.

POPD is then disaggregated into a number of age-sex categories for each time period.

  - 42 age by sex categories are created for use in subsequent modules.
  - Results are aggregated to 8 age by sex categories for the population synthesiser and STM.

While the objective of the previous module was an accurate estimate of population levels in each travel zone, the age-sex module adopts these values and primarily focusses on their distribution across age-sex categories. As such, this process is almost entirely automated.

Iterative Proportional Fitting (IPF) is the approach used to disaggregate travel zone population control totals (POPD) into age-sex categories while maintaining an acknowledged age-sex distribution at the regional level.

Iterative proportional fitting (IFP) concept overview

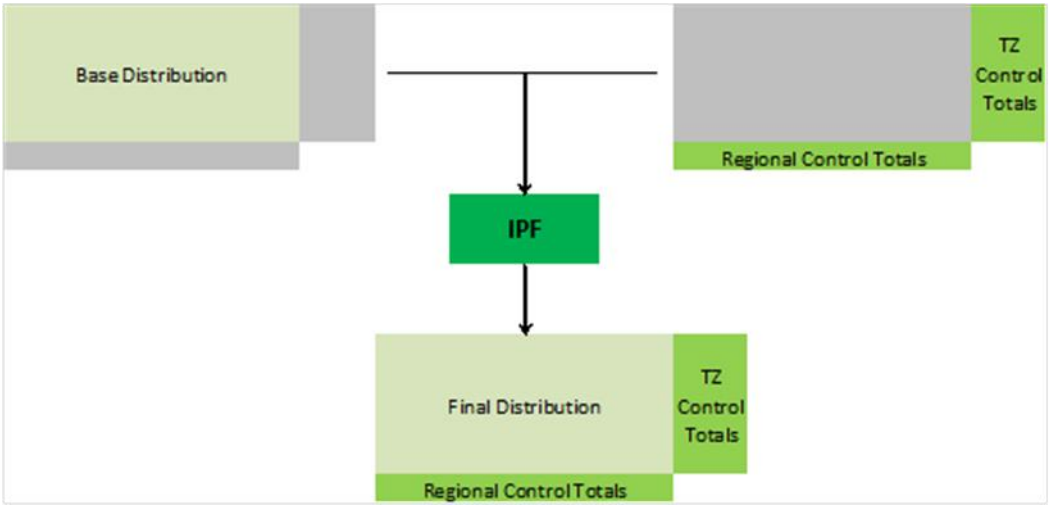
IPF is a statistical method which aligns known totals to an estimated distribution. In the age-sex module, the total population (POPD) of each travel zone is known, having been estimated previously. Similarly, for each control region, the total for each age-sex category is known from DPE population by age-sex projections.

Using these totals, along with an approximated distribution, an iterative process is used to estimate the population in each age-sex category by travel zone. This

process is run for every region in each period. For the base period (currently 2016), the IPF starting distribution is approximated using 2016 census data. Future periods align the final distribution of the preceding period as the starting distribution for the IPF procedure.

The IPF method is outlined below. The approach essentially involves a number of iterations where the distribution is aligned to row totals, then column totals and so on until a convergence criterion is achieved. This results in a final distribution which aligns with the base distribution while matching the row and column control totals.

Figure 21 IPF Procedure



7.2 Module 2 Technical model overview

The following sections provide a brief overview of how the model spreadsheets operate.

Further documentation is included with each spreadsheet.

7.2.1 M02a – Data setup

Brings in relevant data from various sources and structures them for the IPF process.

Inputs

Key Inputs:

- Age by Sex census data by SA1 (obtained via ABS TableBuilder)
- SA1 to TZ11 concordance (obtained from GIS team)

Functions

This module performs two key functions:

- Convert SA1 level census data to TZ11
- Zones with a low total population are likely heavily influenced by ABS randomisation – these zones are adjusted to reflect the age-sex structure of the broader region.

7.2.2 M02b – Iterative Proportional Fitting

Performs IPF process to estimate age and sex breakdowns by Travel Zone over time

Key assumptions

- For future years, the IPF procedure uses the age-sex distribution of the previous period as a starting point.
    - This approach reflects that at small geographic levels, age distribution over time is more heavily influenced by migration than by cohort ageing (e.g. families locating in zones close to schools or young-adults migrating based on access to employment and housing forms)
    - Regional trends in age-sex structure are captured at a travel zone level by the IPF procedure.
- Future growth zones (which have no historical age-sex distribution) are assumed to have a starting distribution which mirrors the distribution of the broader region

Inputs

Key Inputs:

- Travel Zone control totals for ERP in OPD
- Base year ERP totals by Travel Zone
- Regional age-sex control totals
- Base year age-sex distribution by travel zone

The module requires a number of parameter inputs:

- Number of regions – set to 56 under the current BTSLUR structure
- Convergence criteria – iterative process is complete when the sum of the absolute differences (between estimated regional age-sex and regional age-sex control totals) are below this level
- Future year – module will perform IPF up to the specified year
- Base year – specify the base year (currently 2016)
- ERP_OPD Criteria – Specify the share of a Travel Zone's population that has to live in a private dwelling in order for the TZ specific age-sex distribution to be used. Travel Zones which do not satisfy this criteria will be assigned the regional age-sex distribution
- Input data structure – Defines the structure of 'TZ11 ERP' and 'Regional Age Sex Totals' sheets
    - Required for model to read input data correctly
    - Will need to be adjusted if a new regional or TZ structure is adopted

Functions

This module runs the IPF procedure to estimate the ERP in OPD within each age-sex category for the base year and all future periods.

The following is a summary of the estimation process:

- Read input parameter data
- For each region that requires estimation:
    - Read base year control total and age-sex distribution input data
    - Use IPF procedure to estimate final age-sex distribution in base year
    - For each future period:

- Read control total input data and age-sex output from previous period
- Use IPF procedure to estimate final age-sex distribution

7.2.3 M02c – Results and checking

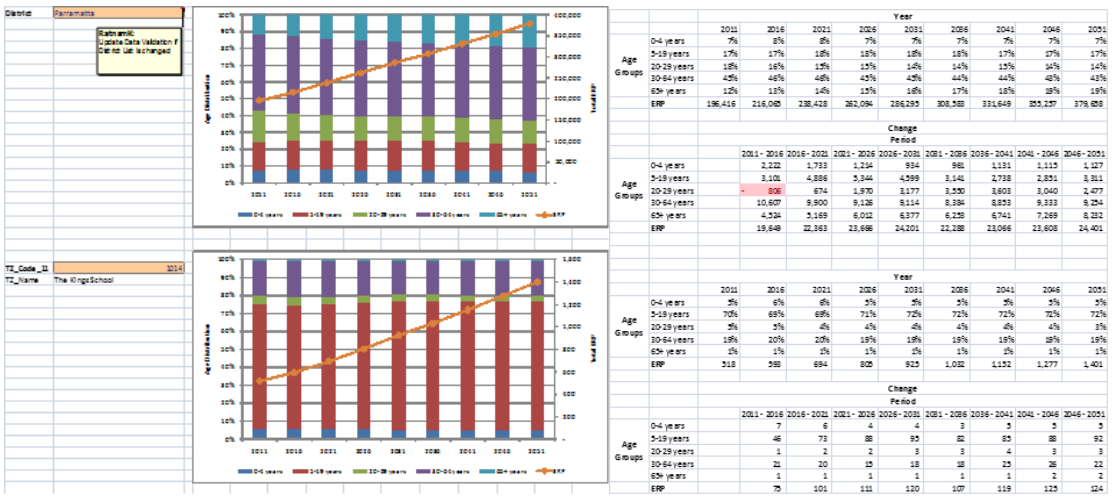
Collates the output from M02b, performs validation checks and provides summaries by regions (currently BTSLUR), district and travel zone.

Functions

This module has four primary functions:

- The ‘TZ ERP Validation’ sheet checks that the final age-sex outputs sum to ERP in OPD control totals for each travel zone and year
    - All cells should equal 0 (displayed as ‘-‘)
- The ‘Regional Validation’ sheet checks that the sum of ERP in OPD within each age-sex category, across all constituent travel zones of each region, is equal to the corresponding regional control totals.
- All cells should be ~0
- The ‘Regional Summary’ sheet aggregates the age-sex output into 5 categories and presents the results for each region and travel zone
    - Select region first then choose travel zone within that region
- The ‘District Summary’ sheet aggregates the age-sex output into 5 categories and presents the results for each district and travel zone
    - Select district first then choose travel zone within that district
    - District definitions can be changed using the ‘District Definitions’ sheet

Figure 22 M02c Output summary

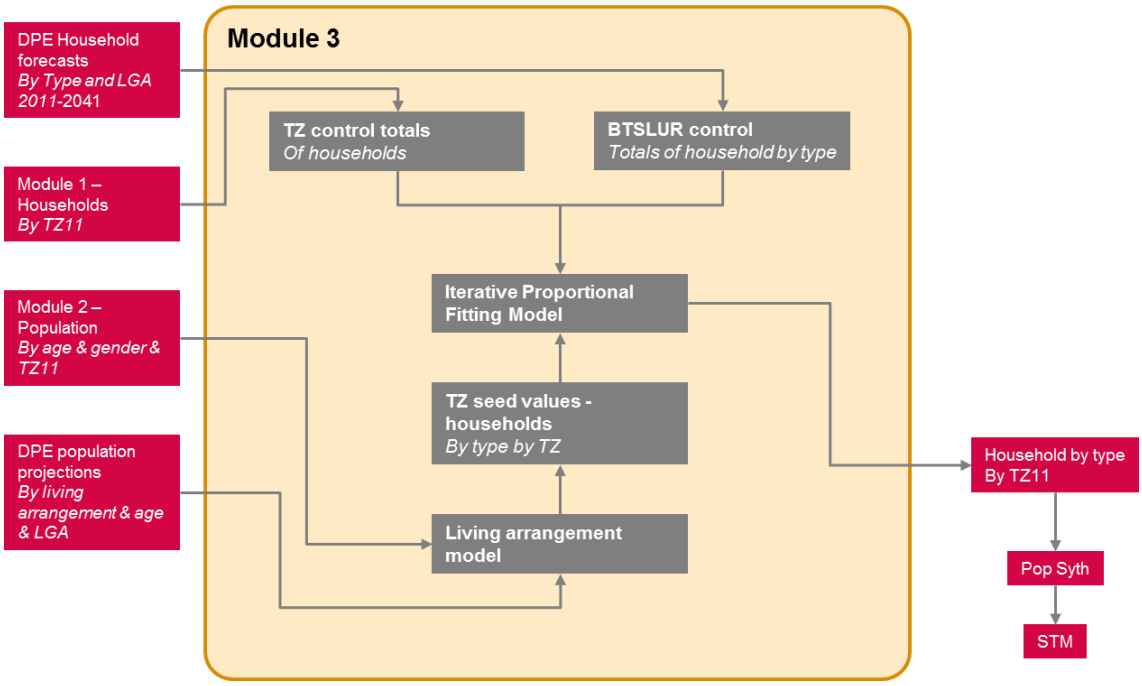


8 Module 3: Household type

8.1 Module 3 Overview

No changes in the method were made to the Household by Type module as compared to TZP2016. The following diagram presents an overview of the module and how it interacts with the TZP framework.

Figure 23 Module 3: household by type overview



The primary function of the Household by Type module is to use the outputs of previous modules and DPE household forecasts and population projections to project household totals by type.

The household types considered are:

- Couples with Children
- Couples only
- Single parent
- Single person
- Other types (which comprises primarily group households).

The inputs into module 3 specify the household totals at the travel zone level and distribution of household types at the adjusted LGA level. Thus, the focus of module 3 is on the distribution of these household types at the travel zone level – this distribution is achieved using the Iterative Proportional Fitting (IPF) approach explained in Module 2 (Section 6).

For each time period the IPF approach requires an approximate distribution of household types at the travel zone level as a starting point, or 'seed value'. These seed values are generated and inferred from the population projections and living arrangement distributions provided by module 2 and DPE, respectively.

8.2 Module 3 technical spreadsheet overview

The following sections provide a brief overview of how the model spreadsheets operate.

Further documentation is included with each spreadsheet.

8.2.1 M03a – Adjusted LGA population

This module converts LGA level population forecasts by age and living arrangement to the Adjusted LGA level.

Inputs

Key Inputs:

- LGA population forecasts by age and living arrangement 2011-2041 (provided by DPE)
- LGA to Adjusted LGA concordance (supplied by TPA)

Functions

This module performs the following function:

- Converts LGA level population forecasts by age and living arrangement to the Adjusted LGA level

8.2.2 M03b – Living arrangement propensities

This module generates population living arrangement propensities by age and household type from the output of M03a.

Key assumptions

To convert the population in occupied private dwellings projected by age and living arrangement to population projection by age and household type the following table is used.

Table 9 Household groups definition

Population in OPDs	Couples with children	Children <15 with 2 parents Child aged 15+ living at home Partnered with children <15 Partnered with children 15+
	Couples only	Partnered with no children at all
	Single parent	Single parent with children <15 Single parent with children 15+ only Living alone
	Single person	Single in a group household
	Group	Single in a group household
Population in NPDs		

Inputs

Key Input:

- Adjusted LGA population forecasts by age and living arrangement 2016-2041 (output from M03a)

Functions

This module performs the following functions:

- Converts population projection by age and living arrangement to population projection by age and household type
- Calculates the Adjusted LGA population living arrangement propensities by age and household type for 2016-2041

8.2.3 M03c – Calculating seed values

Estimates the distribution of households by type at the travel zone level for 2016-2041. This sub-component is scripted in R and technical details are contained in Appendix D.

Key assumptions

- Only persons of age 15 + are considered in the Household by Type categories
- To convert population projections by household type to household projections the following Household-Population factors are assumed for all travel zones.

Table 10 Household – Population factors

Household type	Scale factor
Couples with children	0.5
Couples only	0.5
Single parent	1.0
Single person	1.0
Group	0.449

Inputs

Key Inputs:

- Adjusted LGA population living arrangement propensities by age and household type 2016-2041 (output from M03b)
- TZ11 population projections by sex and age 2016-2041 (output from M02b)
- Group headship data 2016 (2016 Census of Population and Housing)

Functions

This module performs the following functions:

- Converts TZ11 population projections by sex and age to TZ11 population projection by age
- Generates population by age and household type projections 2016-2041

Estimates the TZ11 distribution of households by type for 2016-2041

8.2.4 M03d – Running the IPF

This module estimates the distribution of households by type for 2016-2051.

Key assumptions

- The Adjusted LGA household projections by type are extended to 2056 by assuming the Adjusted LGA distribution of household types remains fixed from 2051

Inputs

Key Inputs:

- TZ11 household projections 2016-2056 (output from M01b)
- Adjusted LGA household projections by type 2016-2051 (supplied by TPA)
- Estimate of the TZ11 household distribution by type 2016-2041 (output from M03c)

Functions

This module performs the following functions:

- Extends the adjusted LGA household projections by type data to 2056
- Performs the IPF procedure using seed values for 2016-2041 from M03
- Performs the IPF procedure for 2046-2051 using seed values from the IPF output of the previous time period

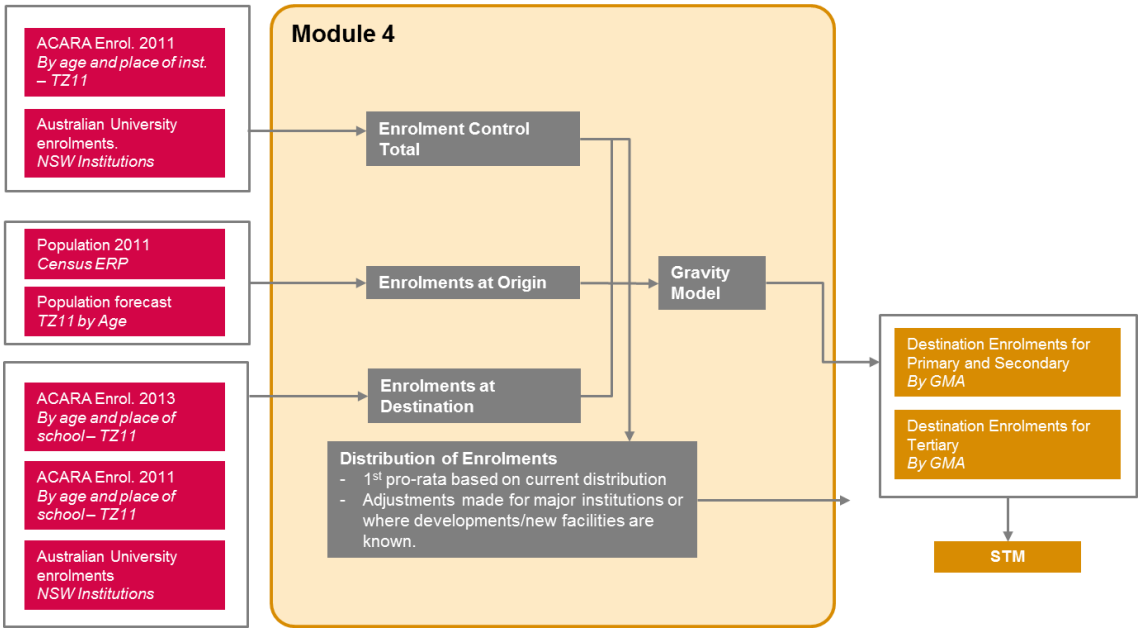
9 Module 4: Enrolments

This chapter provides a summary of the revised student enrolment module which has been reviewed and currently further developed by SGS.

9.1 Module 4 overview

The primary function of the student enrolment module is to use some outputs from previous modules to project primary, secondary and tertiary enrolments. The outputs generated from this module will result in Primary, Secondary and Tertiary enrolments at the place of institution.

Figure 24 Module 4: Enrolment approach overview



9.2 Module 4 technical model overview

9.2.1 Module 4a: Input Data Formatting

This module prepares the input data for the other modules and consists of three separate workbooks: School Data Analysis; Age Control Totals; Census SA1 TZ11 Concordance.

9.2.1.1 M04a_1-School Data Analysis

This module compiles the forecast enrolments for each school.

Input

- DoE School Enrolments Data: 2005-2051
- ACARA Enrolments Data

Functions

- The 'Government Adjusted Enrolments' sheet compares the ACARA and DoE data and adjust DoE enrolments where large discrepancies exist.

9.2.1.2 M04a_2-Age Control Totals

This module reforms the population data in 5-year age categories from M02b into new categories for primary school (5 to 11), secondary school (12 to 17), university (18-25), working age (26-64)

Input

- Census data population single age groups by sex
- Population by age by travel zone (M02b)

Functions

- Disaggregates M02b data into single age groups using the proportions for the wider population and re-aggregates into groups.

9.2.1.3 M04a_3-Census SA1 TZ Concordance

Input

- Census 2016 data by age groups in single years, type of education institution attendees, attendee status at SA1.

Functions

- Aggregates data regarding student place of residence into required age groups
- Adjust for ABS small count randomisation
- Concords SA1 data to TZ

9.2.2 Module 4b: Enrolment Control Total

This module produces the enrolment control totals for residents within the NSW region. These control totals are used in the Module 4c. The IPF method is used to calculate the distribution of age to education enrolment.

Input

- 2016 Census data detailing current age and education status.
- NSW school enrolments (M04a_1)
- Tertiary enrolments data provided by Department of Education and Training and National Centre for Vocational Education Research

Functions

- The 'IPF Output' sheet draws on all other input sheets to produce the enrolment control totals. The IPF method used is the same as used to produce outputs in Module 2. The IPF method is applied to calculate the distribution age groups by education enrolment.

9.2.3 Module 4c: Enrolments at Origin

This module estimates the total number of enrolled residents by travel zone. This module utilises the IPF method to distribute age control totals by institution enrolment. The ERP by age group by travel zone totals are estimated in Module 2 are aligned with the state regional enrolled resident control totals as estimated in Module 4a. This ensures the total number of enrolled persons by school type for each travel zone is equal to the population total for that travel zone as well as being equal to the institution type.

Inputs

- Census data
- M04a_3 – Enrolments by TZM04b control totals
- IPF parameters

Functions

- **"Input Base data":** This sheet requires inputs from M04a_3 and M043_2
- **"Macro Trends":** This is derived from share calculations dependant on the results of M04a. The propensity for each age group to enrol into a specific institution is based on Census data. These propensities are applied to 2016-2051 ERP totals to derive the control totals by age and institution.
- **"Input Control Totals":** Control totals for 2016 come from M04b. For years 2021-2051, institution control totals are trended based on the shares determined in the sheet 'Macro Trends'
- **"Input Distribution":** Reading from the sheet 'Base data'. The distribution is only needed for 2016. The IPF for years 2021-2051 is dependent on the results of the previous year's distribution. Essentially, the Census distribution is used consistently to inform all distributions. This method is similar to the method used in M02b.
- **"Input Parameters":** This sheet requires inputs to the IPF calculations. The user will need to input the number of zones to be calculated.

Key assumptions

- Macro trends are based on Census trends. This assumption is linear but currently the best estimate.
- 2016 Census distribution is the base for IPF. This distribution essentially informs the future years as the IPF calculations draw on the distribution determined in the previous year.

9.2.4 Module 4d: Enrolments at Destination

This module collates the two key datasets and allocates a travel zone to each school listed to determine the base year (2016) distribution of enrolments. The current capacity of each school is currently being analysed. The forecasts given by the Department of Education are linear. These forecasts will be used as a base for determining the capacities of each school.

The Department of Education has specific capacity parameters for schools. They are:

- Primary schools can only allow for a maximum of 1,000 students without any major work being completed to the school site.
- Secondary schools can only allow for a maximum of 2,000 students without any major work being completed to the school site.
- These capacities are further restricted to allow a minimum of 10 sqm of play space/green area per child on the school site.

Inputs

- School capacity forecast from M04a_1

Key assumptions

School capacities are a major assumption in this module. The Department of Education is currently reviewing their capacity forecasts. This data is not currently available. This means significant assumptions around school capacities will be made in this module as well as any development of new schools. This module will be created to allow for Department of Education forecasts to be easily fed into the model once they are available.

9.2.5 Module 4e: Final School Enrolments

This module estimates the small area destination enrolments in each future year out to 2056 using a gravity model. It is programmed in R and uses inputs from the preceding modules.

Inputs

- Capacity of each school by type by travel zone into the future, assumptions are made regarding the approximate locations/size of future schools as detailed in M04d. This is only a ‘capacity’ input and students are only allocated to this capacity if there is sufficient demand estimated by the gravity model.
- Origin enrolled persons by school type by travel zone as detailed in M04c
- Travel time matrix for each travel zone in NSW and each school travel zone as provided by TPA. .

A gravity model in R essentially uses the following conceptual structure estimates the propensity for a student to attend a school. This propensity is calculated from every travel zone to every school.

$$\text{Propensity to attend a school by type} = \frac{\text{Estimated preference}}{\text{Travel time to school}}$$

A different propensity is calculated for each type of student: primary Government, primary non-government, secondary Government and secondary non-government. The estimated preference is not an input; it is calculated in the base year using actual enrolment origin/destination data. This is then used for the next year. The estimated preference adjusts each year if schools reach capacity to reduce the schools 'pull'. This approach results in students attending the closest school (within capacity constraints) adjusted by known preference behaviours (i.e. more prestigious schools will draw a wider catchment).

9.2.6 Module 4f: Tertiary Enrolments

This module estimates tertiary enrolments for the destination of students at both TAFE and University under a Full time and Part time split. First it collates the Tertiary enrolments data and allocates a travel zone(s) to each institution. .

In the previous version TZP2016 the total of enrolled persons in 2011 is then compared to the total from the enrolments data to determine an upscale factor. This is then applied to the origin data for each travel zone to determine the number of enrolments by place of residence for each year from 2011 to 2051. This new total was then re-distributed to each travel zone with a tertiary institution using the base data to estimate the final destination enrolments by place of attendance. Shares for institution by type are then adjusted for future periods to reflect planned redevelopments/expansions where information is available.

This sub-component is scripted in R and technical details are contained in Appendix E.

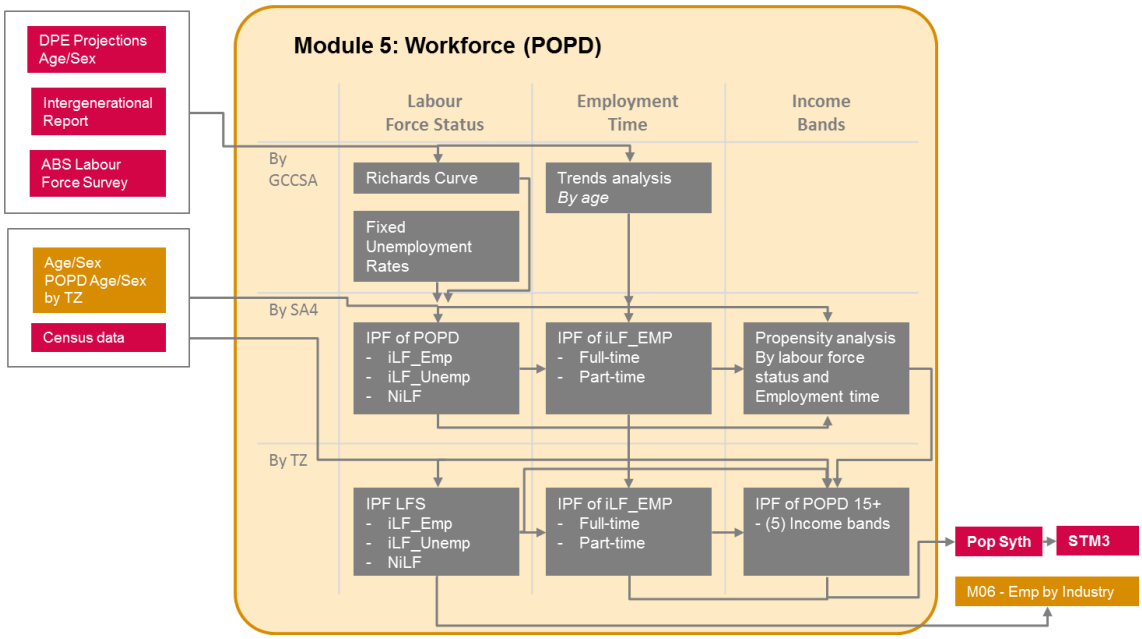
In TZP2016 v1.51 the up scaling outlined above is not adopted. Administrative data from the NCVER counts enrolments, not students. Further, their data indicates that in 2016 only 50% of subject enrolments were delivered in a classroom with the balance including 30% delivered online and a further 11% delivered at the place of work. Additionally, many VET courses are short and a person may enrol in many multiple courses during the year, furthermore, unlike University, enrolments in VET take may place all through the year. In that case the enrolment figure will be above the number of persons enrolled at any one time. For these reasons, it was decided that the census figure provides a closer approximation of the average number of persons travelling to VET institutes on a daily basis than the administrative data available from the NCVER.

10 Module 5: Workforce segmentation

10.1 Module 5 overview

The following diagram presents an overview of the revised Workforce module and how it interacts with the broader TZP framework.

Figure 25 Module 5: Workforce approach overview



The primary function of the Workforce module is to use the outputs of prior modules to project various segmentations of the working age population in a largely automated manner. It essentially does this three times at different geographic scales: Greater Sydney; sub-regions (SA4s), and finally by TZ. The module also breaks down the data in three stages: Workforce Status, Employment status (full-time/part time) and Income Bands. Trends and external data sources are used for high geographies, while a more automated approach is used for smaller geographies.

Specifically the outputs generated by the Workforce modules include:

Segmentation of the working age population (POPD 15+) by workforce status	Segmentation of employed population by employment status (full time/part time)	Segmentation of the working age population (POPD ≥15) by income bands*
• In Workforce and Employed • In Workforce and Unemployed • Not In Workforce	• Employed full-time • Employed part-time	• $0 - $20,799 • $20,800 - $31,199 • $31,200 - $41,599 • $41,600 - $65,000 • ≥$65,000

* It should be noted that the final two income bands generated do not exactly match the bands of the HTS (which are $41,600 - $67,599 and ≥$67,600).

Before each module is described individually the following provides some broader reasoning and processes elements established in this Module:

	Workforce Status	Employment time	Income Bands
By GCCSA	Disaggregation is first undertaken at the metropolitan level for a number of reasons: • The analysis can draw on a wider range of datasets and research which is only available at a more aggregate geographic level. • More complex modelling and analysis can be completed which is not possible for 100s/1,000s of areas. • The data can be easily checked against other key indicators and form a benchmark for subsequent disaggregation. (e.g. projected participation rates can be converted to a total workforce for Sydney and checked against national and state trends and jobs growth)		
By GCCSA	Participation rates estimation focuses on projecting entry/exit rates for each of the 42 age-sex group over time. This captures both changing behaviour (i.e. females having children later) and the evolving age profile of the city. While unemployment fluctuates significantly in the short-term due to economic cycles it is relatively stable in the long term. Therefore, analysis has focused on estimating a reasonable long term (stable) rate.	A wide number of factors contribute to full/part time split. From industry structure, to economic cycles and labour market constraints. Analysis showed the propensity to work part-time is primarily driven by age (e.g. young people studying or older people reducing their hours). Therefore, it has been used to project this attribute over time.	Income distribution is not modelled by GCCSA. This is because labour force survey data at an SA4 level is equally robust for the chosen method (see below).
By SA4	The stage 1 IPF process produces regional level control totals. It is better to complete this process in two stages, rather than jumping straight down to TZ level for a number of reasons: • Additional sub-regional (SA4) data that is not available at a TZ level can be draw into the process • Additional checks and QA can be performed at an aggregate level where results are easier to interpret (i.e. the number of workers in the “Eastern Suburbs” and be checked rather than just TZ 456.)		

	Workforce Status	Employment time	Income Bands
By SA4	This captures regional (SA4) variations in both workforce characteristics and demographic structure. This intermediate step enables the incorporation of the most recent labour force survey data, which is more recent and robust than census data.	This captures regional (SA4) variations in both workforce characteristics and demographic structure. This intermediate step enables the incorporation of the most recent labour force survey data, which is more recent and robust than census data.	Income distribution is complex and similar to full/part time split is influenced by a wide range of factors. Analysis similarly showed that the workforce status and employment time are the most significant determinants (and are also influenced by age-sex structure) and hence were used to project this attribute. As real income growth will cause a shift across income bands, an appropriate rate (currently IGR forecast) is chosen to account for this.
By TZ	Robust metropolitan and sub-regional control totals should now have been established. Therefore, the final IPF process has a much less significant task and simply needs to disaggregate these control totals reasonably across the current and projected population by TZ. Census data is used to capture local variations within each SA4.		

10.2 Module 5 technical spreadsheet overview

10.2.1 Module 05a/b – Participation Rate for Greater Sydney and Rest of NSW

These sheets involve projecting participation rates (PR) by age and sex at a high geographic level (Greater Sydney and Rest of NSW). A cohort based projecting approach, which is consistent with the method which is used both internationally and by the Productivity Commission, has been utilised.

Unlike a direct projection of PR, a cohort approach captures the differing labour market behaviour across generations (caused by factors such as shifting social attitudes or changes to the aptitudes/opportunities of the population).

The cohort approach requires the rates at which cohorts enter or exit the labour market to be modelled. Ignoring net changes to the civilian population of an age cohort (migration, births and deaths), an entry rate represents the proportion of those who are currently not in the workforce that enter the workforce in the next period, while an exit rate represents the proportion of those who are currently in the workforce that will leave the workforce in the next period.

Entry Rate Example:

In the 20-24 year old age group in 1990 there are 200 people in the workforce and 100 who are not in the workforce (participation rate of 66.6%).

If in 1995 there are 250 people in the 25-29 year old age group in the workforce (and 50 not in the workforce), then the entry rate for 20-24 year olds in 1990 was equal to 50%. Alternatively, this represents a PR of 83.3% for the 25-29 year old age group in 1995.

This entry rate could represent students completing university and entering the workforce. Changing attitudes, such as students preferring to study for longer, could result in the entry rate of 20-24 year olds decreasing over time.

A detailed description of the cohort based approach can be found in Technical Paper 3 of the 2005 Productivity Commission Report 'Economic Implications of an Ageing Australia'. A discussion of Richards curves can be found in Technical Paper 2'. These can be accessed via the following link:
http://www.pc.gov.au/inquiries/completed/ageing/technicalpapers

While both entry and exit rates have an upper bound of 1, a lower limit does not theoretically exist. For a given age-sex category, it is therefore appropriate to model the rate which will be positive in future years.

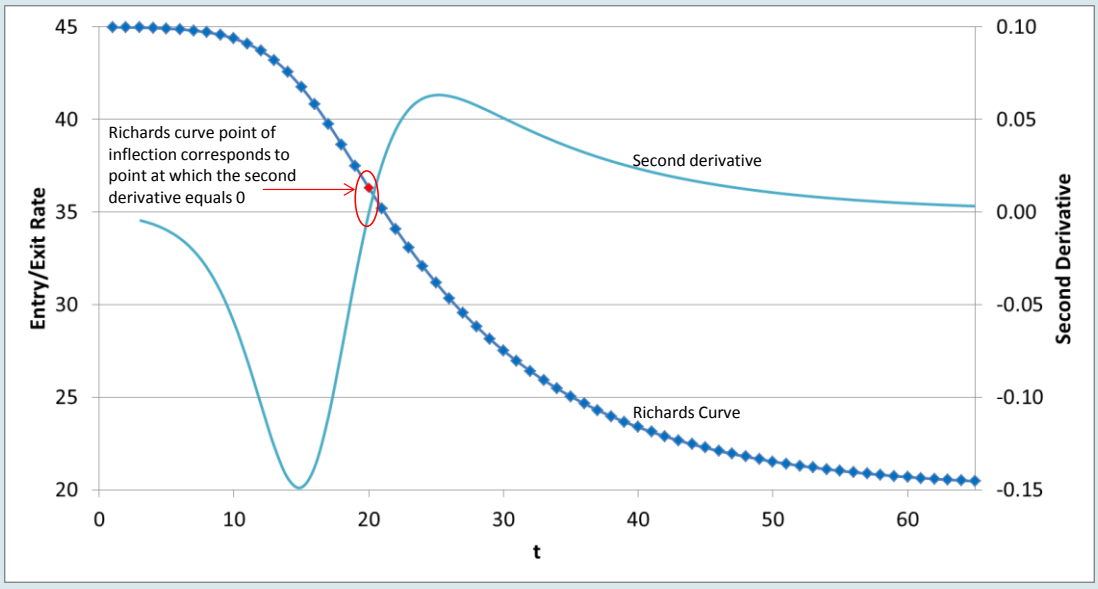
Once the appropriate type of rate is chosen, it is modelled using a Richards curve (generalised logistic function). A Richards curve has desired property of being bounded and, as discussed in the Productivity Commission report, has the advantage of a greater degree of flexibility (e.g. it does not have to be symmetric).

Richards Curves

The functional form of a Richards curve is: $y = c + \frac{a}{(1+b\times e^{-g\times t})^{\lambda}}$

The diagram below presents a Richards curve with parameters:

$$a = 25, b = 0.001, c = 20, g = -0.4, \lambda = 0.2$$



As shown above, the point of inflection of a Richards curve is defined as the point at which its second derivative is zero. This is of significance because if the point of inflection is known, one less parameter needs to be estimated.

Inputs

Key Inputs:

- Labour Force Survey participation rate data - ABS cat 6291.0.55.001 (LM2)

The module requires a number of initial parameter inputs. The 'Input Parameters' sheet will load the relevant data into the individual participation rate projection sheets:

- Number of historical entry/exit rate observations
- First and last years of entry/exit rate observations
- Number of PR observations (this will be 5 years more than the number of entry/exit rate observations, as age cohorts are in 5 yearly groups)
- First and last years of participation rate observations
- Base year – specify the base year (currently 2016)
- Choice of rate to be modelled for each age-sex category
- Choose rate to be modelled based on visual inspection of the 'Entry Rates' and 'Exit Rates' sheets.
- Choose the rate that will be positive in future years
- Maximum PR by sex
- Define participation rate upper bound for each sex (currently based on existing literature which recommends 99% for males and 95% for females)

Once the model has been initialised, a number of inputs are required to model each age-sex category:

- Estimated point of inflection year for entry/exit rate series
- For chosen start and end years, the model will interpolate (or extrapolate) to determine the point at which the second derivative of the series is equal to 0
- Professional judgement as to the most likely point of inflection is required if the second difference series intersects the horizontal axis multiple times
- Select start and end years for Richards curve fitting
- Algorithm will fit curve to only these years of historical data
- Multiple trials may be required to achieve most suitable long term curve
- Starting values for parameters a,b,c and λ must be chosen.
- Multiple trials may be required in order for algorithm to suitable fit

Functions

- The 'Smoothed PR Data' sheet adjusts the input PR data using a Hodrick-Prescott Filter
  - The HP filter is a mathematical algorithm which is used to remove the cyclical component of a time series, resulting in a series that is more sensitive to long-term shocks.
  - The parameter λ, which relates to the penalty for changes in the long-term growth rate, is currently set at 100 (suggested value for annual data)

- On each age group sheet, projected PR of each age-sex category by fitting a Richards curve to historical entry or exit rate data. The following is a summary of the process:
    - Calculate point of inflection based on user inputs
    - Based on chosen years, point of inflection and starting parameter values, use solver algorithm (inbuilt to MS Excel) to compute parameters which minimise the sum of squared errors between the estimated entry/exit rate values and historical data
- The 'Smoothed Projected PR' and 'Projected PR' sheets combine historical PR with the projected entry/exit rates in order to project participation rates by age-sex category
- The 'Summary Charts' sheet provides the means for checking the final PR projections (viewed alongside both actual and smoothed historical PR data)

10.2.2 Module 05c – Unemployment rate for Greater Sydney and Rest of NSW

Module 5c projects unemployment rates by age group for Greater Sydney and the Rest of NSW based on user-defined assumptions regarding their long term ranges for overall (non-age specific) unemployment rates.

Historical unemployment rate data is used to determine the unemployment rate of specific age groups for periods when the regional rate was within a given range.

Unemployment Rate Example:

The overall long term unemployment rate forecast of the IGR is 5%. Based on this forecast, the long term unemployment rates for Greater Sydney could be defined to vary between 4% and 6%.

Examining the period spanning from 1991 to 2015, for periods when the Greater Sydney unemployment rate fell between 4% and 6%, the average unemployment rates for 20-24 year old males and 40-44 year old females were 8% and 4% respectively.

Inputs

- Labour Force Survey data – ABS cat 6291.0.55.001 (LM2)
- Upper and lower bounds of overall unemployment rate forecasts for
    - Greater Sydney
    - Rest of NSW (likely higher than Sydney)

Functions

- The 'Unemployment Rate' sheet identifies the historical years for which overall unemployment was within the user-defined assumption of long term unemployment rate ranges
- Using data from only these identified years, the 'Long Term Unemployment Projection' sheet computes age-specific unemployment rate projections

10.2.3 Module 05d – Workforce Base Data

The purpose of M05d is to synthesise data from various sources and structure the outputs such that they can be used as inputs to the IPF procedures in M05e, M05g and M05h. However, this module also provides a means to review the projections of workforce status and employment time at a metropolitan level (in order to ensure that

projected trends for Sydney and Rest of NSW are reasonable before further disaggregation).

Inputs

Key Inputs:

- SA1 to TZ11 concordance (obtained from GIS team)
- Labour force status by age by SA1 (obtained via TableBuilder)
- Employment time by age by SA1 (obtained via TableBuilder)
- Income band by employment time by SA1 (obtained via TableBuilder)
- Travel zone control totals for ERP in OPD by age (obtained from M02b)
- Labour force survey participation rate and unemployment rate data by age by SA4 (base year and most recent year)
- Historical and projected participation rate and unemployment rate data by age by GCCSA (obtained from M05a, M05b and M05c)
- GMA population in occupied private dwellings by age by sex by GCCSA
- GMA population in occupied private dwellings by age by sex by SA4
- GMA population in non-private dwellings by age by sex by GCCSA
- Historical and projected part-time employment shares by age by GCCSA (obtained from M05f)
- Parameter input: Base year of model

Functions

The first function of M05d is to produce GCCSA control totals and SA4 starting distributions for the stage 1 IPF processes of M05e (workforce status segmentation) and M05g (employment time segmentation). Once inputs are finalised, a macro computes the outputs via the processes described below:

- The model combines the GCCSA participation rate, unemployment rate and part-time employment rate inputs with population (in OPD) by age by sex data to produce GCCSA control totals for each workforce and employment time category
    - This process is replicated for population in non-private dwellings at a GCCSA level for workforce status only. This output is required to determine the overall size of the GMA workforce – used in employment module
- SA4 IPF starting distributions for the base year are computed using participation/unemployment/part-time employment rates from the base year labour force survey in conjunction with SA4 age-sex structures
- SA4 IPF starting distributions for projected years are computed using participation/unemployment/part-time employment rates from the most recent labour force survey in conjunction with SA4 age-sex structures of the relevant projected year

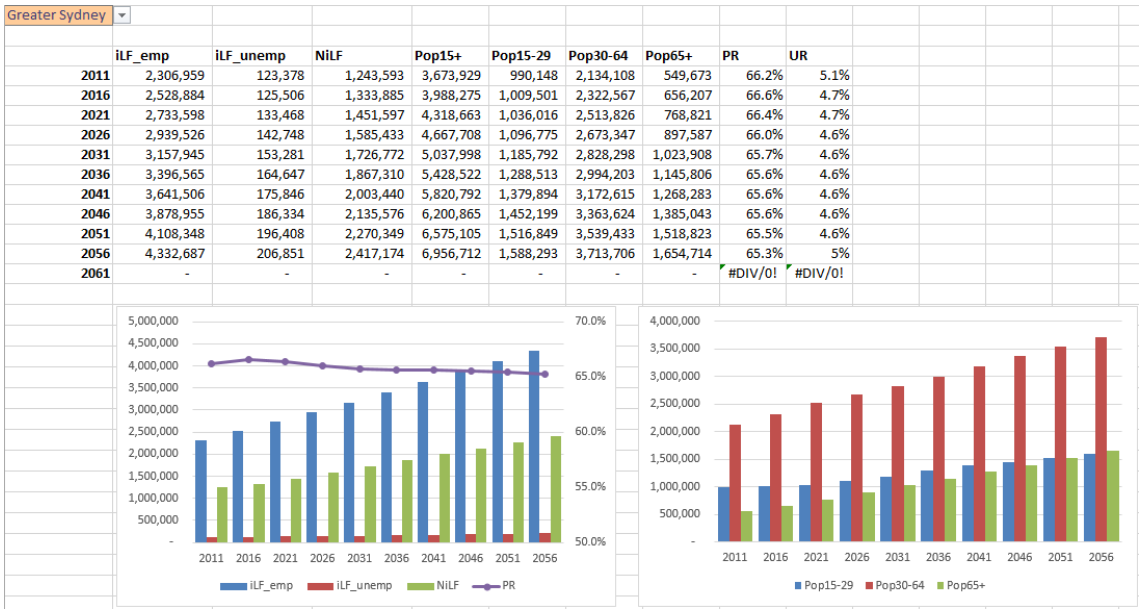
The second function of the module is to estimate the travel zone starting distributions for the stage 2 IPF processes of M05e (workforce status segmentation) and M05g (employment time segmentation). It also computes age-specific income band propensities by TZ (which are used in M05h to calculate IPF process starting distributions). This follows the process outlined below:

    - Convert SA1 level census data to TZ11

- Zones with a low total population are likely heavily influenced by ABS randomisation – these zones are adjusted to reflect the structure of the broader region. This results in the following propensities:
- Workforce status propensity by age group
- Employment time propensity by age group
- Income band propensity by workforce characteristics (employed full-time, employed part-time, unemployed and not in the workforce)
- For each year, IPF starting distributions for workforce status and employment time segmentation are calculated by combining the propensities by age group with the projected age structure of each travel zone
- For each year, the IPF starting distribution for income band segmentation is calculated by combining the income band propensities by workforce characteristics with projections of workforce status (obtained from M05e) and employment-time (obtained from M05h).

Finally, results for Greater Sydney and Rest of NSW can be reviewed as shown below.

Figure 26 M05D Review of workforce segments



10.2.4 Module 05e – Workforce IPF

Performs 2 stage IPF process which first estimates workforce status segmentation by SA4 and then by travel zone over time.

Inputs

Key Inputs (from M05d):

- Greater Sydney and Rest of NSW control totals by workforce segment and year
- POPD by SA4 and year
- Starting IPF process distribution of workforce segment by SA4 and year
- POPD by travel zone and year
- Starting IPF process distribution of workforce segment by travel zone and year

The module also requires a number of parameter inputs:

- Number of SA4s within Greater Sydney and Rest of NSW (15 and 13 respectively)
- Convergence criteria – iterative process is complete when the sum of the absolute differences (between estimated regional workforce segment totals and regional workforce segment control totals) are below this level
- Future year – module will perform IPF up to the specified year
- Base year – specify the base year (currently 2016)
- Input data structure – Defines the data structure of the SA4 and travel zone level input sheets
- Required for model to read input data correctly

Functions

This module first runs the IPF procedure to estimate the ERP in OPD, by SA4, within each workforce category. Using these as regional control totals, a second IPF process then estimates the workforce segments for all constituent travel zones. The estimation process is summarised below:

- Read input parameter data
- For each year and region that requires estimation
    - Read Greater Sydney and Rest of NSW workforce segment control totals, SA4 ERP in OPD control totals and starting distribution data
- Use IPF procedure to estimate final SA4 workforce distribution
- For each year and region that underwent the first stage IPF procedure
    - Read travel zone ERP in OPD control totals and starting distribution data (see above for calculation detail)
- Use IPF procedure to estimate final travel zone workforce distribution.

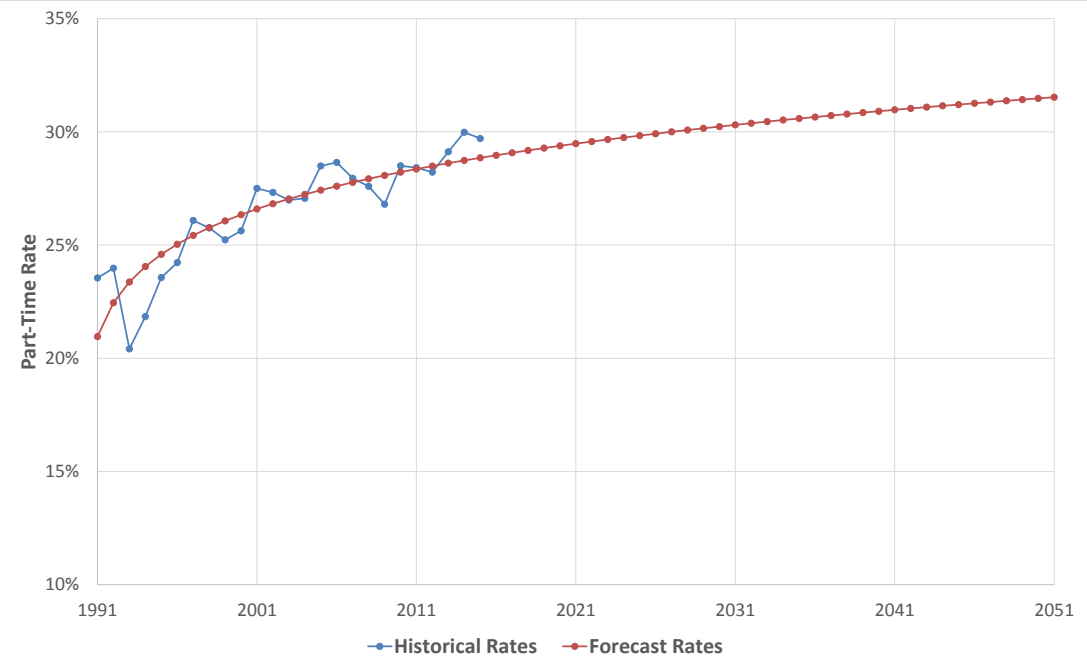
10.2.5 Module 05f – Workforce Part-Time Rates

At a metropolitan level (i.e. Greater Sydney and Rest of NSW), module 05f utilises historical labour force survey data to project age-specific propensities for workers to engage in part-time employment.

The modelling approach chosen is a univariate log-log regression. While a cohort approach, similar to that used in M05a/b was considered, the log-log regression method was selected due to the fact that it achieves reasonable projections with a far greater degree of automation.

Log-log regression

The estimated regression: $\log PR = \alpha + \beta \log t$



The diagram above presents the observed and projected part-time rates for 55-64 year olds in Greater Sydney where α=-0.83 and β=0.07.

Inputs

Key Inputs:

- Labour Force Survey data – ABS cat 6291.0.55.001 (LM1)

The module also requires a number of parameter inputs:

- Data specification inputs detailing the first and last years of historical data and the base projection year
- Part-time employment rate upper limits by age group
- User defined part-time rate upper limits by age group and metropolitan region

Functions

- The 'Parameter Estimation' sheet estimates the parameters α and β for each age group in Greater Sydney and Rest of NSW
- Once the module is run, the 'Final Part-Time Rate Series' sheet complies the historical and projected part-time rates
- The 'Summary Charts' sheet provides the means for checking the final Part-time projections (viewed alongside historical data)

10.2.6 Module 05g – Part-Time Segmentation IPF

Performs 2 stage IPF process which first estimates employment-time segmentation by SA4 and then by travel zone over time.

Inputs

Key Inputs:

- Greater Sydney and Rest of NSW control totals by employment-time segment and year
- Workforce in OPD by SA4 and year
- Starting IPF process distribution of employment-time segment by SA4 and year
- Workforce in OPD by travel zone and year
- Starting IPF process distribution of employment-time segment by travel zone and year

The module also requires a number of parameter inputs:

- Number of SA4s within Greater Sydney and Rest of NSW (15 and 13 respectively)
- Convergence criteria – iterative process is complete when the sum of the absolute differences (between estimated regional workforce segment totals and regional workforce segment control totals) are below this level
- Future year – module will perform IPF up to the specified year
- Base year – specify the base year (currently 2016)
- Input data structure – Defines the data structure of the SA4 and travel zone level input sheets
- Required for model to read input data correctly

Functions

This module first runs the IPF procedure to estimate the workforce in OPD, by SA4 employment-time category. Using these as regional control totals, as second IPF process then estimates the employment-time segments for all constituent travel zones. The estimation process is summarised below:

- Read input parameter data
- For each year and region that requires estimation
    - Read Greater Sydney and Rest of NSW employment-time segment control totals, SA4 workforce in OPD control totals and starting distribution data
- Use IPF procedure to estimate final SA4 employment-time distribution
- For each year and region that underwent the first stage IPF procedure
    - Read travel zone workforce in OPD control totals and starting distribution data (see above for calculation detail)
- Use IPF procedure to estimate final travel zone employment-time distribution

10.2.7 Module 05h – Income Segmentation

Module 05h segments POPD who are of working age (15+) into five income bands. Propensities for individuals to fall within each income band are modelled as a function of their workforce characteristics (employed full-time, employed part-time, unemployed and not in the workforce) and real income growth.

For each workforce and employment-time segment, census data is used to determine the propensity of that segment to fall within eleven census income categories. Assuming a uniform distribution within categories, the eleven bands are disaggregated to form 29 sub-bands.

Combining these propensities with projections of POPD by workforce characteristics (obtained from M05e and M05g) results in a preliminary estimate of POPD within each income band. However, this estimate is based on income levels in the census year 2016, and real wage growth over time will result in a shift across income bands. Using an assumed rate of long-term real income growth (currently set as 1.4% in line with the 2015 IGR); along with the median earnings of each income sub-band, the shifts across bands are calculated. These shifts are then applied to the preliminary estimate to derive a final income band distribution by SA4.

Real Income Growth Example:

Suppose that for a given SA4 in 2016, part-time workers had a 25% propensity to fall within income sub-band 6. In the absence of real income growth, this implies that if there were 100 part-time workers in 2021, 25 of them would fall into income sub-band 6.

However, as real income does in fact grow, these 25 part-time workers will instead shift across income sub-bands - falling within income sub-band 8 rather than sub-band 6.

Similarly, individuals from a lower sub-band may now fall within sub-band 6. Another alternative is that 10 years of real income growth for a particular sub-band (e.g. sub-band 1) is insufficient to cause a shift.

Once income band estimates by SA4 have been derived, these are used as regional control totals in a travel zone level IPF process. This ensures that regional income band characteristics are maintained while still allowing for intra-regional variation based on travel zone specific workforce characteristics and income band propensities.

Inputs

Key Inputs:

- Income by workforce characteristics census data by SA4 (obtained via ABS Tablebuilder)
- Labour force characteristics by SA4 and year
- Working age POPD by travel zone and year
- Workforce characteristics by travel zone and year

The module also requires a number of input parameters;

- Number of SA4s within Greater Sydney and Rest of NSW (15 and 13 respectively)
- Convergence criteria – iterative process is complete when the sum of the absolute differences (between estimated SA4 income band totals and SA4 income band control totals) are below this level
- Future year – module will perform IPF up to the specified year
- Base year – specify the base year (currently 2016)
- Input data structure – Defines the data structure of the SA4 and travel zone level input sheets
    - Required for model to read input data correctly
- Real Income Growth Rate
    - Defines the rate at which individuals will shift across income bands

Functions

- The ‘Income Band Growth’ sheet computes the median income for each of the 29 income sub-bands and applies the assumed real income growth rate to these across time. This determines the shift across income sub-bands
- The disaggregation of the eleven census income categories into the 29 sub-bands (assuming a within-category uniform distribution) is performed on the ‘Adjusted SA4 Census Data’ sheet
- For each year, the preliminary 29 sub-band propensities are adjusted for real income growth and summarised on the sheet ‘Projected sub-band Propensities’
- These 29 sub-bands are then aggregated into the 5 income bands required by the STM and applied to workforce segment projections to derive final estimates of working age population by income band and SA4.
- Travel zone starting distribution is read from M05e (see above for calculation detail) and IPF procedure estimates final income band distribution

10.2.8 Module 05i – Workforce and Segmentation Summary

Collates the output from M05e, M05g and M05h presents them for review.

Functions

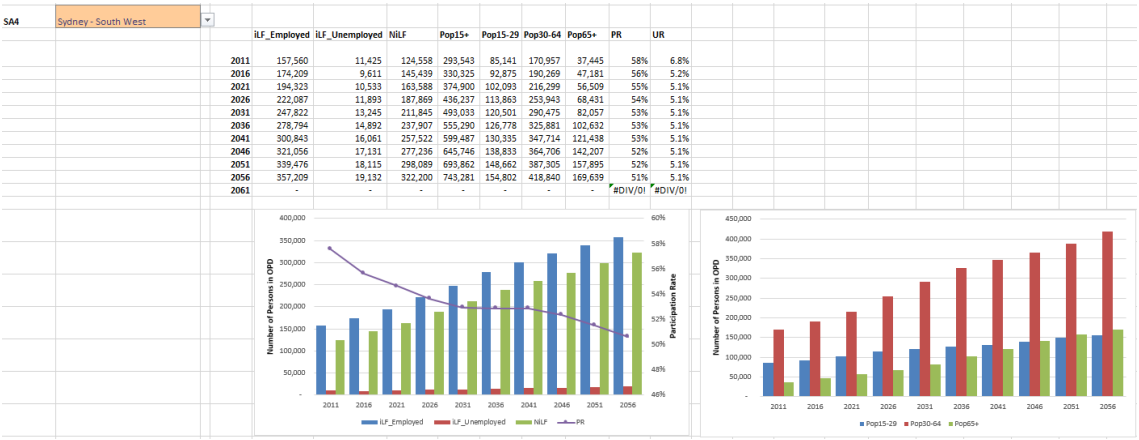
This module provides a means for reviewing the following results:

- Workforce segments (Employed, Unemployed and Not in Workforce) alongside age structure
- Employment-time segments (Full time and Part time) alongside age structure
- Income Bands

Results are reviewed at three geographic levels:

- SA4
- District
    - District definitions can be changed using the ‘District Definitions’ sheet
- Travel zone
    - Select SA4 or district first and then choose from constituent travel zones

Figure 27 M05I Summary output



11 Module 6: Employment by Industry

11.1 Module 6 Overview

The Employment by Industry module has two major functions. First, the quantum of jobs by industry is projected for the GMA region. Following this, the spatial distribution within each industry is projected based on a variety of factors. This spatial projection is itself a two-step process, being estimated first by SA3 and then by travel zone.

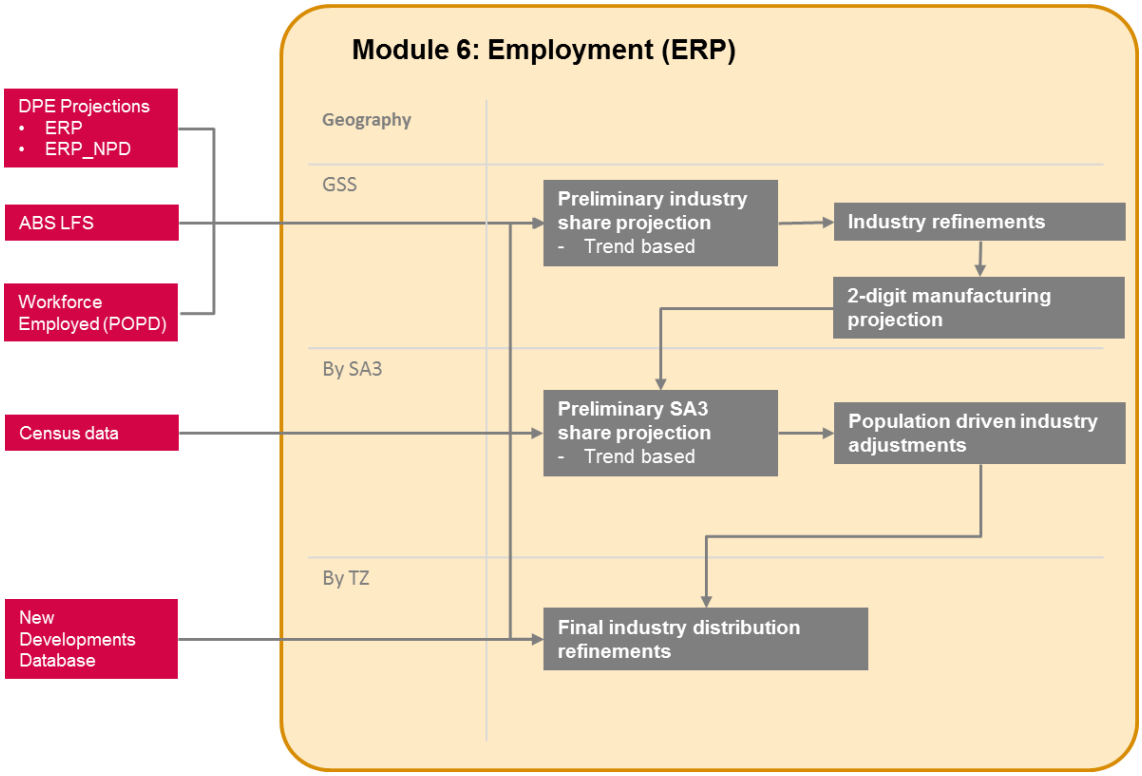
Unlike demographic characteristics, projecting the growth and distribution of employment by industry is more reliant on professional judgement and is influenced by both endogenous and exogenous factors. With this in mind, the design of the modules focuses on the following:

- Automated procedures to produce first-cut estimates (e.g. quantum and distribution of retail employment in GMA)
- Logical, transparent and time-efficient methods by which professional judgement can be applied

An important distinction between module 6 and previous modules is that estimates are not restricted to persons living in private dwellings. This is due to the fact that it is a 'destination' variable and is used as a relative attractor within the STM.

Table 4 lists the outputs generated by the employment module, while the figure below presents an overview of the revised employment module and how it interacts with the TZP framework.

Figure 28 Module 6: Employment projection approach overview



11.2 Module 6 - technical spreadsheet overview

11.2.1 Module 6a – Industry Projection for the GMA region

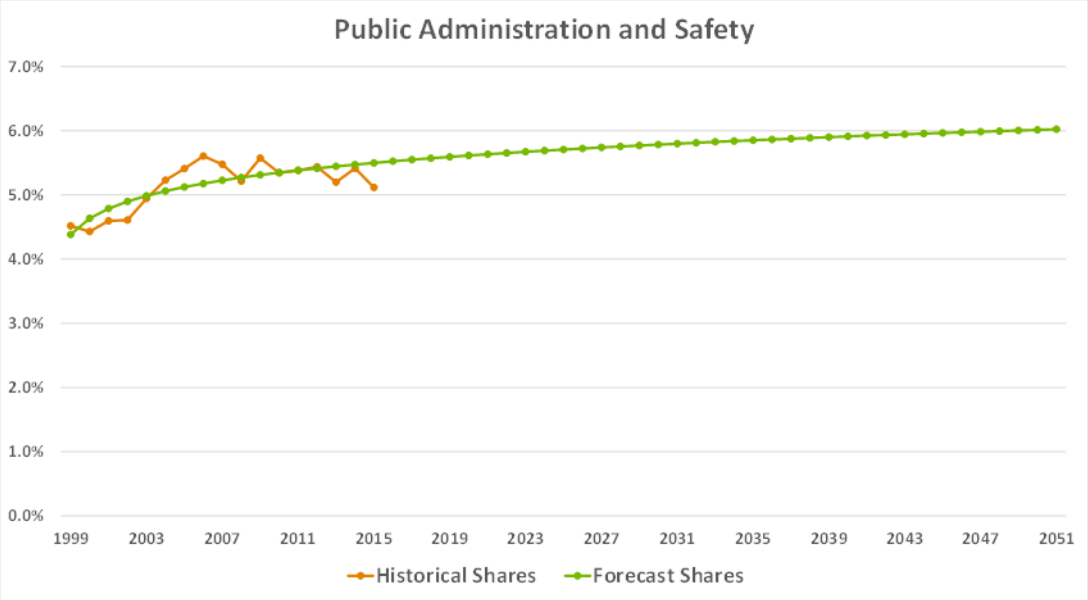
Module 6a combines the outputs of previous modules with historical labour force survey and census data to project the quantum of employment by industry within the GMA region. The projection process has three major components:

- Projected total employment within the GMA region for each period. This involves augmenting the outputs of M05 such that the following factors are captured:
- Flows of workers between the GMA region and the rest of NSW
- Employment of persons who reside in non-private dwellings
- Automated process which produces preliminary share projections based on historical trends
- The user can select from two univariate regression models
- Refinement of industry projections based on a variety of factors (e.g. population to employment ratio or expected growth rate)

Preliminary industry share example

The estimated regression can be of the form: $\log Share_t = \alpha + \beta \log t$ or

$$\log \text{Share}_t = \alpha + \beta t$$



The diagram above presents the historical and projected shares (based on a Log-Log model) of the Public Administration and Safety industry.

Inputs

Key Inputs:

- SA4 Labour Force Survey data – ABS cat. 6291.0.55.001
- Historical ERP by age data – ABS cat. 3218
- Projected ERP by age data - DPE
- 2016 Census data
- GMA workforce (POPD) from M05d

- GMA Population in Non-private dwellings from M01

The model also requires a number of parameter inputs on the 'Preliminary Share Projection' sheet:

- Select preliminary method for projecting the employment share of each industry
- Define the structure of historical data and projections
- Existing adjustments on the individual industry sheets can be cleared if desired

Following this, each industry requires inputs which define how its level is projected:

- Choose measure used to estimate first-cut projection
- Input values which correspond to the chosen measure (e.g. if ERP/Emp ratio is chosen, input projected ratio for each projection year)
- These will differ across industries and are subject to professional judgement
- Select whether to lock projection to first-cut estimate
- Industries which are not locked will have final estimates which are scaled such that the sum of jobs across industries is equal to the benchmark total for the GMA region

Functions

- The 'Annualised Data' sheet annualises the historical LFS data while also using 2016 Census data to discount the employment estimate of the 'Hunter Valley excluding Newcastle' SA4 (as this falls partially outside the GMA region)
- LFS data is defined by place of usual residence. The 'GMA Industry Employment (POW)' sheet adjusts this measure by using 2016 JTW Census data to account for worker flows between the GMA region and rest of NSW. This results in a final estimate of employment by industry within the GMA region
- The 'Parameter Estimation' sheet estimates the parameters α and β for each industry based on both log-log and log-linear models
- These parameters (based on a user selection) are then used to produce initial projections on the 'Preliminary Share Projection' sheet. Clicking 'Run' will load these onto the individual industry projection sheets
- For each industry sheet, review the projections implied by the preliminary shares against the provided indicators and refine if necessary. Charts displaying both level changes and growth rates are also presented, and should be checked to ensure projections follow a smooth path

Industry refinement example

Consider the Health Care and Social industry, presented below:

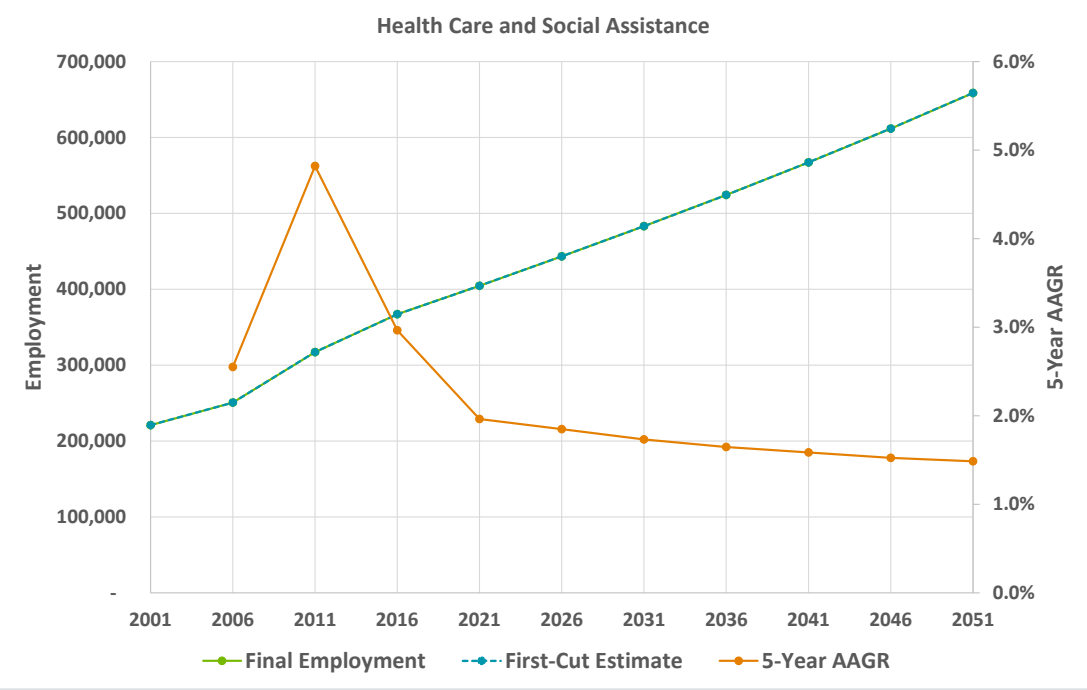
Health Care and Social Assistance	Chosen First-Cut Forecast Measure	Lock to First-cut estimate?												
Year	ERP/Emp	Lock		First-Cut Estimate	Final Employment Estimate	ERP AAGR	Preliminary Share	Share	5-year AAGR	ERP/Emp	ERP_15+/ Emp	ERP_15_64/Emp	ERP_65+/ Emp	ERP_5_24 /Emp
1999	-	216,756	-	216,756	216,756		9.59%	9.59%						
2000	-	214,446	-	214,446	214,446		9.25%	9.25%						
2001	23	221,035	-	221,035	221,035		9.34%	9.34%		22.7	18	15	2.8	6
2002	22	234,805	-	234,805	234,805	0.8%	9.77%	9.77%		21.6	17	15	2.7	6
2003	22	232,152	-	232,152	232,152	0.7%	9.53%	9.53%		22.0	18	15	2.7	6
2004	22	235,393	-	235,393	235,393	0.5%	9.61%	9.61%	1.66%	21.8	18	15	2.7	6
2005	21	243,526	-	243,526	243,526	0.7%	9.74%	9.74%	2.58%	21.2	17	14	2.7	6
2006	21	250,705	-	250,705	250,705	0.8%	9.87%	9.87%	2.55%	20.8	17	14	2.6	6
2007	21	254,567	-	254,567	254,567	1.5%	9.88%	9.88%	1.63%	20.8	17	14	2.7	6
2008	20	273,070	-	273,070	273,070	1.8%	10.28%	10.28%	3.30%	19.7	16	13	2.5	5
2009	19	281,006	-	281,006	281,006	1.8%	10.49%	10.49%	3.61%	19.5	16	13	2.5	5
2010	18	303,307	-	303,307	303,307	1.4%	11.18%	11.18%	4.49%	18.3	15	12	2.4	5
2011	18	317,241	-	317,241	317,241	1.2%	11.43%	11.43%	4.82%	17.7	14	12	2.4	5
2012	18	321,603	-	321,603	321,603	1.4%	11.49%	11.49%	4.79%	17.7	14	12	2.4	5
2013	17	337,758	-	337,758	337,758	1.6%	11.87%	11.87%	4.34%	17.1	14	12	2.4	4
2014	16.6	354,589	-	354,589	354,589	1.7%	12.30%	12.30%	4.76%	16.6	13	11	2.4	4
2015	-	353,723	-	353,723	353,723		12.04%	12.04%	3.12%	-	-	-	-	-
2016	16.5	367,137	-	367,137	367,137	1.5%	12.26%	12.28%	2.96%	16.5	13	11	2.4	4
2021	16.1	404,626	-	404,626	404,626	1.5%	13.43%	12.67%	1.96%	16.1	13	10	2.6	4
2026	15.8	443,431	-	443,431	443,431	1.4%	14.71%	13.11%	1.85%	15.8	13	10	2.7	4
2031	15.4	483,182	-	483,182	483,182	1.3%	16.11%	13.54%	1.73%	15.4	13	10	2.9	4
2036	15.1	524,306	-	524,306	524,306	1.2%	17.65%	13.95%	1.65%	15.1	12	9	2.9	4
2041	14.7	567,238	-	567,238	567,238	1.1%	19.33%	14.36%	1.59%	14.7	12	9	3.0	4
2046	14.4	611,819	-	611,819	611,819	1.0%	21.17%	14.80%	1.52%	14.4	12	9	3.0	3
2051	14.0	658,651	-	658,651	658,651	1.0%	23.19%	15.29%	1.49%	14.0	12	9	3.0	3

(1) (2) (3)

The preliminary share projection results in the industry capturing an unreasonably large share of regional employment (highlighted red). As a population driven industry, the ratio of population to employment is a reasonable indicator. Setting this as the chosen first-cut measure, and selecting values which imply that employment continues to grow slightly faster than population (1), results in a projection which:

- Has a more reasonable share of total employment (2)
- Displays a strong but declining growth rate consistent with recent trend (3)

Additionally, selecting 'Lock' means that the industry will not be scaled.



11.2.2 Module 6b – 2 digit Industry Projections

This module uses historical employment data from census to disaggregate the 1-digit industry projection produced in Module 6a to a 2-digit industry projection. The projection is broken into two steps:

- The share of employment for each 2-digit industry is calculated based on the historical employment data from 1996 to 2016.
- The shares of employment at a 2-digit level are projected using a log-log regression as in Module 6a. These shares are then applied to the 1-digit employment projections as determined in the previous module.

Inputs

Key inputs:

- Historical employment data at the 2-digit level from 1996-2016
- 1-digit industry projection at the 1-digit level produced in M06a.

Functions

- The 'Historical Employment Input' sheet requires inputs of 2-digit employment from 1996-2016
- The 'Industry Focus' sheet requires the user to pick the 1-digit industry to be disaggregated to the 2-digit level
- The '%_shares' sheet projects the share of each 2-digit industry based on a log-log regression.
- Using these projections, the 1-digit employment projection is disaggregated to produce the final 2-digit projection as seen in the 'Summary' sheet.

11.2.3 Module 6c – SA3 Industry Projections

Module 6c distributes the previously determined employment by industry projections to the SA3 level. This is achieved using historical census data in conjunction with information on planned developments and professional judgement. When assessing information on planned developments, which is sourced from the 'New Developments' database compiled by the TPA, there are two cases which should be considered:

- The 'New Developments' in some areas will represent growth that is over and above that which is predicted by projections based on historical data. This will be the case for greenfield or previously undeveloped regions
- The alternative is that these 'New Developments' are already partially captured within the original growth projections. This is because regions with sites that have undergone past urban renewal will be expected to continue exhibiting strong growth through similar renewal in the future.

The module also consists of two sub-modules (described in the next section) which synthesise inputs for M06c. These have been isolated as they are areas which could benefit from additional development.

The projection procedure of M06c is comprised of three stages:

- Preliminary SA3 shares for each industry are estimated using a log-log regression model
    - The projected shares are smoothed in order to avoid trend reversals from 2016 to future years (e.g. share decreasing from 2016 to 2021 and then increasing monotonically)

- An automated process which identifies and adjusts projections in SA3s where initial projections of employment growth are insufficient to service forecasted population growth.
    - This applies to six population serving industries
    - A threshold population to employment ratio (ERP/Emp) is computed in M06c_1. SA3s are adjusted such that this ratio (for growth) in forecasted years is not exceeded.
        - For example, consider an SA3 which is projected to have a significant increase in population but limited growth in Retail Trade. This will result in an ERP/Emp which exceeds the defined threshold, leading the module to adjust the distribution of Retail Trade employment such that the ratio falls to the threshold level
- Refinement of industry distribution (across SA3s) based on professional judgement which should be informed by various presented factors such as:
    - Known developments and growth in employment (i.e. New Developments Database)
    - Forecasted population growth
    - Regions breaking away from recent trends

Projections can then be reviewed through an examination of employment density (projected employment per hectare of employment zoned land) across SA3s. This will provide a high-level representation of how SA3s are expected to evolve relative to their quantity of employment land. This component is an area which will benefit from future development, as improved GIS zoning layers become available (there are currently zoning gaps).

Inputs

Key inputs:

- GMA employment by industry projected from M06a
- ERP projections by SA3 from M01
- ABS Census data at the SA3 level for employment from 1996-2016
- Parameters defining minimum ERP/Emp ratio (M06c_1)
- Concordance of industries to one of three categories (commercial, industrial and other)
- Known estimates of future employment by travel zone (i.e. New Developments Database)

Following this, a range of inputs which will reflect professional judgement based adjustments are required:

- Select industry to which distributional adjustments are being made
    - View ‘Third Cut Summary Charts’ sheet in a separate window while making adjustments
- For SA3s which require an adjustment, input a share adjustment
- Select whether to lock the projection of SA3s. This will prevent the SA3 from bearing any of changes when projections are scaled to match GMA control totals. SA3 projections could be locked if:

- Current projections are accurate and should not vary when subsequent changes are made to other SA3s (e.g. manual adjustments have been made to align an SA3 with known developments)
- Adjustments that are made should not affect certain other SA3s

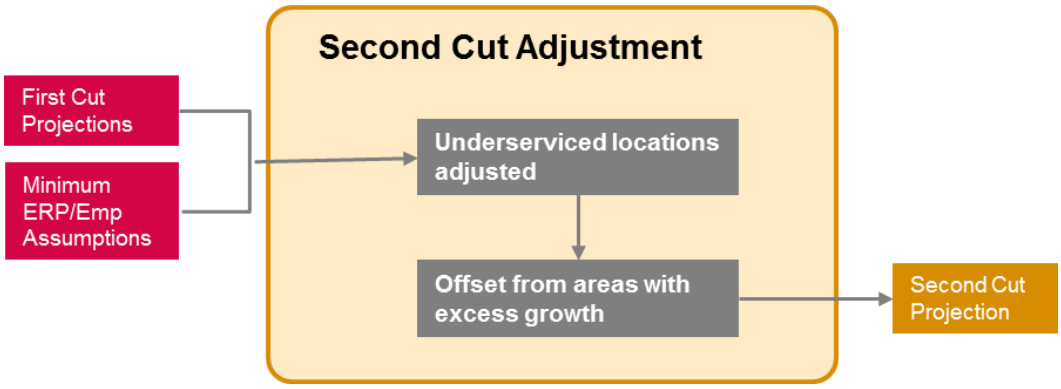
Functions

- The ‘Preliminary Share Projections’ sheet estimates share of employment by industry captured by each SA3 using a log-log regression model. These are then applied to the GMA industry control totals to generate a first cut projection.
- The ‘Second Cut Adjustments’ sheet adjusts the first cut projection to reflect expected population growth in each SA3. For several population serving industries, if the SA3 is projected to have insufficient employment growth, employment will be allocated to that SA3 such that a minimum level of service is provided to these additional residents. To ensure industry totals are maintained, these changes are offset from regions which have a relatively high level of employment growth relative to population growth (see Figure 33).
- The ‘Third Cut Adjustments’ sheet (which should be used while simultaneously viewing the ‘Third Cut Adjustment Charts’ sheet), enables the user to make professional judgement based adjustments to the distribution of employment by industry. Key components of this are:
    - The New_Dev_Flag column indicates SA3s that have a cumulative projection of employment growth which may not fully account for known new developments (i.e. the currently projected employment growth is below or not significantly higher than the growth implied by the ‘New Developments’ database).
    - Manual share adjustments can be specified to adjust the distribution of employment across SA3s (which will propagate to future years once made).
    - Any adjustments are offset against all SA3s in order to maintain industry control totals. The 'Level Lock' input provides the option to prevent selected SA3s from bearing any of the offset (e.g. if a certain SA3 already has a reasonable projection, it should be locked so that it does not bear any of the offset changed which stem from adjustments to other SA3s).
- A review of workforce self-containment should then be conducted on the ‘Employment to Workforce Summary’ sheet.
    - For each SA4 (and constituent SA3), this sheet displays the projected employment/workforce ratio as well as the four fastest growing industries. While different regions will have different levels of self-containment, significant changes to the ratio within a region (SA4) should be examined (as this has the potential to significantly alter traffic flows within the STM).
    - This sheet should only be reviewed after the first round of adjustments have been added to the ‘Third Cut Adjustments’ sheet (e.g. trend smoothing or accounting for known developments)
    - The implied employment densities of each SA3 can then be reviewed on the ‘Final SA3 Employment Densities’ sheet. This provides a summary of employment by broad category (commercial, industrial and total).

Employment density review may or may not be a useful review mechanism at the SA3 level. In some cases it should provide insight regarding how a region is projected to develop compared to others (e.g. projections imply that region 'X' will be as dense in 2021 as region 'Y' was in 2011). However, the incomplete zoning layer published by DPE and differences in the proportion of employment lands across SA3s can make comparisons inappropriate in other cases (e.g. the Sydney – Inner City SA3 has far more commercially zoned land than the Eastern Suburbs – South SA3, making it appear less dense).

Examining employment density will be more useful at the TZ level, and since this dataset would have to be developed for that, we've included it in M06_c as well.

Figure 29 Module 6c: Second-cut projection approach



Third Cut Adjustments Examples

When making manual adjustments, the user should have both the ‘Third Cut Adjustments’ sheet and ‘Third Cut Adjustments Charts’ sheets open side by side.

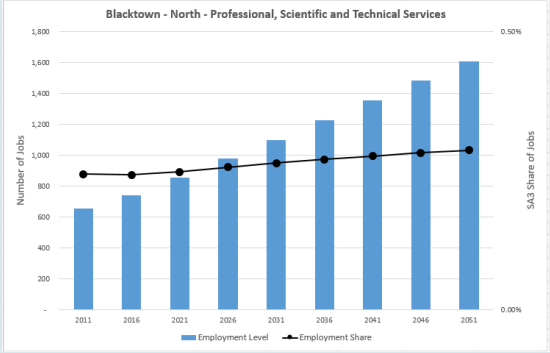
Case 1 – New development indicator

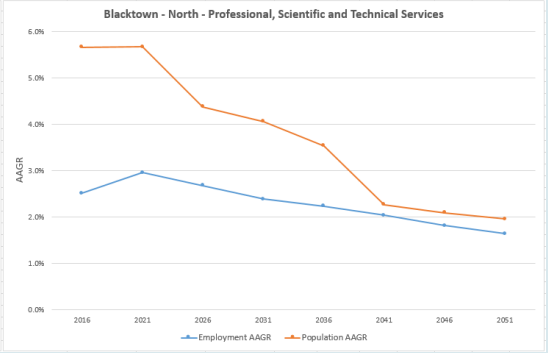
Consider the Professional, Scientific and Technical Services industry in Blacktown - North.

The figures below show both the ‘Third Cut Adjustments’ sheet (above) and ‘Third Cut Adjustments Charts’ sheet (below). The red circle highlights that the expected development occurring in this SA3 will generate significantly more employment than is projected in the second cut (i.e. 650 compared to 202). This is expected in Greenfield regions; as historical data cannot predict future growth.

Final Emp Level 2011	Final Share 2011	Final Emp Level 2016	Final Cumulative Growth 2016	Final Share 2016	Adjusted Second Cut Share 2021	Emp Level 2021	New_Dev Flag 2021	Cumulative New_Dev Growth 2021	Share adjustment 2021	Level Lock 2021	Final Emp Level 2021	Final Cumulative Growth 2021	Final Share 2021
3,557	1.32%	4,331	774	1.42%	1.43%	4,949	0	-			4,949	1,392	1.43%
2,026	0.75%	2,368	342	0.78%	0.79%	2,734	0	420			2,734	708	0.79%
1,314	0.49%	1,530	216	0.50%	0.52%	1,782	0	125			1,782	468	0.52%
1,702	0.63%	1,926	224	0.63%	0.63%	2,171	0	-			2,171	470	0.63%
1,318	0.49%	1,516	198	0.50%	0.51%	1,779	0	-			1,779	461	0.51%
1,030	0.38%	1,166	136	0.38%	0.38%	1,302	1	300			1,302	271	0.38%
7	0.00%	8	1	0.00%	0.00%	9	0	-			9	2	0.00%
915	0.34%	1,097	182	0.36%	0.37%	1,272	0	-			1,272	357	0.37%
4,018	1.50%	4,311	293	1.41%	1.39%	4,820	0	-			4,820	802	1.39%
2,164	0.81%	2,449	285	0.80%	0.80%	2,761	0	-			2,761	597	0.80%
916	0.34%	1,117	201	0.37%	0.38%	1,304	0	-			1,304	388	0.38%
9,343	3.48%	10,574	1,231	3.46%	3.45%	11,923	0	1,800			11,923	2,580	3.45%
1,688	0.63%	1,976	288	0.65%	0.65%	2,251	0	-			2,251	563	0.65%
1,252	0.47%	1,417	165	0.46%	0.47%	1,613	0	-			1,613	361	0.47%
7,562	2.82%	8,843	1,281	2.90%	3.05%	10,535	1	3,670			10,535	2,973	3.05%
805	0.30%	922	117	0.30%	0.30%	1,052	0	-			1,052	247	0.30%
286	0.11%	342	56	0.11%	0.11%	393	0	-			393	107	0.11%
433	0.16%	337	(96)	0.11%	0.10%	347	1	800			347	(87)	0.10%
1,663	0.62%	1,847	184	0.60%	0.58%	2,018	0	-			2,018	354	0.58%
654	0.24%	740	86	0.24%	0.25%	856	1	650			856	202	0.25%
768	0.29%	817	50	0.27%	0.26%	916	1	600			916	148	0.26%
2,125	0.79%	2,405	280	0.79%	0.78%	2,712	0	100			2,712	587	0.78%

Blacktown - North - Professional, Scientific and Techn	2011	2016	2021	2026	2031	2036	2041	2046	2051
Final Emp Level	654	740	856	977	1,098	1,226	1,356	1,483	1,608
Final Share	0.2%	0.2%	0.2%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%
Employment AAGR		2.5%	2.9%	2.7%	2.4%	2.2%	2.0%	1.8%	1.6%
Population AAGR		5.7%	5.7%	4.4%	4.1%	3.5%	2.3%	2.1%	2.0%



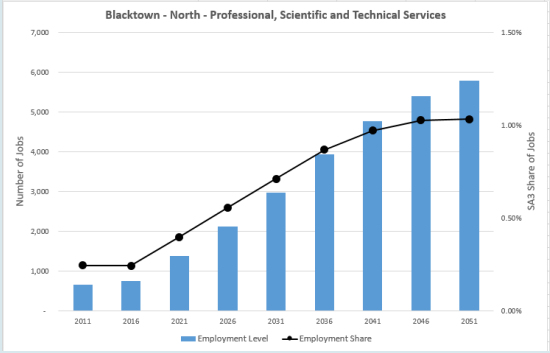


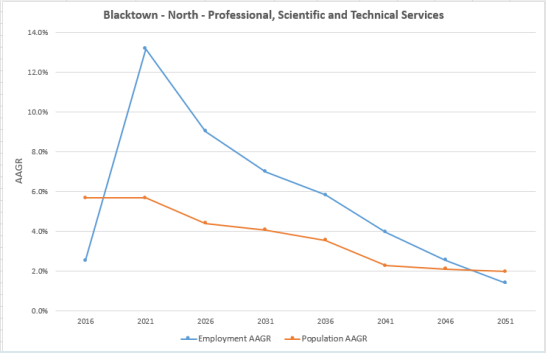
.

Inputting a manual adjustment (circled blue below) alters the share of industry employment that Blacktown – North captures such that it is sufficient to account for the new developments input (circled red below). Entering '1' in the 'Level Lock' column ensures that this projection will not change should changes be made to other SA3s. Comparing the 'Third Cut Adjustments Charts' sheet after the adjustment to prior the adjustment shows the employment share has now increased significantly in this SA3.

Final Emp Level	Final Share	Final Emp Level	Final Cumulative Growth	Final Share	Adjusted Second Cut Share	Emp Level	New_Dev Flag	Cumulative New_Dev Growth	Share adjustment	Level Lock	Final Emp Level	Final Cumulative Growth	Final Share
2011	2011	2016	2016	2016	2021	2021	2021	2021	2021	2021	2021	2021	2021
3,557	1.32%	4,331	774	1.42%	1.43%	4,949	0	-			4,942	1,385	1.43%
2,026	0.75%	2,368	342	0.78%	0.79%	2,734	0	420			2,730	704	0.79%
1,314	0.49%	1,530	216	0.50%	0.52%	1,782	0	125			1,779	465	0.51%
1,702	0.63%	1,926	224	0.63%	0.63%	2,171	0	-			2,168	467	0.63%
1,318	0.49%	1,516	198	0.50%	0.51%	1,779	0	-			1,776	458	0.51%
1,030	0.38%	1,166	136	0.38%	0.38%	1,302	1	300			1,300	269	0.38%
7	0.00%	8	1	0.00%	0.00%	9	0	-			9	2	0.00%
915	0.34%	1,097	182	0.36%	0.37%	1,272	0	-			1,270	355	0.37%
4,018	1.50%	4,311	293	1.41%	1.39%	4,820	0	-			4,812	795	1.39%
2,164	0.81%	2,449	285	0.80%	0.80%	2,761	0	-			2,757	593	0.80%
916	0.34%	1,117	201	0.37%	0.38%	1,304	0	-			1,302	386	0.38%
9,343	3.48%	10,574	1,231	3.46%	3.45%	11,923	0	1,800			11,905	2,562	3.44%
1,688	0.63%	1,976	288	0.65%	0.65%	2,251	0	-			2,248	560	0.65%
1,252	0.47%	1,417	165	0.46%	0.47%	1,613	0	-			1,611	359	0.47%
7,562	2.82%	8,843	1,281	2.90%	3.05%	10,535	1	3,670			10,520	2,958	3.04%
805	0.30%	922	117	0.30%	0.30%	1,052	0	-			1,050	245	0.30%
286	0.11%	342	56	0.11%	0.11%	393	0	-			392	106	0.11%
433	0.16%	337	(96)	0.11%	0.10%	347	1	800			346	(87)	0.10%
1,663	0.62%	1,847	184	0.60%	0.58%	2,018	0				2,014	351	0.58%
654	0.24%	740	86	0.24%	0.25%	856	0	650	0.15%	1	1,374	720	0.40%
768	0.29%	817	50	0.27%	0.26%	916	1	600			915	147	0.26%
2,125	0.79%	2,405	280	0.79%	0.78%	2,712	0	100			2,708	583	0.78%

Blacktown - North - Professional, Scientific and Techn	2011	2016	2021	2026	2031	2036	2041	2046	2051
Final Emp Level	654	740	1,374	2,117	2,966	3,934	4,771	5,405	5,787
Final Share	0.2%	0.2%	0.4%	0.6%	0.7%	0.9%	1.0%	1.0%	1.0%
Employment AAGR		2.5%	13.2%	9.0%	7.0%	5.8%	3.9%	2.5%	1.4%
Population AAGR		5.7%	5.7%	4.4%	4.1%	3.5%	2.3%	2.1%	2.0%





The 'Match New_Dev' function will enable the above adjustments to be made across all SA3s and years simultaneously. The 'Match New_Dev (Current row)' function applies the adjustments to all years for the selected SA3

Industry	Public Administration	Save Adjustments	Match New_Dev
New_Dev Threshold	100%	Clear Adjustments	Match new_Dev (Current row

Case 2 – Employment share smoothing (previous TZP2016 example)

This particular issue could occur if the preliminary share estimate for 2016 (calculated via regression) aligns poorly with the observed 2011 share. Consider the Professional Services industry in the Penrith SA3. The figure below shows a large jump in the employment share between 2011 and 2016 (1). 2016 also displays a significantly higher growth rate compared to other periods (2). An adjustment to the 2016 share of employment will nullify this issue (as the correction will propagate to future years).

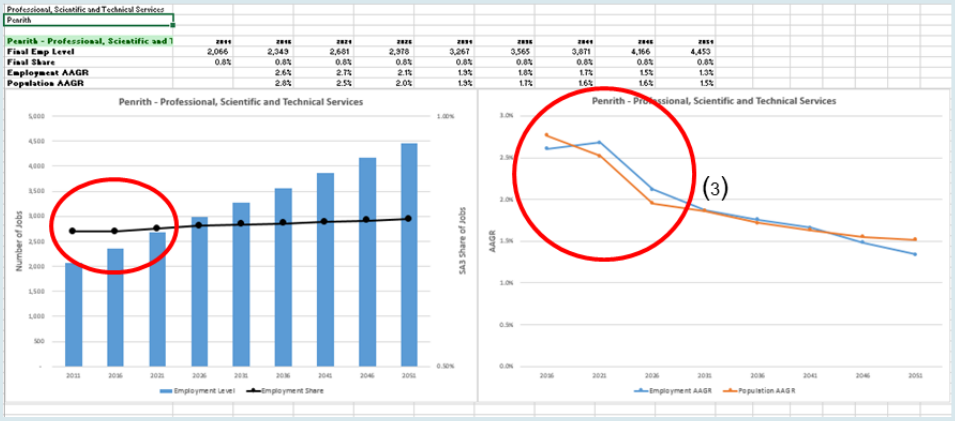
Final Emp Level	Final Share	Adjusted Second Cut Share	Emp Level	New_Dev Flag	New_Dev Level	Share adjustment	Adjusted Emp Level	Level Lock	Final Emp Level	Final Share
2011	2011	2016	2016	2016	2016	2016	2016	2016	2016	2016
3,557	1.32%	1.42%	4,331	0	0		4,331		4,331	1.42%
2,026	0.75%	0.78%	2,368	0	245		2,368		2,368	0.78%
1,314	0.49%	0.50%	1,530	0	75		1,530		1,530	0.50%
1,702	0.63%	0.63%	1,926	0	0		1,926		1,926	0.63%
1,318	0.49%	0.50%	1,516	0	0		1,516		1,516	0.50%
1,030	0.38%	0.38%	1,166	0	50		1,166		1,166	0.38%
7	0.00%	0.00%	8	0	0		8		8	0.00%
915	0.34%	0.36%	1,097	0	0		1,097		1,097	0.36%
4,018	1.50%	1.41%	4,311	0	0		4,311		4,311	1.41%
2,164	0.81%	0.80%	2,449	0	0		2,449		2,449	0.80%
916	0.34%	0.37%	1,117	0	0		1,117		1,117	0.37%
9,343	3.48%	3.46%	10,574	0	850		10,574		10,574	3.46%
1,688	0.63%	0.65%	1,976	0	0		1,976		1,976	0.65%
1,252	0.47%	0.46%	1,417	0	0		1,417		1,417	0.46%
7,562	2.82%	2.90%	8,843	0	300		8,843		8,843	2.90%
805	0.30%	0.30%	922	0	0		922		922	0.30%
286	0.11%	0.11%	342	0	0		342		342	0.11%
433	0.16%	0.11%	337	1	320		337	1	337	0.11%
1,663	0.62%	0.60%	1,847	0	0		1,847		1,847	0.60%
654	0.24%	0.24%	740	0	30		740		740	0.24%
768	0.29%	0.27%	817	0	100		817		817	0.27%
2,125	0.79%	0.79%	2,405	0	0		2,405		2,405	0.79%
1,513	0.56%	0.56%	1,713	0	80		1,713		1,713	0.56%
94,931	35.35%	35.61%	108,746	0	6515		108,746		108,746	35.61%
5,381	2.00%	2.08%	6,353	0	0		6,353		6,353	2.08%
2,645	0.99%	1.01%	3,086	0	0		3,086		3,086	1.01%
2,259	0.84%	0.80%	2,442	0	0		2,442		2,442	0.80%
1,263	0.47%	0.47%	1,430	0	0		1,430		1,430	0.47%
2,037	0.76%	0.75%	2,306	0	0		2,306		2,306	0.75%
2,095	0.78%	0.78%	2,372	0	90		2,372		2,372	0.78%
3,106	1.16%	1.15%	3,515	0	0		3,515		3,515	1.15%
3,117	1.16%	1.16%	3,528	0	0		3,528		3,528	1.16%
3,709	1.38%	1.37%	4,198	0	0		4,198		4,198	1.37%
16,536	6.16%	5.92%	18,086	0	0		18,086		18,086	5.92%
2,389	0.89%	0.88%	2,691	0	0		2,691		2,691	0.88%
4,772	1.78%	1.77%	5,401	0	0		5,401		5,401	1.77%
22,686	8.45%	8.39%	25,623	0	0		25,623		25,623	8.39%
1,560	0.58%	0.58%	1,766	0	0		1,766		1,766	0.58%
2,514	0.94%	0.93%	2,845	0	0		2,845		2,845	0.93%
5,650	2.10%	2.04%	6,221	0	0		6,221		6,221	2.04%
954	0.36%	0.39%	1,182	0	0		1,182		1,182	0.39%
1,656	0.62%	0.61%	1,874	0	0		1,874		1,874	0.61%
537	0.20%	0.24%	728	0	0		728		728	0.24%
1,480	0.55%	0.58%	1,764	0	0		1,764		1,764	0.58%
-	0.00%	0.00%	6	0	0		6		6	0.00%
2,066	0.77%	0.86%	2,632	0	40		2,632		2,632	0.86%
708	0.26%	0.26%	802	0	0		802		802	0.26%

(3)

Inputting a share adjustment (1) to set the final share equal to the first forecasted share, smooths the projection (2) and results in a more realistic series of employment growth rates (3).

Final Emp Level	Final Share	Adjusted Second Cut Share	Emp Level	New_Dev Flag	New_Dev Level	Share adjustment	Adjusted Emp Level	Level Lock	Final Emp Level	Final Share
2011	2011	2016	2016	2016	2016	2016	2016	2016	2016	2016
3,557	1.32%	1.42%	4,331	0	0		4,331		4,335	1.42%
2,026	0.75%	0.78%	2,368	0	245		2,368		2,370	0.78%
1,314	0.49%	0.50%	1,530	0	75		1,530		1,531	0.50%
1,702	0.63%	0.63%	1,926	0	0		1,926		1,928	0.63%
1,318	0.49%	0.50%	1,516	0	0		1,516		1,517	0.50%
1,030	0.38%	0.38%	1,166	0	50		1,166		1,167	0.38%
7	0.00%	0.00%	8	0	0		8		8	0.00%
915	0.34%	0.36%	1,097	0	0		1,097		1,098	0.36%
4,018	1.50%	1.41%	4,311	0	0		4,311		4,315	1.41%
2,164	0.81%	0.80%	2,449	0	0		2,449		2,451	0.80%
916	0.34%	0.37%	1,117	0	0		1,117		1,118	0.37%
9,343	3.48%	3.46%	10,574	0	850		10,574		10,584	3.47%
1,688	0.63%	0.65%	1,976	0	0		1,976		1,978	0.65%
1,252	0.47%	0.46%	1,417	0	0		1,417		1,418	0.46%
7,562	2.82%	2.90%	8,843	0	300		8,843		8,851	2.90%
805	0.30%	0.30%	922	0	0		922		923	0.30%
286	0.11%	0.11%	342	0	0		342		342	0.11%
433	0.16%	0.11%	337	1	320		337	1	337	0.11%
1,663	0.62%	0.60%	1,847	0	0		1,847		1,849	0.61%
654	0.24%	0.24%	740	0	30		740		741	0.24%
768	0.29%	0.27%	817	0	100		817		818	0.27%
2,125	0.79%	0.79%	2,405	0	0		2,405		2,408	0.79%
1,513	0.56%	0.56%	1,713	0	80		1,713		1,714	0.56%
94,931	35.35%	35.61%	108,746	0	6515		108,746		108,848	35.64%
5,381	2.00%	2.08%	6,353	0	0		6,353		6,359	2.08%
2,645	0.99%	1.01%	3,086	0	0		3,086		3,088	1.01%
2,259	0.84%	0.80%	2,442	0	0		2,442		2,444	0.80%
1,263	0.47%	0.47%	1,430	0	0		1,430		1,431	0.47%
2,037	0.76%	0.75%	2,306	0	0		2,306		2,308	0.76%
2,095	0.78%	0.78%	2,372	0	90		2,372		2,374	0.78%
3,106	1.16%	1.15%	3,515	0	0		3,515		3,519	1.15%
3,117	1.16%	1.16%	3,528	0	0		3,528		3,531	1.16%
3,709	1.38%	1.37%	4,198	0	0		4,198		4,202	1.38%
16,536	6.16%	5.92%	18,086	0	0		18,086		18,103	5.93%
2,389	0.89%	0.88%	2,691	0	0		2,691		2,693	0.88%
4,772	1.78%	1.77%	5,401	0	0		5,401		5,406	1.77%
22,686	8.45%	8.39%	25,623	0	0		25,623		25,647	8.40%
1,560	0.58%	0.58%	1,766	0	0		1,766		1,768	0.58%
2,514	0.94%	0.93%	2,845	0	0		2,845		2,848	0.93%
5,650	2.10%	2.04%	6,221	0	0		6,221		6,227	2.04%
954	0.36%	0.39%	1,182	0	0		1,182		1,183	0.39%
1,656	0.62%	0.61%	1,874	0	0		1,874		1,876	0.61%
537	0.20%	0.24%	728	0	0		728		729	0.24%
1,480	0.55%	0.58%	1,764	0	0		1,764		1,766	0.58%
-	0.00%	0.00%	6	0	0		6		6	0.00%
2,066	0.77%	0.86%	2,632	0	40	-0.09%	2,349	1	2,349	0.77%
708	0.26%	0.26%	802	0	0		802		803	0.26%

(1)



(2)

(3)

11.2.3.1 M06c Sub-modules

The two sub-modules of M06 prepare inputs to the primary module. However, these have been identified as areas which could benefit from further development, and have therefore been created separately.

11.2.3.2 M06c_1

One input required for M06c is the minimum level of employment required to service population growth for selected industries (input defined as a maximum ERP to employment ratio). M06c_1 examines historical employment and population data to determine this ratio. This is currently achieved by making an assumption regarding the percentage of SA3s which currently provide an appropriate level of service. For example, if the assumption is that 95% of SA3s provide an adequate level of service (for population serving industries), it implies that 5% do not, and the calculated maximum ERP/employment ratio will be defined as the lowest ratio among this 5% (note, a higher ratio indicates a lower level of service).

Key inputs:

- GMA employment by industry projected from M06a
- ERP projections by SA3 from M01
- ABS Census data at the SA3 level for employment by industry

11.2.3.2.1 M06c_2

A final assessment of projections in M06c involves reviewing the employment density of SA3s. This compares employment projections to the area of currently zoned employment land. M06c_2 computes this area of employment land by both travel zone and SA3.

Key inputs:

- Area overlap of travel zones by planning zone obtained via GIS analysis

11.2.4 Module 6d – TZ Industry Projections

Module 6d completes the distribution of employment by allocating the SA3 industry control totals to a travel zone level. The base distribution is defined by census data (for the base year only). This distribution is then augmented by information on planned developments (i.e. the New Developments database) and professional judgement. The projection process has two primary components:

- Use census data and the New Developments database to compute the base distribution of employment by industry within each SA3
- The data combination of the two inputs acts as an attractor variable which is used to distribute SA3 employment across the constituent travel zones.
- Previously defined manual calibrations can be retained if changes are made to the base shares (e.g. New Developments input is updated)
- Apply manual calibrations to override projections at a travel zone level

Charts presenting the growth of employment (total and by industry) can then be used to review the projections. Additionally, the employment density of travel zones will indicate locations in which too much growth is projected (e.g. highly developed existing CBD zones).

Further details on the employment disaggregation process are provided in Appendix E.

Inputs

Key Inputs:

- ABS Census data at the travel zone level (2016)
- Estimate of available land by travel zone (M06c_2)

- SA3 employment by industry projections from M06c
- Known estimates of future employment by travel zone (i.e. New Developments Database)

Following this, a range of inputs which will reflect professional judgement based adjustments are required:

- On 'TZ Summary' sheet, select SA3 to which travel zone calibrations are to be applied
    - View 'Manual Calibrations' sheet in a separate window
- On the 'Manual Calibrations' sheet, enter a distribution override for desired industries
- Select whether to lock projections of each industry. This will prevent that travel zone from bearing any changes (caused when projections are scaled to SA3 control totals) should other zones be subsequently adjusted.
    - E.g. If industry 'A' in zone 'x' is locked, any changes to industry 'A' in zone 'y' will not result in a redistribution towards/away from zone 'x'.
- Enter a note which describes the motivation for manual changes

Functions

- The 'Base Shares' sheet combines census and New Developments data to create an attractor variable which is used to distribute employment by industry control totals across the constituent travel zones of each SA3.
    - For each period, the final shares of the previous period are used to calculate a first cut projected for each zone
    - For zones which have a base projection lower than the growth implied by 'New Developments' data, the projection is adjusted, the SA3 rebalanced to control totals and new shares computed.
- The 'TZ Summary' sheet presents the currently projected total employment for all travel zones within the selected SA3. Growth rates and density (jobs/hectare) are also shown, and should be used for review.
- The 'Manual Calibrations' sheet will display the employment projections, by industry, for the travel zone that is currently the active cell on the 'TZ Summary' sheet.

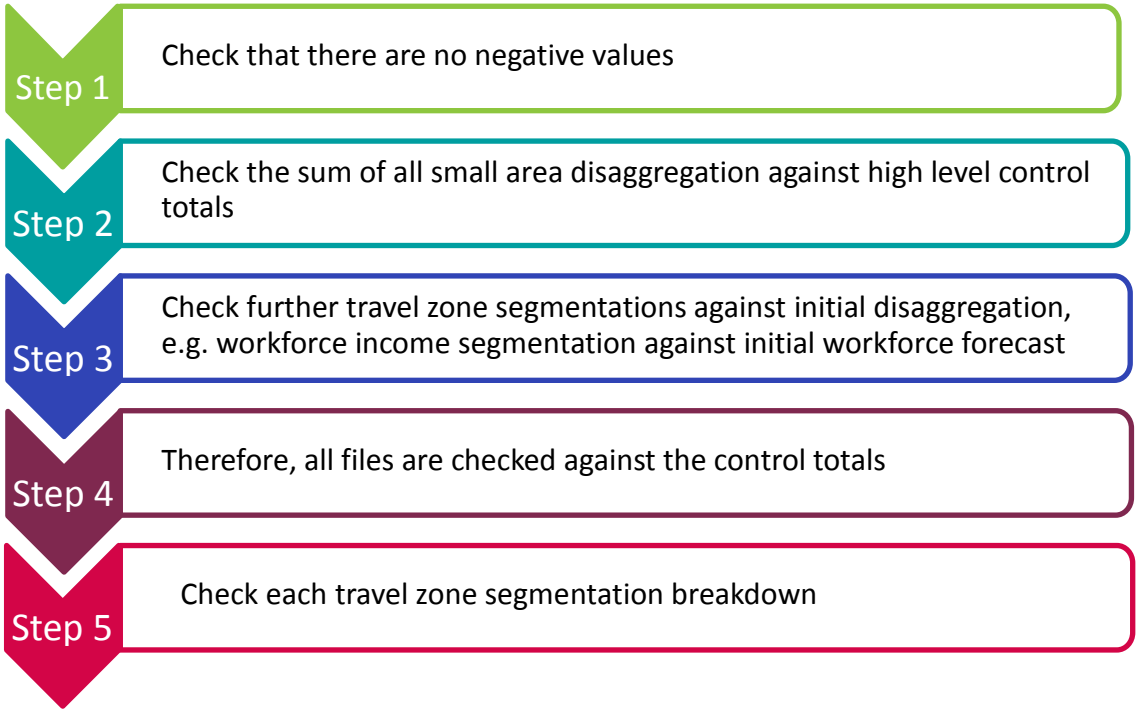
Share override adjustments, based on professional judgement, can be entered here. Once calibrations are saved, projections will balance to SA3 control totals

12 Module 7: Checks and Validation

12.1 Overview

The quality checks module is the last step in the model. This module is an R Script that takes the output from previous modules and performs a number of automated calculations to evaluate quality and consistency within and between the outputs of the various modules. The checks module has three main functions, the negatives checks, control totals checks and validations checks. The negatives check determines whether output files contain any negative numbers. The control totals check which determines whether the DPE control totals for the Sydney GMA area are consistent with aggregated figures in the output files. The validations check further checks for consistency between different sub-geographies.

12.2 The Overall Logic of the Order of the Checks



The checks module is configured to be run from start to finish in its entirety. Each module output file along with global functions is loaded at the start of the module run and then required in each section of the script. The checks module takes an average of two minutes to run.

12.3 Negatives Check

This step of the module checks whether there are any negative numbers in any of 23 tables in the module output files folder listed below.

0.1 Historical Structural Private Dwellings_TZ.csv

0.2 Historical ERP_TZ.csv

1.1 Structural Private Dwellings_TZ.csv

1.2 Occupied Private Dwellings_TZ.csv

1.3 Population Occupied Private Dwellings_TZ.csv

1.4 Population Non Private Dwellings_TZ.csv

1.5 ERP_SA3.csv

2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv

2.2 Age Groups (15-64+) _TZ.csv

2.3 Age by Sex 5 Year Age Groups (15-64+)_SA4.csv

3.1 Household Type_TZ.csv

4.1 Primary and Secondary School students PUR_TZ.csv

4.2 Primary and Secondary School students POI_TZ.csv

4.3 Tertiary Students PUR_TZ.csv

4.4 Tertiary Students POI_TZ.csv

5.1 Workforce_TZ.csv

5.2 Workforce Part Time Full Time_SA4.csv

5.3 Workforce Part Time Full Time_TZ.csv

5.4 Workforce Income Segmentation_TZ.csv

5.5 ERP_NPD EMP_GMA.csv

5.6 Workforce_Employed_SA3.csv

6.1 Employment 2-digit Forecast_TZ.csv

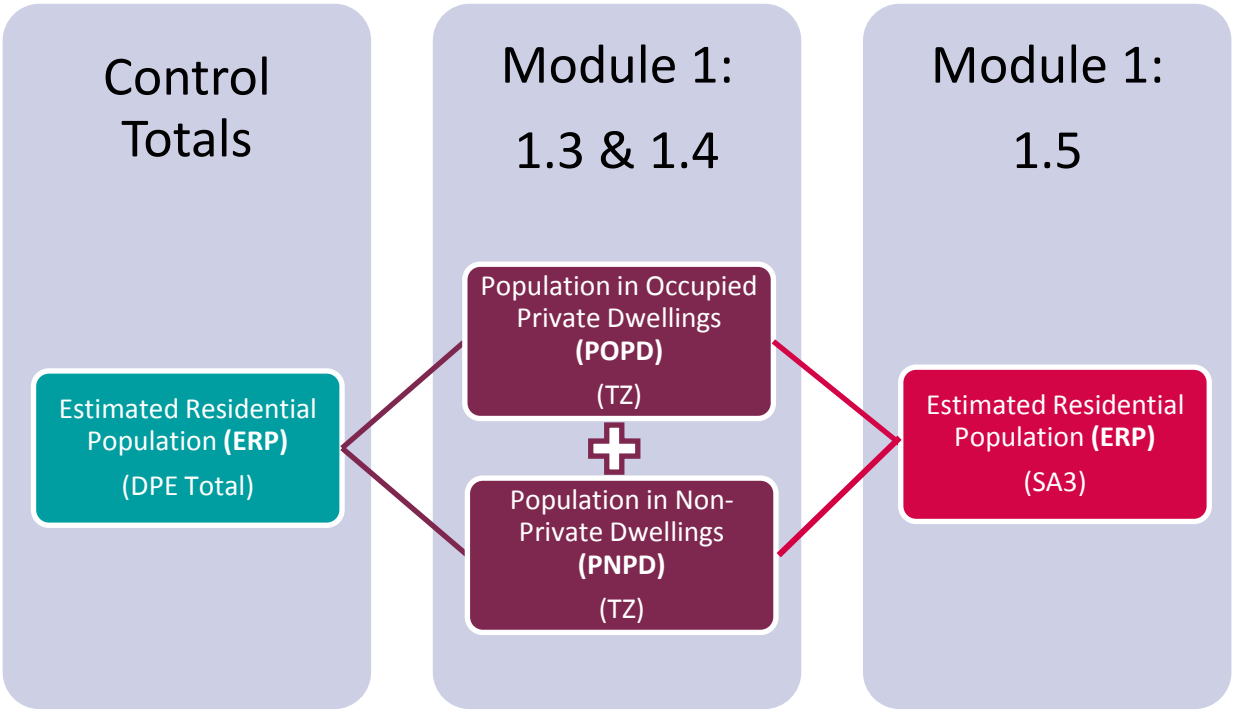
6.2 Employment SA3 Forecast_SA3.csv

12.4 Control Totals

The control totals step of the module checks that figures in the module output files are consistent with the DPE control totals for the Sydney GMA. This validation also checks for consistency between output files of different modules. For example, the workforce for any given area (Module 5) is equivalent to the resident population aged between 15 and 65 (Module 2). This check in this module sum all geographic areas for each file and only check for consistency between the aggregate figures, not for different sub-areas. Each test is outlined below.

12.4.1 Module 1 Population and Dwellings

Figure 30 Graphical representation of control total check for Module 1 output files



This section of the module checks that ERP reconciles to both POPD and PNPD

0.0 ERP (Total) = 1.3 POPD + 1.4 PNPD (Travel Zones)

It then checks whether the numbers in the POPD and PNPD files are equivalent to ERP in the SA3 file.

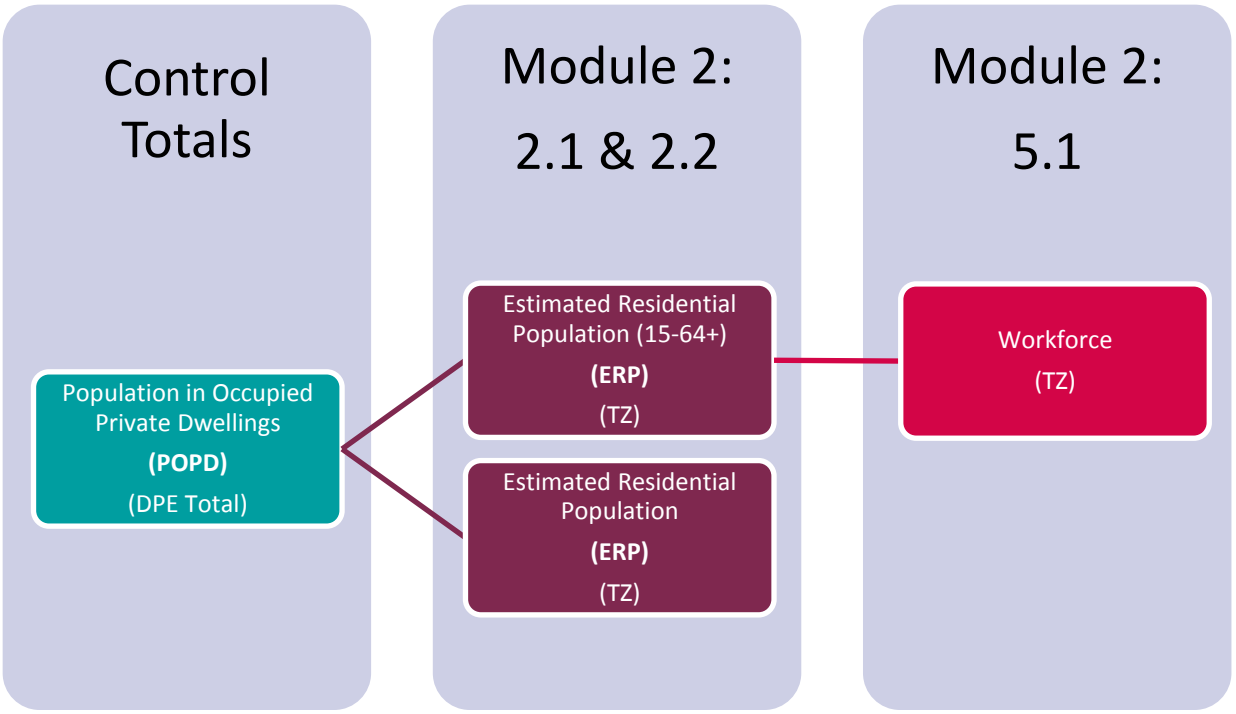
1.3 POPD + 1.4 PNPD (Travel Zones) = 1.5 ERP (SA3)

By comparing the data to a common data set, we have implicitly checked that the estimated residential population on an SA3 level reconciles to the totals given by the Department of Planning and Environment (DPE Total).

The final check for Module 1 requires input from Module 3 and is described in section 12.4.3 below.

12.4.2 Module 2 Age-Sex

Figure 31 Graphical representation of control total check for Module 2 output files



Output files 2.1 and 2.2 for Module 2 are checked against the DPE totals for the population in occupied private dwellings, for the relevant age groups.

0.0 POPD (Total) = Sum of 2.1 ERP (Travel Zones)

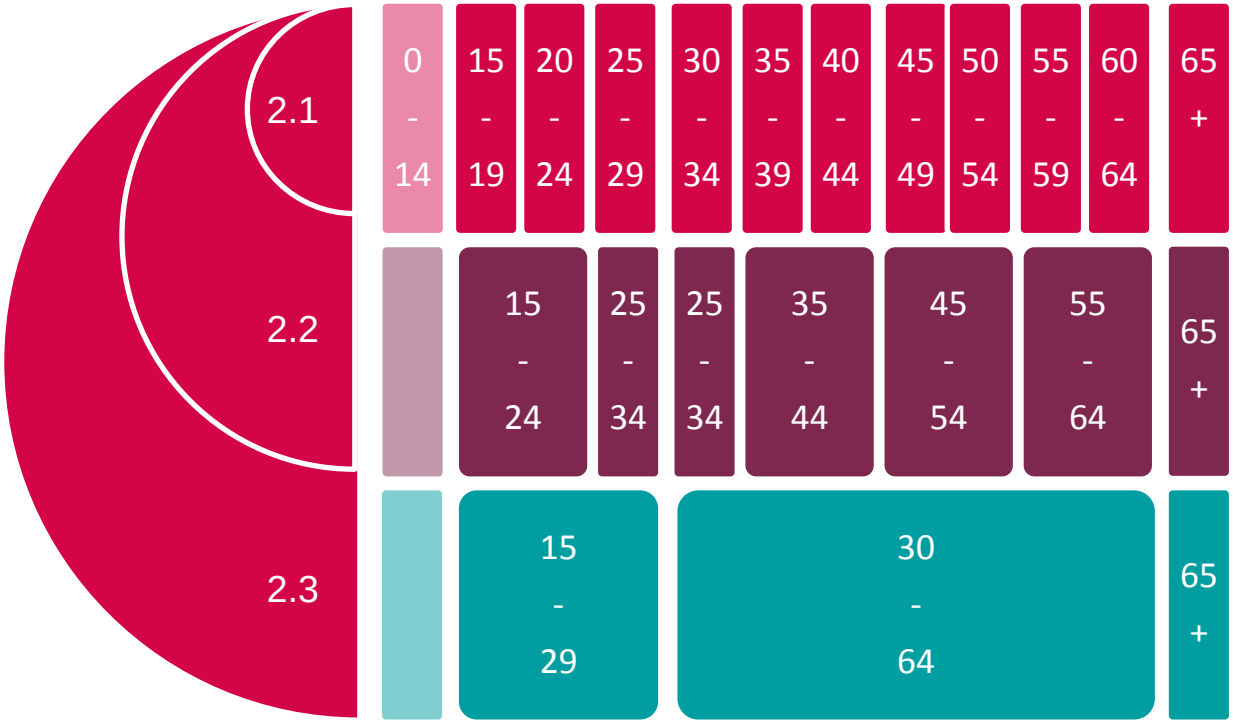
POPD 15+ (Total) = Sum of 2.2 ERP (Age 15+) (Travel Zones)

Output file 2.2 for Module 2 is then checked against the workforce figures in output file 5.1 for Module 5.

5.1 Workforce (Employed + Unemployed + Not looking for work) = 2.1 ERP (Age 15+) (Travel Zones)

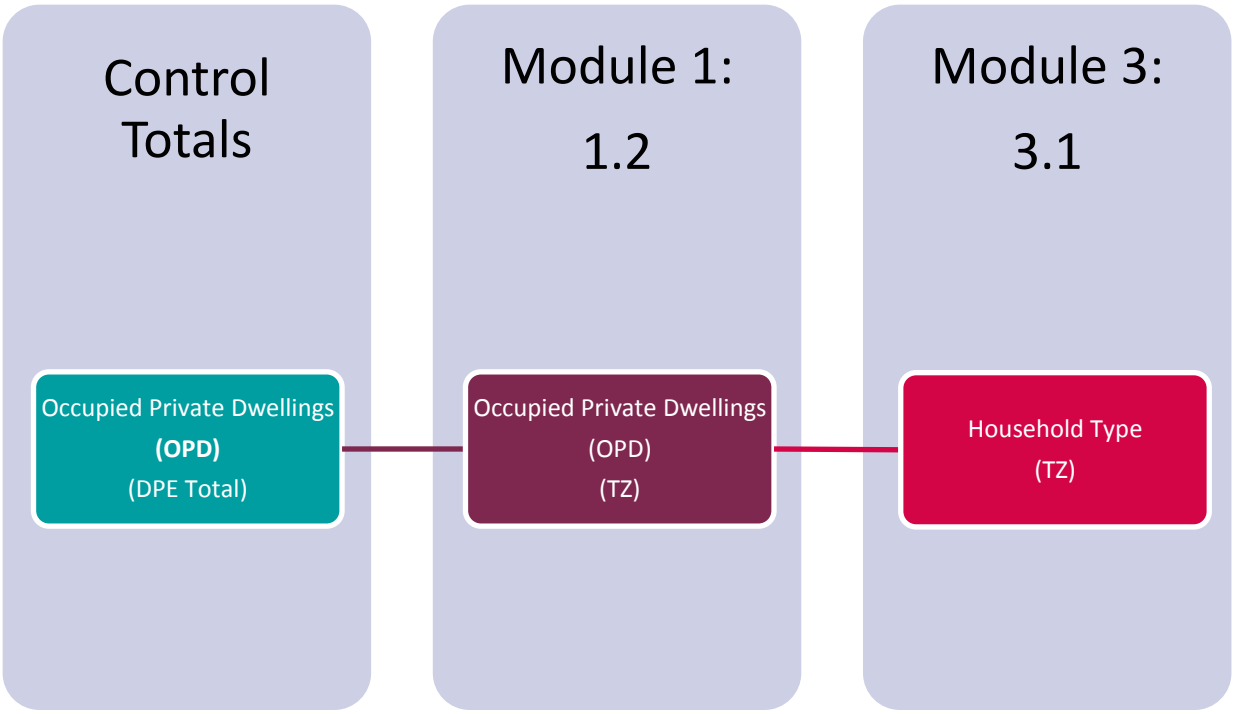
Checks are also completed between Module 2 output files, to check that the sum of the 5 year age brackets in output file 2.1 is equivalent to each of the aggregated age bracket output files 2.2 and 2.3 as outlined in Figure 32 below:

Figure 32 ERP age categories in Module 2 output files 2.1, 2.2 and 2.3



12.4.3 Module 3 Household Type

Figure 33 Graphical representation of control total check for Module 3 output file



Occupied Private Dwellings are sourced from the DPE totals. Output file 1.2 is checked against the DPE control total.

0.0 OPD (Total) = 1.2 OPD (Travel Zones)

Output file 3.1 is checked for equivalency against file 1.2.

3.1 Household Type (With children (single and couple) + No children (single and couple)) (Travel Zones) = 1.2 OPD (Travel Zones)

This also completes the checks for the Module 1 output files.

12.4.4 Module 4 Enrolments

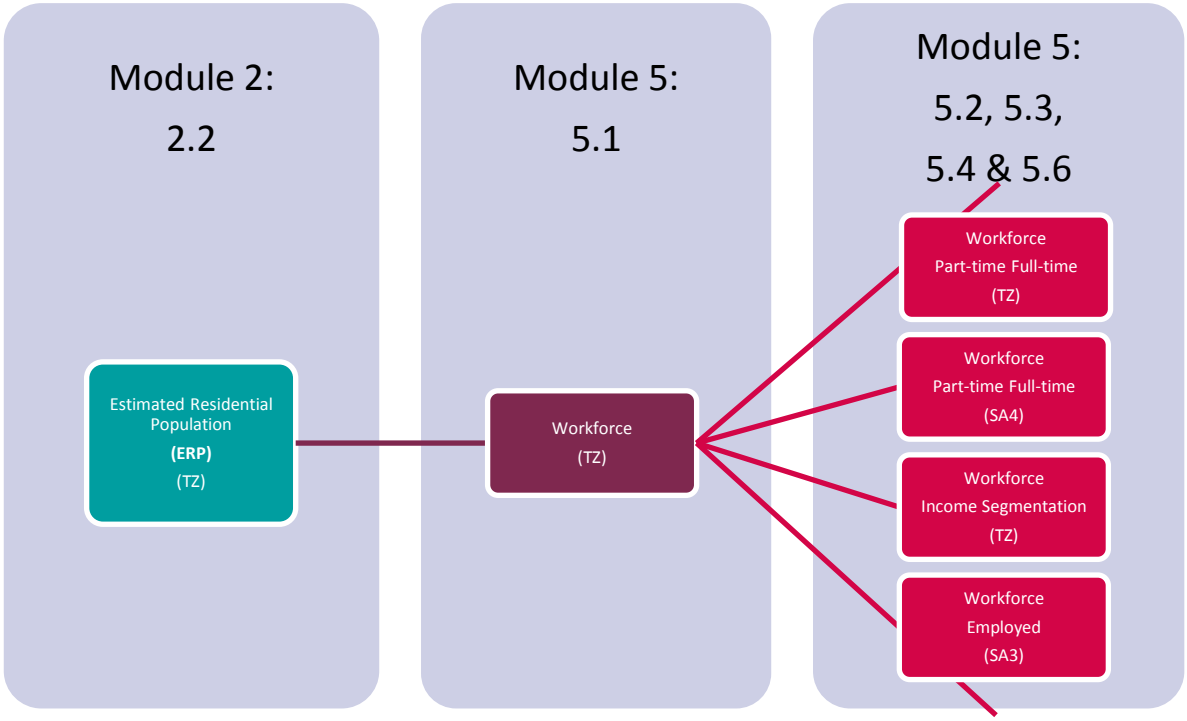
Module 4 is independent of any other Module. It cannot be checked against any DPE control total or population groups in the output files of any other modules. The output files of Module 4 are checked for internal consistency such that the total number of students within the Sydney GMA.

4.1 Primary and secondary school students at place of usual residence (Total) = 4.2 Primary and secondary school students at place of institution (Total)

4.3 Tertiary students at place of usual residence (Total) =
4.4 Tertiary students at place of institution (Total)

Module 5

Figure 34 Graphical representation of control total check for Module 5 output files



Module 5 output file 5.1 is checked for consistency with the ERP figures in Module 2 output file 2.2.

2.2 ERP (Ages 15-64+) (Travel Zones) = 5.1 Workforce (Employed + Unemployed + Not looking for work) (Travel Zones)

The remaining Module 5 output files are then checked for consistency against 5.1.

5.1 Workforce (Travel Zones) = 5.2 Workforce (Part-time + Full-time) (SA4)

5.1 Workforce (Travel Zones)= 5.3 Workforce (Part-time + Full-time) (Travel Zone)

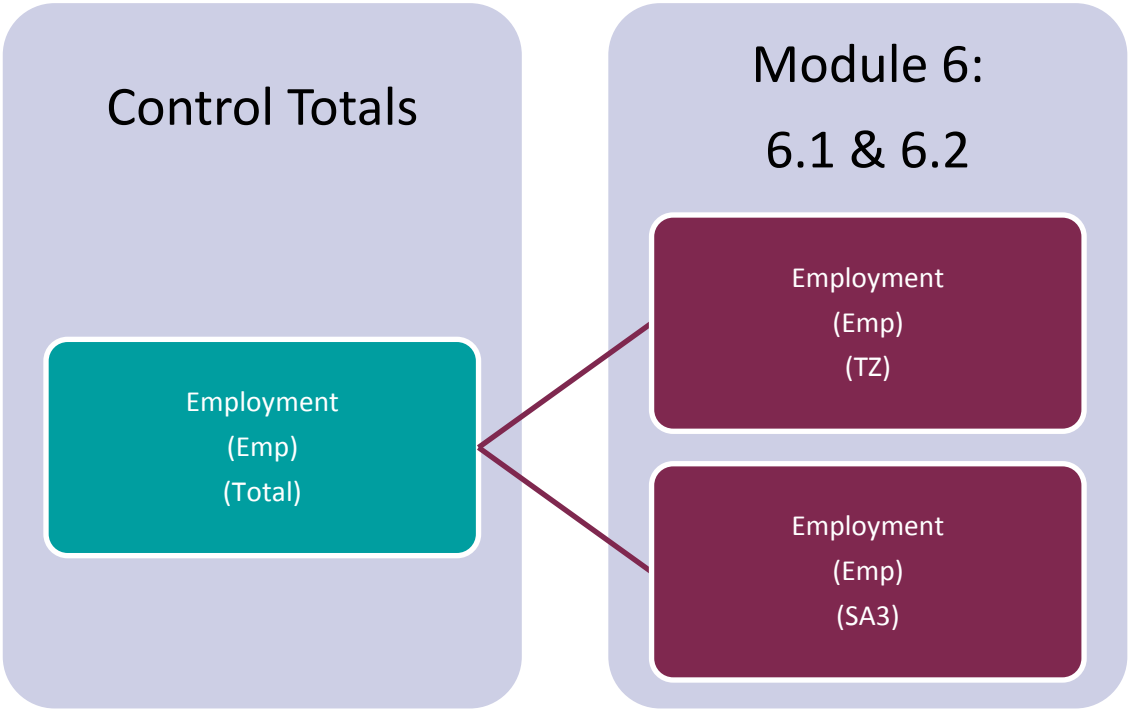
5.1 Workforce (Travel Zones)= 5.4 Workforce (sum of income brackets)

5.1 Workforce (Travel Zones) = 5.6 Workforce (Employed) (SA3)

12.4.5 Module 6 Employment by Industry

Module 6 is self-contained, and independent of any other module.

Figure 35 Graphical representation of control total check for Module 6 output files



6.1 and 6.2 are checked against the control totals.

0.0 Emp (Total) = 6.1 Emp (sum of industry sectors) (Travel Zones)

0.0 Emp (Total) = 6.2 Emp (sum of industry sectors) (Travel Zones)

This completed the checks for Module 6.

12.5 Validations

The validations step of the module conducts further consistency checks between figures in module output files, e.g. check that the number of occupied dwellings does not exceed the number of people in each travel zone. Unlike the control totals step outlined above, this step of the module checks for consistency in reported numbers between corresponding geographic sub-areas (travel zones, SA3 and SA4). The validations follow the same general order as the control totals checks.

Table 11 Description of Validation Checks

Module 1 – Population and Dwellings	
1.1 OPD (TZ) / 1.2 SPD (TZ) <= 1	Checks that the number of occupied dwellings (OPD) does not exceed the total dwelling stock (SPD) in the model for each travel zone.
1.3 POPD (TZ) / 1.2 OPD (TZ) >= 1	Checks that the number of people in occupied dwellings (POPD) exceeds the number of private occupied dwellings in the model for each travel zone.
1.3 POPD (TZ) + 1.4 PNPD (TZ) = 1.5 ERP (SA3)	Checks that the population assigned to private and non-private dwellings in travel zones is equivalent to the estimated resident population in the corresponding SA3.
Module 2 – Age-Sex	
2.1 ERP (Sum of 15-64+) (TZ) = 2.2 ERP (15-64+) (TZ)	Checks that the aggregate of the five year categories in output file 2.1 is equivalent to the pre-aggregated groups in output file 2.1 for each travel zone.
2.1 ERP (Sum of 15-29) (TZ) = 2.2 ERP (15-29) (TZ)	
2.1 ERP (Sum of 30-64) (TZ) = 2.2 ERP (30-64) (TZ)	
2.1 Age by Sex (Sum of 64+) (TZ) = 2.2 Age by Sex (64+) (TZ)	
2.1 ERP (15+) (TZ) = 2.3 ERP (15+) (SA4)	Checks that the 15+ population reported at the SA4 level is equivalent to the aggregate of corresponding travel zones.
2.2 ERP (TZ)= 0.4 ERP (LGA)	Checks that the population reported at the LGA level is equivalent to the aggregate of corresponding travel zones.
Module 3 – Household Type	
3.1 Household Type (with children (Single and Couple) + No children (Single and Couple)) (TZ)= 1.2 OPD (TZ)	Checks that the aggregate of household types with and without children is equivalent to the number of occupied private dwellings in the model for each travel zone.
Module 4 - Enrolments *No further relevant checks as origin and destination geographies differ*	

Module 5 – Workforce Segmentation	
5.1 Workforce (Employed + Unemployed + Not looking for work) (TZ)= 2.2 ERP (Ages 15-64+) (TZ)	Checks that workforce figures are equivalent the population aged over 15 years for each travel zone.
5.1 Workforce (Employed) (TZ)= 5.4 Employees (sum of all income brackets) (TZ)	Checks that the workforce figures are equivalent to the aggregate employees across all income categories for each travel zone.
5.1 Workforce (Employed) (TZ) = 5.2 Employees (Part-time + Full-time) (SA4)	Checks that the workforce figures for travel zones are equivalent to the aggregate of part-time and full-time employees at the corresponding SA4 level.
5.1 Workforce (Employed) (TZ) = 5.3 Employees (Part-time + Full-time) (TZ)	Checks that the workforce figures are equivalent to the aggregate of part-time and full-time employees for each travel zone.
5.1 Workforce (TZ) = 5.6 Workforce (SA3)	Checks that the workforce reported at the SA3 level is equivalent to the aggregate of the workforce for the corresponding travel zones.
Module 6 *No further relevant checks*	

Part 3 – Redistribution Models

13 Introduction

13.1 Background

The land use redistribution models known as Module 08 and Module 09 were developed for generating land use scenario inputs for transport modelling in the STM. The tool has been developed as an extension of TZP2016 v1.51 and inherently adopts the assumptions made in producing these travel zone projections.

13.2 Assumptions

The key assumptions for each model are outlined below:

13.2.1 Module 08: Population Redistribution Model

- The base distribution of population by travel zones is calculated from the current TZP2016 v1.51 figures for POPD. Unless additional information is used, the base case land use assumptions will inherently inform the new scenario outputs.
- The control total forecast for population will remain consistent with the DPE projections. This means assumptions around the natural rate of increase and net migration are inherently consistent with the DPE projections. The implication for the modelling redistribution assumes where there is an increase in persons in a particular area, there must be an equal subtraction elsewhere.

13.2.2 Module 09: Employment Redistribution Model

- The base distribution of employment by travel zones is calculated off the current Travel Zone Projection 2016 v1.51 figures for employment. Unless additional information is used, the base case land use assumptions will inherently inform the new scenario outputs.
- The control total forecast for employment will remain consistent with the DPE projections – total jobs as well as industry mix at the Broad Industry Categories (optional user choice)[8].
- The redistribution of employment to the 33 ANZSIC industries is based on the forecast shares of employment at the travel zone level. This implies the ANZSIC mix of industry is consistent to the base forecast.

[8] See the Travel Zone Projection Technical Guide

14 Redistribution Models

14.1 Objective

The objective of the redistribution models is to allow the user to generate alternative land use scenarios for both population and employment. This is required due to the static nature of forecasting – where each land use scenario in the future is based on the realisation of selected assumptions.

At a high level, the TZP2016 v1.51 dataset produced by Transport for NSW is influenced primarily by Census data, the Department of Planning and Environment population forecasts, NSW State Government policy (e.g. committed greenfield or renewal precincts) and known key infrastructure developments that impact the growth of population and employment[9].

As State Government policy and infrastructure planning considers projects which are constantly evolving, the assumptions around developments that inform the Travel Zone Projections quickly become outdated. New infrastructure projects, which are not considered as part of the base case Travel Zone Projections, require a tailored approach of modelling through the STM. The redistribution models enable the development of new population and employment forecasts which accurately reflect the effects of these infrastructure projects.

Module 08 and 09 allow the user to create scenarios quickly for both population and employment. The models rely on three key factors:

- The user knowing where the change areas are (defined by travel zone),
- The net change in these areas; and
- The base year for redistribution

[9] For a more detailed explanation of inputs, assumptions and methodology, please refer to the Travel Zone Projections 2016 v1.51 Technical Guide.

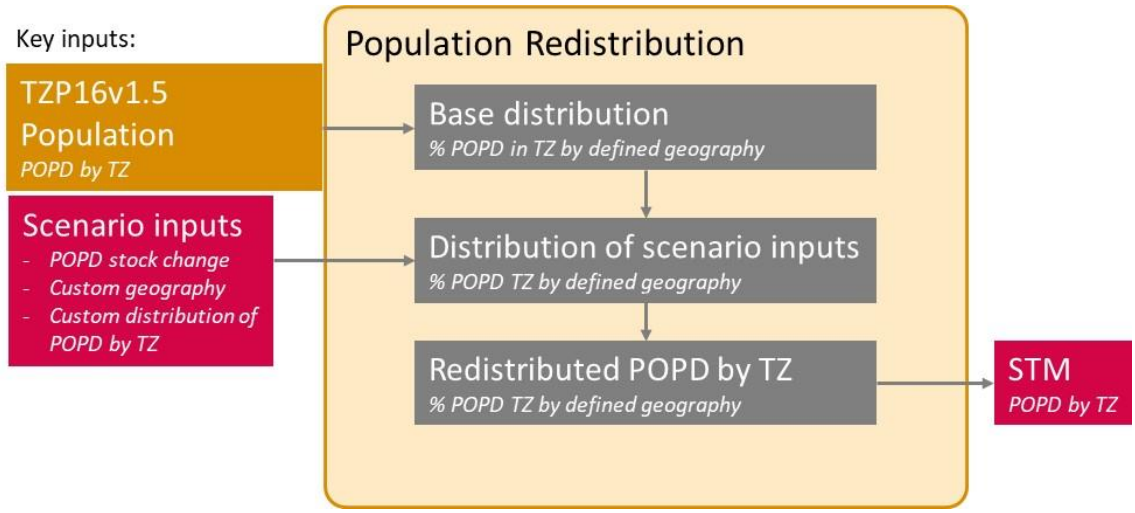
15 Module 08: Population Redistribution Model

The Population Redistribution Model allows the user to develop a land use scenario and produce the subsequent redistribution while keeping total population consistent with the baseline forecasts for the GMA. The model has been built to allow the user to define 'Custom Geographies' by travel zones which allows for greater user control during the development of the scenario, providing the ability to define where the population will be increasing/decreasing relative to the baseline forecasts. This model produces the final redistribution by Population in Occupied Private Dwellings.

15.1 Model Overview

The redistribution method requires the user to know the **net change** between the current land use forecast numbers to the scenario[10]. The following diagram represents an overview of the Population Redistribution Model.

Figure 1. Population Redistribution Model flowchart



The model by default, calculates the distribution of population by travel zone using the TZP2016 v1.51 POPD. This redistribution is calculated using the **growth** between the project year and a user specified base year (for example the growth between 2016 (base year) – 2036(project year)). This default distribution can be adjusted by the user if the distribution by travel zone is known for the project. It is recommended to choose the closest date to the commencement of the project as the base year. This will ensure that the most appropriate TZP2016 v1.51 assumptions influence the redistribution[11].

---

10 For example – TZP2016 v 1.51 POPD forecast for 2036 in Travel Zone A is 4,500. The scenario forecasts a total of 5,000 POPD in Travel Zone A. As an input, the model requires the total number of people that are expected to increase/decrease for the specified geography. In this the example, the user would input "+500" persons in Travel Zone A.

[11] Consider a project which results in land use changes in 2031 and 2036, and that requires an STM evaluation at 2036 only. The most appropriate base year choice will be 2026. Within the scenario-defined change areas, the travel zone base distribution will therefore reflect TZP2016 v1.51 changes that occur between 2026 and 2036 (Additionally. for an area subject to a change in land use, the most significantly affected travel zones from the redistribution will be those which TZP2016 v1.51 forecast to grow most strongly from 2026 – 2036).

15.2 Model run through

15.2.1 Step 1: Define your geography

Firstly, select the 'Input Concordance' sheet. This requires the user to define the change areas the scenario is aiming to target. Depending on the granularity of input provided, the user can select a relevant geography. For example, where adjustments to the population are provided at the LGA level, the user can redistribute at the LGA level. Where a hybrid of geography levels is required, the user can select the Custom Geography input.

The user has the following options:

- Statistical Area 2
    - For example, Ashfield SA2 to increase by 1,000 persons due to additional infrastructure development within this particular SA2. This increase will then have to be offset across other travel zones, defined by their respective SA2 geography.
- Local Government Area
    - For example, Burwood LGA to increase by 1,000 persons due to additional infrastructure development within the Local Government Area. However, the information on the specific locations of the additional persons is not known. Therefore the 1,000 additional persons will be distributed across the entire LGA. There will also need to be a corresponding decrease of 1,000 across other travel zones, defined by their respective LGA geography.
- Precinct (GSC Centres)
    - For example, Condell Park to increase by 1,000 persons due to additional infrastructure development the defined GSC centre. There will also need to be a corresponding decrease of 1,000 across other travel zones, defined by their respective precinct geography.
- GSC Districts
    - For example, Central District to increase by 100,000 persons due to additional infrastructure development and increased infill development. However, the precise locations on the developments are unknown so the additional persons will be distributed across the District. There will also need to be a corresponding decrease of 100,000 across other travel zones defined by their respective GSC District.
- Custom Geography
    - For example, Travel Zone 5 will increase by 1,000 persons due to additional dwelling development. There will also need to be a corresponding decrease of 1,000 persons across other custom geographies[12].

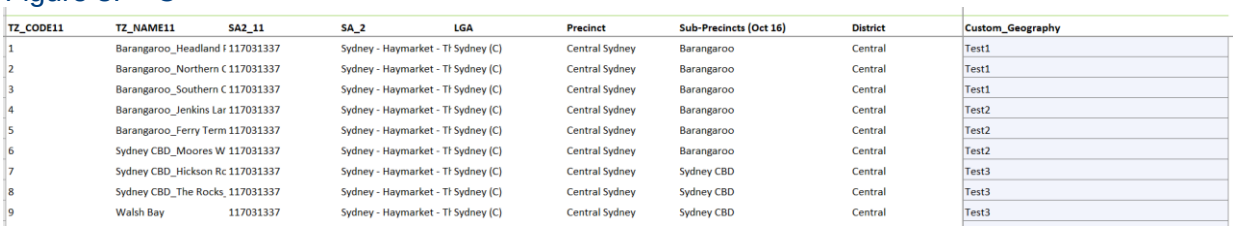
To define a 'Custom Geography', the user assigns a geography name by travel zone in the 'Input Concordance' tab, within the 'Custom Geography' field. The user is

[12] This could be any grouping of travel zones, from a single zone to the remainder of the GMA

required to complete the custom geography fields in the ‘Input Concordance’ tab prior to defining a custom geography in the ‘Input Parameter’ tab (Step 2). See Figure 2**Error! Reference source not found.** as an example. In this example, 3 custom eographies have been developed, defined as Test1, Test2 and Test3 encompassing travel zone 900 to 952.

Figure 2. Custom geography input

Figure 3. C



TZ_CODE11	TZ_NAME11	SA2_11	SA_2	LGA	Precinct	Sub-Precincts (Oct 16)	District	Custom_Geography
1	Barangaroo_Headland	117031337	Sydney - Haymarket - Th	Sydney (C)	Central Sydney	Barangaroo	Central	Test1
2	Barangaroo_Northern C	117031337	Sydney - Haymarket - Th	Sydney (C)	Central Sydney	Barangaroo	Central	Test1
3	Barangaroo_Southern C	117031337	Sydney - Haymarket - Th	Sydney (C)	Central Sydney	Barangaroo	Central	Test1
4	Barangaroo_Jenkins Lar	117031337	Sydney - Haymarket - Th	Sydney (C)	Central Sydney	Barangaroo	Central	Test2
5	Barangaroo_Ferry Term	117031337	Sydney - Haymarket - Th	Sydney (C)	Central Sydney	Barangaroo	Central	Test2
6	Sydney CBD_Moores W	117031337	Sydney - Haymarket - Th	Sydney (C)	Central Sydney	Barangaroo	Central	Test2
7	Sydney CBD_Hickson Rc	117031337	Sydney - Haymarket - Th	Sydney (C)	Central Sydney	Sydney CBD	Central	Test3
8	Sydney CBD_The Rocks_	117031337	Sydney - Haymarket - Th	Sydney (C)	Central Sydney	Sydney CBD	Central	Test3
9	Walsh Bay	117031337	Sydney - Haymarket - Th	Sydney (C)	Central Sydney	Sydney CBD	Central	Test3

ips:

- All cells in this column do not need to be completed – only the areas that will be defined as a change area. If the predefined geographies are more useful, then the user is not required to fill out this column.
- The user must enter **more than one** custom geographical area otherwise an error will occur if only one unique geography is provided in the list. This is because any change must be offset in at least one (or more) other geographies.
- The user must define where the population is to be rebalanced from. Using the example shown in Figure 2, geography Test3 will rebalance any population changes made to Test1 and Test2. See Step 2 for further details.

15.2.2 Step 2: Define the parameters

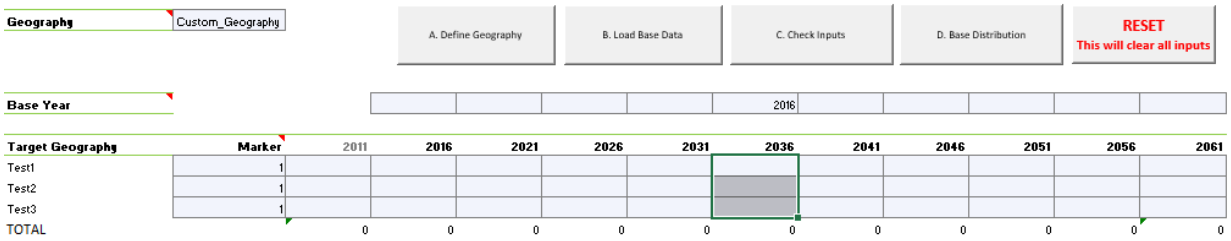
This step requires the user to define the net change of population in the defined change areas. In the ‘Input Parameters’ sheet:

- Click the dropdown menu (Cell reference B9) and choose the geography by which the scenario will be defined.
- Click the button labelled ‘A. Define Geography’. This will create a table with a list of ‘Target Geographies’. The Target Geographies reflect the geography that was selected in the first step of the process. See Figure 3 below for a scenario where three custom geographies have been defined: Test1, Test2, Test3.

For custom geography scenarios, check:

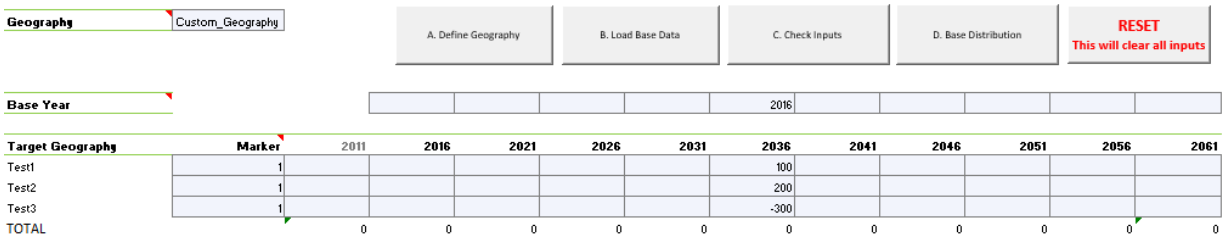
- The same number of travel zones in the customer template are created in the custom geography for the model redistribution; and
- The adjustments for each redistribution year in the ‘rebalanced travel zone’ geography is calculated by a negative sum formula of the adjusted travel zones e.g. =-sum(x1:x5).

Figure3. Defining the ‘Target Geographies’ for the model



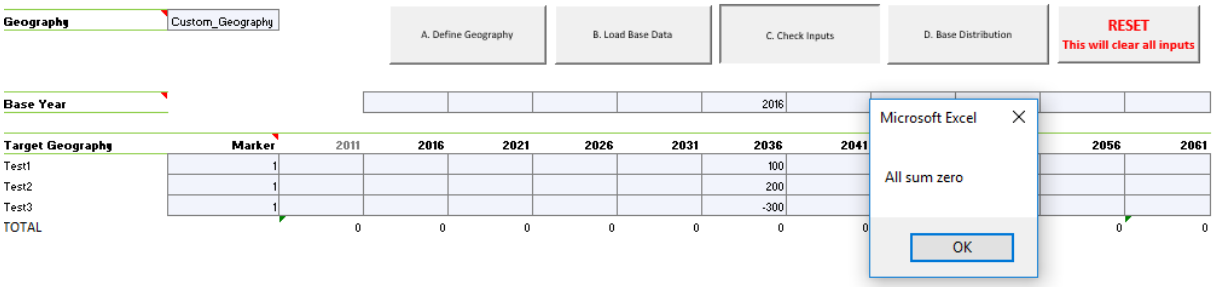
- Choose the base year with which to calculate the growth and click the button ‘B. Load Base Data’. This will load the base data required for the redistribution.
- Mark geographies that are to be adjusted with a 1 in the ‘marker’ column (cell reference B15). For those geographies that are not going to change, leave as blank.
- Input the **net change** by year for each change area. The aggregated change across all areas for the project year/s MUST sum to zero to ensure the GMA control totals reconcile to TZP2016. See Figure 4 as an example where the geographical area Test1 increases by an additional 100 persons, Test2 increases by an additional 200 persons and Test3 decreases by 300 persons.

Figure 4. Input net change by year



- Check that the net changes sum to zero by clicking the ‘C. Check Inputs’ button. If the inputs do not sum to zero, an error box will prompt the user to check the inputs (Figure 5). The user will be required to make amendments such that the net changes sum to zero before continuing the modelling process.

Figure 5. Check inputs



- Click the ‘D. Base Distribution’ button to calculate the base distribution of POPD by travel zone. This may take a few minutes to complete.

From the Input Distribution sheet, check the model is running correctly by Highlighting individual columns M:W:

- To ensure the cell count is equal to the number of Travel Zones (2949); and
- That the sum of each column is equal to the number of geographies used in the redistribution (e.g. based on figure 5, a sum of 300% should be observed).
- Where cells are blank or if the above checks do not reconcile, the user must investigate further before proceeding – cells may need to be filled as 100% to ensure the model runs correctly.

15.2.3 Step 3: Define the distribution of population

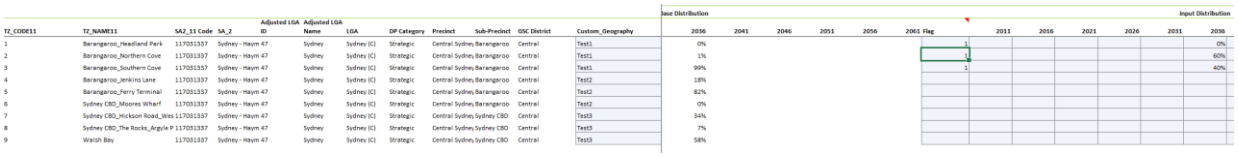
In the 'Input Distribution' tab, the user has the option to alter the base case distribution of population by travel zone, or to define a custom distribution.

To define a custom distribution for a selected geography, enter a value of 1 in the 'Flag' column (Column reference X) and then specify a custom distribution in the respective model year/s (columns "Y" to "AI"). The distribution entered for any geography must sum to 100%. If this does not occur, the redistribution will fail the control total check.

Once the distribution is defined, click the 'E. Redistribute' button. This option should be utilised when there is sufficient information on the distribution of population at the travel zone level.

For example, where there is sufficient information of the distribution within the custom geography such as Test1 which is split over three travel zones and of the additional 100 persons, a 0/60/40 split is known (compared to a 0/1/99 split identified in the default base distribution). See Figure 6. A custom distribution can be input. In this case the user will input 60% into cell.

Figure 6. Define the distribution



TZ_CODE11	TZ_NAME11	SA2_11 Code	SA_2	Adjusted LGA ID	Adjusted LGA Name	LGA	DP Category	Precinct	Sub-Precinct	GSC District	Custom_Geography	Base Distribution 2036	2041	2046	2051	2056	2061	Flag	Input Distribution 2011	2016	2021	2026	2031	2036
1	Barangaroo_Headland Park	117031337	Sydney - Haym 47		Sydney	Sydney (C)	Strategic	Central Sydney	Barangaroo	Central	Test1	0%						1						0%
2	Barangaroo_Northern Cove	117031337	Sydney - Haym 47		Sydney	Sydney (C)	Strategic	Central Sydney	Barangaroo	Central	Test1	1%						1						60%
3	Barangaroo_Southern Cove	117031337	Sydney - Haym 47		Sydney	Sydney (C)	Strategic	Central Sydney	Barangaroo	Central	Test1	99%						1						40%
4	Barangaroo_Jenkins Lane	117031337	Sydney - Haym 47		Sydney	Sydney (C)	Strategic	Central Sydney	Barangaroo	Central	Test2	18%												
5	Barangaroo_Ferry Terminal	117031337	Sydney - Haym 47		Sydney	Sydney (C)	Strategic	Central Sydney	Barangaroo	Central	Test2	82%												
6	Sydney CBD_Moores Wharf	117031337	Sydney - Haym 47		Sydney	Sydney (C)	Strategic	Central Sydney	Barangaroo	Central	Test2	0%												
7	Sydney CBD_Hickson Road_Wes	117031337	Sydney - Haym 47		Sydney	Sydney (C)	Strategic	Central Sydney	Sydney CBD	Central	Test3	34%												
8	Sydney CBD_The Rocks_Argyle P	117031337	Sydney - Haym 47		Sydney	Sydney (C)	Strategic	Central Sydney	Sydney CBD	Central	Test3	7%												
9	Walsh Bay	117031337	Sydney - Haym 47		Sydney	Sydney (C)	Strategic	Central Sydney	Sydney CBD	Central	Test3	58%												

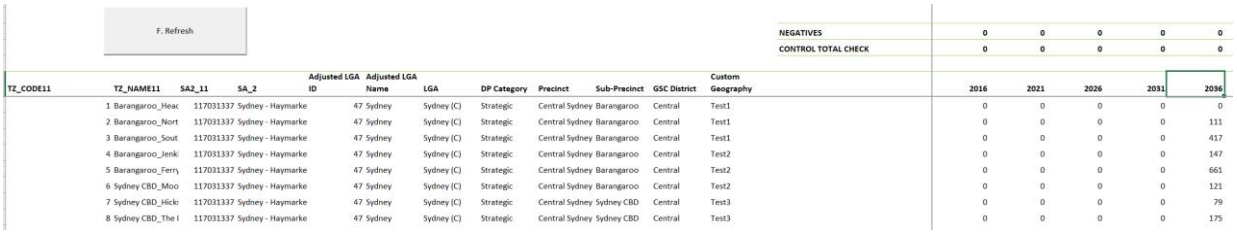
15.2.4 Step 4: Print results

In the 'Output TZ' tab, click to the 'F. Refresh' button (see Figure 7) to print the results. These are the final outputs of the redistribution by travel zone. The same results are generated in the sheet labelled 'Output_TZ_CSV', which is the required output file type for most GIS software.

Check that the 'Negatives' and 'Control total' checks are equal to zero (N8:W9).

For redistributed Travel Zones, compare input data from customer template with the 'Output TZ' tab. A final check should be conducted reconciling Travel Zone adjustments between the output, base case and input parameter spreadsheet.

Figure 7. Print results using the refresh button



Tips:

- The final checks of the ‘Output TZ’ tab in Figure 8 below should return zeros for all years.
- The ‘negatives check’ counts for any Travel Zones that return a **negative** value. If there are negative values, the likely source of the error is due to a larger reduction of total stock in the geography than currently forecast in TZP2016.
    - For example, current stock at 2016 for geographical area ‘X’ is 200 persons. However, the change entered for ‘X’ in the input sheet is -300 persons. This will lead to negative persons in the final redistribution. The user should note this and re-run the model, respecifying the geography the model will use to rebalance the ‘positive’ change in the scenario.
- The control total check is a sum of the final redistribution forecast minus TZP2016 v1.51 for the same year. This should return a zero if the redistribution reconciles to the control total. If a number other than zero is returned, an error has occurred, and inputs should be checked (e.g. aggregate change specified in step 15.2.2). It is important for the control totals to match as the outputs of the modelling from the STM will be compared to the base case. If the control total is not met, then the results will not be comparable to the base case scenario. The user must have justification for not meeting control totals if they are not met (e.g. a large-scale project is expected to increase internal migration to NSW).

Figure 8. Final redistribution check

NEGATIVES	0	0	0	0	0	0	0	0	0	0
CONTROL TOTAL CHECK	0	0	0	0	0	0	0	0	0	0

2011	2016	2021	2026	2031	2036	2041	2046	2051	2056	2061

15.2.5 Resetting the model

The model has a built-in reset button on the ‘Input Parameters’ sheet that clears **all** inputs the user has changed and will essentially return the model to a sanitised state. It is recommended the user use this button if a mistake has been made in the input process. It is also recommended that the user **save** a copy of the model with the following naming convention.

- YYMMDD _Population Redistribution_ProjectXXXXX

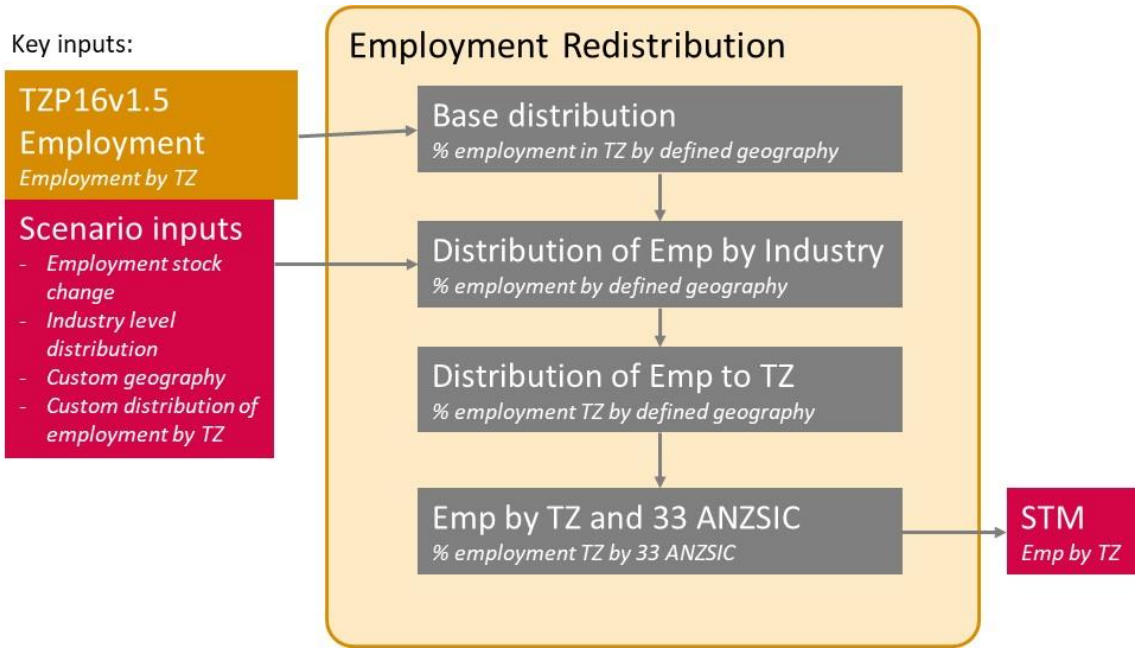
16 Module 09: Employment Redistribution Model

This model allows the user to develop an employment land use scenario and produce a redistribution while keeping total employment consistent with the baseline forecasts for the GMA. It has been built to allow the user to define 'custom geographies' by travel zones. Essentially, this allows for greater control during the development of the scenario, giving the user the ability to define where employment will be increasing/decreasing **relative to the baseline forecasts**. This model produces the final redistribution of employment by the 33 ANZSIC categories used in TZP2016 v1.51.

16.1 Model Overview

The redistribution method requires the user to know the **net change** between the current Land Use forecast numbers to the scenario. The following diagram represents an overview of the Employment Redistribution Model.

Figure 9. Employment Redistribution Model flowchart



The model by default, calculates the distribution of jobs by travel zone using the TZP2016 v1.51 employment at Place of Work. Similar to the population redistribution model, this distribution of employment is calculated using the **growth** between the project year and a user specified base year (for example 2016 – 2036 growth rate). This default distribution can be adjusted by the user if the distribution by travel zone is known for the project. It is recommended to choose the closest date to the commencement of the project as the base year. This will ensure that the most appropriate TZP2016 v1.51 assumptions influence redistribution[13].

[13] Consider a project which results in employment changes in 2031 and 2036, and that requires an STM evaluation at 2036 only. The most appropriate base year choice will be 2026. Within the scenario-defined change areas, the travel zone base distribution will therefore reflect TZP2016 v1.51 changes that occur between 2026 and 2036 (e.g. for an area subject to a negative change, the most significantly affected travel zones will be those which TZP2016 v1.51 forecast to grow most strongly from 2026 – 2036).

16.2 Model run through

16.2.1 Step 1: Define your geography

Firstly, select the 'Input Concordance' sheet. This requires the user to define the change areas the scenario is aiming to target. Depending on the granularity of input provided, the user can select a relevant geography. For example, where adjustments to the jobs are provided at the LGA level, the user can redistribute at the LGA level. Where a hybrid of geography levels is required, the user can select the Custom Geography input.

The user has the following options:

- Statistical Area 2
    - For example, Ashfield SA2 to increase by 1,000 jobs due to additional infrastructure development within this particular SA2. This increase will then have to be offset across other travel zones, defined by their respective SA2 geography.
- Local Government Area
    - For example, Burwood LGA to increase by 1,000 jobs due to additional infrastructure development within the Local Government Area. However, the information on the specific locations of the additional jobs is not known. Therefore the 1,000 additional jobs will be distributed across the entire LGA. There will also need to be a corresponding decrease of 1,000 jobs across other travel zones, defined by their respective LGA geography.
- Precinct (GSC Centres)
    - For example, Condell Park to increase by 1,000 jobs due to additional infrastructure development the defined Greater Sydney Commission centre. There will also need to be a corresponding decrease of 1,000 jobs across other travel zones, defined by their respective precinct geography.
- GSC Districts
    - For example, Central District to increase by 100,000 jobs due to additional infrastructure development and increased infill development. However, the precise locations on the developments are unknown so the additional jobs will be distributed across the District. There will also need to be a corresponding decrease of 100,000 jobs across other travel zones defined by their respective GSC District.
- Custom Geography
    - For example, Travel Zone 5 will increase by 1,000 jobs due to additional employment precinct development. There will also need to be a corresponding decrease of 1,000 jobs across other custom geographies[14].

To define a 'Custom Geography', the user assigns a geography name by travel zone in the 'Input Concordance' tab, within the 'Custom Geography' field. The user is

[14] This could be any grouping of travel zones, from a single zone to the remainder of the GMA

required to complete the custom geography fields in the 'Input Concordance' tab prior to defining a custom geography in the 'Input Parameter' tab (Step 2). See the example presented in Figure 10**Error! Reference source not found.**. In this xample, 3 custom geographies have been developed, defined as Test1, Test2 and Test3 encompassing travel zone 1 to 9.

Figure 10. Custom Geography input

TZ_CODE11	TZ_NAME11	SA2_11	SA2_NAME11	LGA	Precinct	Sub-Precincts (Oct 16)	District	Custom_Geography
1	Barangaroo_Headl	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Barangaroo	Central	Test1
2	Barangaroo_North	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Barangaroo	Central	Test1
3	Barangaroo_South	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Barangaroo	Central	Test2
4	Barangaroo_Jenkii	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Barangaroo	Central	Test2
5	Barangaroo_Ferry	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Barangaroo	Central	Test2
6	Sydney CBD_Moo	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Barangaroo	Central	Test2
7	Sydney CBD_Hicks	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Sydney CBD	Central	Test3
8	Sydney CBD_The F	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Sydney CBD	Central	Test3
9	Walsh Bay	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Sydney CBD	Central	Test3
10	Sydney CBD_Hicks	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Sydney CBD	Central	
11	Sydney CBD_Dawe	117031337	Sydney - Haymarke	Sydney (C)	Central Sydney	Sydney CBD	Central	

Tips:

- All cells in this row do not need to be filled out – only the areas that will be defined as the change area. If the set defined geographies are more useful, then the user is not required to fill out this row.
- It should be noted the user must enter **more than one** custom geographical area. An error will occur if only one unique geography is provided in the list. This is because any change must be offset in at least one (or more) other geographies.

16.2.2 Step 2: Define the parameters

This step requires the user to define the net change of employment in the defined change areas. In the 'Input Parameters' sheet:

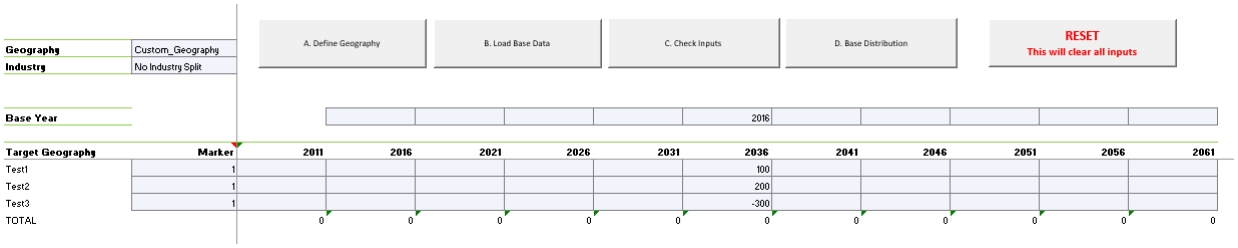
- Click the dropdown menu (Cell reference B9) and choose the geography by which the scenario will be defined.
- Click the dropdown menu (Cell reference B10) and choose one of two modes under which the mode can run. Selecting 'No Industry Split' will model the scenario at total jobs, with the model distributing total jobs to ANZSIC categories at the end of the process[15]. This assumes the Travel Zone Projections distribution of industry will remain consistent for the new scenario. The other option is selecting 'Broad Industry Category'. This will allow the user to model the distribution of employment by the Broad Industry Categories. Essentially this gives more control to user to define what type of job will eventuate from this scenario.
- Click the button labelled 'A. Define Geography'. This will create a table with a list of 'change areas'/'target geographies'. See Figure 11 below for a scenario where three custom geographies have been defined: Test1, Test2, Test3.

For custom geography scenarios, check:

[15] This is the fastest option computationally. It should be used when the focus of the scenario is on total employment rather than industry detail

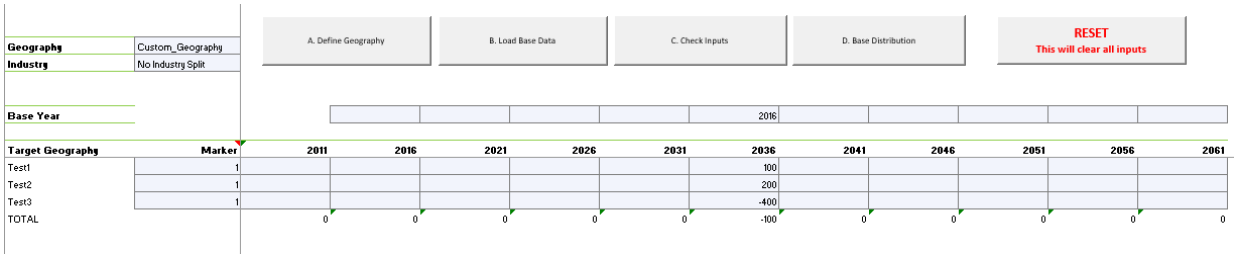
- The same number of travel zones in the customer template are created in the custom geography for the model redistribution; and
- The adjustments for each redistribution year in the ‘rebalanced travel zone’ geography is calculated by a negative sum formula of the adjusted travel zones e.g. =-sum(x1:x5).

Figure 11. Defining the ‘Target Geographies’ for the model



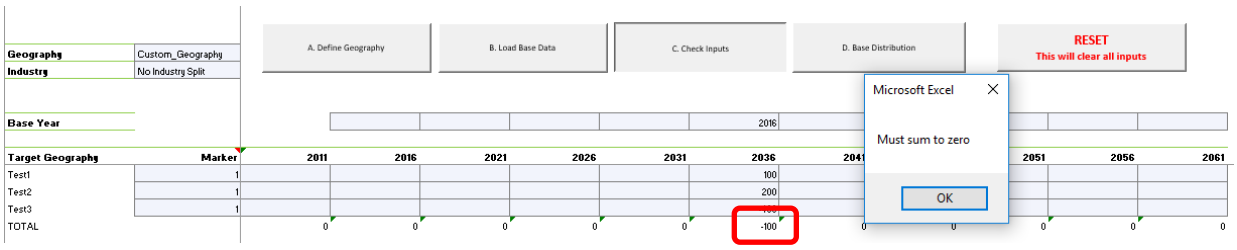
- Choose the base year with which to calculate the growth and click the button ‘B. Load Base Data’. This will load the base data required for the redistribution.
- Mark geographies that are to be adjusted with a ‘1’ in the ‘Marker’ column (cell reference B15). For those geographies that are not going to change, leave as blank.
- Next, input the **net change** by year for each change area. The aggregated change across all areas for a particular year MUST sum to zero to ensure the GMA control totals reconcile to TZP2016. See Figure 12 and Figure 13 below for an example of an incorrect input specification and resulting error message.

Figure 12. Input net change by year



- Check the net changes sum to zero by clicking the ‘C. Check Inputs’ button. If the inputs do not sum to zero, an error box will prompt the user to check the inputs. The user will be required to make amendments such that the net changes sum to zero before continuing the modelling process.

Figure 13. Check inputs.



- Next, click the ‘D. Base Distribution’ button to calculate the base distribution of employment by travel zone.

From the Input Distribution sheet, check model is running correctly by Highlighting individual columns D:M:

- To ensure the cell count is equal to the number of geographies adjusted; and
- That the sum of each column is equal to the number of geographies used in the redistribution (e.g. based on figure 5, a sum of 300% should be observed).
- Where cells are blank or if the above checks do not reconcile, the user must investigate further before proceeding – cells may need to be filled as 100% to ensure the model runs correctly.

16.2.3 Step 3: Define the distribution of employment

In the ‘Input Distribution’ tab, the user can choose between using the base case distribution to distribute the change of employment by industry, or to define a custom industry distribution. The custom distribution is only necessary if the user has chosen to model the scenario at the Broad Industry Category. If ‘No Industry Split’ was chosen, skip the following steps defined in the dot points below and click E ‘Recalibrate’ then ‘F. Distribution’ to redistribute the employment.

- To define a custom industry distribution at the Broad Industry Category for a selected geography, enter a value of 1 in the ‘Flag’ column (Column reference N) and then specify a custom distribution for the model years. Once distribution is defined, click the ‘E. Recalibrate’ button. This option should be utilised when there is sufficient information on the distribution of employment by industry.

- For example, there is sufficient information to know that within the custom geography Test1 that the split of the additional 100 jobs will be 100% within the Broad Industry Category of ‘Population Serving’ jobs. See Figure 14.

- After the model has recalibrated to employment control totals, press the ‘F. Distribution’ button.

Figure 14. Define the industry distribution

		Base Distribution											Input Distribution						
Target Geography	Broad Industry Category	2011	2016	2021	2026	2031	2036	2041	2046	2051	2056	2061	Flag	2011	2016	2021	2026	2031	2036
Test1	1		9%	8%	8%	8%	8%	8%	8%	8%	8%	8%	1						0
Test2	1		88%	88%	88%	89%	89%	89%	89%	90%	90%	90%							
Test3	1		50%	51%	51%	51%	52%	53%	54%	54%	55%	55%							
Test1	2		0%	1%	1%	1%	1%	1%	2%	2%	2%	2%	1						0
Test2	2		3%	3%	3%	3%	3%	3%	3%	3%	3%	3%							
Test3	2		4%	4%	5%	5%	5%	5%	5%	5%	5%	5%							
Test1	3		47%	75%	75%	76%	77%	77%	78%	78%	78%	78%	1						1
Test2	3		7%	7%	7%	7%	7%	7%	6%	6%	6%	6%							
Test3	3		43%	42%	42%	41%	41%	40%	39%	39%	38%	38%							
Test1	4		45%	16%	16%	15%	14%	13%	13%	12%	12%	12%	1						0
Test2	4		2%	2%	2%	2%	1%	1%	1%	1%	1%	1%							
Test3	4		3%	3%	3%	3%	3%	2%	2%	2%	2%	2%							

Tips:

- The distribution entered for any geography must sum to 100%. If this does not occur, the redistribution will fail the control total check.

16.2.4 Step 4: Print results

The final outputs of the redistribution by travel zone will be printed on the ‘Output_TZ’ sheet. The same results are generated in the sheet labelled ‘Output_TZ_CSV’.

Check that the ‘Negatives’ and ‘Control total’ checks are equal to zero (N8:W9).

For redistributed Travel Zones, compare input data from customer template with the ‘Output TZ’ tab. A final check should be conducted reconciling Travel Zone adjustments between the output, base case and input parameter spreadsheet.

Tips:

- The final checks in Figure 15 below should all return zero.

- The negatives check counts for any Travel Zones that return a **negative** value. If there are negative values, the likely source of the error is due to a larger reduction of total employment in the geography than currently forecast in the baseline data.

- For example, current employment at 2016 for geographical area ‘X’ is 200 jobs. However, the change entered for ‘X’ in the input sheet is -300 jobs. This will lead to negative persons in the final redistribution. The user should note this and re-run the model, respecifying the geography the model will use to rebalance the ‘positive’ change in the scenario.

- The control total check is a sum of the final redistribution forecast minus TZP2016 v1.51 for the same year. This should return a zero if the redistribution reconciles to the control total. If a number other than zero is returned, an error has occurred in the redistribution and inputs should be checked (e.g. aggregate change specified in step 4.2.2). It is important for the control totals to match as the outputs of the modelling from the STM will be compared to the base case. If the control total is not met, then the results will not be comparable to the base case scenario. The user must have justification for not meeting control totals if they are not met (e.g. a large-scale project is expected to increase internal migration to NSW).

Figure 15. Final redistribution check

NEGATIVES	0	0	0	0	0	0	0	0	0	0
CONTROL TOTAL CHECK	0	0	0	0	0	0	0	0	0	0
	2011	2016	2021	2026	2031	2036	2041	2046	2051	2056	2061

16.2.5 Resetting the model

The model has a built-in reset button on the ‘Input Parameters’ sheet that clears **all** inputs the user has changed and will essentially return the model to a sanitised state. It is recommended the user use this button if a mistake has been made in the input process. It is also recommended that the user **save** a copy of the model with the following naming convention.

- YYMMDD _Employment Redistribution_ProjectXXXXX

17 Appendix A: New development database

The New Developments Database is a major input into M06c. Figure 23 below shows the projects that are considered in the projection. The top twenty projects (based on overall number of jobs projected by 2056) were assessed by the DPE and adjusted to meet the Department's employment projected expectations for these projects.

Table 12: Summary of new developments database for employment

Project Name	Project Description	Project Area by Travel Zones
Alex Avenue	Alex Avenue Precinct was rezoned for urban development in May 2010. It will provide a Town Centre of up to 25, 000sq.m retail area (TZ 3958). A school will be located within TZ 3959. Another school will be located within TZ 3960.	3958, 3959, 3960
Area 20	This precinct was rezoned for urban development in October 2011. This precinct will include a light industrial zoned area as well as a planned station as part of the North West Rail Link.	3948
Austral / Leppington North	This project is the continued development of the Austral/Leppington North precinct.	3627
Barangaroo	This project is the development of the Barangaroo precinct including the three international towers. This project is expected to be completed by 2016.	1, 2, 3, 4, 5
Bays Precinct	This project includes the redevelopment of the foreshore to include an innovation hub.	153, 806, 828, 830, 832, 835
Box Hill/Box Hill Industrial	The Hills Council Contributions Plan No. 15 (June 2016) provides estimates for jobs within the Box Hill Precinct. The Precinct Plan provides for over 115 hectares of employment land with the potential to generate approximately 17, 800 jobs. Most of the new development will occur within the 11-20 year development timeframe (refer to the contributions plan for more detail).	4426, 4428, 4429, 4430, 4431
Broader Western Sydney Employment Area	The Western Sydney Employment Area (WSEA) is broken up into precincts A-G, with only precinct A released for development at this time. Once fully constructed, WSEA will provide around 97, 000 jobs throughout all the precincts and any additional land covered by the associated WSEA SEPP.	4045, 4961, 4964, 5002, 3626, 3477, 3479, 3482, 4059

Project Name	Project Description	Project Area by Travel Zones
Central to Eveleigh	This project is the redevelopment of the corridor between Central to Eveleigh. The project includes increased neighbourhood shops and some commercial development.	330, 132, 134, 136, 137, 140, 141, 142, 143, 144, 145, 146, 151, 152, 154, 156, 157, 158, 159, 160, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 176, 178, 179, 180, 181, 182, 183, 184, 185, 188, 189, 190, 191, 192, 193, 194, 195, 197, 196, 198, 199, 201, 202, 205, 206, 207, 208, 212, 214, 215, 217, 218, 223, 226, 227, 228, 229, 230, 231, 232, 233, 236, 238, 239, 240, 242, 243, 246, 247, 249, 255, 256, 257, 258, 260, 261, 262, 263, 264, 265, 266, 267, 270, 271, 272, 273, 274, 276, 277, 279, 281, 282, 283, 284, 285, 286, 287, 288, 559, 560, 562, 563, 565, 812
Clemton Park Village Centre/Former Sunbeam Factory Campsie	This project is the construction of a mixed use development comprising a retail component, mini-major tenancy, and supermarket with 290 residential apartments, community centre. This is stage 3 of the redevelopment of the former Sunbeam Factory site.	2517
Defence Land Moorebank - Intermodal Terminal	The defence site at Moorebank was rezoned to facilitate the development of the Moorebank Intermodal Terminal resulting in additional employment in this precinct.	3824
Edmondson Park Town Centre	This project is centred around the development of a Town Centre around the newly built train station of Edmondson.	3239

Project Name	Project Description	Project Area by Travel Zones
Enfield Intermodal Terminal	This project will be used as a storage and handling facility resulting in a small increase of employment within this travel zone.	984
Former Hoxton Park Airport - Development	This project is a result of a rezoning of the former Hoxton Park Aerodrome to a mixed use. The rezoning is expected to result in increased employment.	3712
Freight Transport Warehouse and Distribution Centre - Huntingwood	This project is expected to provide increased employment for this travel zone when complete.	4163
Glenfield to Macarthur	This project is a large development in and around the Campbelltown station. It will include large residential developments as well as commercial buildings.	3213, 3215
Greater Macarthur Investigation Area	This project is a large area release. It is planned to include approximately 33, 000 dwellings and 30, 000 local jobs within the area.	3300, 3305, 3307, 3005, 3007, 3008, 3009, 3010
Horsley Drive Business Park	This project is a development of a business park that will be including warehousing and distribution facilities.	3480
Hunter Economic Zone (HEZ)	This project is the continued development of the industrial park within the Hunter Economic Zone.	6740
Huntingwood West Precinct	This project is an industrial estate development.	4060
Huntlee Development	This project is a development of the Huntlee town centre within Cessnock. It will include an approximately 60ha of employment lands as well as residential development	6701
Liverpool CBD	This project is a rezoning of lands to mixed use. The development will include office and retail floorspace.	3843
Macquarie Centre expansion	This project is an expansion of the Macquarie Centre including the expansion of the shopping centre, four mixed use towers.	1539

Project Name	Project Description	Project Area by Travel Zones
Marrickville Metro	This project is a redevelopment of the centre to include commercial and retail.	303
Marsden Park	This project is a growth centre with an expected residential and employment precinct.	3950, 3956, 3961
Marsden Park Industrial	This precinct was rezoned to industrial and bulky goods retailing.	3962, 3968, 3970
Newcastle CBD	This project includes the rezoning and building of the light rail project along within the Newcastle CBD.	6317, 6351, 6352
Northern Beaches Hospital	This project is expected to be complete by 2018.	2140
NWRL-Kellyville	The North West Rail Link is expected to result in employment around the proposed station	4568
NWRL-Rouse Hill	The North West Rail Link is expected to result in employment around the proposed station	4436
Oran Park	This project is a business park development.	3678
Other developments	These projects are small scale redevelopments sourced from Cordells. This includes small rezoning and mixed use developments expected to provide small levels of employment	5331, 5180, 6613, 5700, 5723, 5736, 5838, 5644, 5647, 6400, 6414, 6427, 6210, 6117, 4434, 4160, 79, 97
Parramatta	This project is the continued development of the Parramatta precinct.	1018, 1019, 1021, 1023, 1024, 1025, 1026, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1062, 1064, 1065

Project Name	Project Description	Project Area by Travel Zones
Penrith Panthers development	This project is a development of a seniors housing development expected to provide an increase in health employment.	4980
Penrith Station Car park	The development is to include business, commercial and light industrial space.	4977
Port Botany Expansion	The expansion of the port will result in increased employment capacities.	401, 402
Potts Hill Redevelopment	The redevelopment is expected to include increased employment lands.	2329
Redfern-Waterloo Development	The redevelopment between Redfern and Waterloo is expected to increase employment capacity.	210, 211, 216, 222
Residential/commercial development at former BATA manufacturing site	The redevelopment of the former manufacturing site is likely to include some employment capacity for retail employment.	423
Riverstone	The Riverstone Growth Centre Precinct was rezoned for urban development in May 2010. It will include a Major Town Centre, industrial land, schools, and neighbourhood centres.	3933, 3935, 3941, 3953
Riverstone West	Riverstone West Growth Centres precinct was rezoned August 2009. It is expected that 16ha will be Business Park, 72 ha will be (IN1) Industrial and 16ha will be (IN2) Light Industrial.	3934, 3937
SICEEP Precinct	The Sydney International Convention, Exhibition and Entertainment Precinct (SICEEP) development includes a convention centre, a theatre, a hotel complex, and 'The Haymarket' residential/ commercial development.	89, 108, 125

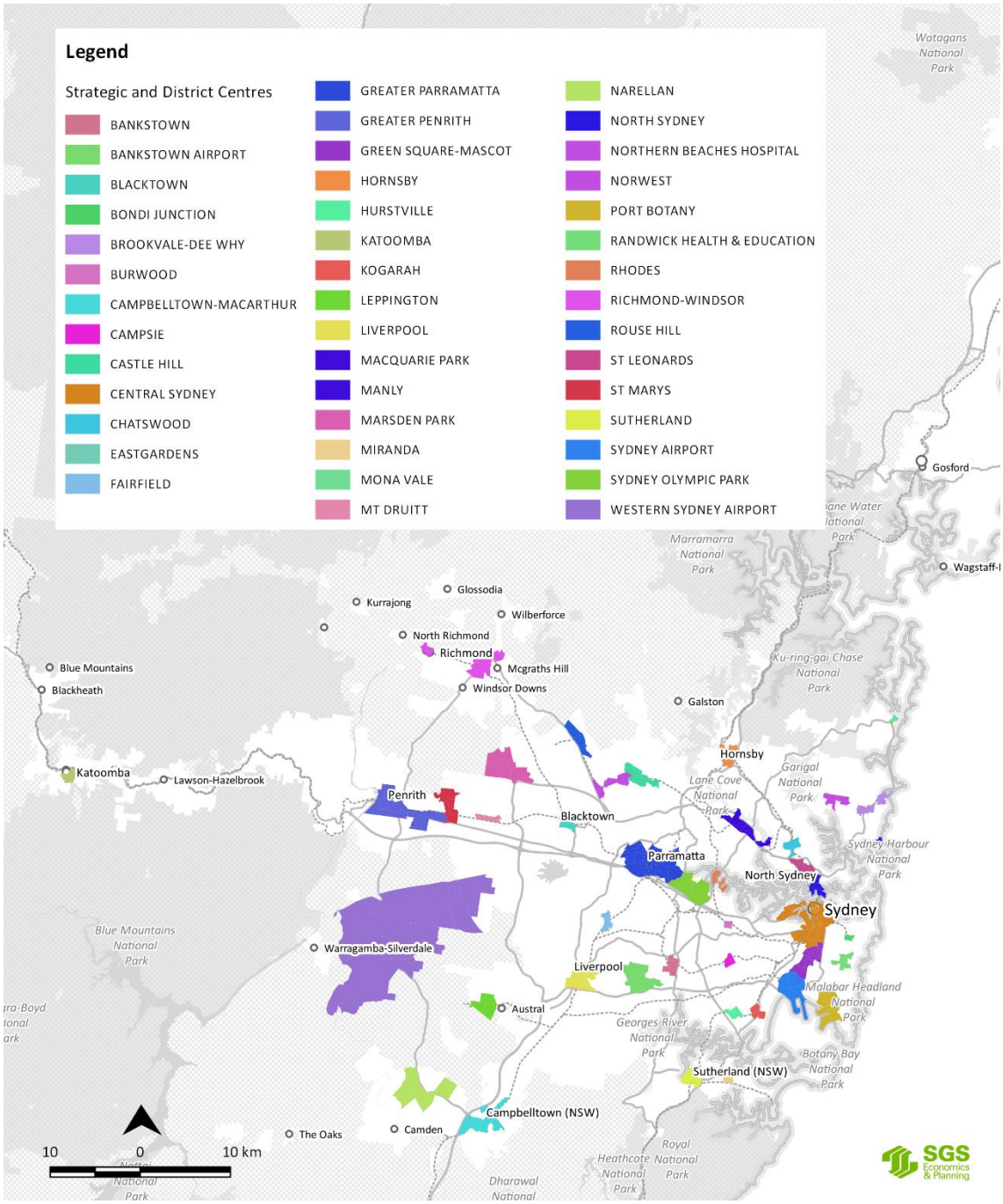
Project Name	Project Description	Project Area by Travel Zones
South West Growth Area	This project will include large residential and commercial development within the South West Growth Area precinct.	3243, 3615, 3616, 3621, 3624, 3629, 3630, 3632, 3633, 3634, 3651, 3653, 3654, 3655, 3656, 3657, 3658, 3659, 3660, 3662, 3663, 3664, 3668, 3669, 3671, 3672, 3674, 3679, 3681, 3683, 3685, 3689, 3691, 3692, 3724, 3478, 3483, 3864, 3865
Southern Employment Lands	This project is a large industrial development for warehousing and distribution centre.	1238
Sydney Metro - Bankstown to Sydenham	This project is the development of the corridor the Sydney Metro will run. It is expected the development of the metro will increase commercial employment capacity along the corridor.	307, 310, 313, 2301, 2305, 2307, 2308, 2311, 2504, 2509, 2510, 2511, 2512, 2514, 2515, 2533, 2541, 2542, 2545, 926, 927, 941, 943, 944, 945
Sydney Metro - Northwest	This project is the development of the corridor the Sydney Metro will run. It is expected the development of the metro will increase commercial employment capacity along the corridor.	4514, 4515, 4516, 4517, 4521, 4532, 4533, 4534, 4536, 4571, 4433, 4438, 3906, 3908, 1805, 1806, 1402, 1406, 1407

Project Name	Project Description	Project Area by Travel Zones
Sydney to Parramatta	This project is the redevelopment of Parramatta road from Sydney to Parramatta. It will include increased commercial floorspace along the corridor.	316, 317, 318, 237, 706, 707, 717, 719, 720, 735, 738, 819, 820, 821, 826, 900, 901, 903, 910, 911, 912, 939, 950, 951, 952, 958, 960, 961, 963, 969, 970, 1300, 1301, 1303, 1304, 1318, 1320, 1321, 1322, 1323, 1324, 1326, 1328, 1333, 1334, 1335, 1220, 1222, 1046, 1045, 1047, 1048, 1068, 1071
Tallawarra Lands	This project is the development of a new water and wastewater infrastructure to service the Tallawarra Lands.	5726
Town Centre and Civic Precinct of Warnervale	Development of the town centre and civic.	5186
Western Sydney Airport and surrounds	This project is the development of Badgery's airport and surrounds. This is expected to provide employment once the airport is developed.	4967, 4968, 4970, 3609
Wolli Creek redevelopment	The redevelopment of Wolli Creek is expected to increase employment capacity due to increased provision of commercial and retail floor space.	2702
Woolooware Bay Town Centre Development	This redevelopment is expected to provide a small increase to the employment capacity within the travel zone.	2914

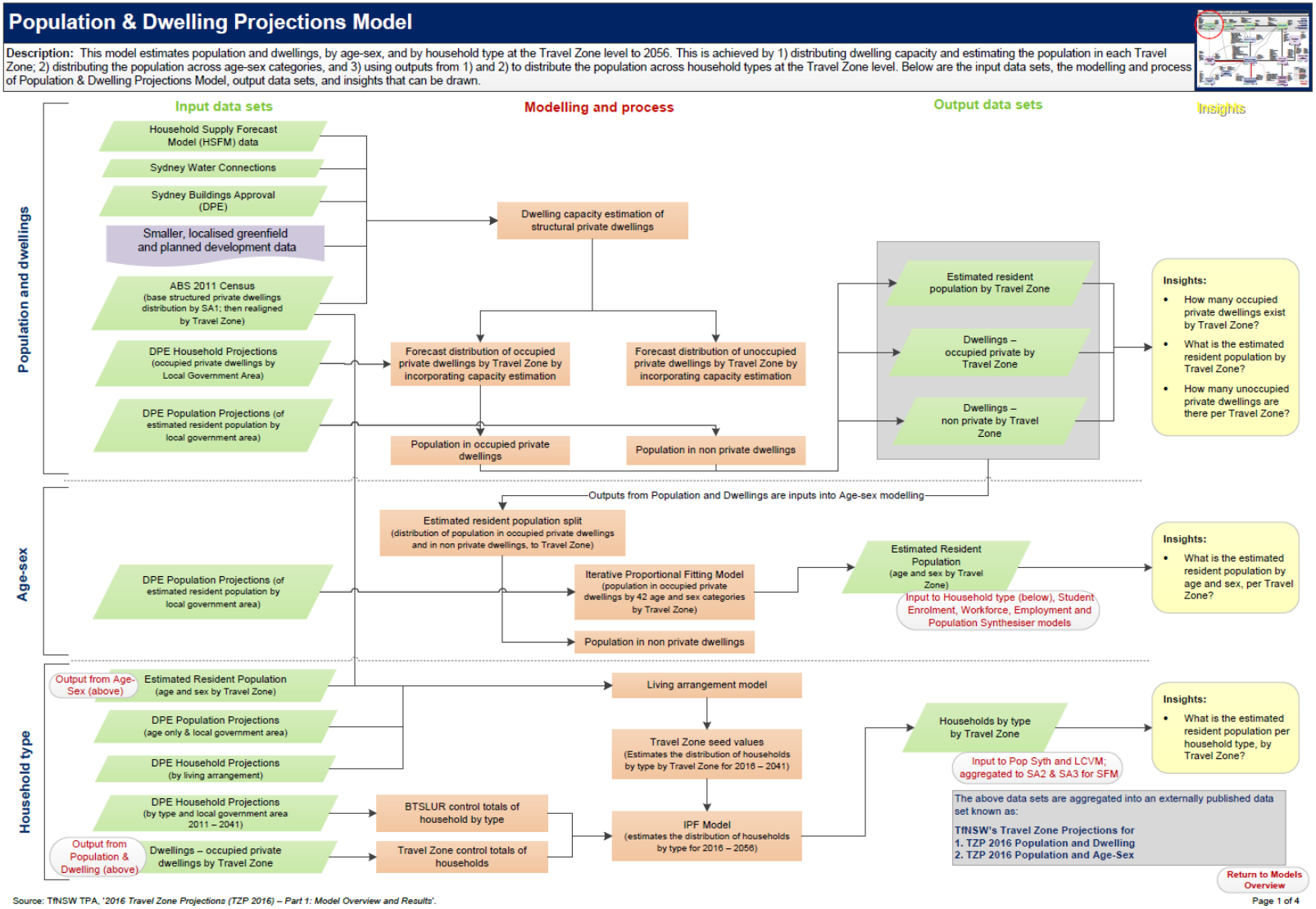
18 Appendix B: GSC Centres

The figure below shows the centres that are defined by the GSC. These centres outlined below are the major strategic and district centres within Greater Sydney.

Figure 36 GSC Centres

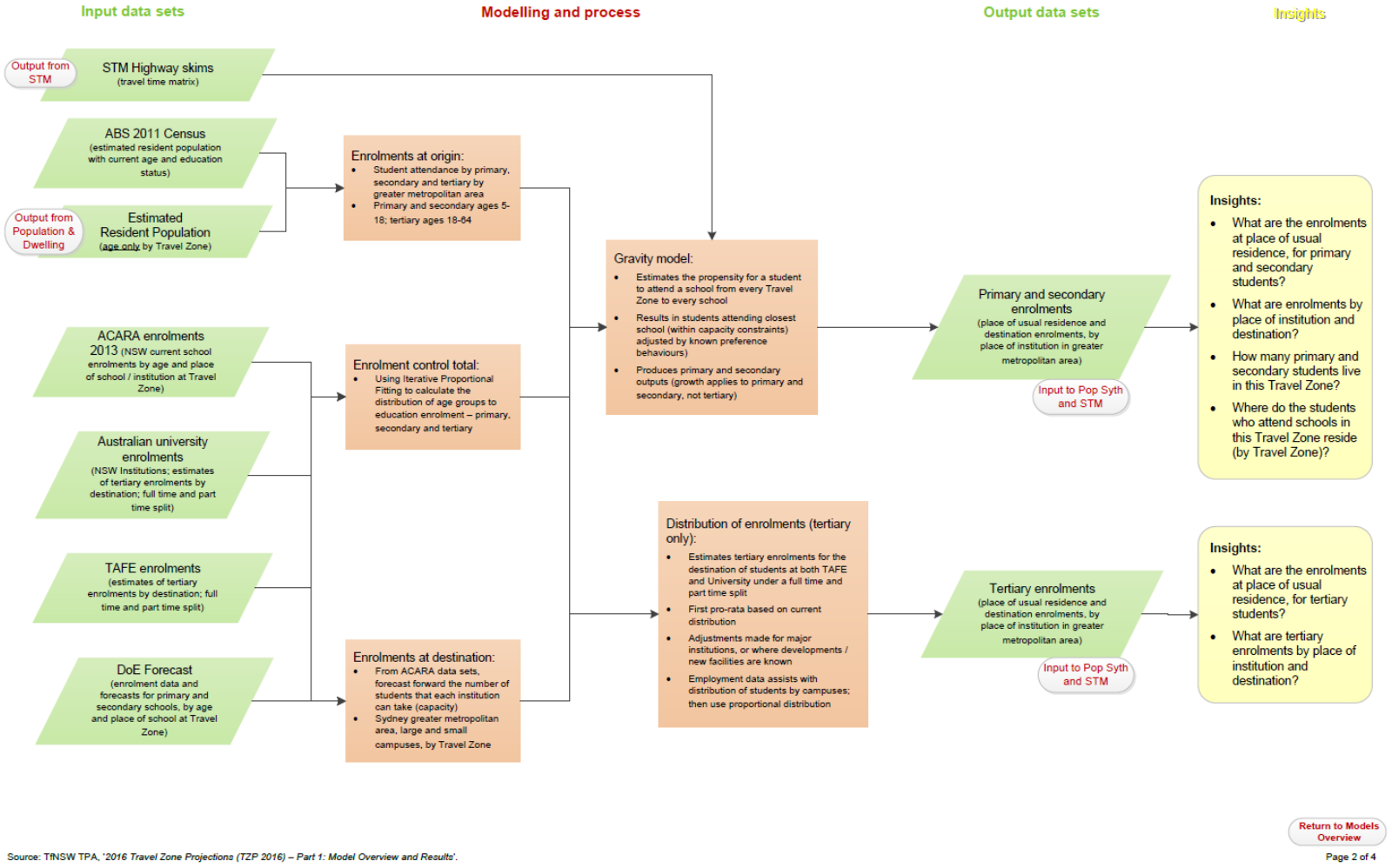


19 Appendix C: Travel Zone Projections 2016: Process Maps



Student Enrolment Projections Model

Description: This model uses outputs from the Population & Dwelling Projections model (Estimated Resident Population – age and sex), to estimate student enrolments for primary, secondary and tertiary education, by place of residence and place of institution at the Travel Zone level to 2056. Below are the input data sets, the modelling and process of the Student Enrolments Projections Model, output data sets, and insights that can be drawn.

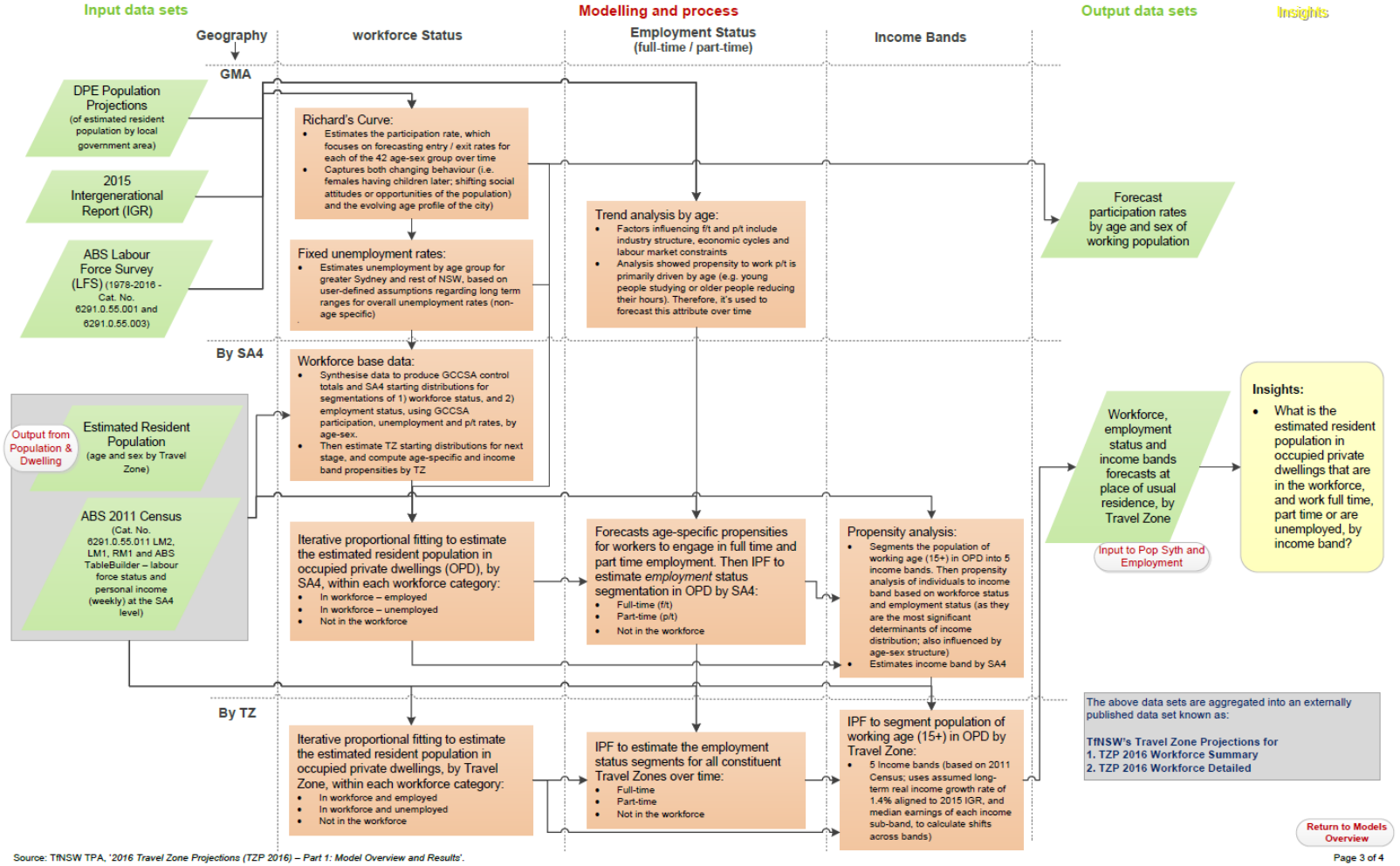


Source: TfNSW TPA, *'2016 Travel Zone Projections (TZP 2016) – Part 1: Model Overview and Results'*.

Page 2 of 4

Workforce Projections Model

Description: This model uses outputs from the Population & Dwelling Projections Model to estimate various segmentations of the working age population, over 3 modelling runs at 3 geographic scales: Greater Sydney, sub-regions of SA4s, at the Travel Zone level to 2056. The population is broken down by workforce status (employed or 'workforce', unemployed and not in the workforce), and employed persons are further broken down by employment status (employed full-time or part-time) and income bands. Below are the input data sets, the modelling and process of Workforce Projections Model, output data sets, and insights that can be drawn.



Employment Projections Model

Description: This model is used to produce employment projections by industry, at the Travel Zone level to 2056. It has 2 major functions: 1) the total employment (persons employed by place of work) by industry is projected for the GMA region, and 2) the spatial distribution within each industry is projected based on a variety of factors, at geographies SA3 and Travel Zone. *Note: As estimating the growth and distribution of employment by industry is reliant on professional judgement and influenced by endogenous and exogenous factors, the modules are designed to automate first-cut estimates, then apply professional judgement to obtain 2nd cut final estimates.* Below are the input data sets, the modelling and process, output data sets, and insights that can be drawn.

Input data sets

Geography

GMA

Output from Population & Dwelling — Estimated Resident Population (age and sex by Travel Zone)

ABS Labour Force Survey (LFS) (1978-2016 - Cat. No. 6291.0.55.001 and 6291.0.55.003)

Output from Workforce — Workforce, employment status and income bands projections at place of usual residence, by Travel Zone

Output from Population & Dwelling — Dwellings – non private by Travel Zone

ABS Census data (historical employment at the 2 digit level from 1996 – 2011 by SA3 and Destination Zone (DZN). Also 2011 Journey to Work (inflow and outflow of workers from ABS TableBuilder) by industry at SA3 level)

By SA3

New developments database (known developments and growth in future employment by Travel Zone. The top 20 projects (based on overall number of jobs forecast by 2056) were assessed by DPE)

By TZ

Modelling and process

Preliminary industry share forecast – Trend based:

- Forecasts the quantum of employment by industry within the GMA region by:
  - 1) forecasting total employment within GMA region for each period to capture 1.1) net inflow of workers from the rest of NSW to the GMA, and 1.2) employment of persons who reside in non-private dwellings
  - 2) Refining industry forecasts based on variety of factors, such as population to employment ratio, or expected growth rate
- Produces 1-digit ANZSIC industry forecast

Industry refinements:

- For each industry sheet, review the forecasts implied by the preliminary shares against the provided indicators and refine the level changes and growth rates if necessary
- For example if a population driven industry captures an unreasonable share of regional employment (shown in the ratio of population to employment), adjust the growth values to result in a forecast that 1) has a more reasonable share of total employment, and 2) displays a strong but declining growth rate consistent with recent trends

2-digit ANZSIC manufacturing forecasting:

- Uses historical employment data from Census (1996 to 2011) to disaggregate the 1-digit industry forecast to a 2 digit industry forecast

Preliminary SA3 share forecast – Trend based:

- Distributes the previously determined employment by industry forecasts to SA3 level by using 1) historical census, 2) planned developments and 3) professional judgement (such as growth in employment in known developments, forecast population growth, regions breaking away from recent trends, impacts on industries)
- Forecasts then reviewed through an examination of employment density (forecast employment per hectare of employment zoned land) across SA3. This provides a high-level representation of how SA3 are expected to evolve relative to their quantity of employment land

Population driven industry adjustments:

- Checks population in SA3 areas is sufficiently serviced by employment

Final industry distribution refinements at SA3 level:

- control total to ensure SA3 is broken down to the Travel Zone level

Preliminary TZ industry forecasting share – Trend based:

- Allocates the SA3 industry control totals to a TZ level by 1) using Census data and new developments database to compute the base distribution of employment by industry within each SA3, and 2) use professional judgement to apply manual calibrations to override forecasts at TZ

Population driven industry adjustments:

- Checks population in Travel Zones are sufficiently serviced by employment

Final industry distribution refinements at Travel Zone level

Output data sets

Employment (persons employed by place of work) projections by industry at Travel Zone level

Input to STM, LCVM, PB-SA-FMM, FMM and SFM

The above data sets are aggregated into an externally published data set known as:

TfNSW's Travel Zone Projections for
1. TZP 2016 Employment Summary
2. TZP 2016 Employment Detailed

Insights

Insights:

- What is the employment (persons employed by place of work) by industry at the Travel Zone level?
- What is the likely future employment per Travel Zone?

Return to Models Overview

Source: TfNSW TPA, '*2016 Travel Zone Projections (TZP 2016) – Part 1: Model Overview and Results*'.

Page 4 of 4

20 Appendix D: Table of Assumptions

Module	Assumption
M01a - Capacity Estimation	HSFM data assumed to have the highest degree of certainty - assigned to priority capacity
	ALU14v4 data assumed to have a lower degree of certainty - assigned to secondary capacity
M01b - Dwelling Forecasting	Structural private dwellings aligned to 'Adjusted LGA' totals (prepared by TPA, using DPE data)
	Occupied private dwellings aligned to 'Adjusted LGA' totals (prepared by TPA, using DPE data)
M01c - Population by Travel Zone	Persons in occupied private dwellings aligned to 'Adjusted LGA' totals (prepared by TPA, using DPE data)
	Within an LGA, relative differences in travel zone household sizes are based on SA1 census data
M02a/b - Age by Gender IPF	SA1 Census data used to inform base distribution of POPD by age by gender at travel zone level
	Due to randomisation of small numbers in census data, travel zones with a base estimate below 20 POPD in any category are assigned the distribution of it's encompassing 'Adjusted LGA'. This is particularly relevant for future growth zones, which have no historical data available
	Base year to run the IPF is 2016

	To ensure that PNPD does not influence the IPF results (i.e. POPD by age by gender), travel zones with an ERP/OPD ratio less than 50% are assigned a seed distribution equal to the encompassing 'Adjusted LGA'
	For each future year, the IPF procedure uses the age-gender results of the previous period as the seed distribution
M03a - Adjusted LGA Population	The DPE population forecast by living arrangements aligned to the 'Adjusted LGA' totals using a concordance between DPE LGA and TPA 'Adjusted LGA' (prepared by TPA, using DPE data).
M03b - Living Arrangement Propensities	ERP by TZ forecast transformed to reflect living arrangement using the Department of Planning and Education estimates on persons by LGA by living arrangement.
M03c - Calculating seed values	Group household propensity/scale factor is calculated using the Family/Household Reference Person Indicator - Relationship in Household figures from the 2016 census.
M03d - IPF	Base year to run the IPF is 2016
	For each future year, the IPF procedure uses the household results of the previous period as the seed distribution

M04a_1	DoE school enrolment forecast concorded to Travel zones. (prepared by TPA, using DoE forecast)
M04a_2	The single age distribution derived from the 2016 Census data is applied to the population forecast by 5 year age groups to obtain the school age groups.
M04a_3	SA1 Census data used to inform base distribution of POPD by schooling age at travel zone level
	Due to randomisation of small numbers in census data, travel zones with a base estimate below 20 POPD in any category are assigned the distribution of it's encompassing 'Adjusted LGA'.
M04b	The 2016 Census age by education attendance is as seed values into the IPF to obtain enrolment control totals by institution type
M04c - Enrolments at Origin	The 2016 distribution of school attendance is applied to the age group forecast from 2021-2056
	Where there is no base data input in the initial distribution of persons, the regional distribution is applied. This accounts for growth areas where there is no precedent of population.
	Base year to run the IPF is 2016
	For each future year, the IPF procedure uses the household results of the previous period as the seed distribution

M04d - Enrolments at Destination	School capacity for primary schools is assumed to 1,000 persons and secondary students assumed to be 2,000
M04e	Capacity of the secondary schools adjusted in future years to assist the model run.
	2011 travel times are used for future years implying there are no improvements to school access in the future
M04f	The spatial distribution of tertiary students across the GMA is assumed to remain consistent across the forecast. This does not account for any future institutions.
M05a/b - Participation Rates	Estimate of annual participation rates by age and gender are defined as the average of the monthly observations (calendar year), which will be centred on June
	Hodrick Prescott filter used to smooth participation rates. Chosen parameter of lambda = 100
	Maximum participation rates imposed. For males this is 99%, and for females 95%
M05c - Unemployment Rates	NSW long-term unemployment rate of 5% assumed (2016 NSW IGR)

	Unemployment rates for each age by gender category are estimated using historical data. For greater Sydney, data points considered are those where the unemployment rate was between 4%-6%. For the Rest of NSW, data points considered are those where the unemployment rate was between 4%-8%
M05d/e - Labour Force IPF	SA1 Census data used to inform base distribution for all labour force segments at travel zone level
	Due to randomisation of small numbers in census data, travel zones with a base estimate below 20 POPD in any category are assigned the distribution of it's encompassing 'Adjusted LGA'. This is particularly relevant for future growth zones, which have no historical data available
	For each future year, the IPF procedure uses the labour force status results of the previous period as the seed distribution
M05d/f/g - Part-Time Segmentation IPF	Estimate of annual part-time rates by age are defined as the average of the quarterly observations (calendar year), which will be centred on June
	Upper limits assumed for part-time rates in Greater Sydney as follows: 60% for 15 - 24 year olds 30% for 25 - 34 year olds 35% for 35 - 54 year olds 50% for 55 - 64 year olds 65% for 65+ year olds

	Upper limits assumed for part-time rates in Greater Sydney as follows: 60% for 15 - 24 year olds 40% for 25 - 34 year olds 50% for 35 - 44 year olds 40% for 45 - 54 year olds 50% for 55 - 64 year olds 65% for 65+ year olds
M05d/h - Income Segmentation	Due to randomisation of small numbers in census data, travel zones with a base estimate below 20 POPD in any income category are assigned the distribution of it's encompassing 'Adjusted LGA'. This is particularly relevant for future growth zones, which have no historical data available
	Assumed real income growth rate of 1.4%. This is the long term gross national income growth rate forecast in the 2015 IGR
	For each future year, the IPF procedure uses the labour force status results of the previous period as the seed distribution
M06a - Industry Forecasting	Census JTW data used to account for labour force flows across GMA border. These patterns are assumed to persist in future years
	PNPD are included in Place of Work forecasts. In Greater Sydney, participation and unemployment rates of PNPD by age are assumed to equal the Greater Sydney average (similarly for Rest of NSW)

M06c - SA3 Industry Forecasting	For six population servicing industries, a maximum ERP/Emp ratio (growth) is specified. These are estimated in M06c_1, and the resulting limits are as follows: Construction: 37 Retail Trade: 34 Accommodation and Food Services:56 Rental, Hiring and Real Estate Services: 245 Education and Training:44 Health Care and Social Assistance:36
M06d - TZ Industry Forecasting	In the absence of additional information, travel zone employment shares (of their encompassing SA3) are assumed to remain the same as the base year (i.e. 2016)
	All travel zone assumptions contained within the 'New Developments Database' (supplied by TPA)

21 Appendix E: R Installation and Module Documentation

21.1 R Module Introduction

In order to reduce processing time, some modules within the model are in the format of an R script in place of an excel spreadsheet. These modules are 3c, 4e and Output Checks. R is a programming language commonly used by statisticians and researchers to perform advanced data analysis. These modules are saved in the same folders as the excel sheets but have a '.Rexec' file path ending.

21.1.1 Software

Specific software requirements are required to run R scripts. The scripts have been written within R version R 3.4.0 and it is recommended that this version is installed as newer versions may require de-bugging.

To install, go to this page, and follow the instructions:

https://cran.r-project.org/bin/windows/base/old/3.4.0/

R studio, the common integrated development environment for R users, does not need to be installed to run the checks. Nevertheless, it useful software for de-bugging and amending R scripts and is free to download here:

https://www.rstudio.com/

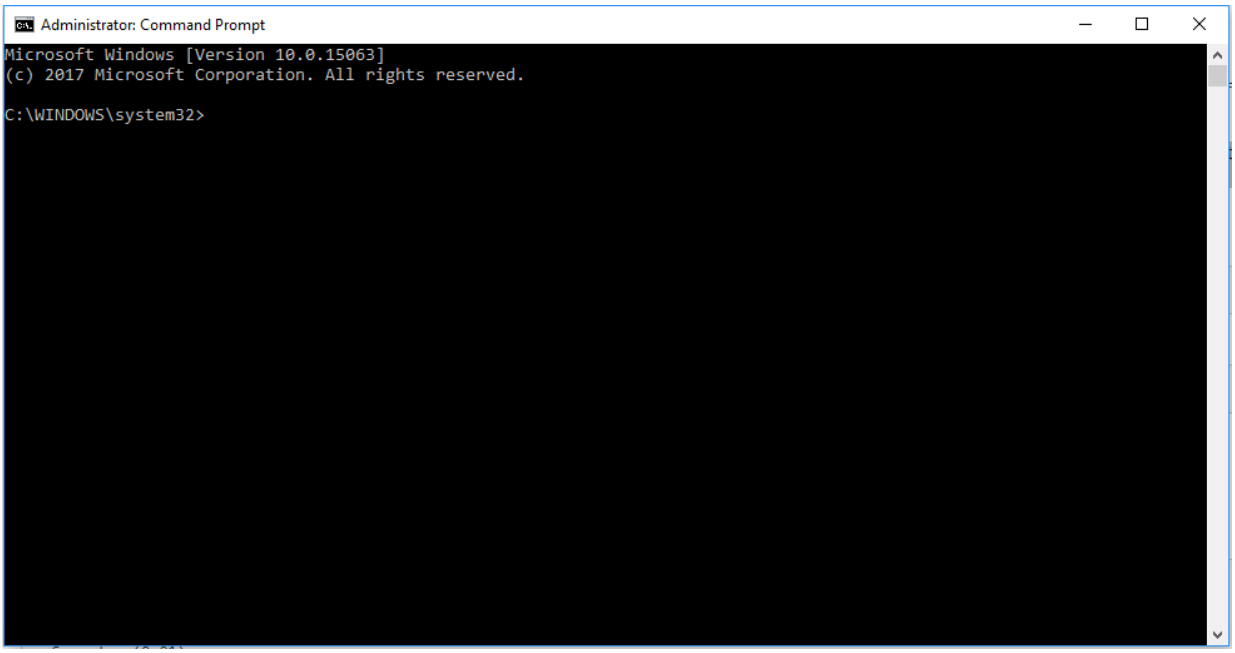
21.1.2 Setup to Run the R Modules (Required Only Once)

After installing R it is necessary to associate the '.Rexec' file path with the installed 'Rscript.exe' program. This can be achieved through the command prompt (Figure 1).

1. Click **Start** and type **cmd** into the search bar. Instead of hitting enter right away, the user should right click on the **'cmd.exe'** search result.

2. The user should choose **Run as administrator** from the context menu, and then click **Yes** on the following pop up window. The windows command line should pop up thereafter.

Figure 16. Command Prompt Window



Within the command line, the user should type:

```
ASSOC .Rexec=RScriptExecutable
```

... then ...

```
FTYPE RScriptExecutable=C:\Program Files\R\R-3.4.0\bin\Rscript.exe%1 %*
```

... while making sure that the path used above leads to the RScript.exe file (this assumes that R 3.4.0 is the version of R that is installed).

21.1.3 R Module Structure Overview

The R modules link to various '.csv' files containing output data from previous modules or module steps. These files are contained within the 'R_input_csv' files within the modules. Neither the file nor csv names should be changed as this will break the links to the R module and cause it to stop working. Similarly, file names and relative locations must also remain the same for the modules to work.

21.1.4 Running the R Modules

The user is able to open and edit the script in programs other than RStudio such as text editor and notepad. This is useful for making minor edits to the script such as changing the directory.

21.1.4.1 Changing the Directory and Library Path

To change the directory, the user should right-click the .REXEC module file, and open with notepad. The user should locate this line at the beginning:

```
directory = file.path("C:", "Users", "jsmith", "Documents")
```

```
libraryPath = file.path("C:", "Users", "jsmith", "Documents", "R", "R-3.4.0"
, "library", sep='')
```

To change the location of the directory in the script, the user should find the directory in which the model has been saved with each part in part in quotes as shown and paste this text inside the brackets of “file.path()”.

21.1.4.2 Running the scripts

Once the setup steps above have been completed, to run the R modules simply double click on the ‘.Rexec’ file. Once complete the command window should close automatically.

21.2 Module 03c – Calculating Seed Values Instructions

This module is an R Script that takes the output from previous components of Module 3 and estimates the distribution of households by type at the travel zone level for 2016-2041.

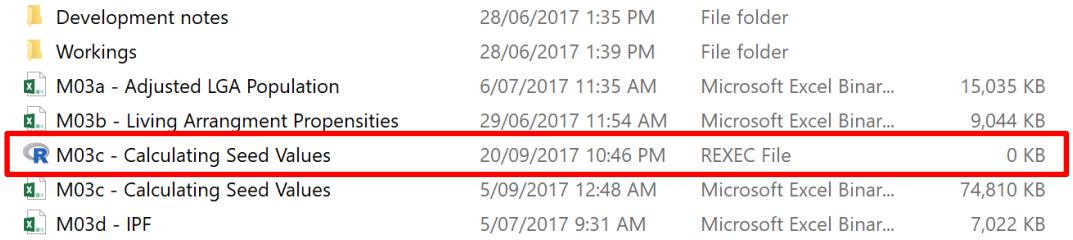
21.2.1 Folder Structure

The R component of the module sits alongside the spreadsheet components in the ‘M03_Households’ folder of the main model structure. The module reads input from the same source as the previous excel version of this module step located in the "O01_TZP Database (CSV)" folder in the main model structure.

21.2.2 Running the Module Component

Once the set-up steps listed in Section 21.1 are complete and the relevant file paths located at the top of the script changed the user should simply double click on the “M03c - Calculating Seed Values.REXEC” file to run the component of the module.

Figure 17. Module 3 folder



This will open the command prompt which will scroll through the script similar and close automatically on completion.

21.2.3 Module Output

Module output saved direct to the ‘O01_TZP Database (CSV)’ folder in the main model structure. The file name of the module output is ‘3.3 IPF Seed Values_TZ’. This file is linked into the subsequent sub-module.

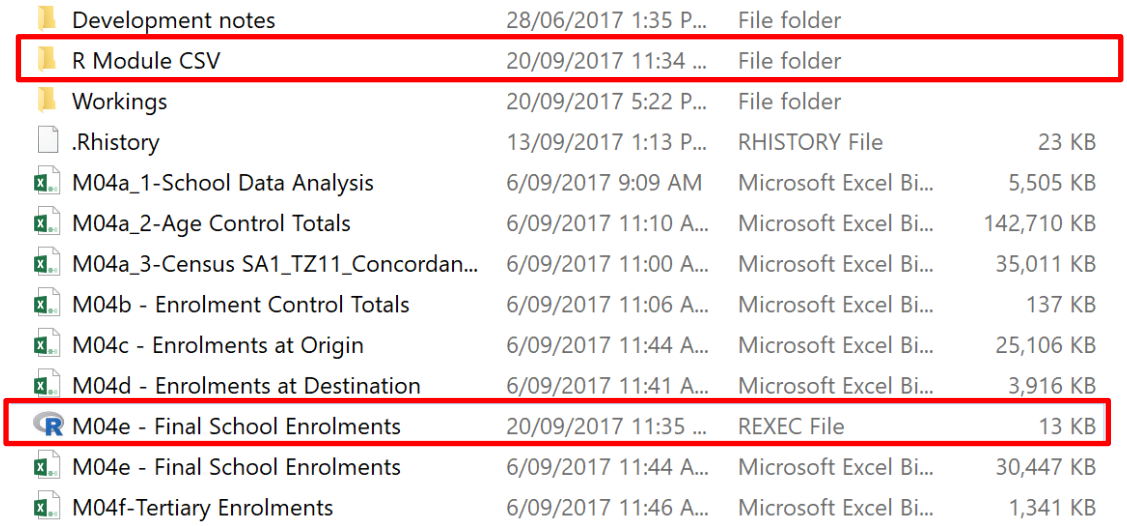
21.3 Module 04e – Calculating Seed Values Instructions

This module is an R Script that takes the output from previous components of Module 4 and estimates the small area destination school enrolments in each future year out to 2051 using a gravity model.

21.3.1 Folder Structure

The R component of the module sits alongside the spreadsheet components in the 'M04_Students' folder of the main model structure. The module reads input from the 'R Module CSV' folder in the module structure (Figure 3). This folder contains the travel time matrix, origin enrolments and destination enrolments (manually populated from sub-modules 03c and 03d).

Figure 18. Module 4 folder



21.3.2 Running the Module Component

The R script links directly to the previous calculated excel modules. Whilst running the module the user should ensure Excel input modules 'M04c – Enrolments at Origin' and 'M04d –Final School Enrolments' are open. Once the set-up steps listed in Section 21.1 are complete the user should simply double click on the "M04e – Final School Enrolments.REXEC" file to run the module component. This will open the command prompt which will scroll through the script similar and close automatically on completion.

21.3.3 Module Output

Module output saved direct to the 'O01_TZP Database (CSV)' folder in the main model structure. The files produced as module output are '4.1 Primary and Secondary School students PUR_TZ' and '4.2 Primary and Secondary School students POI_TZ'.

21.4 Module 07 – Output Quality Checks Instructions

This module is an R Script that takes the output from previous modules and performs a number of automated calculations to evaluate quality and consistency within and between the outputs of the various modules.

21.4.1 Folder Structure

There are two main components in the folder structure, both required to run the module:

1. Outputs CSV: This folder contains the reformatted module-output data required for the checks module to run. The files are in '.csv' format.

2. R Scripts: Contains the R scripts that perform the automated checks.

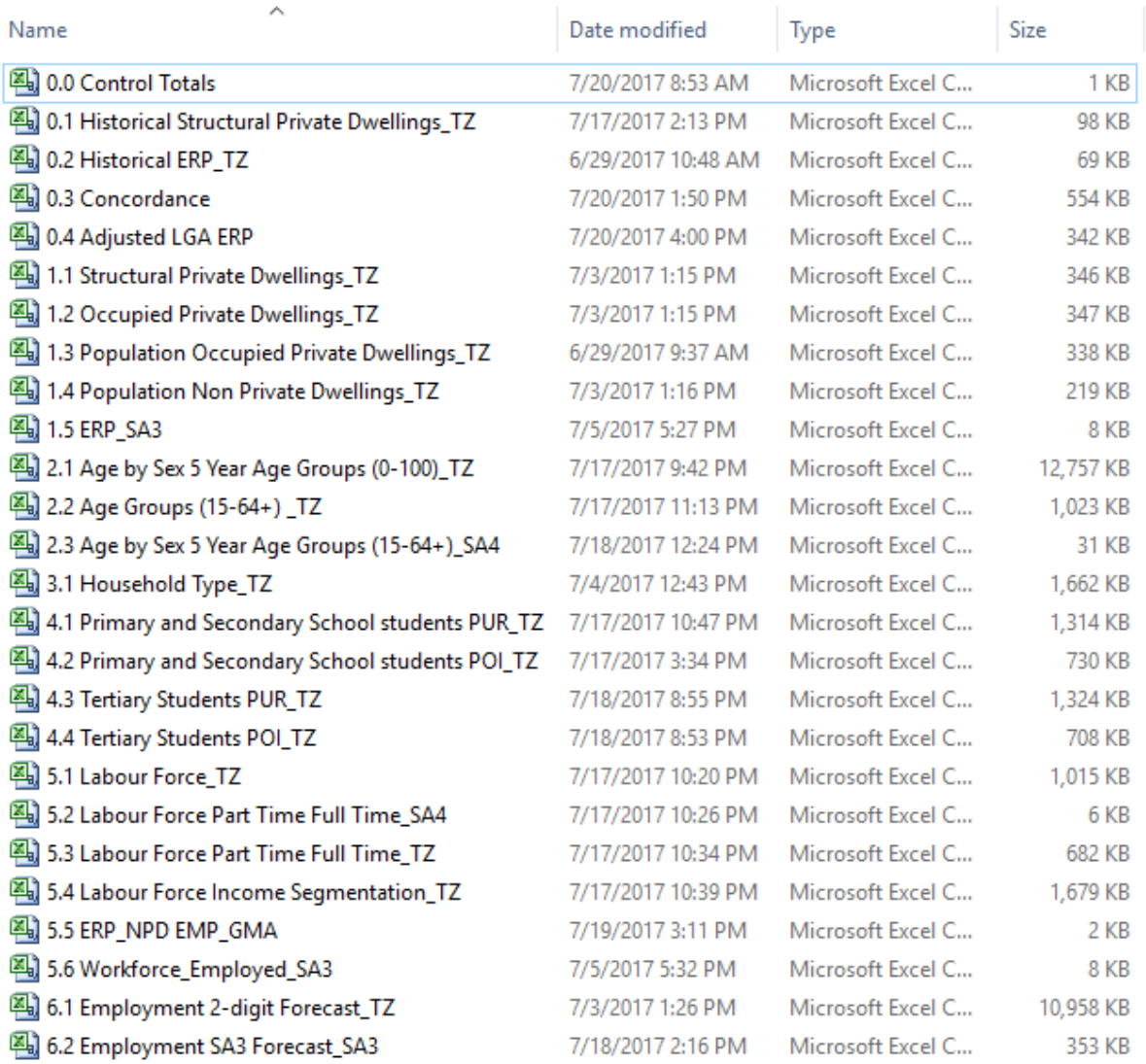
Both folders are described in more detail below.

21.4.1.1 Outputs CSV

The outputs CSV folder contains the output files from the modules and is the source from which the module calculates its checks. The Outputs CSV is located at the following file path. It contains the files shown in Figure 4 below and needs to be updated manually according to the process outlined in section 21.4.2.1 below.

LU16 Forecast -NEW STRUCTURE > O01_TZP Database (CSV) > Outputs CSV

Figure 19. Module output files within the Outputs CSV folder



Name	Date modified	Type	Size
0.0 Control Totals	7/20/2017 8:53 AM	Microsoft Excel C...	1 KB
0.1 Historical Structural Private Dwellings_TZ	7/17/2017 2:13 PM	Microsoft Excel C...	98 KB
0.2 Historical ERP_TZ	6/29/2017 10:48 AM	Microsoft Excel C...	69 KB
0.3 Concordance	7/20/2017 1:50 PM	Microsoft Excel C...	554 KB
0.4 Adjusted LGA ERP	7/20/2017 4:00 PM	Microsoft Excel C...	342 KB
1.1 Structural Private Dwellings_TZ	7/3/2017 1:15 PM	Microsoft Excel C...	346 KB
1.2 Occupied Private Dwellings_TZ	7/3/2017 1:15 PM	Microsoft Excel C...	347 KB
1.3 Population Occupied Private Dwellings_TZ	6/29/2017 9:37 AM	Microsoft Excel C...	338 KB
1.4 Population Non Private Dwellings_TZ	7/3/2017 1:16 PM	Microsoft Excel C...	219 KB
1.5 ERP_SA3	7/5/2017 5:27 PM	Microsoft Excel C...	8 KB
2.1 Age by Sex 5 Year Age Groups (0-100)_TZ	7/17/2017 9:42 PM	Microsoft Excel C...	12,757 KB
2.2 Age Groups (15-64+) _TZ	7/17/2017 11:13 PM	Microsoft Excel C...	1,023 KB
2.3 Age by Sex 5 Year Age Groups (15-64+)_SA4	7/18/2017 12:24 PM	Microsoft Excel C...	31 KB
3.1 Household Type_TZ	7/4/2017 12:43 PM	Microsoft Excel C...	1,662 KB
4.1 Primary and Secondary School students PUR_TZ	7/17/2017 10:47 PM	Microsoft Excel C...	1,314 KB
4.2 Primary and Secondary School students POI_TZ	7/17/2017 3:34 PM	Microsoft Excel C...	730 KB
4.3 Tertiary Students PUR_TZ	7/18/2017 8:55 PM	Microsoft Excel C...	1,324 KB
4.4 Tertiary Students POI_TZ	7/18/2017 8:53 PM	Microsoft Excel C...	708 KB
5.1 Labour Force_TZ	7/17/2017 10:20 PM	Microsoft Excel C...	1,015 KB
5.2 Labour Force Part Time Full Time_SA4	7/17/2017 10:26 PM	Microsoft Excel C...	6 KB
5.3 Labour Force Part Time Full Time_TZ	7/17/2017 10:34 PM	Microsoft Excel C...	682 KB
5.4 Labour Force Income Segmentation_TZ	7/17/2017 10:39 PM	Microsoft Excel C...	1,679 KB
5.5 ERP_NPD EMP_GMA	7/19/2017 3:11 PM	Microsoft Excel C...	2 KB
5.6 Workforce_Employed_SA3	7/5/2017 5:32 PM	Microsoft Excel C...	8 KB
6.1 Employment 2-digit Forecast_TZ	7/3/2017 1:26 PM	Microsoft Excel C...	10,958 KB
6.2 Employment SA3 Forecast_SA3	7/18/2017 2:16 PM	Microsoft Excel C...	353 KB

It is important to note:

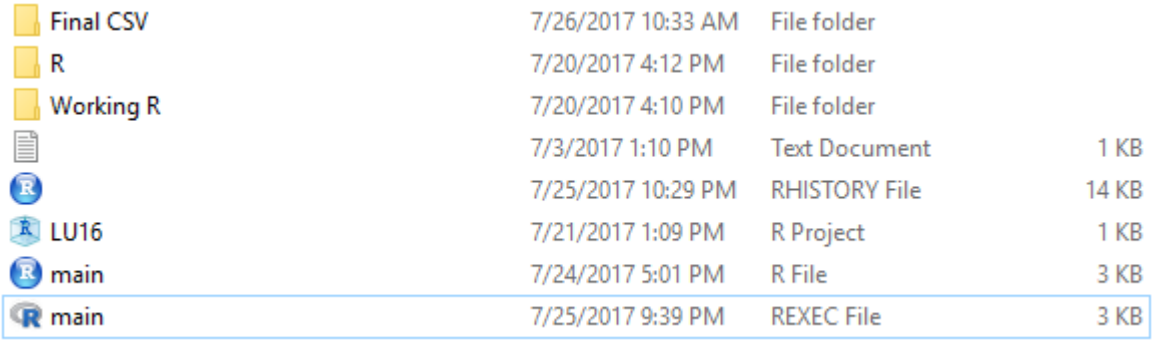
- Adding files to the folder will not automatically add them to any checks, or interfere with the checks;
- The files in this folder follow a naming convention;
    - The first number indicates the module that the data belongs to
    - The letters following the underscore at the end indicate the type of statistical area that the file contains; and
    - Each file has a unique ID in the form of a number #.# where # is one digit.

21.4.1.2 R Scripts

The R scripts are run to calculate the checks on the output files and produce the checks output. R Scripts can be found through the following path. This path contains the items shown in Figure 5 below.

LU16 Forecast -NEW STRUCTURE > O02_Outputs > Checks > R Scripts

Figure 20. Figure 37 Module output files within the R Scripts folder



The R Scripts folder contains the following sub-folders:

- **Final CSV -** The output of the automated checks is saved in this folder with a date and time stamp each time the module is run. The folder will also contain the results of any previous checks that have not been deleted.
- **R -** Where the detailed functions for the checks are kept. These are separate from the main.R file which links to these functions within its script. The structure of Module 7 is modularised in this fashion to make changes easier.
- **Working R -** contains previous versions of the R scripts in the checks. These are for documentation and are not used in the automated checks.

Only the main.REXEC file is used to run the scripts, as outlined below.

21.4.2 File Structure

21.4.2.1 Outputs CSV

The contents of the files in the Outputs CSV folder follow a certain format. This format must be maintained for the R scripts to run correctly. Changing the file names, contents or structure will cause references in the scripts to break. This format is outlined below.

21.4.2.1.1 Column titles

The format of the column title names should not be changed. The format of the column totals in the output '.csv' files is as follows:

- One column contains the geography, for example a travel zone or SA3. This column is titled with the name of the geography, for example TZ or SA3, and ends with _## where ## is two digits for example TZ_11 or SA3_16.
- All columns concerning a year end with #### where #### are the four digits that make up the year.
- Any other columns that are not the geography or a year, do not end with these formats: _## or ####.
- When updating the output files only geographies may be contained in the geographies column. If desired, totals may be added to the bottom of any other column, or any comments,
- Underscores separate different labels, and dots should separate the words within those labels. e.g. "Single.parent_2011" is correct, because "Single parent" is one

piece of information, and “2011” is another. But, “Single” and “parent” are not two different pieces of information, as they combined to make one label.

21.4.2.1.2 Cell contents

The format of the cell contents also requires a specific format as outlined below:

- There should be no commas in any of the numbers in the cells, for example “1,429” should be written “1429”. If the contents of the file are changed, any data copies and pasted into the file should be pasted in plain text format (Right click > Paste Special > Text). This will ensure that no commas are transferred and that the data is not rounded in the ‘.csv’ files which would cause inaccuracies in the results.
- If a cell needs to be left blank, use the character "-", or leave it empty.
- Space at the start or end of cell contents will not affect the outcome.

21.4.2.1.3 Data Source

The data source to manually update each spreadsheet for the check (whilst retailing the column and cell content structure outlined above) is outlined in Table 1 below.

Table 1. Source location for Outputs CSV file data

File name	Source module	Excel sheet/notes
0.0 Control Totals.csv	Sum of final outputs from all modules	
0.1 Historical Structural Private Dwellings_TZ.csv	Global input folder	This does not require updating
0.2 Historical ERP_TZ.csv	Global input folder	This does not require updating
0.3 Concordance.csv	Global input folder	This requires updating if travel zone geographies are changed
0.4 Adjusted LGA ERP.csv	Global input folder	TPA file
0.5 Adjusted LGA OPD.csv	Global input folder	TPA file
1.1 Structural Private Dwellings_TZ.csv	M01b	Final SPD Forecast
1.2 Occupied Private Dwellings_TZ.csv	M01b	Final OPD Forecast
1.3 Population Occupied Private Dwellings_TZ.csv	M01c	Final POPD Forecast

File name	Source module	Excel sheet/notes
1.4 Population Non Private Dwellings_TZ.csv	M01c	Final PNPD Forecast
1.5 ERP_SA3.csv	M01c	Final POPD Forecast Final PNPD Forecast
2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv	M02b	IPF Output
2.2 Age Groups (15-64+) _TZ.csv	M02b	IPF Output
2.3 Age by Sex 5 Year Age Groups (15-64+)_SA4.csv	M02b	IPF Output
3.1 Household Type_TZ.csv	M03d	IPF Output
3.2 Adjusted LGA Household Type.csv	M03d	Adjusted LDA Control Totals
3.3 IPF Seed Values_TZ.csv	M03c	(R automatically generated)
4.1 Primary and Secondary School students PUR_TZ.csv	M04c	Summary_P&S
4.2 Primary and Secondary School students POI_TZ.csv	M04e	(R automatically generated)
4.3 Tertiary Students PUR_TZ.csv	M04f	Origin enrolments
4.4 Tertiary Students POI_TZ.csv	M04f	Output
5.1 Labour Force_TZ.csv	M05e	TZ11 IPF Output
5.2 Labour Force Part Time Full Time_SA4.csv	M05e	TZ11 IPF Output summed to SA4
5.3 Labour Force Part Time Full Time_TZ.csv	M05g	TZ11 IPF Output
5.4 Labour Force Income Segmentation_TZ.csv	M05h	TZ11 IPF Output
5.5 ERP_NPD EMP_GMA.csv	M05d	Final SA4 LF Distribution
5.6 Workforce_Employed_SA3.csv	M05e	TZ11 IPF Output summed to SA3
6.1 Employment 2-digit Forecast_TZ.csv	M06d	Employment forecast

File name	Source module	Excel sheet/notes
6.2 Employment SA3 Forecast_SA3.csv	M06d	Employment forecast

21.4.2.2 R Scripts

The file which contains the R Scripts of Module 07 is “main.Rexec”. To view the contents, right-click “main.Rexec” and open with “Notepad”.

The structure of the file is as follows:

- Comments - The file is broken up by lines beginning with ‘#’ which indicate a comment. These comments are descriptions for the reader’s benefit and do not affect the functionality of the script.
- Running order - R scripts are run from top to bottom. The entire script should be run in order from the beginning to end, as many functions are dependent on previous outputs.
- File directory - How to change this file to change the directory is discussed in section **Error! Reference source not found.** of this manual.

The box below contains the code for the “main.REXEC” file.

```
# For .Rexec only, set the working directory

directory = file.path("/", "Melsvr", "company", "Synergy", "Projects", "2015
", "20150276 BTS Land Use Modelling Advice", "Model Development", "SGS Model
", "LU16 Forecast -NEW STRUCTURE")


setwd(

file.path(

directory,

file.path("O02_Outputs", "Checks", "R Scripts")

)

)


# Measure the time taken

start.time <- Sys.time()


# Load functions in other R files

source("R/Functions.R")

source("R/Negatives Check.R")

source("R/Control Totals.R")

source("R/Validations.R")
```

```
# Load Packages
load_packages()

# Load the Travel Zone Projections
# Note: If files is rerun, both files and files2 must be rerun!
files = load_TZP()
files2 = load_TZP_stacked()

# Negatives Check------------------------------------------------------------
write_csv(cbind(Name = names(files),
Check = "Check for negative values.",
create_output()
),
paste0("Final CSV/Negatives Check ", gsub(":", ".", Sys.time()), ".csv"))

# Control Totals Check-------------------------------------------------------

# Load the DPE totals to check against
DPE = read_csv(
file.path(
directory,
file.path("O01_TZP Database (CSV)", "Outputs CSV", "0.0 Control Totals.csv")
)
)

# Choose limit to account for rounding
total_limit = 10

write_csv(create_output_totals(),
paste0("Final CSV/Control Totals ", gsub(":", ".", Sys.time()), ".csv"))

# Validations Check----------------------------------------------------------

# Load the concordance to check against
CCD = read_csv(
file.path(
```

```
directory,
file.path("O01_TZP Database (CSV)", "Outputs CSV", "0.3 Concordance.csv")
)
)

# Load the adjusted LGAs for one check
LGA = read_csv(
file.path(
directory,
file.path("O01_TZP Database (CSV)", "Outputs CSV", "0.4 Adjusted LGA ERP.csv
")
)
)

# Choose a limit for the difference, to account for rounding errors
limit = 0.1

write_csv(create_output_geo(),
paste0("Final CSV/Validations ", gsub(":", ".", Sys.time()), ".csv"))

# Measure the time taken
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken

message(
paste0("The checks have been completed in ", round(as.numeric(time.taken), d
igits = 2), " minutes.")
)
```

21.4.3 How to Run the Automated Checks

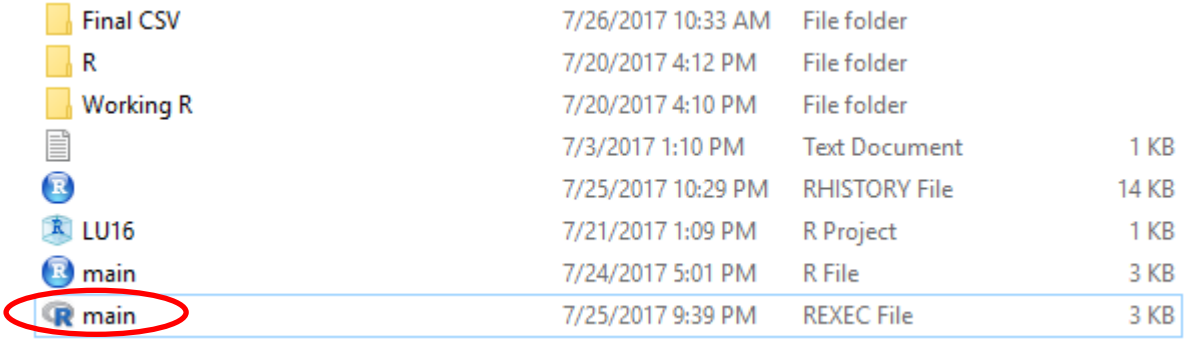
The user should navigate to the folder that contains the Land Use Model (Figure 6).

Figure 21. Module folders

Name	Date modified	Type	Size
A_TZP Model Control	7/21/2017 3:06 PM	File folder	
I01_Global Inputs	7/6/2017 2:13 PM	File folder	
M01_Dwellings & Population	7/20/2017 12:58 PM	File folder	
M02_Age by Sex	7/20/2017 12:58 PM	File folder	
M03_Households	7/17/2017 11:36 AM	File folder	
M04_Students	7/19/2017 1:36 PM	File folder	
M05_Workforce	7/17/2017 10:39 PM	File folder	
M06_Employment	7/18/2017 2:49 PM	File folder	
O01_TZP Database (CSV)	7/18/2017 6:14 PM	File folder	
O02_Outputs	7/7/2017 4:21 PM	File folder	

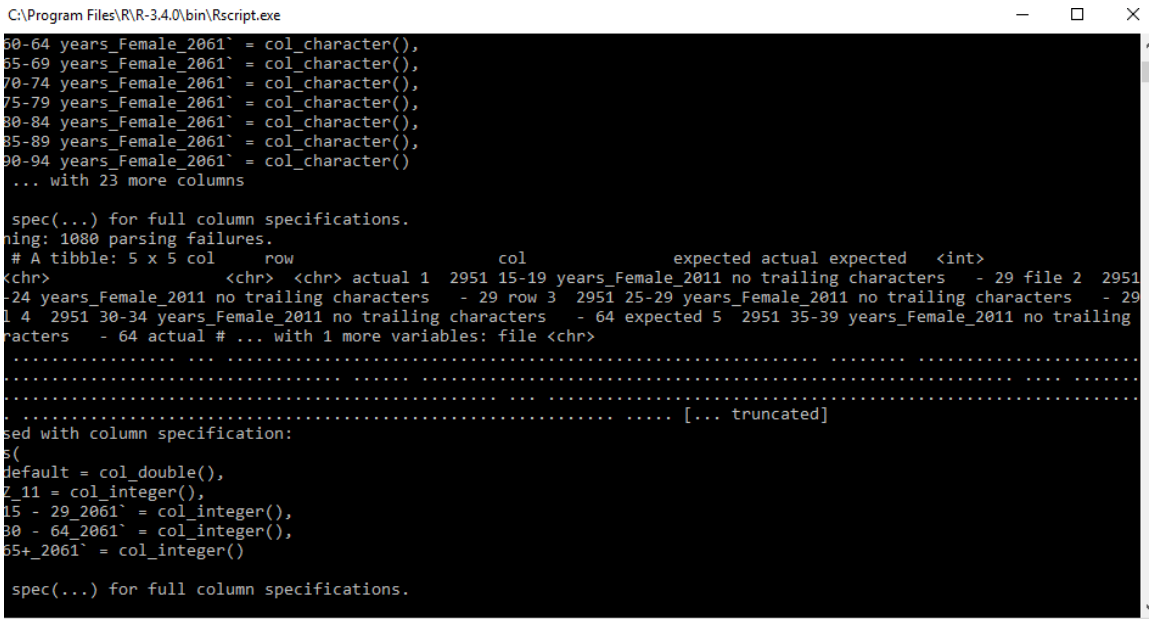
The user should double-click **O02_Outputs**, then **Checks**, then **R Scripts** to navigate to the folder shown in Figure 7.

Figure 22. Module 7 R Script folder



The user should Double-click the file called **main.REXEC** to run the module. The command window shown in Figure 8 will appear and display the script as it scrolls through the module. Once the script has completed the window will close.

Figure 23. Command prompt for checks module

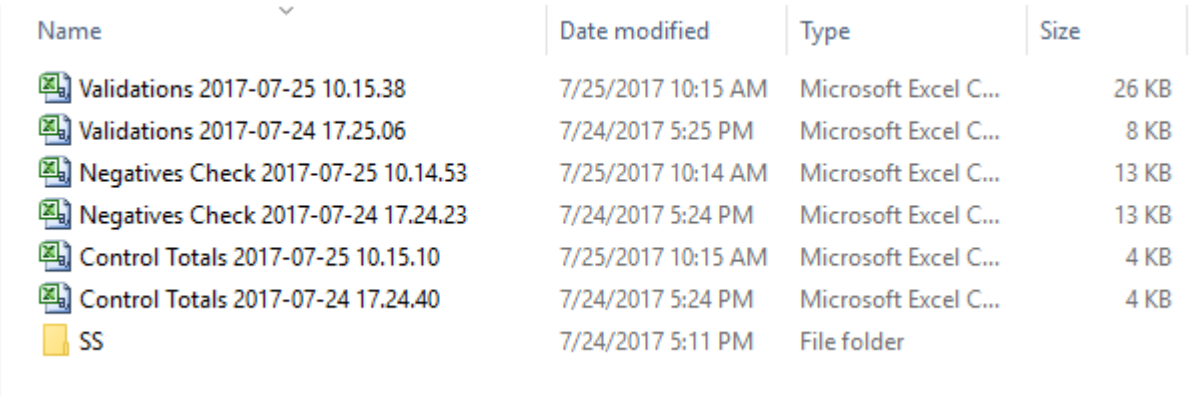


To view the results of the check, the user should navigate to the **Final CSV** folder. This folder should contain at least three timestamped files, the Negatives Check, Control Totals and Validations. The purpose of each is outlined below:

- **Negatives Check** – Checks for negative numbers in module output
- **Control Totals** – Various checks for overall consistency that the sum of figures in module output files is equivalent to DPE control totals.
- **Validations** – Various checks for consistency that the sum of figures in module output files is consistent for each sub-geography.

Additional files will be the results of previous checks (Figure 9).

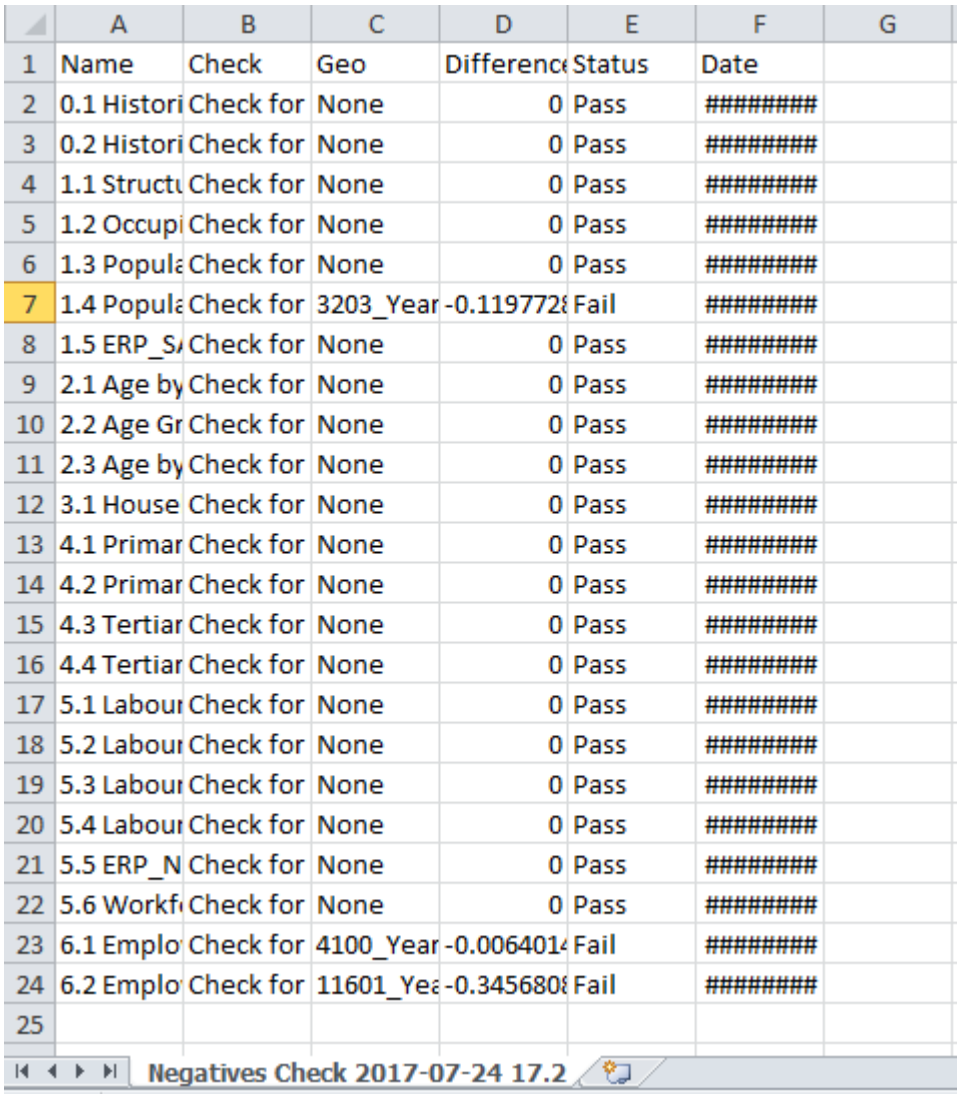
Figure 24. Check results file



21.4.4 Interpreting the Negatives Check File

Upon opening the Negatives Check file, the user will see the following (Figure 10).

Figure 25. Negative checks output



	A	B	C	D	E	F	G
1	Name	Check	Geo	Difference	Status	Date	
2	0.1 Histori	Check for	None	0	Pass	########	
3	0.2 Histori	Check for	None	0	Pass	########	
4	1.1 Structu	Check for	None	0	Pass	########	
5	1.2 Occupi	Check for	None	0	Pass	########	
6	1.3 Popula	Check for	None	0	Pass	########	
7	1.4 Popula	Check for	3203_Year	-0.119772	Fail	########	
8	1.5 ERP_S	Check for	None	0	Pass	########	
9	2.1 Age by	Check for	None	0	Pass	########	
10	2.2 Age Gr	Check for	None	0	Pass	########	
11	2.3 Age by	Check for	None	0	Pass	########	
12	3.1 House	Check for	None	0	Pass	########	
13	4.1 Primar	Check for	None	0	Pass	########	
14	4.2 Primar	Check for	None	0	Pass	########	
15	4.3 Tertiar	Check for	None	0	Pass	########	
16	4.4 Tertiar	Check for	None	0	Pass	########	
17	5.1 Labour	Check for	None	0	Pass	########	
18	5.2 Labour	Check for	None	0	Pass	########	
19	5.3 Labour	Check for	None	0	Pass	########	
20	5.4 Labour	Check for	None	0	Pass	########	
21	5.5 ERP_N	Check for	None	0	Pass	########	
22	5.6 Workf	Check for	None	0	Pass	########	
23	6.1 Emplo	Check for	4100_Year	-0.006401	Fail	########	
24	6.2 Emplo	Check for	11601_Yea	-0.345680	Fail	########	
25							

Negatives Check 2017-07-24 17.2

The user should expand Column A, Column B, and Column F to get the clearest results (Figure 11).

Figure 26. Expanded negative checks output

	A	B	C	D	E	F
1	Name	Check	Geo	Difference	Status	Date
2	0.1 Historical Structural Private Dwellings_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
3	0.2 Historical ERP_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
4	1.1 Structural Private Dwellings_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
5	1.2 Occupied Private Dwellings_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
6	1.3 Population Occupied Private Dwellings_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
7	1.4 Population Non Private Dwellings_TZ.csv	Check for negative values.	3203_Year	-0.119772	Fail	7/24/2017 17:24
8	1.5 ERP_SA3.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
9	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
10	2.2 Age Groups (15-64+) _TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
11	2.3 Age by Sex 5 Year Age Groups (15-64+)_SA4.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
12	3.1 Household Type_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
13	4.1 Primary and Secondary School students PUR_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
14	4.2 Primary and Secondary School students POI_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
15	4.3 Tertiary Students PUR_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
16	4.4 Tertiary Students POI_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
17	5.1 Labour Force_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
18	5.2 Labour Force Part Time Full Time_SA4.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
19	5.3 Labour Force Part Time Full Time_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
20	5.4 Labour Force Income Segmentation_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
21	5.5 ERP_NPD EMP_GMA.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
22	5.6 Workforce_Employed_SA3.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
23	6.1 Employment 2-digit Forecast_TZ.csv	Check for negative values.	4100_Year	-0.006401	Fail	7/24/2017 17:24
24	6.2 Employment SA3 Forecast_SA3.csv	Check for negative values.	11601_Yea	-0.345680	Fail	7/24/2017 17:24

A description of each column is provided below:

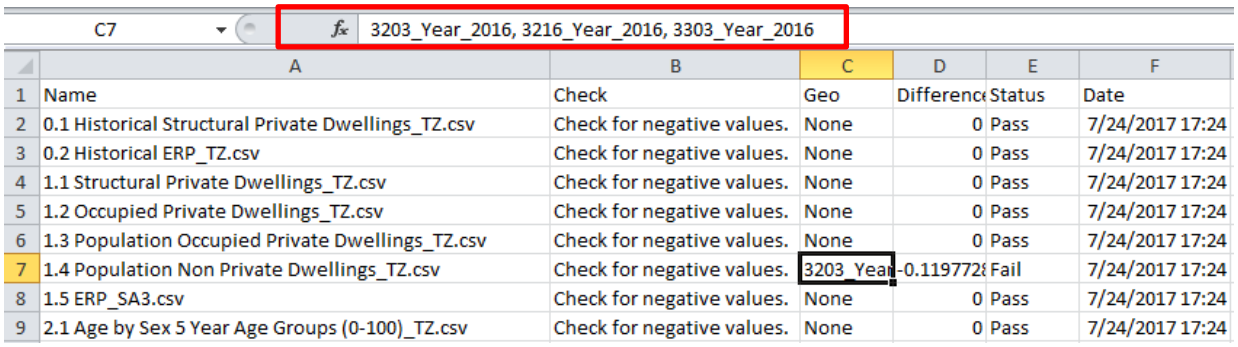
- **Name** – Contains the name of the .csv file containing module output.
- **Check** – Description of the calculation used for the check.
- **Geo** – Location of the negative value in output module. For example, “3203_Year_2016” indicates that the geography with the code 3203, in the year 2016, has the negative value.
- **Difference** – Records the value of the negative number detected.
- **Status** – indicates whether the file listed in Column A has passed the check. If the status says “Fail”, there are negative values in the file.
- **Date** – Records the date the check was run.

21.4.4.1 To check the results of the Negative Check in the Original File

21.4.4.1.1 Reading the checks output

To see all the geographies that have negative values in a file, the user should click on the relevant cell in the ‘Geo’ column. In Figure 9 below the user can see that the file “1.4 Population Non Private Dwellings_TZ.csv” has three negative values, and where those negative values are.

Figure 27. Example of negative checks output



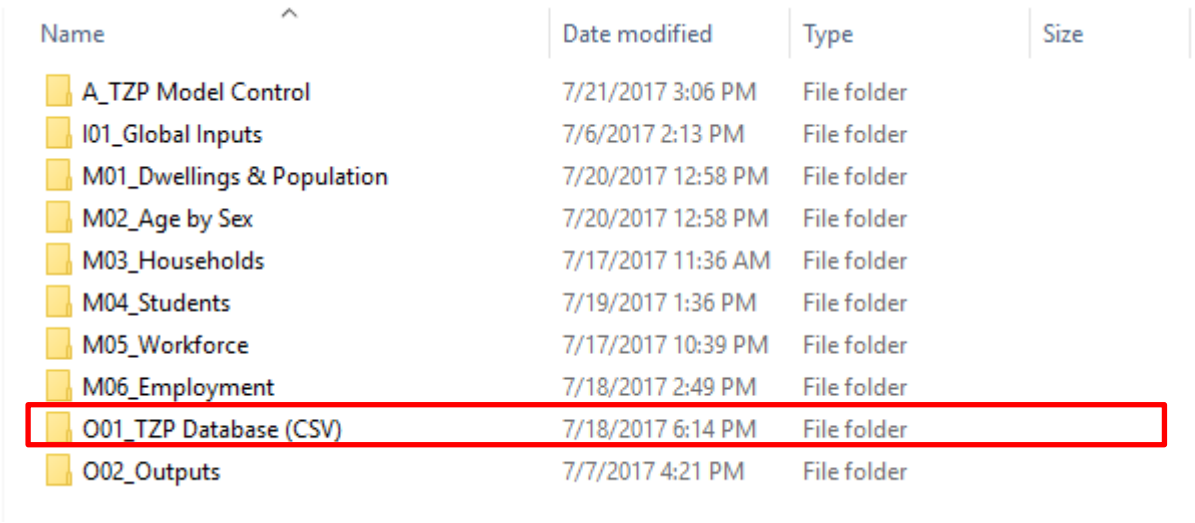
C7 | fx 3203_Year_2016, 3216_Year_2016, 3303_Year_2016

	A	B	C	D	E	F
1	Name	Check	Geo	Difference	Status	Date
2	0.1 Historical Structural Private Dwellings_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
3	0.2 Historical ERP_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
4	1.1 Structural Private Dwellings_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
5	1.2 Occupied Private Dwellings_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
6	1.3 Population Occupied Private Dwellings_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
7	1.4 Population Non Private Dwellings_TZ.csv	Check for negative values.	3203_Year	-0.119772	Fail	7/24/2017 17:24
8	1.5 ERP_SA3.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24
9	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv	Check for negative values.	None	0	Pass	7/24/2017 17:24

21.4.4.1.2 To Check the Results in the Original File

Navigate to the folder that contains the Land Use Model (Figure 13). The user should Double-Click **O01_TZP Database (CSV)**, then **Outputs CSV.**

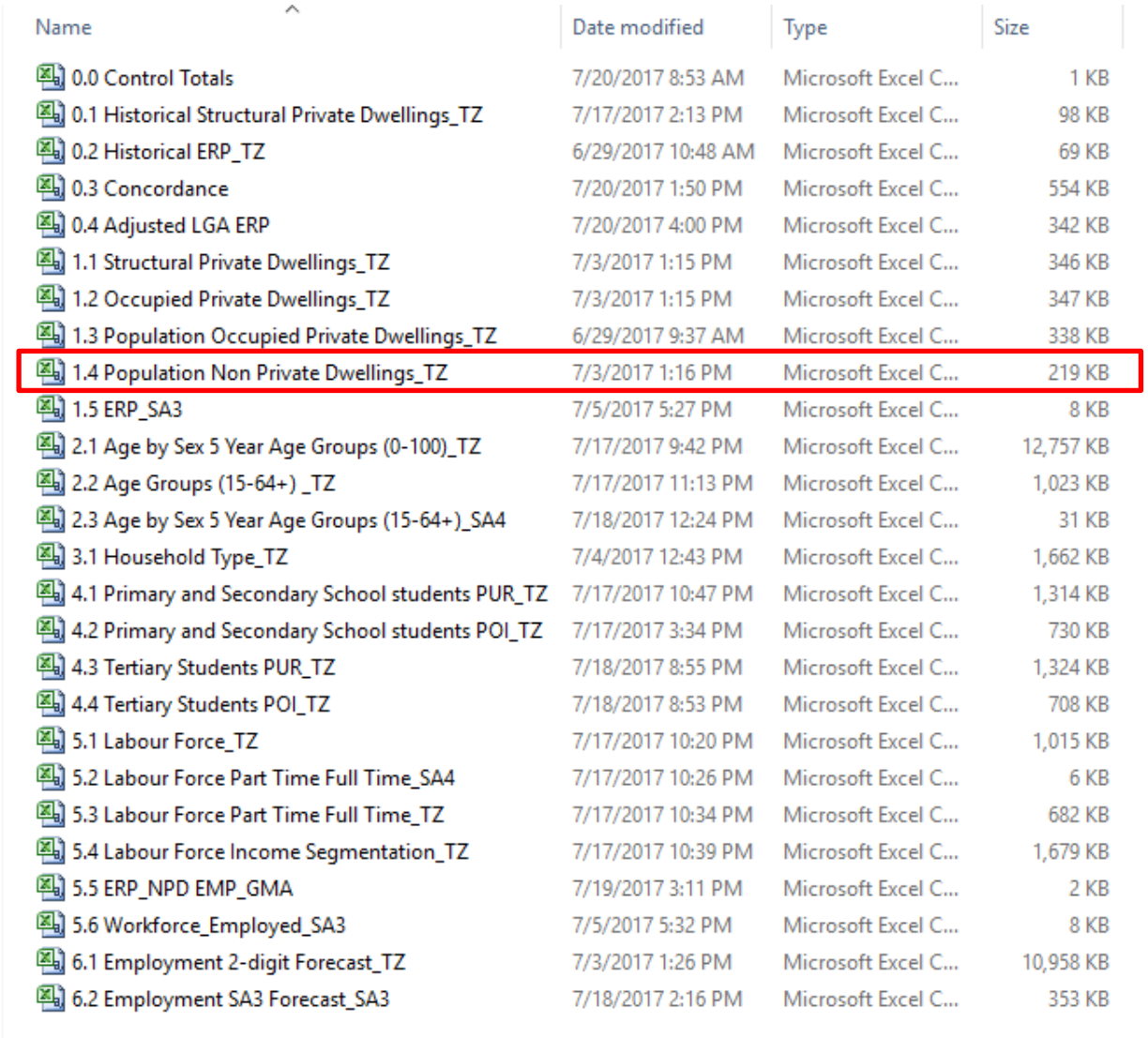
Figure 28. Module folders



Name	Date modified	Type	Size
A_TZP Model Control	7/21/2017 3:06 PM	File folder	
I01_Global Inputs	7/6/2017 2:13 PM	File folder	
M01_Dwellings & Population	7/20/2017 12:58 PM	File folder	
M02_Age by Sex	7/20/2017 12:58 PM	File folder	
M03_Households	7/17/2017 11:36 AM	File folder	
M04_Students	7/19/2017 1:36 PM	File folder	
M05_Workforce	7/17/2017 10:39 PM	File folder	
M06_Employment	7/18/2017 2:49 PM	File folder	
O01_TZP Database (CSV)	7/18/2017 6:14 PM	File folder	
O02_Outputs	7/7/2017 4:21 PM	File folder	

For this example, the user should open up file '1.4 Population Non Private Dwellings_TZ' (Figure 14).

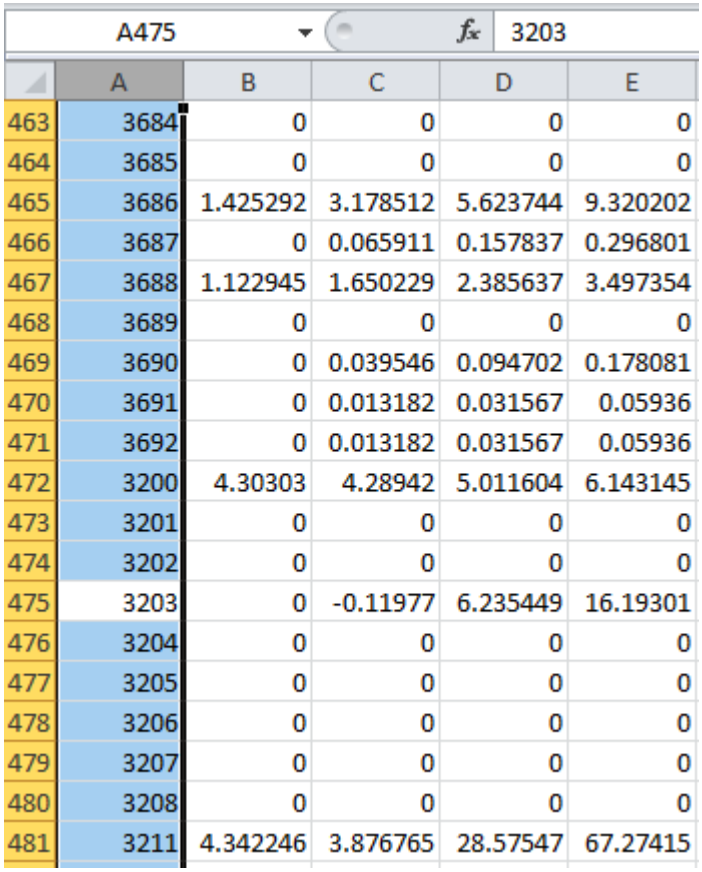
Figure 29. Module output '.csv' files



Name	Date modified	Type	Size
0.0 Control Totals	7/20/2017 8:53 AM	Microsoft Excel C...	1 KB
0.1 Historical Structural Private Dwellings_TZ	7/17/2017 2:13 PM	Microsoft Excel C...	98 KB
0.2 Historical ERP_TZ	6/29/2017 10:48 AM	Microsoft Excel C...	69 KB
0.3 Concordance	7/20/2017 1:50 PM	Microsoft Excel C...	554 KB
0.4 Adjusted LGA ERP	7/20/2017 4:00 PM	Microsoft Excel C...	342 KB
1.1 Structural Private Dwellings_TZ	7/3/2017 1:15 PM	Microsoft Excel C...	346 KB
1.2 Occupied Private Dwellings_TZ	7/3/2017 1:15 PM	Microsoft Excel C...	347 KB
1.3 Population Occupied Private Dwellings_TZ	6/29/2017 9:37 AM	Microsoft Excel C...	338 KB
1.4 Population Non Private Dwellings_TZ	7/3/2017 1:16 PM	Microsoft Excel C...	219 KB
1.5 ERP_SA3	7/5/2017 5:27 PM	Microsoft Excel C...	8 KB
2.1 Age by Sex 5 Year Age Groups (0-100)_TZ	7/17/2017 9:42 PM	Microsoft Excel C...	12,757 KB
2.2 Age Groups (15-64+) _TZ	7/17/2017 11:13 PM	Microsoft Excel C...	1,023 KB
2.3 Age by Sex 5 Year Age Groups (15-64+)_SA4	7/18/2017 12:24 PM	Microsoft Excel C...	31 KB
3.1 Household Type_TZ	7/4/2017 12:43 PM	Microsoft Excel C...	1,662 KB
4.1 Primary and Secondary School students PUR_TZ	7/17/2017 10:47 PM	Microsoft Excel C...	1,314 KB
4.2 Primary and Secondary School students POI_TZ	7/17/2017 3:34 PM	Microsoft Excel C...	730 KB
4.3 Tertiary Students PUR_TZ	7/18/2017 8:55 PM	Microsoft Excel C...	1,324 KB
4.4 Tertiary Students POI_TZ	7/18/2017 8:53 PM	Microsoft Excel C...	708 KB
5.1 Labour Force_TZ	7/17/2017 10:20 PM	Microsoft Excel C...	1,015 KB
5.2 Labour Force Part Time Full Time_SA4	7/17/2017 10:26 PM	Microsoft Excel C...	6 KB
5.3 Labour Force Part Time Full Time_TZ	7/17/2017 10:34 PM	Microsoft Excel C...	682 KB
5.4 Labour Force Income Segmentation_TZ	7/17/2017 10:39 PM	Microsoft Excel C...	1,679 KB
5.5 ERP_NPD EMP_GMA	7/19/2017 3:11 PM	Microsoft Excel C...	2 KB
5.6 Workforce_Employed_SA3	7/5/2017 5:32 PM	Microsoft Excel C...	8 KB
6.1 Employment 2-digit Forecast_TZ	7/3/2017 1:26 PM	Microsoft Excel C...	10,958 KB
6.2 Employment SA3 Forecast_SA3	7/18/2017 2:16 PM	Microsoft Excel C...	353 KB

The user then highlights the first column, and presses **Ctrl+F**, and types 3203 into the search window, then presses **Enter**. The user should see similar output to Figure 15.

Figure 30. Checking negatives location in module output file



A475 | fx 3203

	A	B	C	D	E
463	3684	0	0	0	0
464	3685	0	0	0	0
465	3686	1.425292	3.178512	5.623744	9.320202
466	3687	0	0.065911	0.157837	0.296801
467	3688	1.122945	1.650229	2.385637	3.497354
468	3689	0	0	0	0
469	3690	0	0.039546	0.094702	0.178081
470	3691	0	0.013182	0.031567	0.05936
471	3692	0	0.013182	0.031567	0.05936
472	3200	4.30303	4.28942	5.011604	6.143145
473	3201	0	0	0	0
474	3202	0	0	0	0
475	3203	0	-0.11977	6.235449	16.19301
476	3204	0	0	0	0
477	3205	0	0	0	0
478	3206	0	0	0	0
479	3207	0	0	0	0
480	3208	0	0	0	0
481	3211	4.342246	3.876765	28.57547	67.27415

Here the user can see that in Column C, which is for the year 2016, the PNPD in travel zone 3203 is negative, the same value displayed in the “Difference” column in the Negatives Check file seen previously.

***Note**: the file “0.4 Adjusted LGA ERP” will display the ages 5-9 and 10-14 as dates when opened in Excel. Do not change this format as it loads the age bands as expected in R.

21.4.5 Interpreting the Control Totals File

Upon opening the Negatives Check file, the user will see the following (Figure 16).

Figure 31. Example of control totals output

	A	B	C	D	E	F
1	Name	Check	Geo	Differenc	Status	Date
2	1.3 Population Occupied Private Dwellings_TZ.csv, 1.4 Population Nor	POPD + PNPD (Travel Zones) = ERP (SA3)		0	Pass	24/07/2017 17:24
3	0.0 Control Totals (ERP), 1.3 Population Occupied Private Dwellings_T	ERP (Total) = POPD + PNPD (Travel Zones)		0	Pass	24/07/2017 17:24
4	0.0 Control Totals (POPD), 2.1 Age by Sex 5 Year Age Groups (0-100)_T	POPD (Total) = All ages and sexes summed		0	Pass	24/07/2017 17:24
5	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.2 Age Groups (15-6	Ages 15-65+ = Ages 15- 65+		0	Pass	24/07/2017 17:24
6	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.2 Age Groups (15-6	Ages 15-19 to 25-29 = Ages 15-29		0	Pass	24/07/2017 17:24
7	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.2 Age Groups (15-6	Ages 30-34 to 60-64 = Ages 30-64		0	Pass	24/07/2017 17:24
8	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.2 Age Groups (15-6	Ages 65+ = Ages 65+		0	Pass	24/07/2017 17:24
9	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.3 Age by Sex 5 Year	Ages 15-19 to 65+ = Ages 15-24 to 65+		0	Pass	24/07/2017 17:24
10	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.3 Age by Sex 5 Year	Ages 15-19 to 20-24 = Ages 15-24		0	Pass	24/07/2017 17:24
11	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.3 Age by Sex 5 Year	Ages 25-29 to 30-34 = Ages 25-34		0	Pass	24/07/2017 17:24
12	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.3 Age by Sex 5 Year	Ages 35-39 to 40-44 = Ages 35-44		0	Pass	24/07/2017 17:24
13	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.3 Age by Sex 5 Year	Ages 45-49 to 50-54 = Ages 45-54		0	Pass	24/07/2017 17:24
14	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.3 Age by Sex 5 Year	Ages 55-60 to 60-64 = Ages 55-64		0	Pass	24/07/2017 17:24
15	2.1 Age by Sex 5 Year Age Groups (0-100)_TZ.csv, 2.3 Age by Sex 5 Year	Ages 65+ = Ages 65+		0	Pass	24/07/2017 17:24
16	3.1 Household Type_TZ.csv, 1.2 Occupied Private Dwellings_TZ.csv	With children (Single and Couple) + No children (Single and Couple		0	Pass	24/07/2017 17:24
17	5.1 Labour Force_TZ.csv, 2.2 Age Groups (15-64+) _TZ.csv	Employed + Unemployed + Not looking for work = Ages 15-64+		0	Pass	24/07/2017 17:24
18	5.1 Labour Force_TZ.csv, 5.4 Labour Force Income Segmentation_TZ.cs	Employed = All income brackets summed		0	Pass	24/07/2017 17:24
19	5.1 Labour Force_TZ.csv, 5.2 Labour Force Part Time Full Time_SA4.csv	Employed = Part-time + Full-time (SA4)		0	Pass	24/07/2017 17:24
20	5.1 Labour Force_TZ.csv, 5.3 Labour Force Part Time Full Time_TZ.csv	Employed = Part-time + Full-time (Travel Zone)		0	Pass	24/07/2017 17:24
21	0.0 Control Totals (EMP), 6.1 Employment 2-digit Forecast_TZ.csv	EMP(Total) = All industry employment summed		0	Pass	24/07/2017 17:24
22	0.0 Control Totals (EMP), 6.2 Employment SA3 Forecast_SA3.csv	EMP(Total) = All industry employment summed		0	Pass	24/07/2017 17:24
23	5.1 Labour Force_TZ.csv, 5.6 Workforce_Employed_SA3.csv	Employed = All income brackets summed		0	Pass	24/07/2017 17:24
24	0.0 Control Totals (POPD_15+), 2.2 Age Groups (15-64+) _TZ.csv	POPD 15+ (Total) = All ages 15+, and all sexes summed	NA	1733105	Fail	24/07/2017 17:24

A description of each column is provided below:

- **Name** – Contains the name of the .csv files containing module output. Each control totals check involves two or more files.
- **Check** – Description of the calculation used for the check.
- **Geo** – This column will be blank if the check passes, and display “NA”, meaning not applicable, if the check fails. The column does not display any geographies because the control totals are not reconciled on a geographical level. Instead, they compare the total sum of relevant values of one file to another.
- **Difference** – Records the magnitude of the difference calculation.
- **Status** – indicates whether the file listed in Column A has passed the check. If the status says “Fail”, the total sum of the relevant output file is not equivalent to the control totals.
- **Date** – Records the date the check was run.

***Note**: Check 24, “POPD 15+ (Total) = All ages 15+, and all sexes summed” for files “0.0 Control Totals (POPD_15+), 2.2 Age Groups (15-64+) _TZ.csv”, will fail. This is because the adjusted LGAs are known to have inaccuracies, and may be revised in the future.

21.4.5.1 To Check the Results of Control Totals Check in the Original File

To check the results in the original files, the user may sum columns and place the results at the bottom of the csv. Saving these calculations will not interfere with the checks, so long as the column with the geographies (the first column) remains empty.

For example, in Figure 17, we can see that there is content below the data in file 2.1. This will not interfere with the checks, because nothing is added to the first column:

Figure 32. Checking control total in module output file

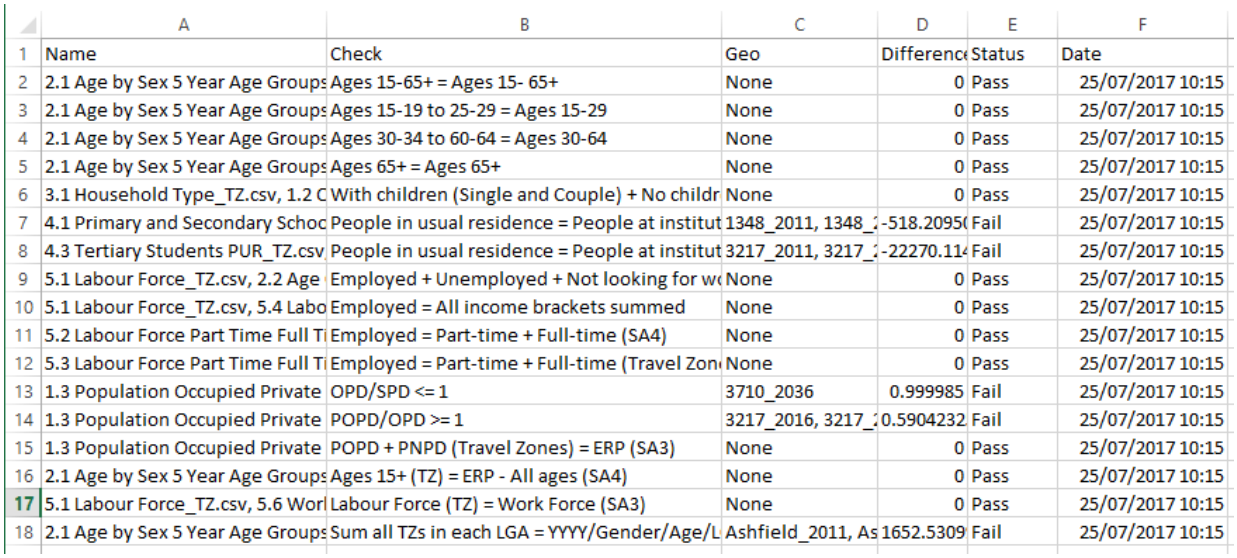
	A	B	C	D	E	F	G
2932	5180	44.94783	39.72495	28.37176	39.95539	48.04883	46.6849
2933	5183	6.92431	18.46717	8.852498	6.173735	10.27052	10.9853
2934	5184	2.583385	5.613992	5.589304	3.582994	0	3.278795
2935	5185	2.377002	6.339476	3.038917	2.119342	3.525701	3.771072
2936	5186	57.15294	82.12751	54.95722	37.83705	18.62302	34.38825
2937	5187	162.9209	162.8473	141.6515	109.728	67.49055	76.94077
2938	5188	15.39527	10.84776	4.340208	15.14995	21.0953	9.226968
2939	5189	70.29144	105.1697	108.7735	104.4534	55.85366	43.73182
2940	5190	33.72886	50.95492	61.46236	47.17293	18.91834	19.93519
2941	5191	4.122203	5.089785	17.22917	16.33497	6.551014	3.269901
2942	5192	62.68392	59.74601	49.77171	37.6798	28.40374	41.00274
2943	5193	17.07145	16.18832	10.74475	9.921806	7.234663	10.83342
2944	5194	73.96745	70.7037	45.8397	45.78089	49.56136	69.03784
2945	5195	15.6011	13.70503	6.814701	7.259836	10.4298	14.80546
2946	5198	138.1754	114.9233	161.8126	163.0441	122.2507	113.0258
2947	5199	9.711358	4.263414	10.61166	9.620749	10.28887	3.423755
2948	5200	61.53872	52.55559	49.90697	62.61558	68.4207	54.20222
2949	5201	26.03383	19.54393	34.81954	46.42368	24.27218	17.62228
2950	5202	40.93691	39.39981	44.38793	50.95367	56.71638	44.40731
2951							
2952					15 - 29	15 - 29	15 - 29
2953					15-24 year	15-24 year	25-34 year
2954					15 - 19	20 - 24	25 - 29
2955							
2956						4477076	

21.4.6 Interpreting the Validations File

The validations files is similar to the control totals with the addition of recording the difference between sub-geographies.

Upon opening the Validations file, the user will see the following (Figure 18).

Figure 33. Validation checks output



	A	B	C	D	E	F
1	Name	Check	Geo	Differenc	Status	Date
2	2.1 Age by Sex 5 Year Age Groups	Ages 15-65+ = Ages 15- 65+	None	0	Pass	25/07/2017 10:15
3	2.1 Age by Sex 5 Year Age Groups	Ages 15-19 to 25-29 = Ages 15-29	None	0	Pass	25/07/2017 10:15
4	2.1 Age by Sex 5 Year Age Groups	Ages 30-34 to 60-64 = Ages 30-64	None	0	Pass	25/07/2017 10:15
5	2.1 Age by Sex 5 Year Age Groups	Ages 65+ = Ages 65+	None	0	Pass	25/07/2017 10:15
6	3.1 Household Type_TZ.csv, 1.2 C	With children (Single and Couple) + No childr	None	0	Pass	25/07/2017 10:15
7	4.1 Primary and Secondary Schoc	People in usual residence = People at institut	1348_2011, 1348_2	-518.2095(	Fail	25/07/2017 10:15
8	4.3 Tertiary Students PUR_TZ.csv	People in usual residence = People at institut	3217_2011, 3217_2	-22270.114	Fail	25/07/2017 10:15
9	5.1 Labour Force_TZ.csv, 2.2 Age	Employed + Unemployed + Not looking for w	None	0	Pass	25/07/2017 10:15
10	5.1 Labour Force_TZ.csv, 5.4 Labo	Employed = All income brackets summed	None	0	Pass	25/07/2017 10:15
11	5.2 Labour Force Part Time Full T	Employed = Part-time + Full-time (SA4)	None	0	Pass	25/07/2017 10:15
12	5.3 Labour Force Part Time Full T	Employed = Part-time + Full-time (Travel Zon	None	0	Pass	25/07/2017 10:15
13	1.3 Population Occupied Private	OPD/SPD <= 1	3710_2036	0.999985	Fail	25/07/2017 10:15
14	1.3 Population Occupied Private	POPD/OPD >= 1	3217_2016, 3217_2	0.5904232	Fail	25/07/2017 10:15
15	1.3 Population Occupied Private	POPD + PNPD (Travel Zones) = ERP (SA3)	None	0	Pass	25/07/2017 10:15
16	2.1 Age by Sex 5 Year Age Groups	Ages 15+ (TZ) = ERP - All ages (SA4)	None	0	Pass	25/07/2017 10:15
17	5.1 Labour Force_TZ.csv, 5.6 Worl	Labour Force (TZ) = Work Force (SA3)	None	0	Pass	25/07/2017 10:15
18	2.1 Age by Sex 5 Year Age Groups	Sum all TZs in each LGA = YYYY/Gender/Age/L	Ashfield_2011, As	1652.5309	Fail	25/07/2017 10:15

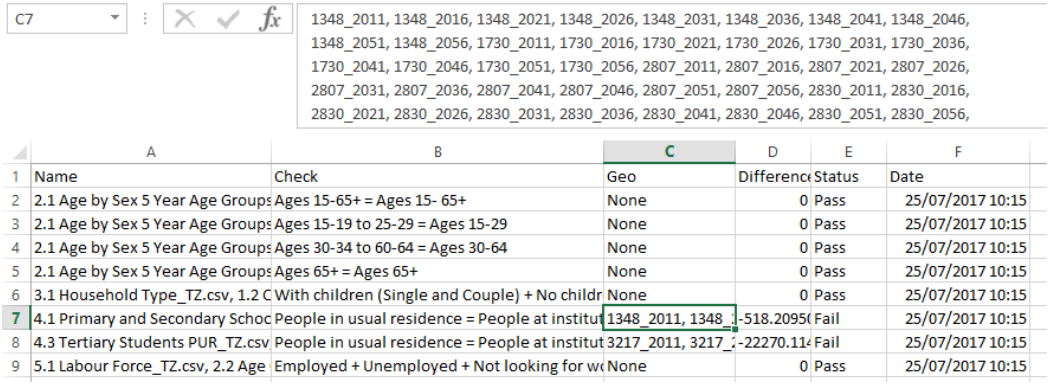
A description of each column is provided below:

- **Name** – Contains the name of the .csv files containing module output. Each validations check involves two or more files.
- **Check** – Description of the calculation used for the check.
- **Geo** – Location of the inconsistency in output modules. For example, “1348_2016” indicates that the geography with the code 1348, in the year 2016, is not consistent between the files.
- **Difference** – Records the magnitude of the difference calculation.
- **Status** – indicates whether the file listed in Column A has passed the check. If the status says “Fail”, the sum of figures for the relevant geographies are not consistent between module output files.
- **Date** – Records the date the check was run.

21.4.6.1 To Check the Results of the Validations Check in the Original File

To see all the geographies that are inconsistent in a file, the user should click on the relevant cell in the ‘Geo’ column. For example, in Figure 19 below the user can see that a number of travel zones in the files “4.1 Primary and Secondary School students PUR_TZ.csv” and “4.2 Primary and Secondary School students POI_TZ.csv” display inconsistencies.

Figure 34. Example of validation checks output



C7

1348_2011, 1348_2016, 1348_2021, 1348_2026, 1348_2031, 1348_2036, 1348_2041, 1348_2046, 1348_2051, 1348_2056, 1730_2011, 1730_2016, 1730_2021, 1730_2026, 1730_2031, 1730_2036, 1730_2041, 1730_2046, 1730_2051, 1730_2056, 2807_2011, 2807_2016, 2807_2021, 2807_2026, 2807_2031, 2807_2036, 2807_2041, 2807_2046, 2807_2051, 2807_2056, 2830_2011, 2830_2016, 2830_2021, 2830_2026, 2830_2031, 2830_2036, 2830_2041, 2830_2046, 2830_2051, 2830_2056,

	A	B	C	D	E	F
1	Name	Check	Geo	Differenc	Status	Date
2	2.1 Age by Sex 5 Year Age Groups	Ages 15-65+ = Ages 15- 65+	None	0	Pass	25/07/2017 10:15
3	2.1 Age by Sex 5 Year Age Groups	Ages 15-19 to 25-29 = Ages 15-29	None	0	Pass	25/07/2017 10:15
4	2.1 Age by Sex 5 Year Age Groups	Ages 30-34 to 60-64 = Ages 30-64	None	0	Pass	25/07/2017 10:15
5	2.1 Age by Sex 5 Year Age Groups	Ages 65+ = Ages 65+	None	0	Pass	25/07/2017 10:15
6	3.1 Household Type_TZ.csv, 1.2 C	With children (Single and Couple) + No childr	None	0	Pass	25/07/2017 10:15
7	4.1 Primary and Secondary Schoc	People in usual residence = People at institut	1348_2011, 1348_	-518.2095(	Fail	25/07/2017 10:15
8	4.3 Tertiary Students PUR_TZ.csv	People in usual residence = People at institut	3217_2011, 3217_	-22270.11	Fail	25/07/2017 10:15
9	5.1 Labour Force_TZ.csv, 2.2 Age	Employed + Unemployed + Not looking for w	None	0	Pass	25/07/2017 10:15

To check the results in the original files, the user should navigate to the relevant geography and year in the module output files and compare the two figures.

21.4.7 Changing the Script

At the top of the 'main.REXEC' script there are some parameters that may be adjusted by the user. The method to change file paths is the same for all R modules and detailed in section 21.1.4.1 above. Parameters that may be changed by the user are detailed in section 21.4.7.1 below.

21.4.7.1 Changing the Limits

There are buffers in the checks to account for rounding of data values. For the control totals, the limit is 10, and can be changed by amending the script below accordingly. For validations the limit is set to 0.2. To amend the script simply replace the number after the 'total_limit' or 'limit' as shown in the boxes below.

```
total_limit = ~~10~~ 20
```

```
limit = ~~0.2~~ 0.5
```

These limits can be changed to include or exclude certain failures depending on what is considered a significant difference - note the context of the limit adopted, 10 refers to 10 people, and 0.1 is a tenth of a person.

The different limits reflect the scale of the control totals i.e. differences are magnified when the total is composed of many smaller differences summed together.

In contrast the validation checks sum figures at smaller scales including geography and year for each individual area, so the potential differences between areas is a lot smaller.

22 Appendix F: TZP2016 v1.51 Population Methodology Adjustment

Using ABS Estimated Resident Population (ABS cat 3218.0) and 2016 Census data for Population in Occupied Private Dwellings (POPD), Population in Non-Private Dwellings (PNPD), Occupied Private Dwellings (OPD) and Structural Private Dwellings (SPD), SGS derived an updated control total for Adjusted LGAs[16] in 2016 that is used as the basis of the update for population, dwellings, workforce, employment and other variables by 2011 Travel Zones (TZ11).

The methodology for the update assumed that the Census information for the year 2016 would be the starting point for the projections with control totals based on DPE information kept constant from 2036 onwards and a smoothed transition between 2016 and 2036.

This update found discrepancies between the assumed split of POPD and PNPD in TZP2016 compared to the split obtained by using 2016 Census data[17].

The analysis of the difference indicated that the previous assumptions for PNPD in 2016 were different to the updated Census data for many LGAs in Sydney.

As a consequence, attempting to maintain the 2036 controls for POPD and PNPD resulted in distortions and issues with the projections at multiple levels of geography post 2036.

Essentially SGS has adjusted the ratio of POPD and PNPD to Adjusted LGA control totals in TZP2016 v1.51 that are reflective of the 2016 Census. This has resulted in changes to control totals for POPD and PNDP post 2036 to keep consistency in the ratios and the projections throughout.

It should be noted that no changes in ERP control totals has occurred post 2036.

The process undertaken to resolve and rectify these discrepancies and the impact on the TZP2016 v1.51 projections is outlined below:

[16] Adjusted LGAs are a custom geography created by SGS to align LGA boundaries with TZs.
[17] Census data extracted at an SA1 (ASGS 2016) level and aligned to 2011 travel zones

.

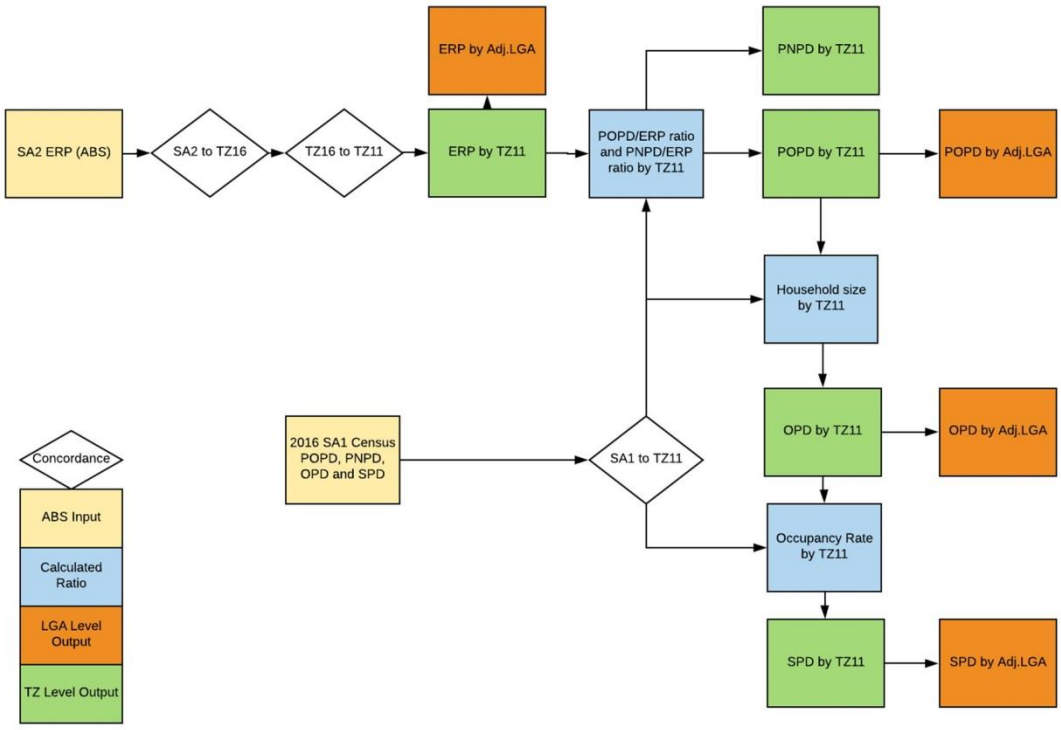
22.1.1 TZP2016 v1.51 Methodology

Rebased 2016 Adjusted LGA control totals were calculated for all variables – ERP, POPD, PNPD and OPD.

Assumptions:

- ERP calculated via ABS Estimated Resident Population (ABS cat 3218.0)
- Travel zone distributions for PNPD and POPD as well as occupancy rates and household sizes were derived using a SA1 16 to TZ11 concordance
- Ratios of POPD/PNPD and occupancy rates and household sizes applied as per ABS Census 2016[18]
- Control totals for POPD and PNPD change post 2036, control totals for ERP and OPD do not change post 2036
- 2016 OPD control totals were updated and then smoothed to meet the existing 2036 control totals using a smoothed household size

Error! Reference source not found. shows a flow chart on how the Adjusted LGA control totals ere calculated.



[18] 2016 Census - Counting Persons, Place of Enumeration (MB) SA1 by DWTD Dwelling Type; Counting: Persons Location on Census Night and 2016 Census - Counting Dwellings, Place of Enumeration (MB) SA1 by DWTD Dwelling Type; Counting: Dwellings Location on Census Night

This methodology allows consistency with the 2016 Census through the projection period. The table below shows how control totals change after 2036.

Table 13: Variables after 2036

Variable	Control totals change after 2036?
ERP	No
POPD	Yes
PNPD	Yes
OPD	No

22.1.2 Comparison

This distribution was then applied to ERP by travel zone to get travel zone POPD, PNPD, OPD and SPD that aligns with ERP. These values were then aggregated to Adjusted LGA to get the rebased 2016 control totals.

Rebasing 2016 control totals across multiple variables resulted in many areas having significantly different POPD/PNPD splits when compared to the previous version as shown in Table 14.

Table 14: comparison of 2016 and 2011 control total vs census values

Sydney LGA	Control Totals vs Census (2016)		
	POPD	PNPD	ERP
TZP2016 v1.51	198,800	30,101	228,900
TZP2016	208,900	8,850	217,760
2016 Census	198,950	50,560	222,717

N.B CENSUS ERP IS CALCULATED DIFFERENTLY THAN A SIMPLE SUMMATION OF POPD AND PNPD.

The control total for Sydney LGA PNPD is lower than in the Census due to the method used to derive travel zone population and dwelling estimates that aligns with ERP and prioritises the allocation of POPD.

The TZP2016 projections of POPD and PNPD were provided to TPA by DPE at an LGA level. TPA then converted this information to Adjusted LGA before providing to SGS.

A review of this data shows that the ratios of POPD and PNPD relative to ERP were relatively constant across forecast years between 2016 and 2056.

Table 15 shows a comparison of the ratios in TZP2016 versus TZP2016 v1.51. The data shows a strong alignment with the exception of Sydney LGA which has a lower ratio of POPD in TZP1.5 and therefore a bigger proportion of PNPD as indicated in Table 14 above. At a Sydney GMA level the overall difference at 2016 is less than 1%.

A review of Census data for Sydney LGA through table builder shows that when removing overseas visitors that the percentage of residents in an Occupied Private Dwelling is approximately 85% which aligns the POPD ratio used in TZP2016 v1.51 for the Sydney LGA.

Table 15: Comparison of TZP2016 and TZP2016 v1.51 POPD as % of ERP

Adj LGA Code	Adjusted LGA Name	2016	2056	TZP2016 diff 2016-2056	TZP2016 v1.51 Constant Ratio	TZP2016 v TZP2016 v1.51 at 2016
1	Ashfield	95.95%	92.79%	3.16%	95.76%	0.20%
2	Auburn	97.37%	96.73%	0.64%	96.29%	1.08%
3	Bankstown	98.66%	97.72%	0.94%	98.20%	0.46%
4	Blacktown	99.00%	97.78%	1.22%	98.96%	0.05%
5	Blue Mountains	98.28%	96.42%	1.86%	96.53%	1.75%
6	Botany Bay	98.81%	98.05%	0.77%	96.44%	2.38%
7	Burwood	96.72%	95.06%	1.66%	97.43%	-0.71%
8	Camden	99.08%	97.30%	1.78%	99.26%	-0.18%
9	Campbelltown	98.98%	97.45%	1.53%	98.56%	0.42%
10	Canada Bay	98.99%	98.16%	0.83%	98.95%	0.05%
11	Canterbury	99.24%	98.72%	0.52%	99.09%	0.15%
12	Cessnock	97.47%	95.39%	2.08%	96.07%	1.41%

13	Dungog	98.84%	97.68%	1.16%	98.49%	0.35%
14	Fairfield	99.13%	97.72%	1.41%	98.93%	0.20%
15	Gosford	98.47%	96.69%	1.78%	98.01%	0.47%
16	Great Lakes	97.74%	94.83%	2.91%	96.71%	1.03%
17	Hawkesbury	98.27%	96.91%	1.36%	97.94%	0.34%
18	Holroyd	99.12%	98.28%	0.83%	98.82%	0.30%
19	Hornsby	98.19%	95.97%	2.22%	98.16%	0.03%
20	Hunters Hill	91.38%	84.53%	6.84%	88.77%	2.61%
21	Hurstville	98.85%	97.57%	1.28%	98.81%	0.04%
22	Kiama	98.49%	96.10%	2.39%	98.17%	0.32%
23	Kogarah	99.14%	98.29%	0.85%	98.27%	0.87%
24	Ku-ring-gai	98.27%	97.09%	1.18%	97.69%	0.58%
25	Lake Macquarie	98.34%	96.35%	2.00%	98.21%	0.14%
26	Lane Cove	98.46%	97.37%	1.08%	96.81%	1.64%
27	Leichhardt	97.89%	96.88%	1.02%	97.76%	0.13%
28	Liverpool	98.86%	97.54%	1.32%	98.78%	0.08%

29	Maitland	99.09%	97.78%	1.31%	98.40%	0.69%
30	Manly	99.22%	98.98%	0.23%	97.06%	2.15%
31	Marrickville	97.43%	96.39%	1.04%	97.32%	0.11%
32	Mosman	98.59%	97.27%	1.32%	98.22%	0.38%
33	Newcastle	97.48%	96.01%	1.46%	95.92%	1.55%
34	North Sydney	98.54%	98.00%	0.54%	97.00%	1.54%
35	Parramatta	97.85%	96.67%	1.18%	97.29%	0.56%
36	Penrith	98.65%	97.33%	1.32%	98.20%	0.45%
37	Pittwater	98.58%	96.69%	1.89%	98.48%	0.10%
38	Port Stephens	98.75%	97.33%	1.41%	97.31%	1.44%
39	Randwick	95.69%	94.78%	0.91%	93.81%	1.87%
40	Rockdale	98.58%	97.28%	1.30%	97.67%	0.90%
41	Ryde	97.39%	96.16%	1.23%	96.88%	0.51%
42	Shellharbour	99.08%	97.37%	1.71%	98.82%	0.26%
43	Shoalhaven	97.48%	94.99%	2.48%	96.69%	0.79%
44	Singleton	97.90%	96.38%	1.52%	95.52%	2.38%

45	Strathfield	98.37%	97.23%	1.14%	98.48%	-0.11%
46	Sutherland Shire	99.05%	97.89%	1.17%	98.54%	0.51%
47	**Sydney**	**95.93%**	**95.10%**	**0.84%**	**86.85%**	**9.08%**
48	The Hills Shire	99.22%	98.40%	0.82%	99.14%	0.09%
49	Warringah	98.67%	97.33%	1.34%	98.18%	0.49%
50	Waverley	98.58%	98.05%	0.53%	95.94%	2.64%
51	Willoughby	98.92%	98.03%	0.88%	96.35%	2.57%
52	Wingecarribee	97.25%	94.39%	2.86%	96.28%	0.97%
53	Wollondilly	99.39%	98.32%	1.06%	98.72%	0.67%
54	Wollongong	98.13%	96.67%	1.46%	97.07%	1.06%
55	Woollahra	97.76%	97.30%	0.46%	96.46%	1.30%
56	Wyong	98.60%	97.05%	1.54%	98.20%	0.39%
	Sydney GMA	98.33%	97.08%	1.26%	97.40%	0.93%

23 Appendix G: TZP2016 v1.51 Employment Adjustments

Following the release of TZP2016 v1.5 to selected clients some problems with the allocation of employment were uncovered. This initial issue centred around the airport travel zones (433, 434 and 435).

After discussion with officers at RMS that uncovered this issue it was decided to further investigate the allocation of employment. A rule was proposed, with analysis confined to TZs where there were large revisions (>1,000) for 2016 between TZP2016 v1.3 and v1.5 and where the TZ was from an ABS DZN that had been split.

This rule led to the identification of 35 potential TZs. Further analysis identified 17 of these as requiring adjustment. In order to ensure that employment changes by industry did not affect control totals offset and identical adjustments were needed in other travel zones within the same SA2. The result was 28 TZs that were adjusted.

Travel Zones affected

TZs 208 and 209 - Central station/Ultimo adjustment

- The initial split in TZP2016v1.5 had 1394 persons employed in TZ208. This TZ is basically the entrance to Central station and comprises trainlines only along with a small bus depot. It was decided to adopt an estimate closer to TZP2016v1.3 and put only 59 employed people in this TZ and move the balance to TZ 209 where the estimate of employment increased from 900 to 1418. As with most adjustments the proportional adjustments were implemented at the industry level, moving employment shares by industry from 1 zone to the other.

TZs 433, 434 and 435 - the airport adjustment

The initial split of the airport TZs in 2016 had 7360 persons employed in TZ433, 4286 in TZ434 and 6831 in TZ435, the airport runway. The meshblock boundaries which define the airport TZs were concluded to be causing problems and it was decided to use the TZ16v1.3 split and return to the old airport boundaries which make more sense.

As a consequence total airport employment was adjusted so that employment in TZ433 became 9553, employment in TZ434 became 8479 and employment in TZ435 became 447 persons. Table 1 provides details for the split before an after adjustment for total employment for the period 2011 to 2036. Adjustment were made out to 2056 but are not included in order to make the tables more readable.

Table 1: Employment and employment shares - airport TZ

Initial employment - TZP2016v1.5						
TZs	2011	2016	2021	2026	2031	2036
433	10366	7360	7829	8043	8410	8746
434	6238	4286	4664	4824	5076	5313
435	328	6832	7690	8022	8503	8969
Total	16932	18478	20183	20889	21988	23027
Revised - adjusted employment (TZP2016v1.51)						
433	10366	9553	10434	10799	11367	11904
434	6238	8479	9262	9586	10090	10567
435	328	446	487	505	531	556
Total	16932	18478	20183	20889	21988	23027

As with most adjustments the proportional adjustments were implemented at the industry level, moving employment shares by industry from 1 zone to the other.

Parramatta CBD adjustments

Several adjustments were made around the Parramatta CBD where a combination of issues with the concordance and issues with the input data were uncovered.

1. Adjustments to TZ 1023 (St Patricks Primary School) and TZ1025 (Parramatta Stadium)

The employment split in TZP2016v1.5 between TZ1023 and TZ1025 was almost the reverse of the previous version of the database. After investigation it was concluded the new split was unreasonable and the old concordance was used to split the DZN data with 34% of the total going to TZ1023 and 66% to TZ1025. Again, this adjustment was done at the industry level for all years out to 2056.

Table 2: Adjustments to TZs 1050 and 1053

Initial employment - TZP2016v1.5						
TZs	2011	2016	2021	2026	2031	2036
1023	668	1548	1618	1803	1944	2101
1025	1297	774	804	885	969	1041
Totals	1965	2322	2421	2689	2913	3142
Revised - adjusted employment (TZP2016v1.51)						
1023	668	790	823	914	990	1068
1025	1297	1533	1598	1774	1922	2074
Total	1965	2322	2421	2689	2913	3142

2. Adjustments to TZ 1051 (Parramatta Park) and TZ1053 (St Johns Cathedral)

Investigation indicated a large revision to employment in TZ 1051 from the TZ16v1.3 estimate. It was discovered that the cause of this was the underlying data - DZN data from ABS census, which had over 3,000 persons employed in it. It was concluded this was a misclassification problem by the ABS and after investigation this employment was moved to TZ1053 using the ratio between employment in this two TZs in the TZ16v1.3 database.

Table 2 provides details fo the split before an after adjustment for total employment for the period 2011 to 2036. Adjustment were made out to 2056 but are not included in order to make the tables more readable. Again, adjustments were made at the 33 industry level of the TZP database.

Table 3: Adjustments to TZs 1050 and 1053

Initial employment - TZP2016v1.5						
TZs	2011	2016	2021	2026	2031	2036
1051	115	3457	3882	4107	4569	4910
1053	7661	5522	5997	6406	7116	7660
Total	7776	8979	9879	10513	11686	12570
Revised - adjusted employment (TZP2016v1.51)						
1051	115	132	145	154	172	185
1053	7661	8847	9734	10359	11514	12385
Total	7776	8979	9879	10513	11686	12570

In moving from the TZ11 to TZ16 geography a further adjustment was made with TZ 1053 being split to TZ1050 and TZ1053. This split is made at an industry level using the ratio for employment in the TZ11 to TZ16 concordance book for the split of TZ1053.

TZs 2829 and 2831 - the Lucas Heights adjustment

Travel Zones 2829 (Lucas Heights North) and 2831 (Lucas Heights Research Laboratories) are split from the same DZN. The concordance used in TZP2016v1.5 allocated 90% of employment to TZ2829, while in TZP2016v1.3 employment was split 50/50. And this split was adopted in the revision.

TZs 3685 and TZ 3689 - The Oran Park adjustment

Travel zones 3685 and 3689 along with 4 other travel zones are split from one DZN. 3685 contains the Oran Park Town Centre, which in 2016 contained a majority of the employment of the two travel zones, this split will change over time as TZ3689 starts to develop including the construction of schools and a local centre as Catherine Fields (part) is developed. As a result in 2016 a majority of employment was proportioned to TZ3685, with that shifting in 2021 to a 75% (3685) to 25% (3689) split across the major industries of employment.

TZs 3934 and 3935 - the Riverstone adjustment

Table 4 provides details of the split of total employment between TZ3934 and 3935 around Riverstone in TZP2016 v1.3 and TZP2016v1.5. Analysis indicated little change in information provided in mapping data, with most employment apparently still in TZ3935 where an industrial park is located. In comparison, the only employment visible from googlemaps for TZ 3934 is a sewerage treatment plant.

Table 4: Employment in TZs 3934 and 3935

	Employment - TZP2016 v1.3		Employment - TZP2016 v1.5	
TZ	No.	- % -	No.	- % -
3934	21.6	1.0%	1293.7	60.3%
3935	2047.5	99.0%	852.8	39.7%
Total	2069.1	100.0%	2146.5	100.0%

Based on the mapping information is was decided to revise employment, again using the split adopted in TZP2016v1.3 which resulted in moving eployment from TZ3934 to 3934. The initial and revised estimates of employment in 2016 and the projected employment out until 2036 only, are shown in table 54.

Table 5: Adjustments to TZs 3934 and 3935

Initial employment - TZP2016v1.5						
TZs	2011	2016	2021	2026	2031	2036
3934	20	1294	2120	3077	4916	5879
3935	1900	853	1280	1574	1909	2377
Total	1920	2147	3400	4651	6826	8255
Revised - adjusted employment (TZP2016v1.51)						
3934	20	65	103	141	207	250
3935	1900	2082	3297	4510	6619	8006
Total	1920	2147	3400	4651	6826	8255

TZs 4710 and 4711 - the Windsor adjustment

Table 6 provides details of the split of total employment between TZ4710 and 4711 around Windsor in TZP2016 v1.3 and TZP2016v1.5.

Table 6: Employment in TZs 4710 and 4711

	Employment - TZP2016 v1.3		Employment - TZP2016 v1.5	
TZ	No.	- % -	No.	- % -
4710	1600	50.0%	160	5.1%
4711	1600	50.0%	2974	94.9%
Total	3199	100.0%	3134	100.0%

Visual analysis of information from maps indicated there was little to like about both allocations (see figure 1). TZ4711 (South Windsor Industrial area) is apparently designated for light industry and there is little unused space. In contrast, Windsor station is apparently residential and there would not be expected to be much employment there except for retail, education, health and other services.

On this basis it was decided to allocate about 500 jobs there in 2016, which is 16% of the total jobs of the combined TZs and to apply that split in all years until 2016. Table 7 shows the initial and adjusted split for total employment for these TZs from 2011 to 2036.

Table 7: Adjustments to TZs 4710 and 4711

Initial employment - TZP2016v1.5						
TZs	2011	2016	2021	2026	2031	2036
4710	1554	160	165	167	172	181
4711	1554	2974	3013	3027	3097	3237
Total	3108	3134	3178	3194	3269	3419
Revised - adjusted employment (TZP2016v1.51)						
4710	1554	500	507	510	522	545
4711	1554	2634	2671	2684	2748	2873
Total	3108	3134	3178	3194	3269	3419

Figure 1: Map of Windsor travel zones

TZs 5801 and 5803 - the Albion Park

Table 8 provides details of the split of total employment between TZ5801 and 5803 around Albion Park in TZP2016 v1.3 and TZP2016v1.5. Once again, visual analysis of the travel zones was undertaken and it was decided on this basis to move the distribution of employment back to something like that from the TZP2016v1.3 database.

Table 8 Employment in TZs 5801 and 5803

	Employment - TZP2016 v1.3		Employment - TZP2016 v1.5	
TZ	No.	- % -	No.	- % -
5801	149	9.0%	1337	88.6%
5803	1509	91.0%	172	11.4%
Total	1658	100.0%	1509	100.0%

The distribution of employment (from 2011 to 2036 only) before and after adjustment, is shown in table 9.

Table 9: Adjustments to TZs 5801 and 5803

Initial employment - TZP2016v1.5						
TZs	2011	2016	2021	2026	2031	2036
5801	132	1,337	1,487	1,585	1,709	1,839
5803	1,334	172	189	199	213	229
Total	1,466	1,508	1,676	1,783	1,922	2,067
Revised - adjusted employment (TZP2016v1.51)						
5801	132	136	151	160	173	186
5803	1,334	1,373	1,525	1,623	1,749	1,882
Total	1,466	1,508	1,676	1,783	1,922	2,067

TZs 6323 and 6326 - The Kotara Station adjustment

The adjustments to TZ 6323 (John Hunter Hospital) and TZ 6326 (Kotara Station_North) follows the same pattern as that of the adjustments to TZ 1051 (Parramatta Park) and TZ1053 (St Johns Cathedral). Analysis of the input data identified a large increase in the employment assigned to one DZN which included TZ6326. This TZ is largely a nature reserve. A similar decline in employment was registered in the John Hunter Hospital precinct between TZP2016v1.3 and TZP2016v1.5 for 2016.

The simplests adjustment possible was chosen, with the updated data for TZP2016 v1.5 reallocated based on the shares of employment recorded in the previous version of the database. A visual inspection of maps resulted in a conclusion that the ABS had miscoded employments in this area. The distribution of employment over the 2011 to 2036 range is shown in table 10.

As a final consideration, in moving from TZ11 to TZ16 geography TZ 6323 is split. In this case the concordance ratio was applied to the totals of TZ6323 after adjustment to create a new or adjusted TZ6324.

Table 10: Adjustments to TZs 5801 and 5803

Initial employment - TZP2016v1.5						
TZs	2011	2016	2021	2026	2031	2036
6323	6088	4654	4660	4759	4988	5188
6326	609	3038	3038	3110	3257	3389
Total	6697	7692	7698	7868	8245	8577
Revised - adjusted employment (TZP2016v1.51)						
6323	6088	7015	7021	7176	7519	7822
6326	609	677	677	692	725	754
Total	6697	7692	7698	7868	8245	8577

TZs 6372 and 6373 - The Stockton adjustment

In contrast, the problems identified with the distribution between TZ 6372 (Newcastle Golf Course) and TZ6373 (Stockton) were traced back to the concordance. As table 11 indicates, once again it appears the concordance was allocating data in almost the opposite proportions to the previous version of the model.

Table 11: Employment in TZs 6372 and 6373

	Employment - TZP2016 v1.3		Employment - TZP2016 v1.5	
TZ	No.	- % -	No.	- % -
5801	357	17.0%	1830	93.2%
5803	1745	83.0%	133	6.8%
Total	2102	100.0%	1963	100.0%

After a visual inspection of mapping information, it was decided to use the previous split of employment and the estimates were adjusted. Table 12 provides the estimates for the 2011 to 2036 period.

Table 10: Adjustments to TZs 5801 and 5803

Initial employment - TZP2016v1.5						
TZs	2011	2016	2021	2026	2031	2036
6372	329	1830	1864	1921	2009	2099
6373	1607	133	133	137	142	149
Total	1936	1963	1997	2058	2152	2248
Revised - adjusted employment (TZP2016v1.51)						
6372	329	334	340	350	366	382
6373	1607	1629	1658	1708	1786	1866
Total	1936	1963	1997	2058	2152	2248

TZs 6744, 6745, 6746, 6747, 6748 and 6749 - The Singleton adjustments

The process adopted to investigate the revisions to the database uncovered a number of issues around Singleton. Further investigation, in particular the visual inspection of mapping information with TZ boundaries overlaid, indicated a considerable divergence between the employment allocation and what was on the ground according to the details revealed in the maps. For example, mapping information indicated that two large opencut coal mines were just inside the North Western boundary of Wattle ponds, a fact clearly missed in the Wattle ponds. In this case, the updated database (TZP2016v1.5) looked to be a better representation of the data. In contrast, a visual inspection of maps overlaid with TZ boundaries indicated that the McDougalls Hills area was home to a substantial industrial park and employment there was missed in both TZP2016v1.3 and TZP2016v1.5.

Table 11: Initial and adjusted employment totals and shares (2016)

		Employment - TZP2016 v1.3		Employment - TZP2016 v1.5		Moderated totals	
TZ_Code	TZ_Name	No.	- % -	No.	- % -	No.	- % -
6744	Wattle Ponds	472	5.0%	1997	21.1%	1247	13.2%
6745	Singleton Heights	53	0.6%	5	0.1%	53	0.6%
6746	McDougalls Hill	472	5.0%	302	3.2%	1302	13.7%
6747	Dunolly	1573	16.7%	144	1.5%	144	1.5%
6748	Hunterview	53	0.6%	5	0.1%	53	0.6%
6749	Singleton	6807	72.2%	7028	74.1%	6683	70.5%
Total	Singleton region	9429	100.0 %	9481	100.0%	9481	100.0%

There was no simple way to obtain reasonable employment estimates and the revision to the 2016 employment numbers shown in the last 2 columns were based mainly on the details apparent from aerial images along with some subjective adjustments to obtain a reasonable distribution.

24 Appendix H: Employment Allocation Model

24.1.1 Employment Allocation Model

In TZP v1.51, the Journey to Work (JTW) dataset was used to provide the basis for small area employment projections. The JTW dataset allocates employment data (ABS Census 2016) obtained at a Destination Zone (DZN) level to Travel Zones (TZ).

This dataset was created by SGS Economics and Planning and supplied to Transport for NSW.

Due to issues inherent in census data, such as undercount or poor self-reporting accuracy, the JTW data is recalibrated to align with the Labour Force Survey (adjusted to reflect Place of Work (POW) rather than Place of Usual Residence (PUR) at a Greater Metropolitan Area (GMA) level.

The following information provides a brief description of the methodology used to convert DZN data to TZ data, as found in the JTW dataset. Further technical analysis can be found here.

24.1.2 Building the conversion tables

To obtain the estimated number of jobs within a TZ, estimates are initially prepared at the ABS mesh block level which is a lower geography than TZ. Mesh blocks usually align with the TZ geography and are designed by the ABS to represent a small area of relative demographic homogeneity.

Where a TZ and Mesh Block boundaries align, TZs are a summation of, of the constituent mesh blocks. Where ABS mesh blocks do not align exactly with a travel zones, spatial area within mesh blocks overlapping with travel zones are used to estimate the size of the travel zone.

The ABS publish estimates of the count of dwellings at the mesh block level (ABS 2074.0), Employment estimates however, are more complex, as this data is not available. To obtain these estimates of employment at a Travel Zone level, it is necessary to build a model that utilises Mesh Block land use categories, in combination with employment by ANZSIC category and occupation (i.e. DZN data).

The allocation process involves separating jobs according to ANZSIC category and Occupation (Blue Collar, White Collar – Low Skilled, White Collar – High Skilled).

These jobs are then allocated to a Mesh Blocks within the DZN according to their associated land use category (i.e Residential, Parkland, Commercial, Primary Production, Industrial etc).

The process involves a “50/30/20” hierarchical split, as detailed in Table 16. 50% of jobs are allocated to the first set of Mesh Block Categories, the following 30% of jobs are allocated to the second set and the final 20% are allocated to the final set.

A process is then undertaken to identify yes or no if that Category and Occupation is to be associated with the respective meshblock category (through a 1 or 0 identifier as shown in Table 16. Once this meshblock allocation is complete, a meshblock to TZ concordance then creates employment at a TZ level.

Example: 50% of Blue Collar Jobs in Example Industry #2 are allocated to Industrial Mesh Blocks. The next 30% are distributed across Industrial Mesh Blocks a second time. The final 20% are allocated evenly across Mesh Blocks with Commercial, Industrial Agricultural and Other land use categories.

Table 16: example of employment allocation model (example industry)

	Example Industry #1			Example Industry #2		
	Blue Collar	White Collar (Low Skilled)	White Collar (High Skilled)	Blue Collar	White Collar (Low Skilled)	White Collar (High Skilled)
1st Allocation	50%	50%	50%	50%	50%	50%
Residential						
Commercial		1	1		1	1
Industrial	1	0		1	0	
Education		0			0	
Hospital/Medical		0			0	
Agricultural		0			0	
Transport		0			0	
Parkland		0			0	
Water		0			0	
Other		0			0	
2nd Allocation	30%	30%	30%	30%	30%	30%
Residential	1	1	1		0	
Commercial	1	1	1		1	1
Industrial	1	1	1	1	0	
Education	1	0			0	
Hospital/Medical	1	0			0	
Agricultural	1	0	0		0	
Transport	1	0			0	
Parkland		0			0	
Water		0			0	
Other		0			0	
3rd Allocation	20%	20%	20%	20%	20%	20%
Residential	1	1	1		0	
Commercial	1	1	1	1	1	1
Industrial	1	1	1	1	1	1
Education	1	1	1		0	
Hospital/Medical	1	1	1		0	
Agricultural	1	1	1	1	1	1
Transport	1	0			0	
Parkland		0			0	
Water		0			0	
Other	1	1	1	1	1	1

Source: SGS Economics and Planning, 2018