

Ministry of Transport

National Transport Model – Project Monty

Updated Model Development Report

Reference: 273004-00-01

B | June 2024



This report takes into account the particular instructions and requirements of our client. It is not intended for and should not be relied upon by any third party and no responsibility is undertaken to any third party.

Job number 273004-00

Arup New Zealand Limited | NZBN 68 314 038

Arup New Zealand Limited
Level 2, Hayman Kronfield Building
15 Galway Street
Auckland 1010
New Zealand

[arup.com](https://www.arup.com)

Document Verification

Project title Project Monty
Document title Updated Model Development Report
Job number 273004-00
Document ref 273004-00-02
File reference 273004-00-02

Revision	Date	Filename	230912 Interim Model Development Report.docx		
A	12/09/23	Description	Draft release for Ministry of Transport review.		
			Prepared by	Checked by	Approved by
		Name	Michael Byrne Patrick Manser	Bruce Johnson	Liz Halsted
		Signature	<i>Michael Byrne</i> <i>P. Manser</i>	<i>B. Johnson</i>	<i>L. Halsted</i>
B	04/12/23	Filename	231204 Interim Model Development Report.docx		
		Description	Final release for Ministry of Transport.		
			Prepared by	Checked by	Approved by
		Name	Michael Byrne	Patrick Manser	Bruce Johnson
		Signature	<i>Michael Byrne</i>	<i>P. Manser</i>	<i>B. Johnson</i>
C	04/06/24	Filename	240604 Updated Model Development Report.docx		
		Description	Draft release for review by Wolfgang Scherr		
			Prepared by	Checked by	Approved by
		Name	Patrick Manser	Michael Byrne	Bruce Johnson
		Signature			

Issue Document Verification with Document ☒

Executive Summary

The Ministry of Transport (the Ministry) plays a vital role in shaping New Zealand’s transport system. It acts as a trusted, impartial adviser to Government whilst providing leadership across the sector. In order to improve its ability to plan for and evaluate changes to the transport system, in 2019, the Ministry started a project to develop a national transport simulation capability known as Project Monty (Monty). The planning and development of this project has been conducted jointly by Arup and the Ministry through a co-development partnership model.

Monty aims to create a digital representation of people’s movement in New Zealand through the modelling of individual agents and their decisions. Through this, transport infrastructure and policy planning can be directly tested in a statistically robust manner, forming key evidence for assessments. Monty achieves this through a combination of activity- and agent-based modelling methodologies, orchestrated through a complex series of statistical models and simulation systems. With this approach, Monty represents one of the first transport modelling systems of its kind globally, providing a unique opportunity for decisionmakers to interrogate the transport system both now and in the future in a more granular and integrated way.

The demand generation component of the model has undergone successful external peer review with the network simulation component to follow within the year pending further improvements. Professor Miller’s review concluded that:

“Overall, the current *Monty* activity- and agent-based travel demand forecasting system is a well-designed and well-implemented model system that conforms well to current industry standards with respect to both design and performance. It represents a robust balance between theoretical considerations and practical issues of data availability, computing requirements, etc. It is well-grounded in best-practice experience with such models in both North America and Europe.”

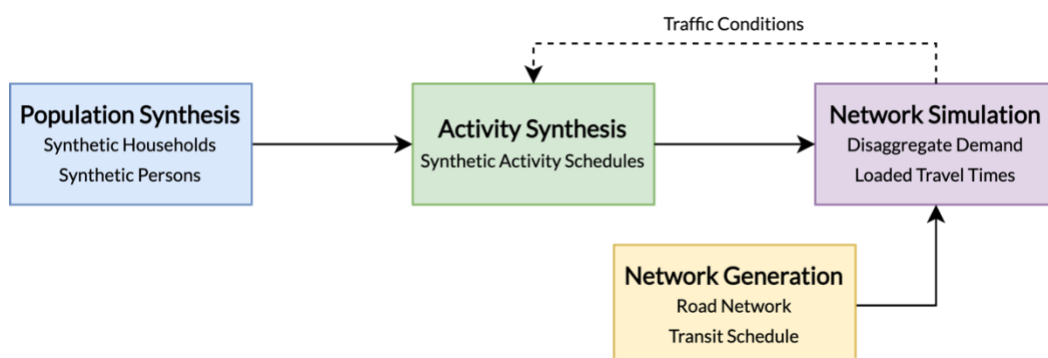
This report provides an overview of the structure of Monty and its current developmental status including the validation of outputs, future priorities and potential opportunities.

Monty’s Structure, Performance and Validation

Monty is a system of models and feedback loops. There are three major tasks undertaken by these models:

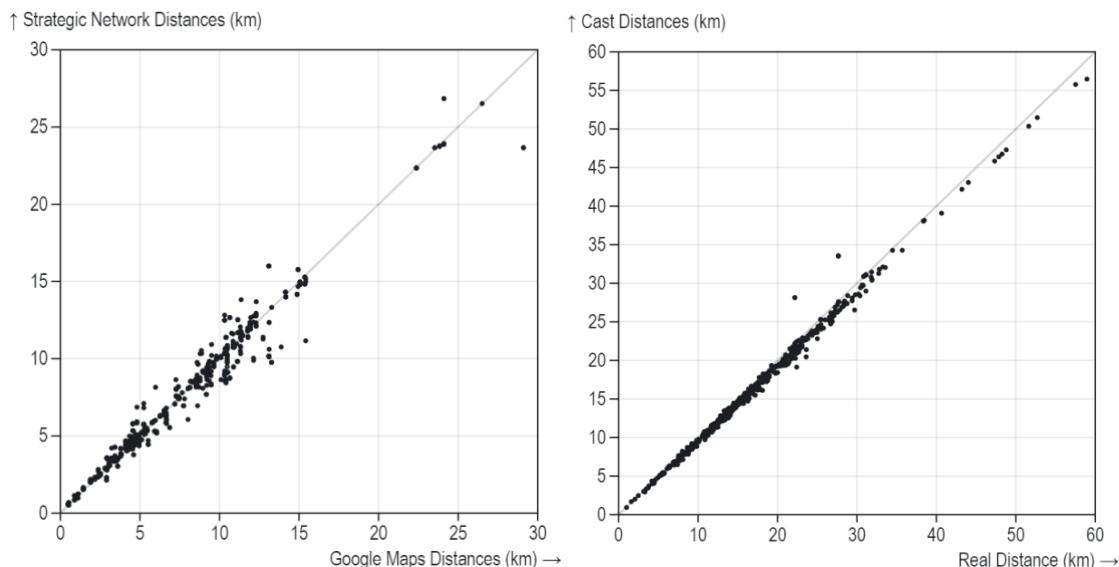
- generate a representative population of agents for New Zealand (each with their own combination of plausible attributes like age, gender, job and household structure);
- provide this representative population of agents with plausible activity schedules that mimic patterns of activities seen in real life; and
- have all these agents interact with one another in a realistic manner across the transport system as they attempt to fulfil their scheduled activities.

Monty achieves this in 4 models:



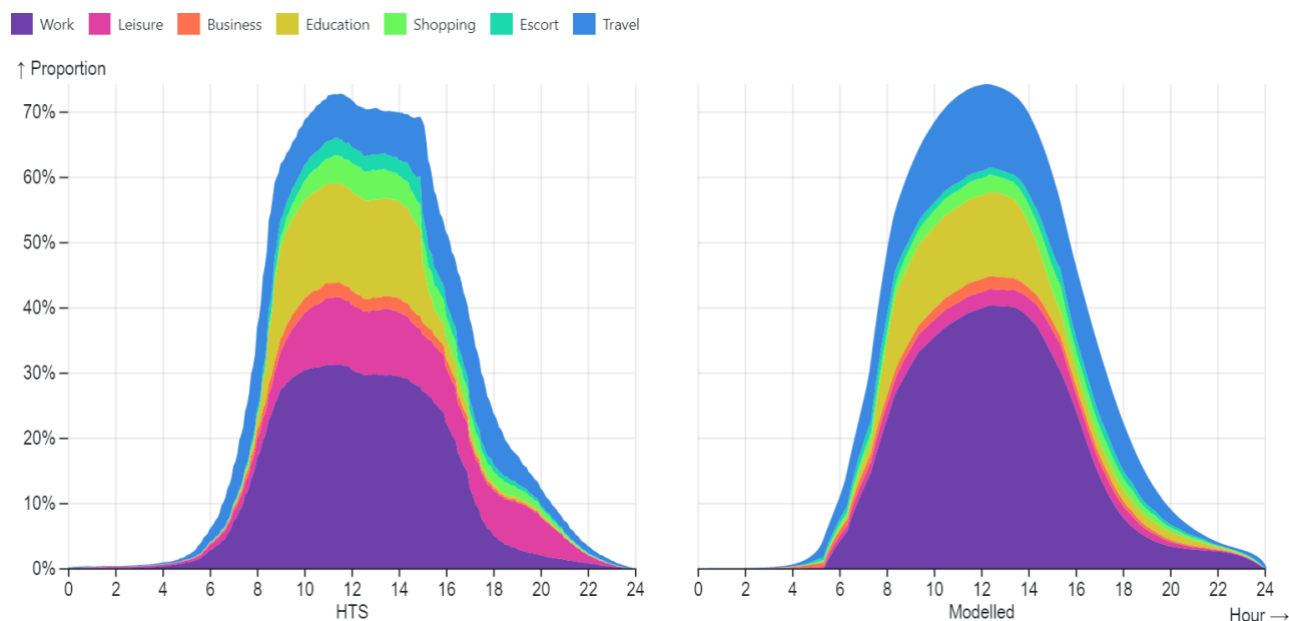
Transport Supply (Network Generation). Monty provides a strategic representation of New Zealand’s entire road and public transport network to support its simulation. The road network is built using OpenStreetMap as a foundation, augmented with relevant information to support a robust simulation of vehicle movements across a variety of urban and regional contexts. The public transport system is based on open timetable feeds released

by various public transport authorities throughout the country. Capability exists to modify these inputs in order to represent alternative networks, such as those representing future years. The current implementation has, in combination with the network simulation, demonstrated capability to produce good validation performance against observed data such as Google Maps as shown in the figure below:



The development team has several priorities in the next year which will further enhance the supply component of Monty including reducing the number of manual steps in the model pipeline, developing a functional representation of the car-ferry across the Cook Strait, improving the representation of public transport fares and coverage, and improving the representation of active modes within the model.

Demand Generation (Population and Activity Synthesis). Monty models travel behaviour by simulating individual people as agents within a synthetic world. In order to facilitate this, the system includes means for generating such agents from demographic assumptions and providing them realistic activity schedules. A sophisticated pipeline of stochastic models has been employed to achieve this, undergoing rigorous validation against real demographic and travel survey data to ensure a realistic model response. The final outputs of this complex process have demonstrated good validation against both the observed population taken from Census, other Statistics NZ data and activity patterns taken from the Household Travel Survey (HTS). The figure below shows a comparison of the culmination of Monty’s activity-based model pipeline against the HTS share of activity types being undertaken throughout the day.

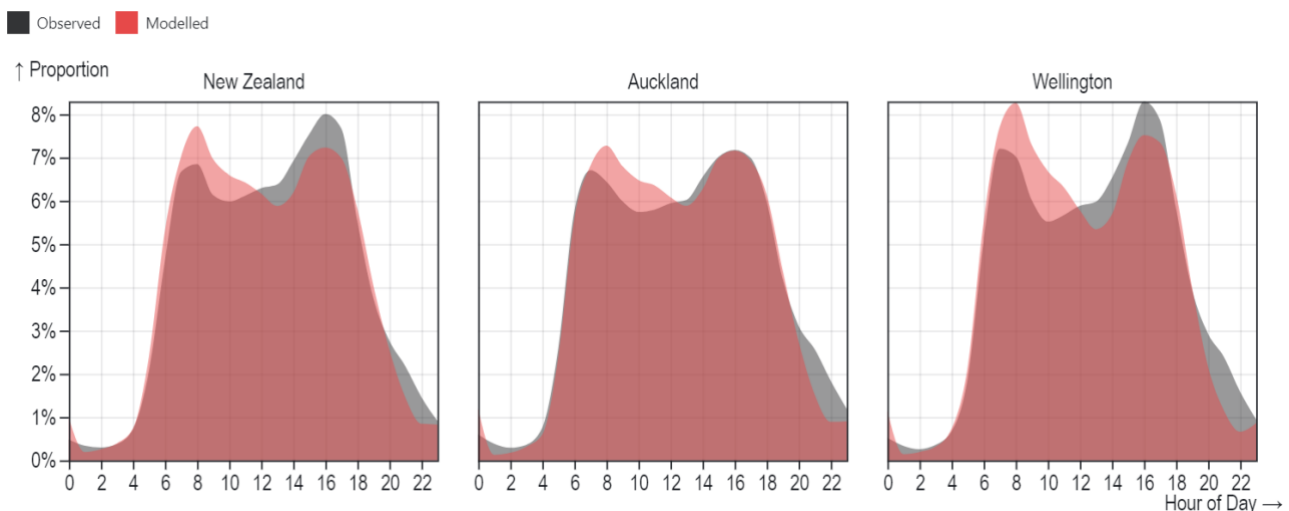


Monty’s AcBM is primarily based on behavioural models that express individual preferences. This enables the capability to not only to produce a base year scenario, but also facilitates predictions about potential future changes such as demographic shifts, changes in land use, or induced travel due to infrastructure investments.

The following limitations of Monty's current AcBM implementation have been identified (by the team and peer review) as most critical to be addressed in future development efforts:

- A range of sensitivity tests to better understand and develop the model's response to changes (elasticities) such as impacts to travel demand due to travel time improvements or pricing.
- Implementation of more behavioural variables that influence activity behaviour such as income or the types of private vehicles available to agents.
- Use of land-use data and facility sampling to improve the ability for the activity model to provide effective locations for activity's based on building sizes and as well as shopping or other leisure activities.

Network Simulation. With both the transport supply and demand generated, Monty is capable of putting these elements together and simulating all the agents undertaking their activity schedules on the transport network. From this process, agents learn how best to complete their tasks by experimenting with different modes of transport, different travel routes and modified departure times. The result is an optimised view of how the population of New Zealand travels across a typical weekday. Much like the demand generation, this component of Monty has also undergone validation against observed data in order to understand accuracy of response. As with the activity while there are some aspects of the model validation that require further investigation. Overall, however, the simulated output provides a good validation against the observed traffic counts, as see below for National, and in Auckland and Wellington.



The current implementation of the Agent-Based Model within Monty uses the open-source software framework MATSim. The adopted scoring coefficients are within a reasonable range when comparing them to known econometric values of travel time available in literature and, overall, there is strong correlation between the model outputs and observed travel behaviours.

Additional functionality is desired to enhance the current traffic simulation capabilities of Monty and the team are working through these over the coming months. Improvements include:

- The ability for agents to be more flexible in changing the timings of their activities.
- More complex treatment of parking costs at some activity locations.
- Improved modelled interactions between cars and busses, including more accurate representation of dedicated bus lanes and dynamic traffic lanes.
- Improved representation of active modes.

Model validation of the ABM outcomes showcases a wide range of statistical comparisons, including mode share, road volumes, public transport patronage and travel times. With minor exceptions, most metrics fell within expected performance benchmarks for strategic models of this scale, especially when considering that the existing focus of development has concentrated on representing areas with the highest population densities

across the country. Development of Monty is actively ongoing, and is likely to consider the following tasks as part of future calibration efforts:

- *Calibration of more regions.* The overall ambition of Monty is to provide a high-quality transport representation for the entire country. This requires additional ongoing calibration efforts in regions outside of Auckland and Wellington to improve validation performance in these areas. Work is currently ongoing to acquire more data to help facilitate such efforts.
- *Public transport patronage.* Existing public transport validation performance has revealed issues in the representation of relatively extreme passenger flows within the city centres of both Auckland and Wellington. Work is currently being planned concerning a more sophisticated representation of public transport travel that will aim to account for these local effects more accurately.
- *Traffic flow speeds.* Monty currently over-estimates traffic speeds under certain circumstances, linked to the way the network is generated from OpenStreetMap. Ongoing work has identified the primary source of this issue, with ongoing calibration of network capacities aimed at addressing this directly in the next release of the model.
- *Visitor demand.* Some specific road segments in Wellington showed significantly lower volumes in the model compared to observed count data. Further investigation revealed that this was due to missing air traffic visits, i.e., people who travel to the airport and then leave Wellington by plane or vice versa. A future work package will likely focus on adding this type of demand directly to the model in areas with large airports to address instances like this. A longer-term ambition is to also introduce domestic air travel mode representation within the model.

Development of Monty by Arup and the Ministry is ongoing, but much has been achieved by the model development team to date. Monty has already been tested on a select set of policy case studies, including:

- Investigations into potential responses to complex road pricing schemes in Auckland and Wellington. This included combined implementation of cordon-based and area-based tolls with time-sensitive pricing.
- Simulation of short- and medium-term road closure impacts due to the 2023 North Island flooding incidents, focusing on demographics affected and magnitude of delay induced.
- Implementation of future infrastructure in Auckland, including a second harbour crossing, congestion charging and other congestion related measures, to investigate demand for the crossing.

The ongoing priority of the model development team is to improve the existing network generation pipeline. As of the time of writing, this has already further improved Monty's validation performance as relevant to New Zealand's transport model development guidelines and paves the way for greater adoption of the ecosystem by model practitioners. Additionally, the model development team is increasing emphasis on sharing progress and application opportunities with broader stakeholder groups as the platform matures.

Contents

1.	Introduction	1
1.1	Background	1
1.2	This Report	2
1.3	Terminology	3
2.	Model Overview	6
2.1	Core Process	6
2.2	Components	7
2.3	Data Sources	8
2.4	Validation	9
3.	Transport Supply	10
3.1	Road Network	11
3.2	Public Transport System	17
3.3	Other Components	20
3.4	Discussion	21
4.	Demand Generation	22
4.1	Population Synthesis	23
4.2	Activity Synthesis	33
5.	Network Simulation	59
5.1	Agent-Based Model Overview	60
5.2	Co-Evolutionary Learning Algorithm	61
5.3	Model Configuration and Parameters	62
5.4	Freight	68
5.5	Skim Matrix Generation	69
5.6	Computational Resources	70
5.7	Validation	71
5.8	Discussion	90
6.	Outlook	92
6.1	Future Year Scenarios	92
6.2	Early Applications	93
6.3	Project Roadmap	93
	References	95

Tables

Table 1: General modelling concept terms	3
Table 2: Time periods	4
Table 3: Monty road classification and capacities	14
Table 4: Hourly parking fare regression model coefficients	16

Table 5: Adopted GTFS feed summary	18
Table 6: Sample households and persons from the HTS	25
Table 7: 2018 control total variables	27
Table 8: Population synthesis control variables	28
Table 9: 2018 base population accuracy at a country-wide level	29
Table 10: Individual regression tests against selected control totals by SA2	29
Table 11: Example outcomes of the long-term location choice models	37
Table 12: Workplace model coefficients	39
Table 13: School model coefficients	40
Table 14: Tertiary institution model coefficients	40
Table 15: Example outcomes of the activity generation step	41
Table 16: Preferences for performing specific activities within non-mandatory tours	42
Table 17: Mode choice model coefficients	45
Table 18: Intermediate activity location coefficients	46
Table 19: Simulated transport modes in MATSim	62
Table 20: Simulation scaling parameters	63
Table 21: MATSim scoring coefficients	65
Table 22: Zonal accessibility variance between skim matrix updates	75
Table 23: Traffic count performance measures	83
Table 24: Average road journey speeds for selected regional councils	89
Table 25: Upcoming work packages	94

Figures

Figure 1: Regional council boundaries	5
Figure 2: Statistical area 2 boundaries in Auckland	5
Figure 3: Monty's main behavioural concepts	7
Figure 4: Monty's core components	8
Figure 5: Road network generation pipeline	12
Figure 6: Modelled journey distance validation	13
Figure 7: Geometry simplification before and after comparison	13
Figure 8: Lane kilometres in New Zealand by road class	15
Figure 9: Average link length per road class	16
Figure 10: Comparison of modelled and observed hourly parking cost	17
Figure 10: Public transport system generation pipeline	18
Figure 12: Comparison of real and casted public transport route	19
Figure 12: Modelled public transport route distance validation	20
Figure 13: Population synthesis process	24
Figure 14: Household and person scatter plots by SA2	30
Figure 15: Total households based on number of dependent children	30
Figure 16: Total families with dependent children by age of youngest child	31
Figure 17: Labour force distributions by age	32
Figure 18: Student status distributions by age	32
Figure 19: Activity synthesis process	34

Figure 20: Sequential model steps in activity synthesis	34
Figure 21: Example activity schedule for a full-time worker	41
Figure 22: Sub-model steps within activity generation	42
Figure 23: Intermediate activity location choice	43
Figure 24: Daily time budgets	46
Figure 25: Mean daily trips by age group	49
Figure 26: Trip count distribution by labour force status	50
Figure 27: Trip count distribution by studying status	51
Figure 28: Work tour count distribution by labour force status	52
Figure 29: Non-mandatory tour count distribution	52
Figure 30: Activity count distributions by labour status	53
Figure 31: Trip length distribution	54
Figure 32: Trip length distribution by regional council	54
Figure 33: Trip length distribution by purpose	55
Figure 34: Origin-destination commuting/enrolment values by zone	55
Figure 35: Activity start time distribution	56
Figure 36: Activity start time distribution by activity type	56
Figure 37: Mean activity durations by trip purpose	57
Figure 38: Daily activity profiles	57
Figure 39: Agent-based simulation process	60
Figure 40: Iterative co-evolutionary algorithm	61
Figure 41: MATSim replanning annealing strategy example	68
Figure 42: Freight kilometres travelled by trip destination regional council	69
Figure 43: Score convergence	72
Figure 44: Mode share convergence	73
Figure 45: Zonal accessibilities before and after fourth iteration	76
Figure 46: Trip-based mode share for New Zealand	76
Figure 47: Distance-based mode shares for New Zealand	77
Figure 48: Trip-based mode shares by income group	77
Figure 49: Trip-based mode shares for select regional councils	78
Figure 50: Trip-based mode shares for select trip purposes	79
Figure 51: Distance distribution by mode	79
Figure 52: Directional traffic volume comparisons for New Zealand (left: daily volumes; right: hourly volumes)	80
Figure 53: Daily directional traffic volume comparisons for selected regional councils	81
Figure 54: Time-period specific traffic volume comparisons for New Zealand	82
Figure 55: Time-period specific traffic volume comparisons for Auckland	82
Figure 56: Time-period specific traffic volume comparisons for Wellington	82
Figure 57: Hourly traffic proportion comparison	84
Figure 58: Hourly traffic profile comparison for randomly selected count sites	84
Figure 59: Daily directional traffic volume comparisons for heavy vehicles New Zealand	85
Figure 60: Hourly heavy vehicle traffic proportion comparison	85
Figure 61: Auckland bus passenger boardings and alightings comparison	86
Figure 62: Wellington bus passenger boardings and alightings comparison	86

Figure 63: Hourly bus passenger profile comparison	87
Figure 64: Auckland rail passenger boardings and alightings comparison	87
Figure 65: Wellington rail passenger boardings and alightings comparison	88
Figure 66: Hourly rail passenger proportion comparison	88

Appendices

Appendix A	97
Demand Generation Peer Review	97

1. Introduction

1.1 Background

The Ministry of Transport (the Ministry) plays a vital role in shaping New Zealand's transport system. It acts as a trusted, impartial adviser to Government whilst providing leadership across the transport sector. To improve its ability to plan for and evaluate changes to the transport system, the Ministry has commissioned a research project known as Project Monty (Monty) to develop a national transport simulation capability.

This initiative began as a set of recommendations responding to the Ministry's ability to understand how New Zealand's transport system is affected by change. Several reviews had concluded that there was a need for a national model to better understand the movement of goods and people across the country, both for the present and into the future. At the time, there was no such national model – only independent models covering smaller jurisdictions in isolation and without connection or consistency in assumptions. The aggregated impact of multiple policies and investments was often difficult to quantify due to persistent integration, resilience and knowledge sharing issues.

Monty was conceived in direct response to this need. It would provide the required national transport simulation capability that enables a consistent and robust approach to transport infrastructure and policy planning. Beyond simply forecasting road volumes and public transport patronage however, the Ministry also stipulated that Monty be capable of exploring the 'who' and 'why' of people's travel choices in greater detail than the existing ecosystem facilitates. The ability to understand these choices and the broader circumstances of the households' people inhabit would enable more direct analysis of equity and enable a more informed policy development approach.

The Ministry partnered with Arup to scope and deliver Monty using an agile, co-design approach that has relied heavily on co-development with emphasis on broader stakeholder engagement. From the planning phases of the engagement, several design principles were defined to guide the model's development process, structure and features:

- *Build internal capability at the Ministry.* This would ensure that future dependence on third-party suppliers for model development and application expertise of Monty is minimised.
- *Maximise the use of open-source data and software to develop the model.* This would help avoid vendor lock-in and costly long-term contracts and licenses associated with proprietary modelling software. A consequence of this is that Monty would be more flexibly shared with other parties in the future.
- *Embody a modular and flexible design, with the ability to accept changes to input data and output requirements as needs adapt.* This would ensure continued relevancy of Monty to the Ministry as the broader industry evolves over time.
- *Maximise the use of accessible software code and version control.* This would guarantee that complex input preparation, modelling and output processing methodologies remain reproducible, repeatable and efficient.
- *Pragmatically consider complex problems.* This would allow for simplistic representations of complicated choice dynamics and behaviours in cases where more accurate approaches are impractical.
- *Alignment with the broader transport sector.* Where practical, model inputs and assumptions should align with those adopted by the broader planning apparatus to enable a cohesive view of the transport system across multiple agencies.
- *Prioritise development of new features based on the importance of modelling use cases.* This will ensure that the simulation capability continuously demonstrates practical value to its users.

Both the Ministry and Arup adopted a staged approach to the development process which is currently ongoing. A proof-of-concept was developed by January 2020 to test the intended technical approach, data availability and compute environments available. From April 2020 to December 2020, this proof-of-concept was scaled to a national representation with a multi-modal transport network. In April 2021, the actual development of Monty began, leaning on insights produced during the proof-of-concept phase. A combined activity- and agent-

based modelling paradigm was ultimately selected as the core simulation methodology, the definition and justification of which are described in subsequent chapters.

1.2 This Report

The development of Monty is an ongoing process, with many of its components undergoing rapid evolution even at the time of writing. This report intends to outline the current structure of the model and the motivations behind design decisions that have been adopted by both the Ministry and Arup (collectively referred to within this report as the model development team). This has been presented in the context of questions Monty is expected to help answer for the Ministry regarding infrastructure and policy scenario testing.

This report also provides a detailed view of Monty's current simulation capabilities and its validation performance, both regarding its ability to accurately represent people's activities (through population and activity synthesis) as well as network performance (through assignment). As previously mentioned, the development of Monty is an ongoing process, and this presentation of outcomes merely serves as a snapshot of the model's current capabilities in time. Where relevant, the model development team have endeavoured to describe ongoing or future work efforts that may alter or impact the material contained in this report. The report is structured as follows:

- **Chapter 1 Introduction** - provides background context for Monty as well as an overview of this report. It also provides definitions for specific terminology that will be used throughout descriptions of the model's structure and performance.
- **Chapter 2 Model Overview** - contains a broad overview of Monty's structure, covering the adopted modelling paradigms and the specific model components that are used to apply those paradigms within the context of a transport simulation. This chapter also describes common data sources that are frequently used throughout the model development and scenario testing process.
- **Chapter 3 Transport Supply** - describes how aspects of the transport network such as road geometry and attributes, the public transport schedule and associated travel costs and fares are represented within Monty. Of note is the use of open data sources to generate a facsimile of current transport infrastructure provision. This is used to assist with model validation and scenario testing.
- **Chapter 4 Demand Generation** - describes the population and activity synthesis processes used by Monty. These turn available demographic data and projections into fully realised simulated individuals with plausible activity schedules. This chapter also provides a detailed accuracy evaluation of these processes, comparing generated outcomes against available and trusted data sources.
- **Chapter 5 Network Simulation** - describes the process to model all the interactions between the individuals created as part of demand generation and their use of the transport network, forming a detailed view of network performance. This chapter also provides an overview of Monty's validation performance for this component relative to observed road volumes, public transport patronage and travel times.
- **Chapter 6 Outlook** - summarises the current state of Monty's ongoing development process, highlighting key areas of ongoing and future work.

As a transport model with an intended use case of evaluating government infrastructure and policy decisions, peer review of the adopted methodology and implementation is a crucial part of the development process. **Appendix A** presents an external peer review for the demand generation component of Monty, accompanied by a response from the project team.

As development is still ongoing, an external peer review of the network simulation component has yet to be undertaken as of the time of writing for this report. However, such a review will be organised within the next year upon completion of several development tasks that will heavily influence the accuracy of the transport network simulation component. An updated model development report will be produced at this time incorporating the findings of the external review.

The report reflects the outcomes of a multi-year project. Given the amount of material that has been collected by the development team in this period, some aspects are out of scope for this document. Most notably, this excludes the learnings along the way or developments options that did not make it into described model version. Monty follows a novel transport modelling approach and one of the key success factors was allowing the

exploration of certain paths that were retrospectively not successful. The second crucial aspect not detailed by this report is the IT infrastructure behind the model. Monty is an enormously complex software tooling and data ecosystem, and a robust orchestration is essential. A significant effort has been made to build this orchestration framework, including processes to ensure end-to-end reproducibility (inputs through software components to model outputs). An overview of the model's IT infrastructure and data management environment was presented at the European Transport Conference in September 2023 (Mackay, 2023).

1.3 Terminology

This report uses specific terminology related to modelling concepts, transport simulation scenarios, region systems, time periods and network performance metrics to contextualise various outcomes whilst describing the model development and validation process. For clarity, these are outlined in the subsequent subsections.

1.3.1 Modelling Concepts

Several transport modelling terms are used frequently when describing aspects of Monty's modelling methodology and implementation. High-level definitions for these terms are provided in Table 1, with reference to relevant chapters and sections of this report that provide further detail where appropriate.

Table 1: General modelling concept terms

Term	Definition
Activities	Activities are tasks performed by agents within Monty's simulation. This includes but is not limited to working, studying, shopping, or accompanying someone to a destination. Each activity undertaken by an agent is associated with a place, starting time and duration. For example, a work activity may take place in a specific office, starting at 8:30am for a duration of eight hours.
Activity Schedule	An agent's activity schedule is the sequence of all activities to be undertaken by that agent during Monty's simulation. Such activity schedules are created during demand generation (see Chapter 4) and must be fully consistent temporally and spatially. In other words, activities assigned to an agent as part of their schedule cannot overlap and it must be physically possible for the agent to travel between their scheduled activities.
Activity-Based Modelling (AcBM)	Activity-Based Modelling (or AcBM) is the demand generation paradigm adopted by Monty to create all the agents required to run an agent-based simulation (see Agent-Based Modelling). This governs the process of synthesising a population of agents and assigning each of them a plausible activity schedule. This concept is discussed in further detail within Chapters 2 and 4. Monty uses both activity-based modelling and agent-based modelling together when simulation scenarios.
Agent	An agent is a synthetic person created as part of the demand generation process with attributes such as age, gender, employment, associated household etc. In this case, 'synthetic' refers to the fact that each agent is generated using a probabilistic process, with no direct association to any real-life persons. This allows models like Monty to simulate behaviour at the individual level without resorting to the use of highly sensitive, individual-specific demographic records. As will be described in Section 4.1 in further detail, each agent is a completely artificial individual. However, when taken in aggregate, a population of agents can be said to be closely representative of a real-life population of people.
Agent-Based Modelling (ABM)	Agent-Based Modelling (or ABM) is the simulation paradigm adopted by Monty to represent interactions between generated agents (see Activity-Based Modelling) within the transport network. From this, network performance metrics and other statistics can be generated from scenario tests. This concept is discussed in further detail within Chapters 2 and 5. Monty uses both agent-based modelling and activity-based modelling together when simulating scenarios.
Travel	Travel is the act of agents moving from one activity to the next in their activity schedule. The convenience of travelling depends upon several factors such as network connectivity, available modes of travel, selected route, departure time and interaction with other agents in the simulation. Travelling may also incur costs to an agent such as fares or parking costs.

1.3.2 Scenarios

This report will often refer to a *2018 base* scenario. This is specifically defined as a scenario that represents 24 hours of travel across New Zealand on an average weekday (not including holidays or school holidays) in 2018. The 2018 year is used for this scenario because it is the year for which Monty was calibrated against available observed data. The motivation for this base year selection is expanded upon throughout subsequent chapters of this report. Other scenarios distinct from the 2018 base scenario will be explicitly described as necessary.

1.3.3 Region Systems

When referring to specific parts of New Zealand this report primarily uses two region systems unless otherwise stated – *regional councils* and *statistical area 2s (SA2s)*:

- *Regional councils*. New Zealand is divided into 16 regional council areas corresponding to jurisdictions presided over by local governments and territorial authorities. In certain instances, the areas of Nelson, Marlborough and Tasman are combined for reporting purposes. Figure 1 shows these boundaries as defined in 2018 (Statistics New Zealand, 2018).
- *SA2s*. Statistics New Zealand (Stats NZ) has defined various statistical boundary systems for the purpose of reporting. SA2 is one of these boundary systems, representing suburbs or parts of suburbs with approximately 2,000 to 4,000 residents. Figure 2 shows these boundaries as defined in 2018 for the central Auckland area (2,260 SA2s exist for the entirety of New Zealand) (Statistics New Zealand, 2018).

1.3.4 Time Periods

As will be described in subsequent chapters, Monty is a 24-hour transport simulation with no discrete time periods. Agent behaviour and interactions are modelled on a second-by-second basis throughout a given scenario test. Nonetheless, despite the continuous nature of the simulation it is useful to be able to refer to parts of the day in a standard manner for the purposes of analysis and commentary. Unless otherwise stated, Table 2 provides definitions for specific time periods and their abbreviations as they are referenced within this report.

Table 2: Time periods

Time Period	From	To
Morning peak (AM)	7am	9am
Inter-peak (IP)	9am	3pm
Evening peak (PM)	3pm	6pm
Off-peak (OP)	6pm	7am

1.3.5 Metrics

Outcomes of Monty scenario tests are represented through metrics that have precise definitions. Travel will often be characterised in terms of *kilometres travelled* or *hours travelled*. This is the number of people or vehicles travelling multiplied by the distance or time spent travelling for a particular period. For example, 400 vehicle hours travelled on a stretch of road during the morning peak means that over the two-hour morning peak period (see Table 2), vehicles spent a cumulative 400 hours traversing the identified piece of infrastructure. Vehicle kilometres/hours travelled will be used to refer to private traffic on the road network, whilst passenger kilometres/hours travelled will be used to refer to public transport-related travel. These will be abbreviated VKT, VHT, PKT and PHT respectively.

Both congested and crowded conditions also have precise definitions within the context of this report:

- *Congestion*. Measures of congestion represent the degradation of road performance that accompanies sources of delay. Within Monty, road performance is a function of volumes relative to a road's capacity based on characteristics like the number of lanes and purpose. As volumes rise towards a road's capacity, congestion generally worsens as represented by higher ratios of volume to the road's capacity.
- *Crowding*. Each type of public transport vehicle has a defined capacity as specified through a given scenario's inputs. In Monty, agents cannot access a public transport vehicle once this capacity is reached, with the level of crowding characterised by the percentage of people on a vehicle relative to this capacity.

Chapter 5 provides further context surrounding the mechanisms that are used to represent congestion and crowding.

Figure 1: Regional council boundaries

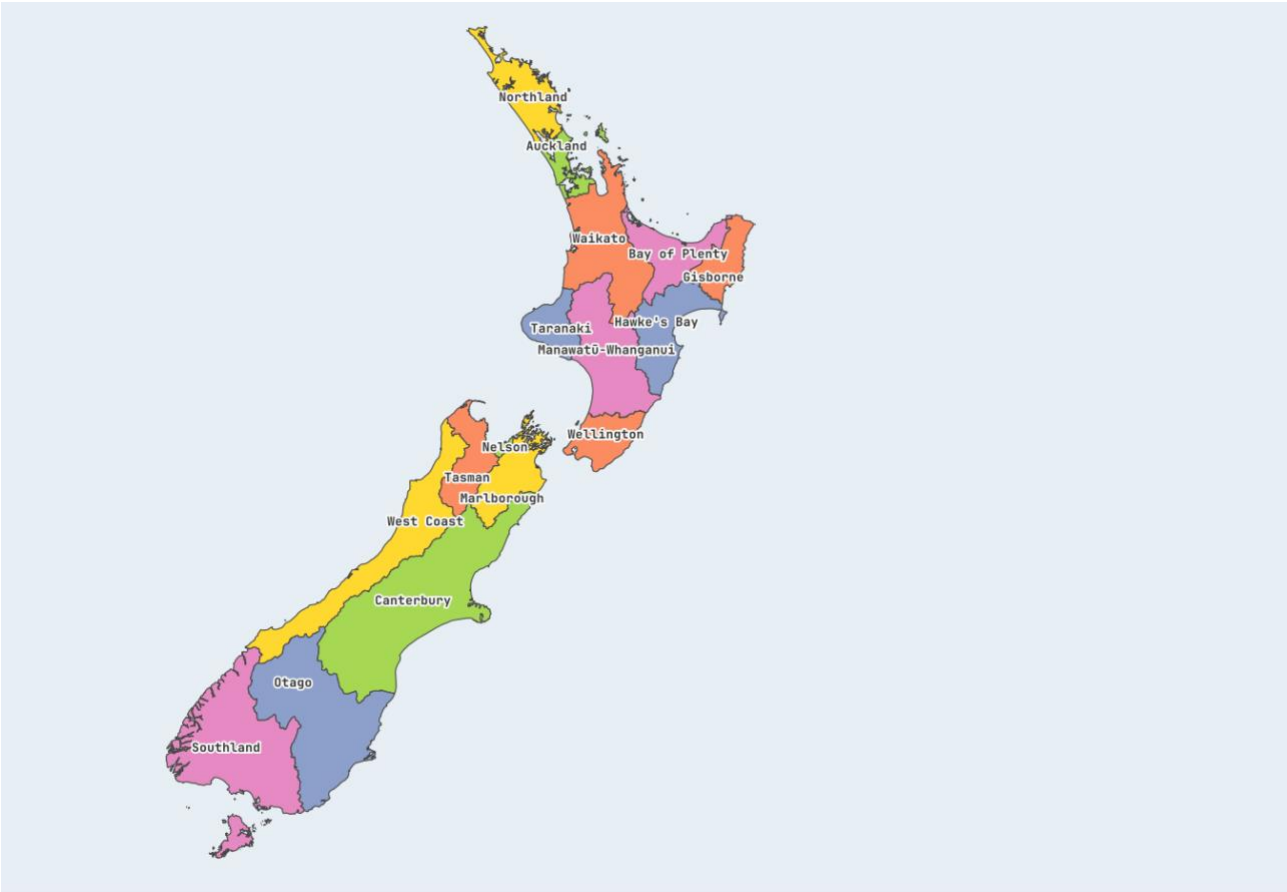
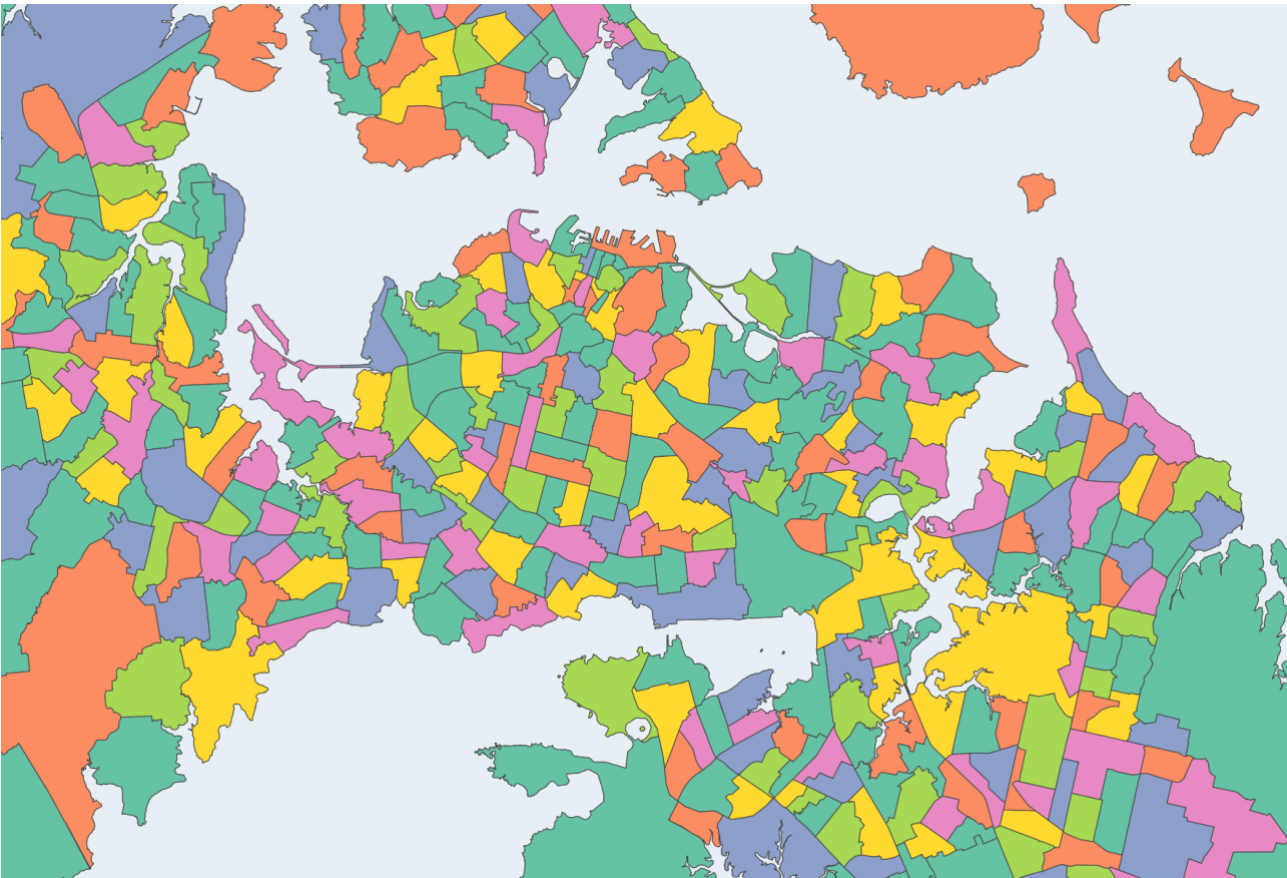


Figure 2: Statistical area 2 boundaries in Auckland



2. Model Overview

As outlined in Section 1.1, Monty is intended to provide a national-scale transport simulation capability for the Ministry. In achieving this goal, it needs to be able to realistically represent travel behaviour seen across a typical weekday throughout New Zealand, covering private vehicle, public transport and aspects of active transport travel. In addition, the Ministry required that Monty be well-placed to interrogate granular aspects of the travel decision-making process such that analyses of equity could take place.

These requirements resulted in the development of a model structure that utilised both AcBM and ABM techniques simultaneously. Section 2.1 will explore this decision in detail, explaining how these modelling paradigms have been adopted and combined to serve Monty's requirements. Section 2.2 will then cover the individual model components that have been created by the model development team to implement the AcBM and ABM methodologies in practice. Following this is Section 2.3, which provides a high-level overview of various data sources used throughout the model development and validation process. Finally, Section 2.4 discusses how the accuracy of the model was evaluated using both internally and externally defined criteria.

2.1 Core Process

Monty aims to create a digital representation of people movements in New Zealand through the modelling of individual agents and their decisions. In real life, people have things to do throughout their day and often they need to travel to facilitate this. People consider the trade-offs of travelling at a certain time or via a certain means, optimising how they get from place to place based on prior experience and constraints. For example, someone may be a frequent car driver except for their commute to work, where peak congestion makes driving too frustrating. Monty aims to capture this decision-making process, representing people as agents, each with their own plausible activity schedules to carry out. The idea behind this approach is that you can:

- generate a representative population of agents for New Zealand (each with their own combination of plausible attributes like age, gender, job and household structure);
- provide this representative population of agents with plausible activity schedules that mimic patterns of activities seen in real life; and
- have all these agents interact with one another in a realistic manner across the transport system as they attempt to fulfil their scheduled activities.

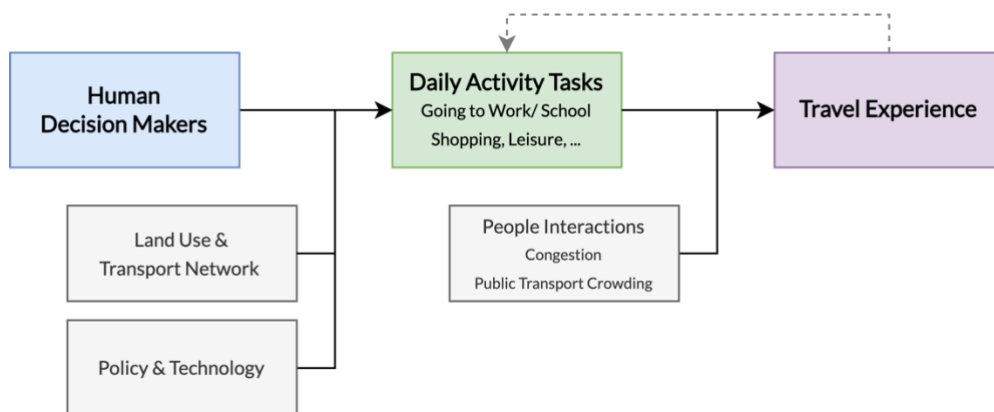
With these components together, you have a means of simulating not only where people are travelling from and going to, but also why they might be making specific decisions and what factors may affect certain groups of people more than others. Figure 3 summarises this high-level modelling philosophy adopted by Monty through a different lens. It divides the modelling process into three main behavioural concepts, providing examples of external influences at each stage. These are described as follows:

- *Human Decision-Makers.* At the centre of the simulation are individual humans (represented as agents within the simulation), each of which has their own characteristics that influence their desires and needs. For example, agents under the age of 18 will (in most cases) be predominantly concerned about their ability to travel to and from primary and secondary school by some means. Those aged older will demonstrate a much broader variety of concerns depending on whether they are working, completing tertiary study or some combination of archetypes.
- *Daily Activity Tasks.* Just as humans have basic desires and needs, this manifests in the requirement to perform certain tasks over the course of a given day. Those studying must access their place of study, just as those working must commute. Some types of activity are required to start at very specific times whilst others (like shopping) may be more flexible. As such, activity patterns can be influenced through several means including individual socio-demographic attributes, household relationships as well as accessibility to opportunity and available travel options. Additionally, activity patterns can also depend on advances in technology and policy (e.g., autonomous driving or flexible working arrangements).
- *Travel Experience.* Building on this need to undertake activities, individuals need to travel from one activity to the next whilst considering planned activity starting times. This act of travelling forms a *travel*

experience that is influenced by interactions with other individuals as well as the capacity constraints of the transport system. For example, a road segment might often be congested in the morning peak, leading to severe journey delays for travellers taking this route. This can have a major impact on the individual’s routing decisions and even their chosen activity schedule in certain instances. As a result, some individuals may end up changing their route in response or attempt to travel at a different time. As such, the process is iterative, with individuals learning and adapting their patterns over time.

The aim for Monty is to accurately capture this decision-making hierarchy for agents within a given scenario.

Figure 3: Monty’s main behavioural concepts



With this background in mind, Monty can be characterised more precisely as an *activity-based and agent-based multi-modal transport model, covering a typical weekday of the entire country of New Zealand*.

- *Activity-based* means that travel demand in the model is expressed through the modelling of individual agents that represent people. These agents are provided a schedule of activities they need to perform throughout the modelled day, giving Monty a high-resolution representation of activity behaviour both spatially and temporally at the level of an individual.
- *Agent-based* means that each agent within the simulation interacts with their surroundings and with each other during the network simulation, learning and adapting based on their experience. The outcome is a disaggregate travel demand, with each agent having assigned travel routes that are optimised with respect to their activity schedule and emergent network conditions.
- *Multi-modal* means that agents can choose between multiple travel mode alternatives whilst carrying out their day subject to constraints. For Monty, these alternatives include private vehicles (differentiating between car drivers and passengers), public transport (buses, trains and selected ferry services) and active modes (walking and cycling). Chapter 5 provides further details surrounding Monty’s representation of travel.
- *Typical weekday* refers to an average non-holiday and non-school-holiday weekday. As such, Monty does not model seasonable variance seen throughout a given year.
- *New Zealand* is the geographic coverage of Monty. This includes the road network, most public transport services as well as its entire population with over 4.5 million persons as of 2018.

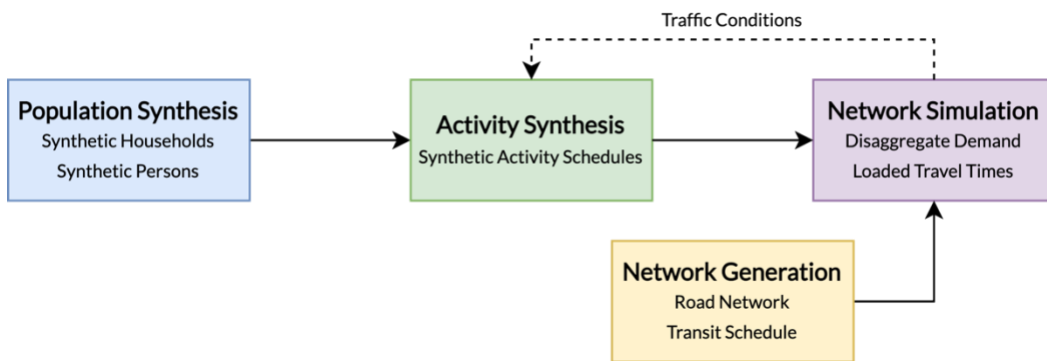
Subsequent chapters of this report will detail what aspects of the transport network and decision-making representation are not directly represented within the model.

2.2 Components

In practice, Monty is comprised of several discrete components that each correspond with parts of the process shown in Figure 3. This is shown more directly in Figure 4, with the components shown in the top row being equivalent to the three behavioural concepts described in Section 2.1. Additionally, a fourth component is introduced to cover the transport network supply aspect of Monty. Each component has the following respective responsibilities:

- *Network generation.* This is responsible for generating the transport network supply for a given scenario. This includes a detailed road network representation as well as public transport timetables for the entire country. This process is explained in further detail within Chapter 3.
- *Population synthesis.* This is responsible for creating agents and their attributes (such as age, gender, employment status etc.) for the entire country of New Zealand. The resulting population of agents is representative of New Zealand’s real population in aggregate, without any individual agent corresponding with a real-life person. This process is explained in further detail within Section 4.1.
- *Activity synthesis.* This is responsible for assigning agents generated during population synthesis with a synthetic activity schedule. Thus, after both population synthesis and activity synthesis have been used, the result is a population of agents that each possess a plausible set of demographic attributes as well as a schedule of activities to perform during the modelled day. This process is explained in further detail within Section 4.2.
- *Network simulation.* This is responsible for facilitating interactions between all agents and their activity schedules, considering factors such as network capacity constraints. Through the simulation, agents can modify their activity schedules and decision-making surrounding route choice and mode choice. This process is explained in further detail within Chapter 5.

Figure 4: Monty’s core components



2.3 Data Sources

The ongoing development and validation process for Monty has used many different data sources that are available both openly and under data sharing agreements organised through the Ministry. The following list outlines those of primary importance but is not exhaustive:

- *2018 Census.* 2018 census tables as collected and released by Stats NZ (Statistics New Zealand, 2020) were used for various purposes throughout the demand generation and network simulation components. In particular, demographic statistics surrounding age, gender, education, and employment status were used to inform the creation of the 2018 base scenario population. Journey to work statistics were used to calibrate sub-components of the activity synthesis process, including work location choice.
- *Household Travel Survey.* The Ministry conducts a continuous household travel survey (HTS) that collects multiple days of travel information from a representative sample of the population (Ministry of Transport, 2020). Several years’ worth of this data was used to produce a view of typical demographic attributes and travel behaviour patterns to assist with demand generation.
- *OpenStreetMap.* OpenStreetMap (OSM) is a community-driven mapping service that includes detailed spatial road network data (OpenStreetMap contributors, 2022). This data was used as the foundation for Monty’s road network representation.
- *General Transit Feed Specification Data.* Individual New Zealand jurisdictions produce their own public transport timetables using the General Transit Feed Specification (GTFS). These timetables were used as the foundation for Monty’s public transport network representation in the 2018 base scenario.

Other sources were used to inform validation of the model, including public transport ticketing data extracted for Auckland and Wellington as well as the national Traffic Monitoring for State Highways dataset prepared by Waka Kotahi (Waka Kotahi NZ Transport Agency, 2020).

2.4 Validation

Transport model validation is the process of evaluating a model's ability to produce both reliable outcomes and those that reflect reality at a given point in time:

- Reliability in this instance refers to the model's sensitivity to input conditions. On balance, very small changes to model inputs (such as a different random seed or a single modified road in a city) should not change scenario outcomes significantly. This would indicate that the model is unstable, making the outcomes of testing difficult to interpret.
- A model's ability to reflect reality is a key aspect of building confidence in its forecasting abilities. This means that it can recreate observed conditions surrounding decision-making and network performance using a standard set of inputs.

In the case of Monty, reliability was evaluated using tests surrounding model convergence. This process will be explored in more detail within Chapter 5. Evaluation of the network simulation was undertaken using metrics outlined in Waka Kotahi's *Transport Model Development Guidelines* (Waka Kotahi NZ Transport Agency, 2019) where relevant and appropriate. This is the authoritative resource governing the expected performance of strategic transport models that are fit for purpose. Demand generation was evaluated using guidelines developed by the model development team – this emerging field does not yet have externally-governed validation criteria to adopt.

3. Transport Supply



Monty requires a broad range of inputs that ultimately define any scenario that is to be assessed. A subset of these inputs is more specifically referred to as *supply-side* inputs, or the *transport supply*. These concern anything related to the representation of the transport network itself, as well as associated costs such as fare prices or tolls. For Monty, this primarily includes the following:

- *Road network.* The spatial representation of the network adopted for scenario testing, including accompanying attributes for individual links to represent driving and relevant public transport behaviour correctly (such as speed limits, capacities etc.). The road network representation also includes the means for defining tolling schemes.
- *Public transport system.* The spatial representation of each route within a particular public transport schedule, as well as the timings of stopping patterns for individual services within that schedule. The public transport system also includes the concept of vehicle capacities associated with each route (allowing for the modelling of different public transport vehicle types) as well as fare structure.

Section 3.1 provides an overview of the road network presentation within Monty in detail. It discusses how the 2018 base scenario road network was generated from open data sources, and how alternative scenarios are constructed from this foundation for testing. Section 3.2 provides a similar overview of the public transport system representation within Monty. Finally, Section 3.3 will discuss aspects of Monty's transport supply representation that are not directly covered under the road network and public transport system umbrellas. Monty's approach to active transport representation is discussed here, as well as specific approaches to unique situations such as inter-island ferry travel.

3.1 Road Network

3.1.1 Road Geometry and Attributes

At present, Monty uses a *strategic* representation of the road network as the basis for its simulations. This means that most, but not all roads that exist are present within Monty. Instead, emphasis is placed on representing key roads that, in aggregate, provide the ability to accurately understand accessibility and key movements between areas. Some lower-order roads, such as cul-de-sacs in local residential streets and entrances to service areas do not fulfil this purpose and are thus excluded from the transport simulation. This is a universal approach adopted across strategic transport model development¹, largely for two reasons:

- A dense representation of every road that exists can cause computational resource bottlenecks when running large transport simulations. In other words, making the network very detailed can cause scenarios to run too slowly.
- A detailed network representation can also make it more difficult to create alternative network infrastructure scenarios. High levels of detail in the transport network demand equally high levels of detail for any subsequent modifications to that network. Consider the need to represent a new greenfield suburban area as part of a future scenario. If the original strategic network included detailed representations of every road, the same level of detail would be expected for the new greenfield area for consistency even if such information does not exist.

Strategic road networks for transport models are usually adapted from prior models developed for a particular region. In Monty's case, no publicly available open-data, strategic road network existed on the scale of the whole country, meaning it needed to be generated from available data sources. In order to adhere to the project's design principles, the primary data source for this task was Open Street Map (OSM) – a community-maintained open geographic database containing many spatial features, including the road network (see Section 2.3).

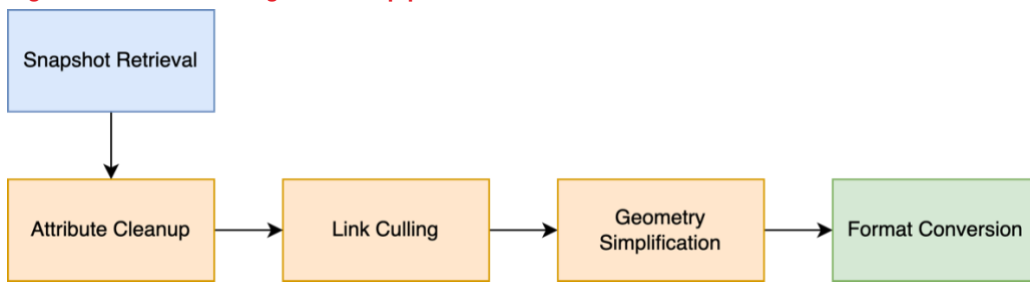
OSM provides the ability to save a snapshot of the database at any point in time, allowing the data to be directly manipulated for a variety of purposes. A snapshot of New Zealand was captured in early 2022, forming the foundation of the subsequent strategic road network generation task for the 2018 base scenario. A historical snapshot of the database in 2018 was not used, as OSM contributions can improve the quality and coverage of the database over time. The 2022 snapshot contained significantly better coverage of the road network than what was available in 2018. As a trade-off, the 2022 snapshot contained road infrastructure that did not exist in 2018. In these instances, newly built major road infrastructure was manually corrected back to its 2018 state. This includes for example the Waikato Expressway, the Transmission Gully, or the Southern Motorway in Christchurch.

For the purposes of Monty, the OSM database snapshot contains the spatial geometry of all roads in New Zealand, as well as accompanying attributes such as:

- The classification of the road type (the “*highway tag*” in OSM). This is used to indicate the importance of the road as part of the overall road network hierarchy, taking on values such as ‘*motorway*’ or ‘*residential*’.
- Direction of travel (one-way or two-way orientation).
- Posted speed limit and number of lanes.

¹ Unlike traditional strategic transport models, Monty's road network representation differs in that centroid connectors are not included. Traditional strategic models use these artificial links in order to load demand onto the network from zones. Because of Monty's highly granular demand generation process (see Chapter 4), agents join the network based on the nearest link to their activities.

Figure 5: Road network generation pipeline



This data was processed using a pipeline with the steps shown in Figure 5. Each step is described as follows:

- *Attribute cleanup.* As a community-driven data source, OSM cannot be completely relied upon without verification. Beyond the correction of errors, information can be stored within the database in a manner not compatible with the transport simulation. For example, roads can often be associated with more than one speed limit. Sometimes, this corresponds with instances where time-specific speeds are enforced but other times this is simply an error. Further, speeds can often be provided in a mixture of units, resulting in nonsensible speed limits if taken at face value.

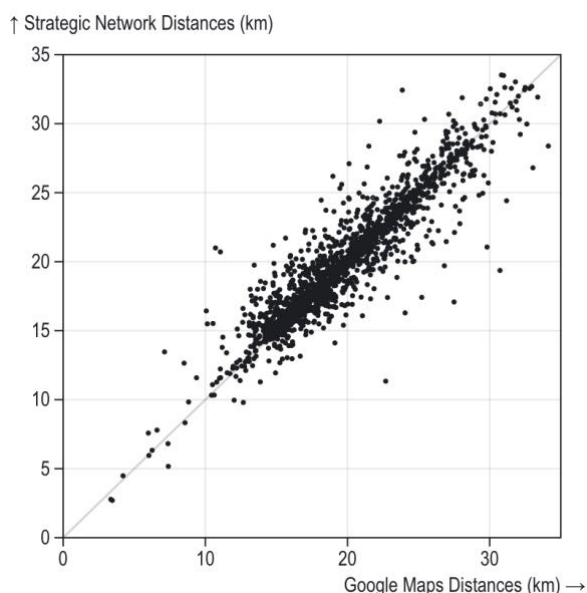
Cleanup procedures were developed to catch situations like this, providing systematic means of replacing ambiguous or incorrect entries with corrections. A further item addressed in this stage was the simplification of road classifications to a hierarchy more relevant to transport models (the existing OSM hierarchy can be varied and subjective).

- *Link culling.* Following attribute cleanup, link culling represents the first of two stages in the road network generation pipeline that act to transform the unmodified OSM database into a *strategic* road network for transport modelling. In this step, the aforementioned ‘key’ roads are kept whilst others are removed to simplify the network representation.

For certain types of road classification, the link culling step is programmatic. For instance, OSM often includes detailed representations of road infrastructure inside carparks referred to as service roads. These types of roads hold no immediate value for the network simulation and are thus removed by filtering for the relevant OSM link classification data. Other road infrastructure like residential roads are treated in a more nuanced fashion. Whilst it is likely true that these lower-order roads are not necessary within the simulation, in certain cases they can act to provide key connectivity that, in their absence, may compromise people’s ability to carry out certain journeys. An iterative approach was developed to systematically cull specific types of lower-order roads based on their broader ability to provide connectivity to the network. For example, cul-de-sacs by definition provide little additional connectivity. Through graph analysis of the underlying road network structure, this type of infrastructure was culled whilst keeping other lower-order roads of greater importance.

The culling process was verified repeatedly during iteration by comparing shortest-path journey distances between the culled network and the original network for randomly generated journeys. Large deviations in journey distance would indicate that a key piece of road network connectivity was inadvertently removed from the network and would likely result in unrealistic simulated travel behaviours. This process was used to refine the methodology adopted in performing link culling. Figure 6 shows an example output of this verification process for the 2018 base scenario network. This compares around 1,700 modelled journey distances for random origin-destination pairs compared to the equivalent journey distances as provided by Google Maps. Strong agreement ($R^2 = 0.9$) and no bias towards shorter routes (Slope = 1.01) between the two provides confidence in the integrity of the link culling process.

Figure 6: Modelled journey distance validation



- **Geometry simplification.** OSM's primary purpose is to provide spatial data directly to users as an alternative to a service like Google Maps. It is thus important for items such as road geometry to match the shapes seen in real life closely, with heavy detail spent representing curves and complex shapes visually. As an example, a bend in a highway might be represented with many different individual link segments to approximate the shape of a road in space. Such detail is not necessary within the transport simulation as it is not relevant to the calculation of optimal routes and representation of congestion. In fact, this level of detail can be actively detrimental – for instance, the many small links that make up a roundabout can cause inaccurate levels of congestion due to the nature of the traffic simulation (see Chapter 5 for further details).

For Monty, road geometry was simplified to represent curves with fewer individual segments by merging road segments together when applicable. Complex intersection geometries were also simplified in certain instances. This results in a significant reduction in the overall number of road segments in the network, leading to overall improved performance in terms of simulation run times and file sizes. However, ending up with a too simplified network with very long road segments is also not beneficial as the routing algorithms in Monty look for the closest nodes around the trip start and end locations. For this reason, a maximum road segment length of 3km has been introduced. Figure 7 shows what this looks like in practice, comparing the network before and after geometry simplification.

Figure 7: Geometry simplification before and after comparison



- **Format conversion.** The final step in road network generation is to serialise the data generated from the previous stages into the specific format required by the simulation itself.

3.1.2 Road Classification

For a realistic representation of road journey durations and congestion delays, it is crucial to define meaningful capacities and free speeds for each road segment. The capacity constrains the number of vehicles that can realistically travel across a certain segment per hour. It is usually higher for wide roads without many intersections (for example motorways) and lower for small roads with obstacles such as on-road parked vehicles. This means that the capacity of a road heavily depends on its design and its surroundings.

As mentioned above, OSM classifies roads into multiple highway categories including for example motorways, primary, or residential roads. These road classes are used as a starting point and then further refined in Monty. Table 3 shows the resulting road classes and their parameters. They were defined through an iterative process of varying road parameters and simulating in the simulation framework (see Chapter 5). The validation included journey durations, travel speeds during peak hours, and route distances (see Subsection 5.7.7).

Motorways and trunk roads (usually *State Highways*) are the highest-order roads in Monty. On- and off-ramps are categorised into two types, normal ramps and demand-managed ramps. The latter include traffic signalling during peak hours to manage the number of vehicles entering a motorway. All other road types are classified into an urban and a rural class, depending on a geo-spatial join of the network with a spatial dataset. Lastly, roundabouts are its own category as they have shown non-realistic bottleneck behaviour in Monty's network simulation framework when the capacity is too low.

Based on the introduced road classification, the network characteristics are defined using the following parameters (see Table 3):

- *Multi-lane capacity*: The number of vehicles per lane that can travel across a road segment per hour. For example, consider a motorway that has given a lane capacity of 1,400 vehicles per hour and lane. A two-lane road with this classification would have a total capacity of 2,800 vehicles per hour.
- *Minimal capacity*: The minimal road capacity regardless of the number of lanes. Usually, it equals the single-lane capacity. In the distinctive example of a roundabout, the minimal capacity is defined as 1,600 vehicles per hour independent from the roundabout's surrounding network. As mentioned above, the capacity is set particularly high to prevent the simulation from creating unrealistic bottlenecks.
- *Speed factor*: Average travel speeds are usually lower than the legal speed limit. For example, pedestrian crossings or stop signs at intersections create flow interruptions, leading to reduced average speeds over a journey. This is often referred to as free-flow speed or the speed someone could realistically travel without any traffic. The free-flow speed is usually lower than the speed limit, which is captured in Monty with a speed factor applied to the speed limit. For road types without major external disruptions such as motorways, this factor is close to 1.0. Lower-order roads have a smaller factor as they have higher intersection densities.
- *Fallback speed*: The speed limit for roads with a missing entry in the OSM speed attribute. This particularly concerns lower-order roads, which are typically not as well maintained in OSM as major roads such as State Highways.

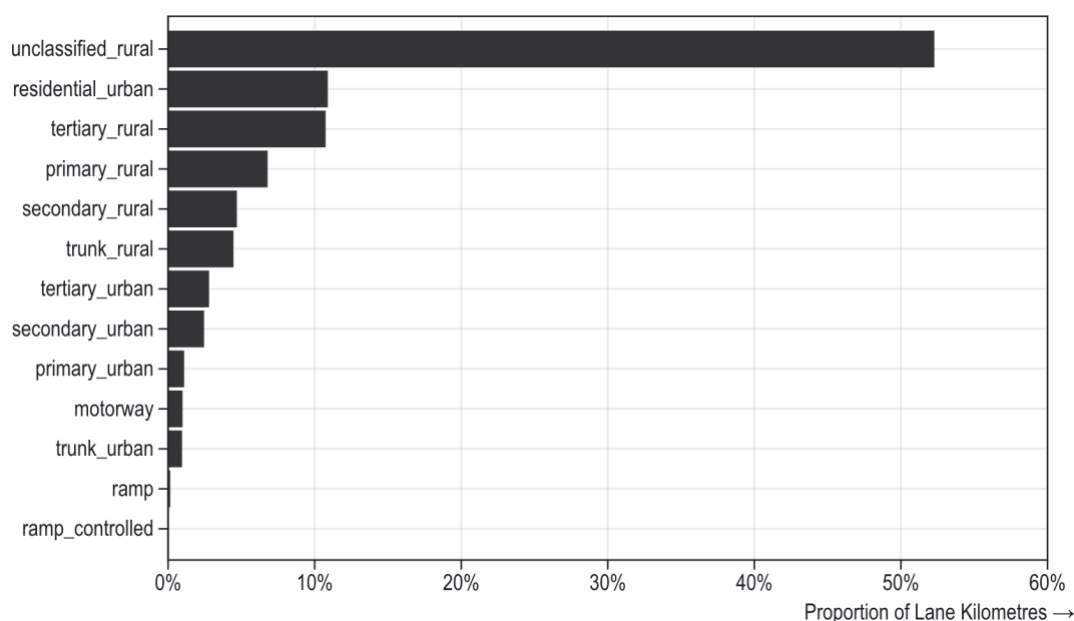
Table 3: Monty road classification and parameters

Monty Road Class	Description	Lane Capacity	Min. Capacity	Speed Factor	Fallback Speed
motorway	Major urban motorways	1,400 veh/h	1,800 veh/h	0.95	100 km/h
ramp	On/off ramps motorways	800 veh/h	1,200 veh/h	0.60	80 km/h
ramp_controlled	Traffic-signalled ramps	600 veh/h	750 veh/h	1.00	30 km/h
trunk_urban	Highways in urban setting	1,500 veh/h	1,800 veh/h	0.90	60 km/h
trunk_rural	Highways in rural setting	1,500 veh/h	1,800 veh/h	0.95	80 km/h
primary_urban	Urban main roads	700 veh/h	900 veh/h	0.85	50 km/h
primary_rural	Rural main roads	800 veh/h	1'000 veh/h	0.85	80 km/h
secondary_urban	Urban intermediate roads	500 veh/h	750 veh/h	0.85	50 km/h
secondary_rural	Rural intermediate roads	600 veh/h	850 veh/h	0.85	80 km/h

tertiary_urban	Urban local roads	500 veh/h	650 veh/h	0.70	50 km/h
tertiary_rural	Rural local roads	600 veh/h	850 veh/h	0.70	60 km/h
residential_urban	Residential roads	500 veh/h	600 veh/h	0.60	50 km/h
unclassified_rural	Low-order rural roads	600 veh/h	800 veh/h	0.80	80 km/h
roundabout	Roundabouts	1,000 veh/h	1,600 veh/h	1.00	50 km/h

Figure 8 shows the share of total lane kilometres of different road classes across New Zealand. The lane kilometres are calculated as the sum of a road's lanes with its length. With New Zealand's large proportion of rural areas, it is not surprising that low-order rural road account for by far the highest share of lane kilometres. Over 50% of all roads in New Zealand are unclassified rural. Residential roads have the highest share of roads in urban settings with an over proportion of around 10%. Urban motorways and trunk roads (usually *State Highways*) account for less than 10% of the road network overall.

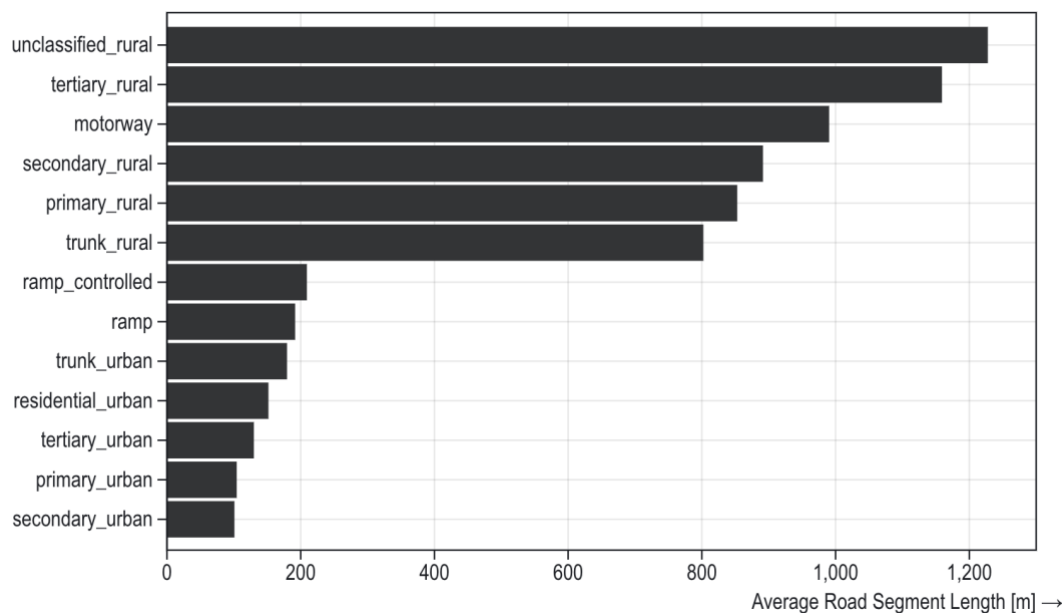
Figure 8: Lane kilometres in New Zealand by road class



Additionally, Figure 9 shows the average link length by road class. As mentioned above, a geometry simplification algorithm simplifies and merges road segments whenever applicable. However, road segments must be kept separated at intersections to allow in- and outgoing traffic through the intersection node. This means that the average size of links reduces as the density of intersections increases. This is captured well in Figure 9, showing larger average lengths for roads in rural settings with the exception of urban motorways. Those road classes are typically located in an environment with low intersection density. In urban settings with high intersection densities, the average length of road segments is much smaller, ranging between 100 m and 200 m.

The average link lengths confirm the importance of dividing the roads into urban and rural settings as their environment is substantially different, resulting in different speed factors and capacities.

Figure 9: Average link length per road class



3.1.3 Parking

Parking is a significant constraint when travelling by car. It requires time and effort to find parking and may be costly in specific areas at certain times. Parking in dense urban areas is often more expensive than regional areas as parking supply requires valuable space. For this purpose, Monty implements a parking cost model that is a function of population and job density, with a geographic aggregation that uses SA2 regions. This model has been used to impute likely parking cost in locations where observed data is not available.

To fit coefficients for the parking cost model, empirical parking cost data was collected from various sources. Average hourly parking rates were then calculated for 85 individual SA2 zones based on this information. A linear regression model was then developed as a predictor of parking cost, with several different model structures tested for accuracy. Table 4 shows the final developed model coefficients associated with the best predictive performance in terms of Root Mean Squared Error (RMSE). This includes:

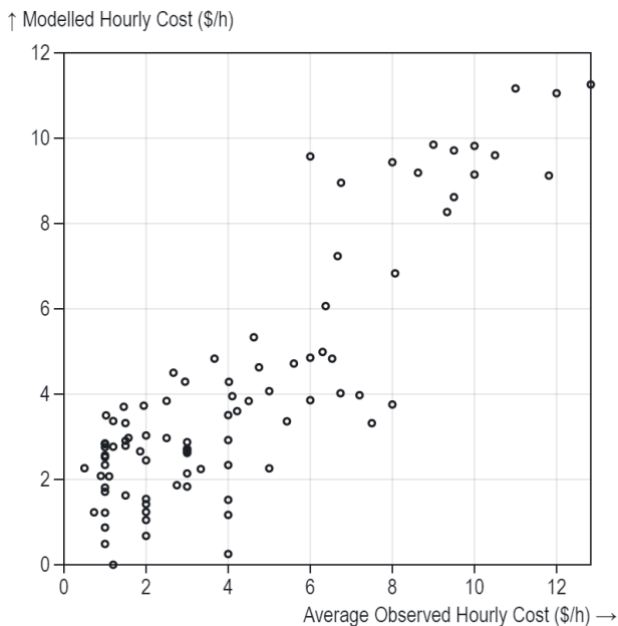
- An intercept.
- Log-transformed population density.
- Log-transformed job density.
- A constant shift if a zone is not located within the Auckland CBD. As this value is negative, it can be interpreted that the Auckland CBD is associated with more expensive parking on average whilst keeping other factors like population and job density constant.

Table 4: Hourly parking fare regression model coefficients

Variable	Coefficient	Standard Error
Intercept	-4.1240	1.80
Population density (log.)	0.3480	0.11
Job density (log.)	1.0423	0.15
Outside Auckland CBD	-3.9734	0.57

Figure 10 compares the predicted zonal parking rates against average observed cost where the data is available. There is high variation in observed parking rates and the model can capture the broad trends well. A few zones under-predict parking cost in the model – these are primarily high-income suburbs within Auckland and Wellington that have low densities but are otherwise very urban and expensive to park in.

Figure 10: Comparison of modelled and observed hourly parking cost



Section 4.2 will describe the activity-based demand model used to generate individual activity schedules for each agent within a given scenario. The parking cost model is leveraged within this system to assist in decisions such as activity location choice. Parameters can also be directly manipulated to represent alternative scenarios that could involve changes in parking costs, such as within future year scenarios. Chapter 5 will describe the dynamic network simulation that takes the generated activity schedules and represents behaviour on the network within the simulation. The parking cost model is also utilised within this component, tracking individual vehicles through the simulation and charging parking fees for parking the car at certain activity locations based on the duration of the stay. This will have a direct impact on simulation outcomes such as mode choice.

3.1.4 Tolls

Another key element of an accurate road network representation is tolling. Monty incorporates custom functionality that allows for the flexible implementation of road charging, including cordon-based, area-based, and distance-based schemes that can vary by time and user characteristic.

The 2018 base scenario implements simple cordon-based charging points that replicate the existing charging schemes present across the network at the time, including infrastructure such as the Auckland Northern Gateway toll road. Pricing within the model reflects pricing at the time, with differentiated charges applied to private vehicle and freight users within Monty. Chapter 5 discusses how these charges ultimately impact the simulation of route and mode choice within the model.

3.2 Public Transport System

3.2.1 Routes and Schedule

Monty's transport simulation represents public transit services discretely. In other words, every public transport service is represented with an individual vehicle that is allocated a unique route and stopping schedule within the network. This means that the model's input requirements surrounding public transport service provision are highly detailed, with specification of movements at the individual service level.

For the 2018 base scenario, available GTFS data was used as the foundation for Monty's public transport routes and schedules. This is a standardised data format used for describing public transport timetables. Its primary purpose is to be ingested by applications such as Google Maps to provide journey planning services, but its fine-grained nature can also be leveraged in replicating the timetable within a transport model like Monty. Typically, such a data source will describe all public transport services active during a set period of time including their precise route geometry, stopping locations and stop departure times. Given Monty's

reliance on this dataset, the accuracy of the model's public transport representation depends directly on the quality of the available GTFS feeds (including their in-built compensation for road congestion).

Figure 11: Public transport system generation pipeline

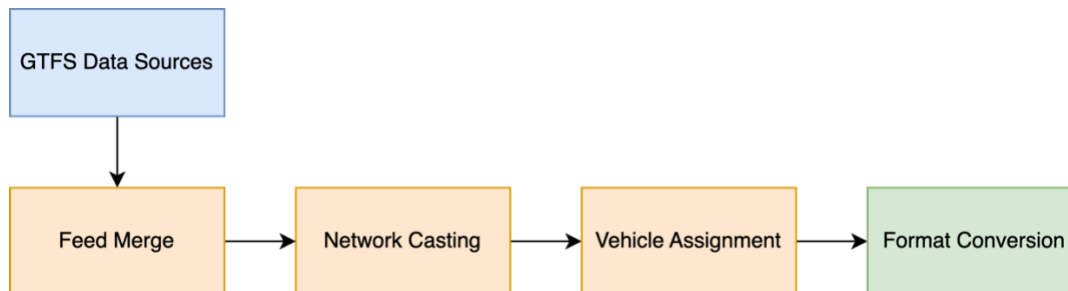


Figure 11 shows the data transformation pipeline used to convert these GTFS data sources into a format ready for scenario testing within Monty. Each step is described as follows:

- *Feed merge.* There are many public transport sub-networks under several jurisdictions within New Zealand. Due to this, there was not a single GTFS set that covered the entire country. Therefore, a process was devised to take available GTFS feeds from each jurisdiction and merge them into a single feed that aimed to best capture the representative day simulated by Monty (see Section 2.1).

As mentioned previously, GTFS data often represents the state of a given public transport timetable over a period of time. For the purposes of this task, the timetable for a given location was extracted for the date at which observed patronage data was available for the purposes of model validation if present. When not present, days conforming to the non-holiday and non-school-holiday requirement as close to March 2018 as possible were arbitrarily chosen. Table 5 summarises the GTFS feeds used for Monty's 2018 base scenario.

Table 5: Adopted GTFS feed summary

Source	Coverage	Modes	Date	Description
Auckland Transport GTFS	Auckland Region	Ferry, Bus, Rail	14 March 2018	SeaLink Pine Harbour, Pavlovich Transport Solutions, SkyBus, AT Metro, Richies, Belaire Ferries, Fullers Ferries, New Zealand Bus, Go Bus, Bayes Coachlines, Tranzit Group, 360 Discovery, Howick and Eastern, Mahu City Express, AT Metro Bus, Waiheke Bus Company.
Metlink GTFS	Wellington Region	Ferry, Bus, Rail, Streetcar	4 March 2018	NZ Bus, Tranzit Group, Mana Coach Services, Uzabus, Transdev, East by West Ferries.
Metro GTFS	Canterbury Region	Ferry, Bus	5 March 2018	Red Bus, Go Bus, Ritchies Transport Holdings, Black Cat.
Bay-Bus GTFS	Tauranga, Rotorua, Western Bay Regions	Bus	13 December 2018	Bayhopper bus networks, Cityride Rotorua, Murupara to Rotorua, Regional Tertiary / Commuter Services - Rotorua, Tauranga, Whakatane and Murupara, Potaka - Opotiki, Bethlehem Coachlines for school bus routes, Town Connector (Katikati, Waihi Beach and Waihi), Mamaku Village Connector.
Busit GTFS	Waikato Region	Bus	7 January 2019	GoBus, Waipawa Bus.
Intercity GTFS	New Zealand	Ferry, Bus, Rail	5 March 2018	Fullers operations, Trans Alpine Rail, Inter-Islander Ferry.
ORC GTFS	Otago Region	Bus	1 April 2019	Orbus Dunedin, Orbus Queenstown.

Network casting. There is a need to associated public transport routes as described in the merged GTFS dataset with the generated transport network geometry described in Subsection 3.1.1. For example, each public transport route needs to be assigned a sequence of links within Monty's network to guide vehicles servicing that route. GTFS often contains either detailed route geometries or at the very least the locations

of individual stops, however this information is not yet connected with the transport network representation as present in Monty.

Network casting is the process by which public transport routes within GTFS data are associated with the transport network through approximation. The goal is to replicate a given public transport network as closely as possible with the given network geometry. Ideally, the exact route is used. In cases where this is not possible, shortest-path algorithms and accompanying heuristics are used to re-route public transport services along nearby parts of the network. A complicating matter is that certain roads that public transport modes like buses use may not exist in the strategic network after network simplification. To remove this complication, the network casting is applied before simplifying the network to a strategic representation. When simplifying the network, the algorithm ensures that no roads are removed that are used by public transport. Figure 12 shows an example comparison of a real and network casted public transport route, highlighting where differences can arise between the two representations.

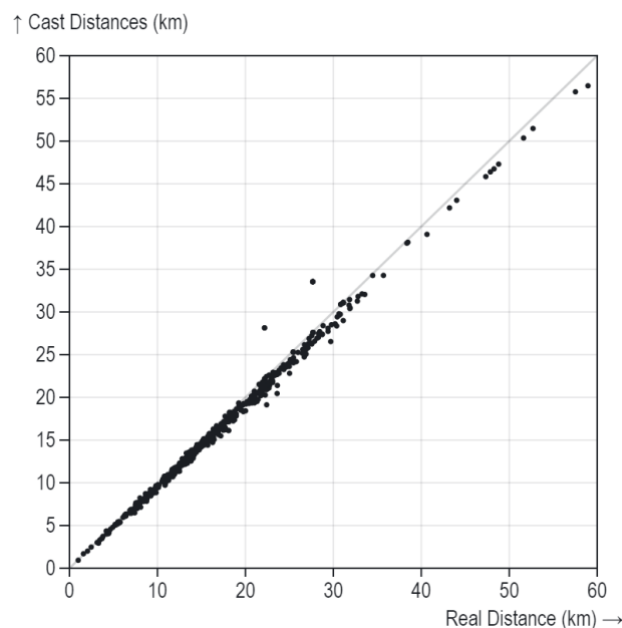
This task of routing a public transport route is of particular importance for future year supply representations, in which road projects and public transport projects are often not fully coordinated. Monty’s network casting algorithm provides a tool that ensures consistency across the entire transport supply system.

Figure 12: Comparison of real and casted public transport route



As was the case with geometry simplification (see Subsection 3.1.1), network casting is a process that must be verified. Generated public transport routes must resemble their real-life counterparts as closely as possible. Unexpected route deviations can cause services to be longer or shorter than expected, distorting their attractiveness to agents within the network simulation. Figure 13 shows such a comparison between modelled and observed public transit journey distances where such data is available, providing confidence in the network casting process.

Figure 13: Modelled public transport route distance validation



- **Vehicle assignment.** Monty requires an understanding of public transport vehicle capacities to represent effects such as crowding on public transport services. To facilitate this, every individual service within the public transport schedule requires an associated vehicle type from which capacity information can be derived. Capacities for existing public transport vehicles were taken from public sources where available, with a particular focus on representing accurate capacities for rail services. Many different bus vehicle types exist in reality – these were simplified to represent *small*, *standard*, and *large* bus sizes based on the broad range operating across New Zealand. All routes were then mapped to their closest representative vehicle according to available information. For instances where size of vehicle was unknown (commonly the case outside of Auckland and Wellington), larger vehicle sizes were assigned to avoid over-representation of crowding. This was justified since this was more likely to occur in regional areas where public transport crowding is less of an issue compared to dense, urban areas.

3.2.2 Public Transport Fares

Monty currently employs a simplified means of representing public transport fares. Agents undertaking public transport journeys within the network simulation are charged based on distance travelled. For the 2018 base scenario, this is set at approximately \$0.35 per kilometre based on fare zone structure seen in Auckland and Wellington as well as typical trip length. Implementation of a more representative ticketing system with discrete zone prices is a priority item for the model development team.

3.3 Other Components

3.3.1 Active Transport

Current performance limitations associated with our use of MATSIM mean that, unlike private vehicle and public transport travel, active transport is not assumed to take place directly on the transport network within Monty. Instead, agents are ‘teleported’ from their origin to their destination, with the distance travelled calculated as a factored version of the straight-line distance between points. Chapter 5 provides further context and justification for this representation of active modes. Given this, there is no direct transport network infrastructure to be represented within Monty at this time regarding active transport specifically. Prototyping efforts are currently being explored to provide explicit routed representations of cycling within Monty which would also include interaction with vehicular traffic. This has required cycling infrastructure to be properly represented within the model’s transport network. As this research matures, future versions of Monty will likely incorporate this functionality directly.

3.3.2 Inter-Island Ferry Travel

Inter-island ferry travel – especially that involving taking your private vehicle – represents an important transport connection for a specific subset of travellers. Correctly modelling the decision-making and routing behind car-ferry interactions is difficult. As a temporary compromise, car-ferry and freight journeys are simulated within Monty using a synthetic road link that approximates the time taken to take the ferry over the Cook Strait (as well as the capacity of this connection). This allows for such journeys to occur in the interim whilst the model development team investigates more robust ways of implementing this kind of journey directly within the simulation.

3.3.3 Air Travel

Monty does not, at this time, directly represent domestic or international air travel. As such, no components of the transport network representation within Monty concern domestic flight routes and similar items. Travel demand from flights is indirectly produced through private vehicle and public transport trips that originate and terminate at airports. This will be discussed in further detail within Chapter 4.

3.3.4 Other Modes

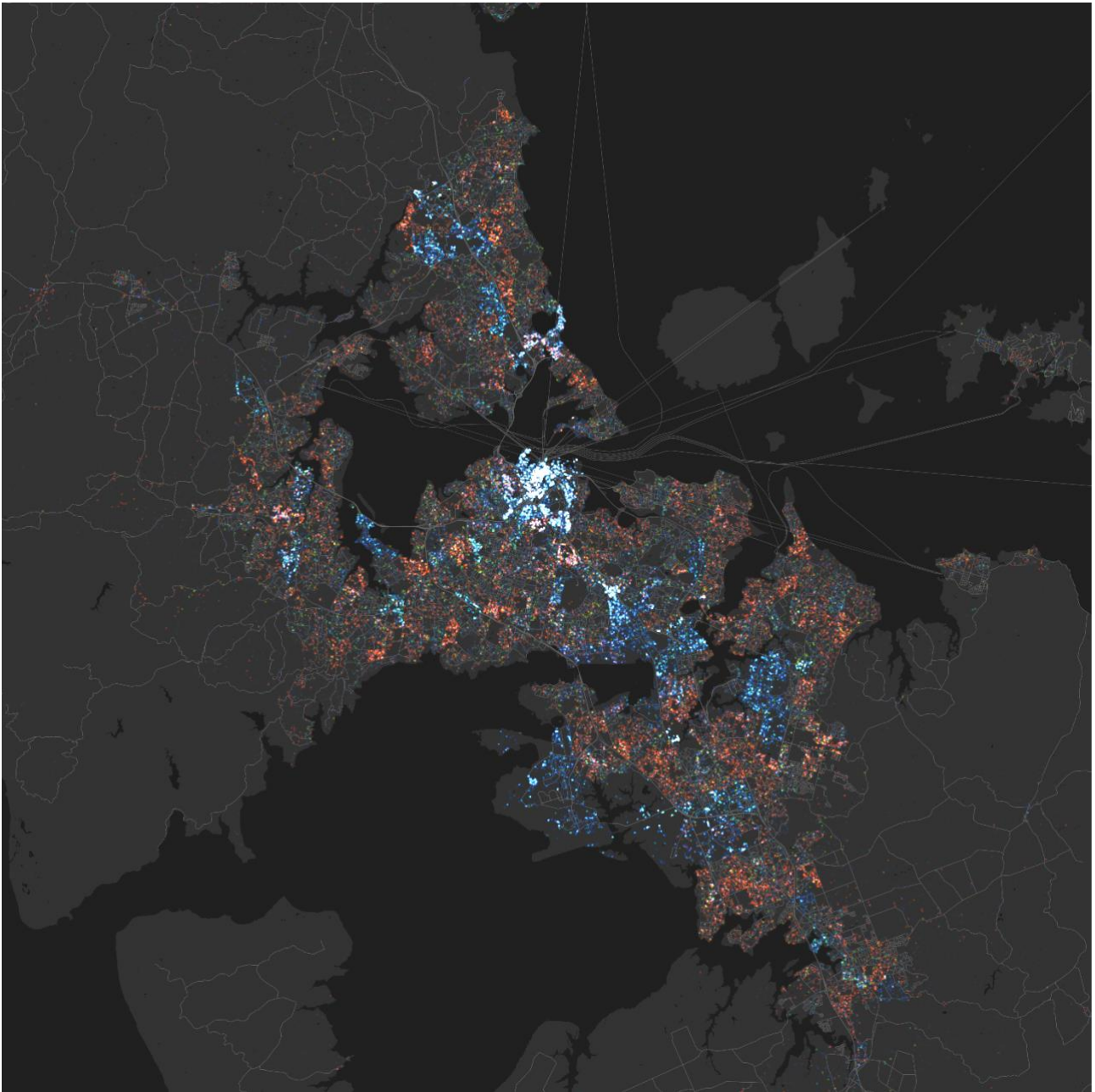
There are many other modes of transport that Monty does not represent at this current time, including school bus services, private coaches (e.g., those servicing the tourism industry), taxis and ride share, as well as e-scooter and bike share. As the model matures, modes such as these will likely be included in the default transport supply.

3.4 Discussion

The supply-side inputs of Monty include various components such as a detailed multi-modal network representation with roads and public transport services, as well as associated tolls and fares for all New Zealand. The current implementation has, in combination with the network simulation, demonstrated capability to produce good validation performance (Chapter 5). However, the development team has several priorities in the next year which will further enhance the supply component of Monty:

- *Inter-Island ferry travel.* As described in Subsection 3.3.2, ferries are currently represented as road links with travel times that approximate the timetable of ferries, and the agents are allowed to traverse the road links at all times. In reality, ferries only operate at specific times of the day leading to high peaks in traffic when they arrive at their destination dock. Current investigations explore the options to introduce a timetable-based ferry service in Monty.
- *Public transport fares.* A distance-based implementation of public transport fares requires a strong simplification of usually complex zonal fare systems. A desired future path for the development team is to explore the feasibility of applying a zonal fare system in the context of Monty's simulation framework (MATSim).
- *Active Transport.* Modes such as walking or cycling are not routed through the strategic network since they need a more detailed representation. After having implemented a more robust software pipeline to generate the network, the development team aims to further enable this functionality.

4. Demand Generation



Demand generation concerns the process of turning basic demographic information such as population statistics into a synthetic population of agents. Each of these agents needs to have their own generated activity schedules to be undertaken during the representative modelled day within the network simulation. Each part of this process is carried out across separate model components: *population synthesis* and *activity synthesis* respectively. Both components are described in subsequent sections.

4.1 Population Synthesis

Each Monty scenario requires a population. This population consists of synthetic agents assigned to synthetic households. The construction of this population must ensure that the attributes of these agents and the composition of the synthetic households they inhabit are statistically representative of the demographic conditions relevant to the scenario. In the case of the 2018 base scenario, the population should be representative of demographic distributions seen across New Zealand in 2018. For clarity, this section discusses how the project has generated the 2018 base scenario, but alternative scenarios can and have been generated using this methodology by modifying specific input assumptions (e.g., replacing actual population figures with future demographic projections).

The task of generating a synthetic population ultimately presents an optimisation problem. Using available data sources, agents and their households need to be generated such that:

- Each agent's combination of attributes could feasibly occur in real life and in the correct proportions. For example, it would be unrealistic to generate an agent who is simultaneously five years old and employed full time on an income of \$60,000 a year.
- Each household's composition of individual agents is representative of real household structures. For example, it would be inaccurate to generate households comprised solely of children and teenagers.
- The aggregated demographic statistics of the synthetic population for each sub-region across the country need to match those expressed in the 2018 census. For example, no matter how the people and households are generated, there still needs to be the correct total number of teenagers or fully employed people within a given zone.

This section outlines the methodology adopted for Monty in generating its 2018 base population (as mentioned previously, alternative populations can also be created using this methodology). First, the high-level process is described. This is followed by a discussion of the data sources used as inputs to the process and the associated analysis and cleaning undertaken. Finally, generation of the population itself is discussed along with validation measures.

4.1.1 Methodology Overview

The adopted methodology for base year population generation utilises the open-source *PopulationSim* tool. This is a framework that takes a *seed sample* of the existing population and expands that to match a set of *control totals*, thus creating a synthetic population of agents. These two inputs are:

- *Seed sample*. A statistically representative collection of real persons and households that will be used as the basis for expansion into the full, synthetic population. This is used for determining what constitutes a realistic person and household structure.
- *Control totals*. Tables outlining the distribution of relevant variables such as age, gender, and labour force status across regions. This is used to ensure that the final synthetic population expresses the correct proportions of specific person and household attributes across geographies.

The seed sample provides a clear picture of what realistic persons and households look like. However, it is not possible to create a synthetic population solely from this source as it is impossible to acquire data of this type for everyone in the country at a particular point in time.

The control totals provide a clear picture of how many people and households of certain attributes should exist across particular regions. However, it is similarly impossible to create a synthetic population solely from control totals as there is often no linkage between the independent variables available. For example, it may be separately known for a particular zone how many residents are between 20-30 years old, working part-time and in a particular income band. However, it is generally not possible to construct a set of realistic persons out of this data alone. This is complicated by the multi-dimensional nature of the problem as the number of relevant variables grows and realistic household structures are considered.

Together, both data sources provide a basis for synthesising a population through the creation of new agents and households that adopt their attributes and structure from the seed sample whilst being allocated in such a manner as to satisfy the control total constraints. A process diagram for this is shown in Figure 14.

Figure 14: Population synthesis process



For Monty, the seed samples were derived from confidential New Zealand HTS data collected from the 2003/04 to 2017/18 financial years. Control totals were derived from publicly available census data collected in 2018 by Stats NZ. Outputs of this process consist of synthetic household and agent tables that detail the locations of residence and associated attributes of each of the generated households and agents.

4.1.2 Seed Sample

The seed sample dataset is comprised of two tables: *seed households* and *seed persons*. Each person in the seed persons table is linked to a corresponding household in the seed households table. Subsequent subsections describe the structure of these tables and how they were derived from HTS data.

4.1.2.1 Seed Households

The seed households table serves several purposes within the PopulationSim pipeline:

- It associates each sample household with a given geographic region.
- It associates each sample household with a weight. This weight indicates how many households that sampled household would represent statistically. This helps PopulationSim ensure that the distribution of household structures remains realistic throughout the expansion process.
- It associates each sample household with relevant variables that will be distributed in the generated synthetic population (for example, car availability).

This table is derived from the HTS data collected on surveyed households after undergoing a cleaning process. The following steps were undertaken to achieve this:

1. Surveyed households were filtered such that only those with a valid annualised weighting were included in the final seed households table. Invalid survey weightings arise when the surveyed household is associated with an incomplete record (e.g., the survey was only partially completed by that household). This cut the total number of households available for the seed households table from approximately 54,600 to 33,400.
2. The HTS associates each surveyed household with a meshblock location (the smallest geographic unit for which statistical data is collected in New Zealand). This was aggregated up to a simplified regional council scheme with the Marlborough, Nelson and Tasman regions further aggregated into a single region. This matched the same aggregation scheme used by the Ministry when weighting households and persons from the HTS.

The purpose of the seed data is to essentially provide ‘template’ persons and household structures that can be distributed throughout the country based on the control total distributions. It is thus counterproductive to assign an overly restrictive geographic unit to each sample household. This could result in specific regions not having any associated seed samples, precluding the ability to synthesise agents and households for those regions. The simplified regional council geographic unit was chosen to avoid this issue whilst still offering enough geographic stratification to capture location-specific demographic trends.

The existing annualised weightings associated with each remaining surveyed household were re-weighted such that they instead represented a weighting over the entire 14-year survey period employed. As a simplified example, if a household in a particular year had a weighting that meant it represented 14% of a given simplified regional council area, this was adjusted down to 1% for the final seed households table. In practice, each year was weighted by the population of the regional council in that year to account for growth over time. Upon

completion of this process, the sum of all household weights in the final table was made to match the total number of actual households in 2018 exactly.

A car availability variable was derived based on the household survey response regarding the number of vehicles available at the time of the response.

It is noted that no variables related to household size and structure are present in the seed households table. These are instead inferred through the linked seed persons for each seed household as described in the next subsection.

Overall, after these modifications the seed households data provides a total of approximately 33,400 households and 86,000 persons respectively. Table 6 shows the distribution of these surveyed households and persons across the simplified regional council region system.

Table 6: Sample households and persons from the HTS

Simplified Regional Council	Seed Households	Seed Persons
Northland	1,391	3,472
Auckland	6,079	17,784
Waikato	3,006	8,024
Bay of Plenty	2,274	5,494
Gisborne	912	2,386
Hawke's Bay	1,310	3,461
Taranaki	1,322	3,268
Manawatū-Whanganui	1,586	3,894
Wellington	3,252	8,331
Nelson/Marlborough/Tasman	1,579	3,735
West Coast	917	2,105
Canterbury	6,345	16,210
Otago	2,094	4,979
Southland	1,327	3,199
Total	33,394	86,342

4.1.2.2 Seed Persons

The seed persons table serves the purpose of linking each surveyed person with an associated seed household. Each seed person is also associated with several variables of relevance to the Monty simulation, namely age, gender, labour force status and student status. Like the seed households table, this table is also derived directly from HTS persons data. The following steps were undertaken to achieve this:

1. Surveyed persons were filtered such that only those belonging to the households present within the seed households table were retained. This cut the total number of persons available for the seed persons table from approximately 97,100 to 86,300.
2. Age and gender variables were taken directly from the HTS data. In a small number of instances responses were missing for these attributes (0.05% of person records were missing a valid age and 0.02% were missing gender). The missing values were randomly imputed using existing census distributed based on labour force status. This was deemed an acceptable infilling method given the low proportion of missing values for these variables.
3. Labour force status (whether someone is employed full-time, part-time, is unemployed or not in the labour force) was derived based on a combination of HTS question responses regarding whether the respondent was working full-time, working part-time, studying, 'keeping the hole' etc. Due to the question structure, it was impossible to determine a likely labour force status for 10.7% of the surveyed persons. For these individuals, a randomly imputed labour force status was assigned to each using a distributed based on the 2018 census labour force status proportions for the simplified regional council area in which those persons

resided. For consistency with broader statistical practice in New Zealand, all individuals below the age of 15 were assumed to not be part of the labour force at all.

4. Student status (whether someone is studying at a school or undertaking tertiary studies full-time or part-time) was also derived based on a combination of HTS question responses. Earlier HTS years did not distinguish between school or tertiary study – for these respondents, student status was estimated using a combination of existing census distributions and local constraints around age.

4.1.3 Control Totals

The control totals table is used to denote how many instances of particular demographic variables are expected to be seen across regions. These variables align with those extracted for the seed sample dataset (see Subsection 4.1.2) as well as providing more basic statistics, for example, how many households and residents are expected for particular areas. This information was derived entirely from publicly available 2018 census data, which was available at the granularity of Statistical Area 1 (SA1) boundaries.

The census-derived control total variables were aggregated to SA2 boundaries. Trying to expand the seed samples using SA1-based control totals presented an optimisation problem too difficult to solve, with the Auckland regional council area alone possessing almost 9,500 SA1 zones compared to only 563 SA2 zones. SA2s provide a pragmatic balance between ensuring that the final population is geographically detailed enough to properly represent local travel behaviour accurately without overwhelming the population synthesis process. Control total variables were also aggregated to the simplified regional council scheme described previously. What proceeds is a list of variables extracted from census data to form the control totals:

- Age, gender, labour force status and student status were taken directly from the Stats NZ 2018 census SA1 dataset without modification.
 - Age was segmented into five groups² broadly representing different life stages. Like the issues encountered when trying to use a spatial region system with too much detail, trying to optimise the population to the accuracy of single age years or smaller bands results in generation difficulties.
 - Student status was only segmented into full-time, part-time, and not studying in contrast with the seed variable equivalent which also distinguishes between school students and tertiary students. This is because the census data from which the control totals are derived does not make this distinction itself.
- Household car availability was derived from the same SA1 dataset based on responses to the number of vehicles present at the surveyed household on census night. The number of households in each SA2 with at least one vehicle was calculated as a proportion of all households in that SA2. This figure was then converted into an integer value of the number of ‘car available’ households within that SA2.

Table 7 shows all the control total variables employed as part of the population synthesis process aggregated across the entire country. It is worth noting that the census night total population figure differs from the June 2018 estimate of 4.9 million estimated residents as released by Stats NZ in 2018. For the purposes of Monty’s development, the census night resident figure has been adopted rather than the estimated resident population. This is because Monty makes use of many other census-derived statistics, thus the use of estimated resident population ensured many different demographic variables are all ultimately taken from the same single source.

² 0-4, 5-19, 20-44, 45-64, 65+ years old.

Table 7: 2018 control total variables

Metric	Total	Proportion
Total People	4,698,213	-
By Age		
• 0-4 years old	294,789	6%
• 5-19 years old	930,198	20%
• 20-44 years old	1,565,085	33%
• 45-64 years old	1,193,136	25%
• 65+ years old	715,005	15%
By Labour Force Status		
• Full Time	1,890,637	40%
• Part Time	553,566	12%
• Unemployed	151,005	3%
• Not In Labour Force	2,103,005	45%
By Student Status		
• Full Time	983,038	21%
• Part Time	109,041	2%
• Not Studying	3,606,134	77%
Total Households	1,653,525	-
By Car Availability		
• Available	1,416,531	86%
• Never Available	236,994	14%

4.1.4 Population Generation

As mentioned in Subsection 4.1.1, PopulationSim is used to transform the seed and control tables outlined in Subsection 4.1.2 into the synthetic population required by Monty. Levers available to influence this process include:

- Variables within the seed tables to control for explicitly with control totals as opposed to uncontrolled variables. For example, consider seed person data that specifies each survey respondent's ownership of a bicycle for example. There may not be equivalent data in the census to create a corresponding control total variable for such an attribute. If there is still a desire for such a variable to be expressed in the synthetic population, then it can be left 'uncontrolled' and its occurrence in the final output will completely depend upon how PopulationSim distributes people and households based on other controlled variables.
- The level of spatial detail used to control each variable. Smaller geographies constrain the optimisation process, helping ensure that local distributions of demographic attributes are accurate. This demands more precision from the control data however and can create situations where it is impossible to generate a valid synthetic population.
- The relative importance of each control variable relative to other control variables. There is no point in creating a synthetic population with the exact right number of school students if the total population of the country is inaccurate. Importance weightings help prioritise optimisation efforts across all the present variables.

The Monty process does not use any uncontrolled variables. Table 8 shows the geography and relative importance adopted for each of the control variables. These values were determined through an iterative

calibration process aimed at maximally improving validation performance across several dimensions. The following should be noted:

- Total households across each SA2 were ultimately assigned a very low importance weighting as the equivalent persons total variable was sufficient in ensuring that the total number of households and people were accurate at an SA2 level.
- Student status was controlled at a simplified regional council level rather than an SA2 level. Attempting to control this variable at the more detailed SA2 level resulted in less accurate outcomes. Further, it should be noted that the total number of full-time school and tertiary students were controlled for together as the distinction between these two categories was not present in the control total data as previously mentioned.

Table 8: Population synthesis control variables

Variable	Geography	Importance
Household Total	SA2	10
Persons Total	SA2	400000
Gender	SA2	400000
Age	SA2	200000
Labour Force Status	SA2	200000
Household Car Availability	SA2	100000
Student Status	Simplified Regional Council	10000

4.1.5 Validation

Accuracy of the generated synthetic population was determined through several mechanisms:

- Verification of country-wide variable totals such as population, households, and their respective sub-categorisations against equivalent totals from the 2018 census.
- Scatter plots comparing the regional accuracy of individual variables against equivalent 2018 census values, quantified with linear regression coefficients and R^2 results.
- Comparison of household structure (i.e., number of people and other attributes) between those present in the synthetic population and those enumerated within the 2018 census.

Table 9 shows the accuracy of each controlled variable at the New Zealand-wide level within the synthetic population. The individual totals are extremely accurate when considering the country as a whole and this accuracy extends further to the more granular regional council level, with the maximum deviation for any one variable being 83 people. Despite the low importance set for the household variable control, the output of the population synthesis process shows no difference between the synthetic population and census values across regional councils.

Figure 15 shows the difference between the synthetic and census totals for households and persons by each individual SA2. The SA2 region system represents a detailed geographic representation, with population counts ranging between 0 and 5,000 persons per SA2. Accuracy is also very high at this level – linear regression coefficient and R^2 values across these two variables achieve a very precise result up to four decimal places in this case. Table 10 expands upon this by summarising the scatter plot outcomes for all variables included in the synthetic population at the SA2 level. There is a high degree of confidence that every variable is being well-represented at this level of geographic detail when compared directly to the census.

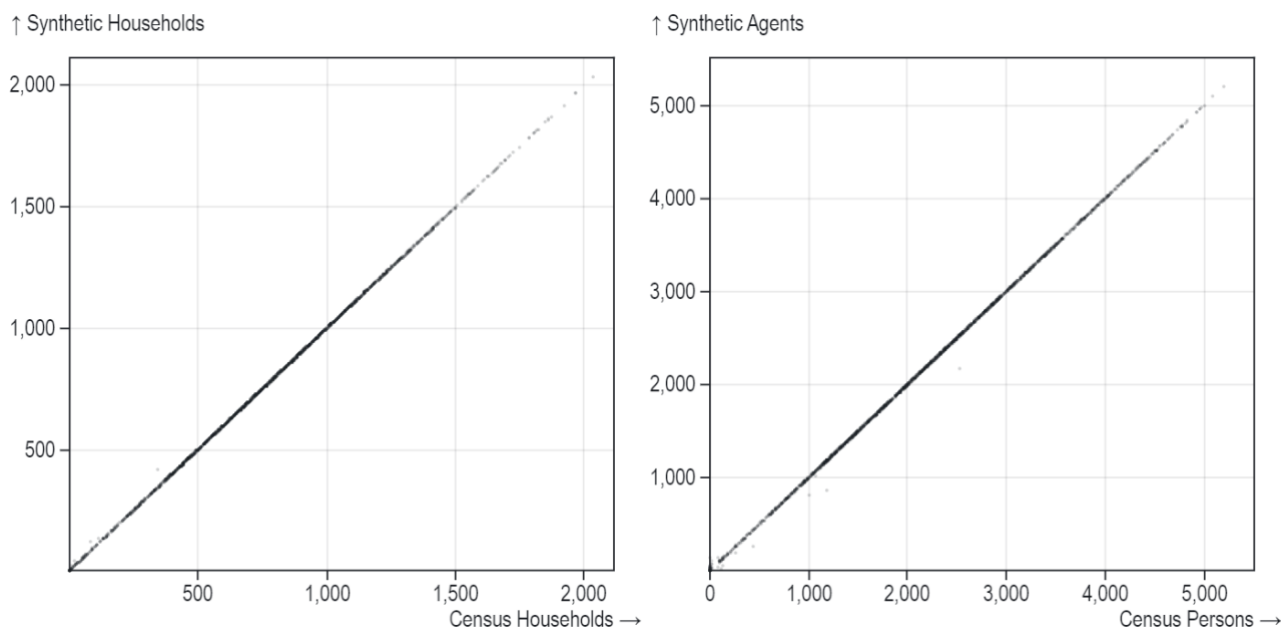
Table 9: 2018 base population accuracy at a country-wide level

Variable	Absolute Diff.	Proportional Diff.
Household total	0	0.0000%
Person total	-83	-0.0018%
Males	-46	-0.0020%
Females	-37	-0.0016%
0-4 years old	-13	-0.0044%
5-19 years old	-23	-0.0025%
20-44 years old	-26	-0.0017%
45-64 years old	-12	-0.0010%
65+ years old	-9	-0.0013%
Full time employed	-19	-0.0010%
Part time employed	-10	-0.0018%
Unemployed	-4	-0.0026%
Not in the labour force	-50	-0.0024%
Household car always available	2	0.0001%
Household car never available	-2	-0.0008%
Studying full time	-54	-0.0055%
Studying part time	-7	-0.0064%
Not studying	-22	-0.0006%

Table 10: Individual regression tests against selected control totals by SA2

Variable	Coefficient	R ²
Household total	1.0000	1.0000
Person total	1.0000	1.0000
Males	1.0001	1.0000
Females	0.9999	1.0000
0-4 years old	1.0002	0.9999
5-19 years old	0.9995	0.9999
20-44 years old	1.0000	0.9999
45-64 years old	0.9999	1.0000
65+ years old	1.0002	1.0000
Full time employed	1.0001	1.0000
Part time employed	0.9997	1.0000
Unemployed	0.9999	0.9999
Not in the labour force	0.9993	0.9998
Household car always available	0.9999	1.0000
Household car never available	1.0003	1.0000
Studying full time	0.9875	0.9906
Studying part time	0.9769	0.9520
Not studying	0.9989	0.9992

Figure 15: Household and person scatter plots by SA2



It is not sufficient to simply verify that the total numbers of agents expressing a particular demographic metric in a region are accurate. Another key component of population synthesis accuracy is how well household composition – i.e., the types of people who make up each household – is represented. It is not desirable for households to have unrealistic compositions (e.g., households solely composed of children). This can significantly influence the decision-making that individuals in a household ultimately end up making as their plans are generated during subsequent activity synthesis. Unrealistic household compositions can thus contribute to unrealistic travel plans.

One of the most consequential aspects of household composition is whether a household contains dependent children (those under the age of 18), and if so, how many dependent children are present. Figure 16 compares the proportion of households across New Zealand with specific numbers of dependent children present between 2018 census data and the generated synthetic population. Accuracy of the synthetic population is very high across this metric, with strong agreement against observed census data across a variety of different household sizes. It is worth noting that PopulationSim samples household compositions from the inputted HTS data. This indicates that the HTS sampling methodology employed when selecting candidate households is robust.

Figure 16: Total households based on number of dependent children

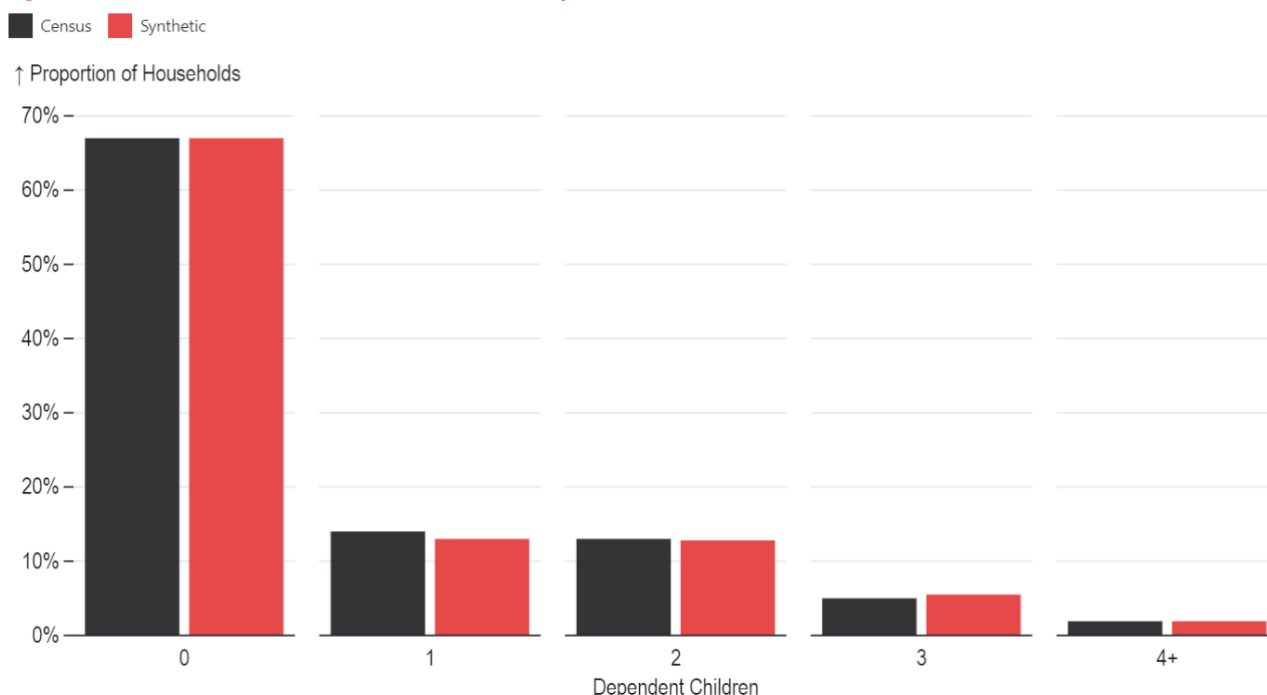


Figure 17 similarly compares the proportion of households across New Zealand with their youngest dependent child of a certain age between 2018 census data and the generated synthetic population. Across this metric, the differences across the available age bands are 3% or less, with especially good matching across very young children. The synthetic population slightly overestimates 10-14-year-olds whilst underestimating the number of 5-9 and 15-17-year-old children. For a statistic that is not directly controlled by the population synthesis optimisation process, this is a very close outcome which builds confidence in the representation of children-present households.

Figure 17: Total families with dependent children by age of youngest child

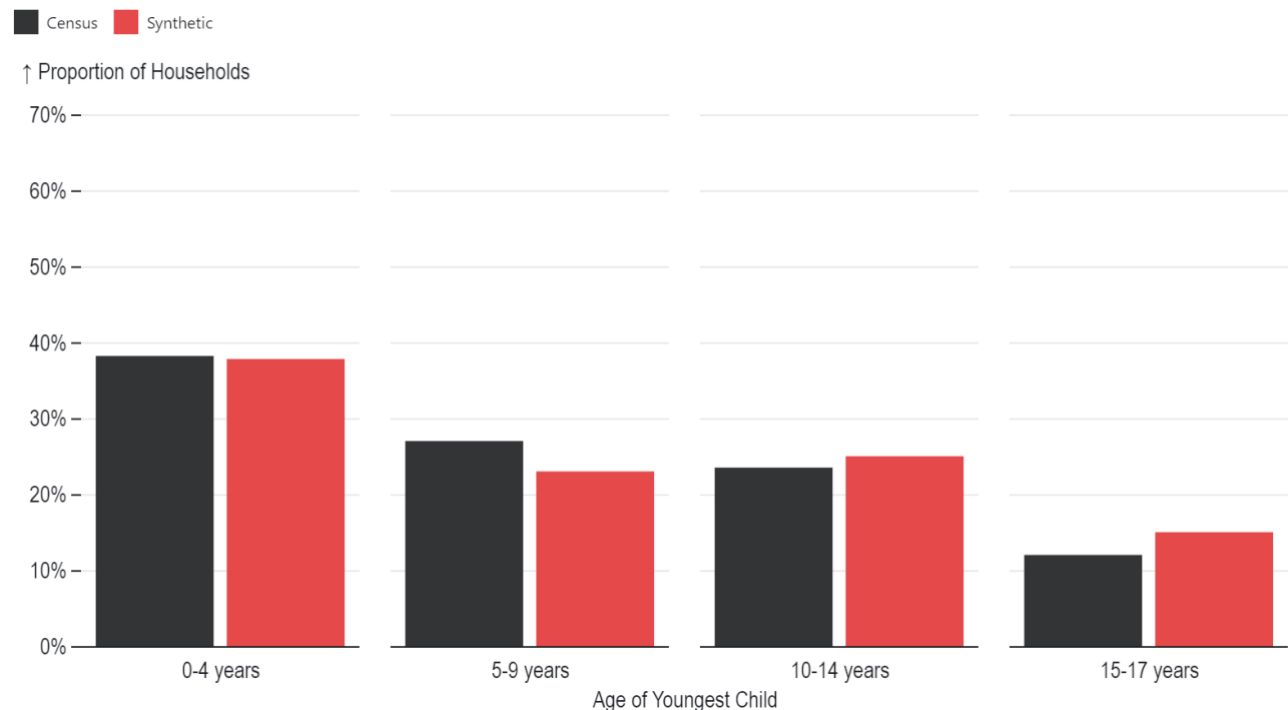


Figure 18 and Figure 19 show the cross-variable relationships between age and two selected attributes: labour force status and studying status. The aim is to demonstrate that individual persons have been assigned realistic socio-demographic characteristics across multiple combined attributes.

- Figure 18 specifically compares labour force distributions between observed census proportions and the synthesised population, split by age group. As expected, those between the ages of 20 and 64 are associated with the highest incident of full-time work. Close agreement is demonstrated across the distributions between observed people and modelled agents in this instance.
- Figure 19 examines a similar comparisons of student status distributions between census proportions and the synthesised population, split by age group. Again, the modelled population demonstrated expected characteristics with close agreement against the census. Most persons attending school are between 5 to 20 years of age. The percentage of tertiary students is highest just above the age of 20.

4.1.6 Discussion

As it stands, the implemented population synthesis generation process for the base modelled year is of sufficient quality to create a statistically representative set of households and persons for activity synthesis across the required variables. Alternative future populations can be created by replacing the census-derived control totals with appropriate projected population statistics such as those within the national population projections (2022 – 2073) released by Stats NZ.

The current process does not, however, incorporate explicit modelling of household income, ownership of specific types of vehicles, or the occupations of the working population. These variables are treated specially:

- *Household income.* The network simulation within Monty is currently not impacted by a person's household income. The model development team is currently investigating implementation options for incorporating such functionality in the future, allowing for more accurate price sensitivity across different demographic cohorts. The quality of household income data within the HTS was of insufficient quality

and coverage to include within the core population synthesis process. As such, household incomes are currently assigned on an SA2-basis using household income distribution data sourced from the 2018 census. Such an approach has proved sufficient for supporting the analysis of broad trends, but a more sophisticated modelling process is being developed to prepare for the situation where income is connected to person-based price sensitivities directly.

- Vehicle ownership.** The existing population synthesis process assigns car availability to each household not a specific number (or associated types) of vehicles. The Monty network simulation process does not currently take into consideration an agent’s vehicle type (except for freight) for any routing or pricing-related decision-making. It is possible that vehicle-specific policies may be implemented in the future, resulting in the need to differentiate between such populations. However, this is currently not necessary and such assignment of specific vehicle types of households and agents is carried out separate to population synthesis depending on the reporting needs of a given scenario test.
- Worker occupation.** A worker’s occupation is a desired agent attribute but is currently not implemented within Monty. Such information is of particular importance in order to allow for the realistic modelling of commuter flows and has been shown to be a crucial factor for people when deciding to work from home. There is a goal to integrate such an attribute into the population synthesis process relatively soon in combination with addressing the ability to model future working from home behaviour.

Figure 18: Labour force distributions by age



Figure 19: Student status distributions by age



4.2 Activity Synthesis

Each agent within the synthetic population must be allocated a daily schedule of activities. The population synthesis step (outlined in Section 4.1) focused solely on the generation of agents and households with specific demographic attributes. The outputs of this process are then fed into an AcBM which is responsible for generating activity schedules for those synthetic agents. An activity schedule consists of a sequence of activities of different types (such as work, education, shopping etc.), each with their own start times and locations (home, work, shops etc.).

Monty's AcBM has been constructed to provide synthetic activity schedules that are consistent with behaviours observed within the HTS. More specifically, the model has been calibrated such that statistical metrics calculated for behaviourally similar groups of people are consistent with empirical observations made through the survey. As a concrete example of this, the HTS demonstrated an observed average of 3.7 trips per individual a day across average week-days – a characteristic intended to be replicated by the AcBM.

The model comprises of two major components, each of which adopts utility-based principles to ensure outcomes match existing behaviour under 2018 base conditions and are appropriately responsive to changes in demographics, land use and transport supply when modelling alternative scenarios:

- *Long-term decision modelling.* Employed agents choose their workplace locations depending on several factors including their home locations, the availability of jobs and the transport supply (taking into consideration auxiliary factors such as congestion levels). Students (separated into primary, secondary, and tertiary sub-groups) make choices regarding school or university locations, again based on several external factors including transport supply.
- *Daily activity preference modelling.* With long-term choices confirmed, a daily activity schedule is generated based on each agent's status (such as their employment status or student status) and preferences (derived from their demographic attributes and household composition), time budgets and available local locations. This includes the sequencing and type of activities undertaken (e.g., home to work to shopping to home), the location of non-work and non-studying activities, as well as individual activity timings.

This section outlines the specific methodology adopted for Monty to simulate activity-based demand. First, a high-level overview of the individual modules comprising the AcBM are described in Subsection 4.2.1. The subsequent two sub-sections provide a deeper insight into the methodology used for long-term location choice modelling specifically (Subsection 4.2.2) and daily activity preferences (Subsection 4.2.3). The section concludes with a detailed comparison of modelled results with observed data as reported in the HTS (Subsection 4.2.6) as well as a discussion of potential next steps (Subsection 4.2.7).

4.2.1 Model Overview

Monty's AcBM generates activity-based demand for each agent within the synthetic population. At its core, it introduces a sequence of choice models combined with a temporal algorithm that considers individual time budgets. It builds on and extends the methodology of a nation-wide model of Switzerland that is applied by Swiss Federal Railways³. This style of methodology has been inspired by various econometric AcBMs that are applied predominantly in Northern America⁴.

Monty's AcBM construction is informed not only through its need to replicate base year travel demand (i.e., travel behaviour as it was observed in 2018) but to also react to specific levers when generating future scenarios. These levers include:

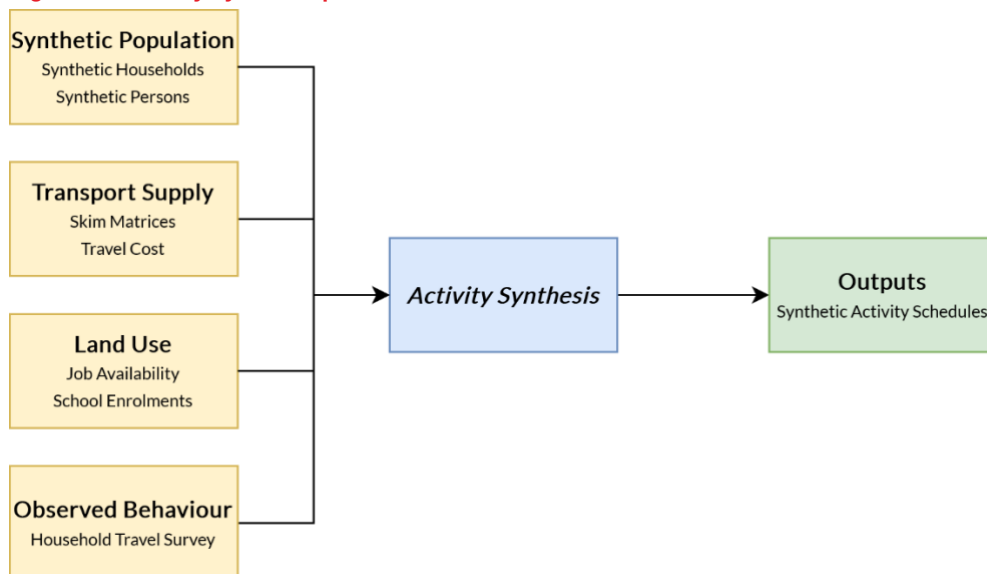
- Changes in population such as population size and household locations in reaction to changes in land use, (such as densification) or demographic shifts over time.
- Behavioural changes influenced by emerging trends such as uptake of micro-mobility, automated driving technologies, working from home or online shopping.

³ See (Sherr, Manser, Joshi, Frischknecht, & Métrailler, 2020). The Swiss AcBM simulates approximately 9 million agents.

⁴ See (Castiglione, Bradley, & Gliebe, 2015) for an overview, and (Bowman & Ben-Akiva, 2001), (Miller & Roorda, 2003), or (ActivitySim, 2022) for specific implementations.

- Infrastructure changes across both the road and public transport system affecting the level and quality of the transport supply and accessibility in turn.
- Changes to the cost of travel caused by external factors and/or policy changes such as road pricing, fuel prices or public transport fares.

Figure 20: Activity synthesis process



To ensure that the Monty AcBM is responsive to all the above-mentioned factors, the model considers four fundamental inputs as shown in Figure 20. These are described as follows:

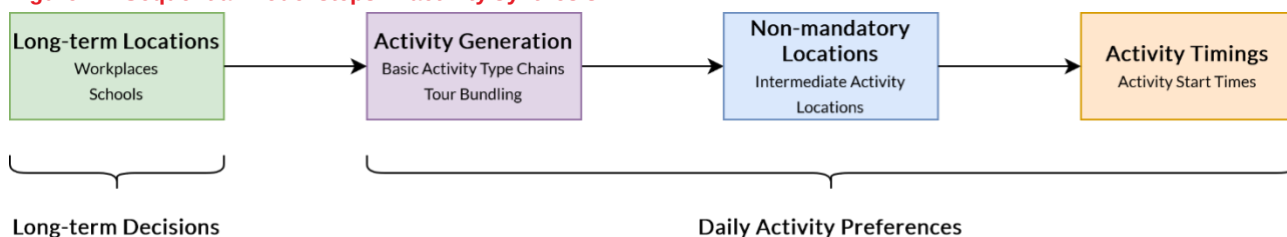
- *Synthetic population.* This represents the demographic attribute distributions as well as the detailed spatial distribution of the population. By changing such attributes when representing a future or alternative population, the emergent activity patterns generated by the AcBM are expected to change accordingly.
- *Transport supply.* The generation of activity schedules is dependent on the transport supply present, including indicators such as congested travel times and assumed vehicle operating cost.
- *Land use.* Land use data (as distinct from the synthetic population) is used to capture the impacts of job and enrolment availability for areas in which those activities typically take place. This data is also used to govern the perceived attractiveness of locations of other activity types like shopping and leisure.
- *Observed behaviour.* Data surrounding observed behaviour is used to drive the utility specifications that are ultimately used to express each agent's individual activity preferences.

The subsequent subsection outlines the decision sequence as applied in Monty's AcBM constructed to replicate base year demand and react to the input levers. This is followed by a general overview of the procedure used to calibrate and validate each decision step. Finally, a general description of multinomial logic models as applied within the Monty AcBM is provided.

4.2.1.1 Decision Sequence

Figure 21 depicts the sequential steps used to generated activity schedules for each agent within Monty.

Figure 21: Sequential model steps in activity synthesis



The model begins with long-term location choice models and progressively adds information to each agent's generated activity schedules from this starting point. The following steps are undertaken:

- *Long-term locations.* Each agent that is currently within the labour force or studying in some capacity is assigned a workplace and/or a school/university location. Driving this process are joint mode and location choice models. This methodology facilitates the simultaneous consideration of transport supply factors such as mode-specific travel times as well as land-use factors such as job/enrolment availability.
- *Activity generation.* The number, type and bundling of *mandatory activities* (work or studying activities) into out-of-home tours for each agent is determined. A sequence of multinomial logit models are applied to generate activity chains, which take into consideration socio-demographic variables as well as a number of accessibility and travel cost measures. This results in the generation of basic activity chains that (at this stage) lack detailed information about timings and *non-mandatory activity* (non-work and non-studying activities) locations which are added in the subsequent steps.
- *Non-mandatory locations.* The location of intermediate activities such as leisure or shopping (if any) are generated for each agent. These are set to take place either between adjacent to mandatory activities (e.g., home to work to *shopping* to home) or in a tour with solely non-mandatory activities (e.g., home to *shopping* to home). They are determined through application of sequential mode and location models. If mandatory activities are present, a so-called *rubber-band extension* ensures that those mandatory activities act as anchor points for the choice of an intermediate location. For example, the location of a shopping trip taking place after work should depend upon both the workplace location and the location of the household that the agent is returning to after shopping.
- *Activity timings.* With non-mandatory activities and their locations generated, the duration and start times of all activities to be performed by an agent within the 24-hour modelled period are also generated. The adopted algorithm for governing this process aims to reproduce empirically reported distributions whilst considering realistic time budgets (i.e., the total time available to spend on an activity). Assigned timings depend upon activity type as well as the socio-demographic characteristics of a given agent. Time budgets are used to ensure that the accumulated sum of all scheduling decisions fit into a single day.
- *Initial travel modes.* A multinomial choice model assigns an initial mode to each trip within an agent's generated activity schedule. The mode decision will ultimately be revisited within the co-evolutionary assignment model implemented within network simulation based on learning from the agents interacting with each other (Chapter 5).

The result of this sequence is a daily activity schedule for each synthetic agent that is responsive to changes of the agent's characteristics, transport infrastructure, and surrounding land-use across all decision steps.

4.2.1.2 Calibration and Validation Procedure

For each AcBM component shown in Figure 21, the following procedure was undertaken to calibrate and validation relevant parameters and selected model structures:

1. *Extraction and preparation of the empirical data.* This involved the identification of relevant attributes in the HTS (and in certain cases the census) data, mapping and simplifying the responses into the terminology and categories relevant to the Monty AcBM, and (if necessary) filtering of invalid or incomplete responses to avoid distorting comparisons. For example, activity types within the HTS are labelled under the '*tractiv*' attribute. For usage of these responses within Monty, the original categories of '*social visit*' and '*sport and exercise*' were grouped in a single category of '*leisure*'. In this particular example, persons with a response of '*unknown value*' were completely removed from the dataset.

Specification of the utility functions. Most decision steps used throughout the Monty AcBM are applications of a logit model. Logistic regression models can use utility principles to express behavioural preferences towards a certain alternative given a set of explanatory variables. For example, the presence of children in a household might be an explanatory variable for a high likelihood in needing to drop someone off or pick someone up (i.e., those activities primarily focused on travelling with another individual). All utility specifications in Monty were determined by iteratively making a hypothesis about a possible set of explanatory variables, estimating the coefficients within the utility specification, and then adding or rejecting explanatory variables based on their statistical robustness and calculated significance.

Coefficient estimation within utility specifications. Maximum likelihood estimation routines were applied to find each specified model's coefficients using openly available software packages⁵. This required enumeration of alternatives in each choice set. The number of alternatives depended on the decision variable and ranges from a very low number (e.g., going to work or not) to over 10,000 alternatives (in the case of a joint mode and location choice model). As mentioned in the previous item, this step was conducted multiple times until a specification with a statistically robust set of coefficients was found.

Application of the estimated coefficients to the synthetic persons. For each agent and decision, the simulation routine was first used to calculate the individual utility for each alternative. Based on these individual utilities, Monte Carlo draws were simulated to make a discrete choice for a certain alternative. For example, an agent may have been scored a 70% chance of needing to go to work. If the random number as drawn from the Monte Carlo simulation was lower than 0.7, then that agent decided to go to work in that instance.

Validation of agent choices against all available empirical data sources. This included HTS as well as information collected by Stats NZ through the census (such as commuter matrices). A detailed overview of the validation metrics used for the Monty AcBM are described in Subsection 4.2.6.

4.2.1.3 Multinomial Logit Models

As mentioned, and used in traditional transport modelling approaches, many decision steps in Monty's AcBM use multinomial logic (MNL) models. In such logistic regression models, preferences of individual towards specific alternatives (e.g., doing sports versus going shopping) are statistically defined. These models are constructed based on the assumption that individuals are most likely to choose the alternatives presented to them with the highest utility. This requires the calculation of a utility term U for each individual decision-maker n and alternative j . The full set of considered alternatives for a specific decision is the choice set C . Mathematically, this results in the following formula:

Equation 1

$$U_{j|n} = V_{j|n} + \epsilon_{j|n} = ASC_j + \sum \beta_{j|n} \cdot k_n + \epsilon_{j|n}$$

where V is the deterministic component and ϵ is a disturbance term. The deterministic component captures an alternative-specific constant (ASC) and the weighted summation of the observable attributes k (including such factors such as a person's age for example). The constant and the coefficients are estimated using maximum likelihood estimation routes and available survey data. The disturbance term in an MNL assumes a Type-I extreme value distribution.

Based on the utilities for each alternative in the choice set, the probability for choosing a specific alternative is defined as:

Equation 2

$$P_{j|n} = \frac{e^{U_{j|n}}}{\sum_c e^{U_{c|n}}}$$

4.2.2 Long-Term Location Modelling

The long-term location choice models applied within the Monty AcBM are designed to find a workplace or a school/university location for each synthetic agent who is required to undertake the relevant working or studying activities. A person's choice of these types of activity locations does not usually change for a long time period (i.e., over multiple years) hence their designation as 'long-term' in nature. The decisions surrounding these locations depend on numerous factors such as the availability of jobs and enrolments across regions, the connectivity of the transport system and the cost of travelling.

Table 11 shows a small portion of the generation synthetic agent table after long-term location choice has taken place. Attributes such as *agent ID*, *age*, *residence SA2*, *labour force status* and *student status* were generated during population synthesis. It shows five randomly selected agents:

⁵ See Apollo (Hess & Palma, 2019) and Biogeme (Bierlaire, 2020).

- The first agent is 74 years old, does not work and does not go to school. For this reason, no workplace and no school have been assigned to this agent.
- In contrast, the second agent is 37 years old and has both a full-time job and is in tertiary training. Consequently, this agent needs both a long-term workplace location as well as a location for their tertiary training.
- The remaining agents are either in the labour force or are currently attending a school. Long-term locations are generated for them accordingly.

Table 11: Example outcomes of the long-term location choice models

Agent ID	Residence SA2	Age	Labour Force Sts.	Workplace SA2	Student Status	School SA2
582580	126000	74	-	-	-	-
2793363	218800	37	Full time	217400	Tertiary (Part-Time)	217300
791077	132800	10	-	-	School (Full-Time)	124500
3528581	254300	56	Part time	251700	-	-
3404347	249700	5	-	-	School (Full-Time)	248700

To enable more accurate location choice characteristics, these long-term location choice models are set up as joint travel mode and location choice models to ensure that a wide range of relevant factors are included in the decision process. The model specification process, including the utility formulation, is detailed across subsequent subsections.

4.2.2.1 Joint Transport Mode and Location Choice Model Specification

In a joint travel mode and location choice model, the utilities for all possible combinations of modes and locations are calculated. For Monty, modes are differentiated car drivers, car passengers, various public transport modes as well as walking and cycling. The spatial resolution of all location choice models are the SA2-level, with precise point-coordinates determined post-location choice using a specific allocation procedure specified in Subsection 4.2.4. Mathematically, the following formulation was used to derive the utility V for each alternative given a household residence i :

Equation 3

$$V_{jm|i} = \theta \cdot \log A_j + ASC_m + \sum_k^{\mathcal{K}} \beta_{k,mode} \cdot k_{ijm}$$

where:

- j is the long-term location zone,
- m is the transport mode,
- i is the household residence zone,
- A represents the attraction of the zone (e.g., as expressed through the number of available jobs or school enrolments) and θ represents the logarithmic size term,
- ASC represents the alternative-specific (mode-specific in this instance) constant, and
- $\mathcal{K}$ represents the set of variables to consider for each mode. This set includes, at minimum, mode-specific travel times between i and j . In the case of private vehicles for instance, this also includes items such as travel cost and parking cost.

In Monty's long-term location choice models, the log-size term θ of the attraction is set to 1. This requires the assumption of a one-to-one relationship between the number of agents seeking opportunities at a long-term location and the number of opportunities available (e.g., the number of agents seeking a job is equal to the number of jobs available).

The mode-specific variables in $\mathcal{K}$ are extracted from skim matrices (origin-destination matrices of travel times and other simulation statistics, described in Chapter 5) and includes the following variables:

- *Private vehicle modes* consider the maximum travel time experienced during the morning and evening peak periods for each origin-destination pair. This assumes that the majority of workers need to travel during these peak periods. Additionally, private vehicles consider distance-based cost of travelling and hourly parking cost at the destination.
- *Public transport modes* consider the average in-vehicle time between two zones as well as a distance-based ticket fare price. Additionally, the AcBM takes into consideration the time needed to access the first stop and egress from the last stop of the journey to the ultimate destination. Average waiting times as part of the utility specification have been tested but did not end up being statistically significant for the long-term location model structures.
- *Active modes* use a calculation of travel time based on the average car network distance divided by an assumed average speed of travel. This was set at 10 km/h for cycling and 3 km/h for walking based on existing standard practice for such modes.

The model coefficients including ASC_m and all $\beta_{k,mode}$ were derived by applying a maximum likelihood estimation procedure.

4.2.2.2 Additional Shadow Utility Terms

To enable the possibility of calibrating the location choice models towards certain control totals, two additional utility terms were added to the utility specification. This resulted in the following utility formulation:

Equation 4

$$V_{jm|i}^* = V_{jm|i} + \lambda_j + \lambda_{ij}$$

with λ_j expressing a term to treat the attraction of a destination zone as a capacity constraint. This is commonly referred to as a *doubly constrained location choice model*. λ_i is derived by applying an iterative fitting algorithm which is implemented within the simulation routing, with the fitting targets being the zonal attractions A_j . The algorithm aims to minimise the gap between number of agents choosing a certain destination and number of jobs/enrolments that are available in this destination.

λ_{ij} is able to control outcomes like commuter flows on the origin-destination level. It is calculated using the following iterative process:

Equation 5

$$\lambda_{ij,iter} = \log \frac{count_{ij} + 1}{model_{ij,iter-1} + 1}$$

with the $count_{ij}$ being an empirically observed trip matrix (such as that available from the “travel to work and education” data collected by the 2018 census). Typically, results were seen to converge after three iterations when cumulatively adding the resulting $\lambda_{ij,iter}$ to the terms from the previous iterations. λ_{ij} is subject to the constraint $\lambda_{ij} < 0.3 \cdot logsum_{ij}$ to prevent the model from over-fitting.

4.2.2.3 Workplace Location Choice Model

HTS data from the years of 2015-2020 was used to calibrate the workplace location choice model (excluding latter parts of 2020 affected by the COVID-19 pandemic). Considering all persons who reported a work address as well as adding persons with a reported work trip and a corresponding address, the processed dataset contained 8,844 observed persons. In total, 1,582 SA2 zones had at least one observed work commuter. The reported commuter mode shares for the entire country of New Zealand in this subset of HTS respondents was 82% car driver, 7% car passenger, 5% public transport, 2% cycling and 4% walking.

For the zonal attractions A , the number of jobs per SA2 zone as recorded in the 2018 census was used, totalling 2.4 million jobs. The regional councils of Auckland and Wellington expectedly accounted for a large proportion of all available jobs, at 34% and 11% respectively.

Table 12 shows the estimated coefficients that resulted by applying the utility specification in Equation 3 to the described observed dataset from the HTS in combination with job attractions data from the census. The model was estimated using the Apollo R package (Hess & Palma, 2019), with some variables transformed either logarithmically or using a box-cox transformation.

For the public transport mode, travel time was separated into in-vehicle time (*ivt*) and access-egress time (*accegre*) which represents the time needed to walk to a stop. A global *ASC* per mode was introduced to represent general mode-specific preferences. An additional *ASC* was applied to agents who did not have access to a car within their household. This constant shift shows a clear preference away from cars towards other modes as designed. The dis-utility for walking is higher than for cycling and the dis-utility for accessing a public transport stop is higher than in-vehicle time as expected. Overall, all coefficients demonstrate reasonable and expected behaviour, with the model demonstrating an adjusted R^2 value of 0.53.

Table 12: Workplace model coefficients

Variable	Car Driver	Car Passenger	Public Transport	Cycle	Walk
<i>ASC</i>	-(ref.)	-2.9382	1.1048	-2.6711	2.0666
<i>ASC</i> shift no car in hh.	-	1.9859	2.6230	2.4631	2.7605
β_{tc} (log.)	-0.6184	-0.6184	-0.6184		
$\beta_{tt,car}$ (linear)	-0.0408	-0.0351			
$\beta_{pc,car}$ (box cox*)	-0.1525	-0.1525			
$\lambda_{pc,car}$	0.4508	0.4508			
$\beta_{ivt,pt}$ (linear)			-0.0162		
$\beta_{accegr,pt}$ (log.)			-1.3901		
$\beta_{tt,active}$ (box cox*)				-0.1519	-0.6354
$\lambda_{tt,active}$				0.7199	0.3064

*($x^\lambda - 1$)/ λ

All values are statistically significant at the 99% confidence level.

Another applicable sensibility check for models of this kind can be found in inspecting the value of travel time (VTT) savings per travel mode. This measure can be defined as the willingness of an agent to exchange money in order to reduce travel time⁶. The coefficients from Table 12 result in an estimated VTT of \$10.90/h for car drivers, \$5.90/h for car passengers, and \$6.10/h for public transport users. These values are consistent with officially published values (Waka Kotahi NZ Transport Agency, 2020). Based on this source, the VTT for car drivers is \$10.30/h, for car passengers \$7.70/h, and for public transport users it ranges from \$6.20/h (seated) to \$8.70/h (standing) when adjusted to 2018 values accounting for inflation.

4.2.2.4 School and University Location Choice Models

The same procedure as described for the workplace location choice model was applied to drive decisions surrounding the locations of schools and tertiary institutions for each agent that was studying. In this decision process, the Monty AcBM differentiates between three different student types: *primary* (below 13 years), *secondary* (age 13-18 years) and *tertiary* (university students).

The zonal attractions *A* were defined by the number of school enrolments as collected by the Ministry of Education (Ministry of Education, 2018). This dataset enumerates 511,169 primary school enrolments, 297,270 secondary school enrolments, and 253,143 tertiary university enrolments.

Locations for primary and secondary students were estimated jointly within the same choice model. Public transport and car passenger modes were grouped into one mode alternative to account for the fact that some school bus services were not part of the openly available public transport timetable used as the basis for the 2018 base scenario simulation. After processing available HTS data (the same range used for the workplace

⁶ Mathematically, VTT can be defined by $\delta V / \delta tt \div \delta V / \delta tc$.

location choice model), 1,908 observed students remained within the dataset. The observed mode shares were 2.1% car drivers, 74.4% car passengers, 4.4% cycle and 19.1% walk.

The resulting estimated coefficients are shown in Table 13. The adjusted R^2 for this estimation equals 0.326. For pupils under the age of 16 years, car driver is not included as a suitable mode alternative to be considered due to licensing restrictions. This was modelled by introducing negative constant shift for those students (effectively negative infinity). Shifts in the travel time coefficients β_{tt} indicate age-specific behavioural preferences that were found to be statistically significant.

Table 13: School model coefficients

Variable	Car Driver	Car Passenger/PT	Cycle	Walk
ASC	-1.0994	-(ref.)	-0.7830	1.7489
ASC shift no car in hh.	0	-	0	1.6768
ASC shift age <16	$-\infty$	-	0	0
β_{tc} (log.)	-0.8337	-0.8337		
β_{tt} (linear)	-0.0838	-0.1779	-0.2420	-0.0742
β_{tt} shift age 9-12		0.0629	0.0916	0.0249
β_{tt} shift age 13-15		0.0887	0.1409	0.0333
β_{tt} shift age > 16		0.0933	0.1531	0.0352

All values are statistically significant at the 99% confidence level.

The HTS as processed for the tertiary university model contained 598 students. The resulting estimated coefficients are shown in Table 14. The adopted model specification includes alternative-specific constant shifts for households without any cars, for tertiary students above the age of 30 and for full-time workers. Both persons above 30 years of age and full-time workers have negative constant shift for the walking mode, meaning that they were observed to walk less compared to the rest of the population keeping all other variables constant. In a similar method to the workplace model, other explanatory variables are used, some of which include distance-based travel costs, mode-specific travel times, as well as parking cost (p_c) for private vehicles and access/egress times for public transport.

Table 14: Tertiary institution model coefficients

Variable	Car Driver	Car Passenger	Public Transport	Cycle	Walk
ASC	-(ref.)	-1.7847	-0.4360	-2.5066	2.3769
ASC shift no car in hh.	-	1.5568	1.7121	1.9234	2.1816
ASC shift age > 30	-	-0.7845	-0.5537		-2.5758
ASC shift full-time worker	-				-2.3775
β_{tc} (log.)	-0.7506	-0.7506	-0.7506		
$\beta_{tt,car}$ (linear)	-0.0966	-0.0966			
$\beta_{pc,car}$ (linear)	-0.0567	-0.0567			
$\beta_{ivt,pt}$ (linear)			-0.0233		
$\beta_{ivt,pt}$ shift age > 30			-0.0375		
$\beta_{accegr,pt}$ (linear)			-0.0835		
$\beta_{tt,active}$ (linear)				-0.1279	-0.0720
$\beta_{tt,active}$ shift age > 30				0	0.0331

All values are statistically significant at the 99% confidence level.

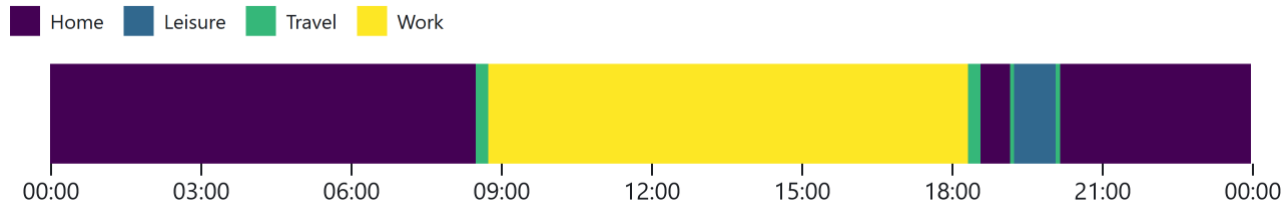
4.2.3 Daily Activity Preference Modelling

Daily activity preferences express the desire to perform activities throughout Monty's simulated day. For example, a person might go to work during the day and join a dance class in the evening. Another might have

the preference to play football in the evening or go shopping. The result is an activity schedule based on individual preferences and time budgets which is fully consistent in time and space.

A possible example of an activity schedule for a typical full-time worker is shown in Figure 22. The green areas between activities indicate trips. This agent starts working at around 9am. The workplace is not far away from their household, so the trip is relatively short. At approximately 6pm, the work activity concludes, and the agent returns home. In the evening, the person leaves home one more time to perform a leisure activity. The plot shows the full integrity of the schedules – each trip is driven by the desire or need of an individual person to perform a specific activity at a certain place and time.

Figure 22: Example activity schedule for a full-time worker



To generate such schedules for the whole synthetic population, each agent must undertake choices surrounding the number and type of activities they are meant to perform, the sequence of those activities, the locations of non-mandatory activities (locations for mandatory activities have already been chosen through long-term location choice modelling), as well as the start and end time of each activity. The methodology for the decision-making around the number and type of activities assigned to each agent is explained in subsequent subsections. This is followed by an explanation of the activity timing assignment process.

4.2.3.1 Activity Generation

During the process of activity generation, decisions are generated for each agent regarding the number and type of activities they will undertake during the simulated day. Additionally, the activities are bundled into *out-of-home tours* – a sequence of activities between two home activities (e.g., home to work to shopping to home). Each agent’s activity schedule can contain multiple out-of-home tours.

This results in activity chains that are not yet filled with information about the location and timing of activities, only the activity types. An example of a possible outcome of the activity generation step for an agent is shown in Table 15. This table shows an agent who performs two out-of-home-tours:

- The first is a work tour, meaning that this tour (per definition) must contain at least one work activity.
- The second tour is a non-mandatory tour, meaning that it does not contain work or education activities. This person has decided to perform a leisure activity in their non-mandatory tour.

This presents a very simple example of a schedule. More complex schedules may contain multiple activities per tour, including intermediate activities between home and work/education as well as work-based sub-tours (e.g., work to *lunch* back to work). As mentioned, there is no information generated regarding the locations of non-mandatory locations or activity timings at this specific model stage.

Table 15: Example outcomes of the activity generation step

Agent ID	Tour ID	Tour Type	Trip ID	Activity Type
2793363	1	Work	1	Work
			2	Home
	2	Non-Mandatory	1	Leisure
			2	Home

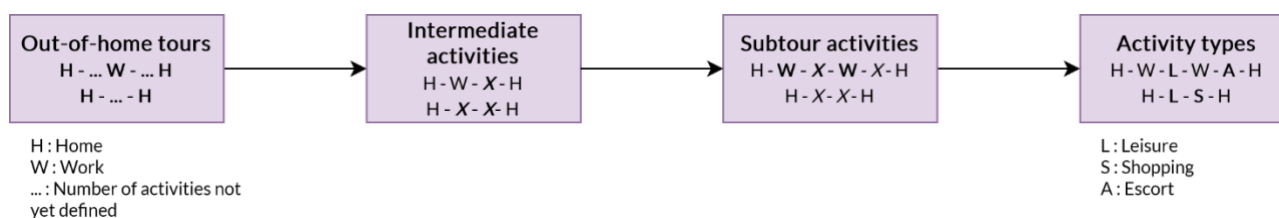
Methodologically, the procedure applies a sequence of MNL models. All choice models depend on various factors such as socio-demographic attributes, household attributes or accessibility measures. The sequence of sub-model steps within activity generation is shown in Figure 23 and are outlined as follows:

1. The *out-of-home tour* model governs how many times an agent leaves their home during the day for a specific purpose. Monty’s AcBM differentiates between three different tour types: work, school and non-

mandatory tours (the former two of which are grouped into mandatory tours). Work or school tours refer to tours where the individual performs at least one of these respective activities during an out-of-home tour. A non-mandatory tour is a tour without either work or education-related activities. A daily activity schedule may contain multiple tours of multiple types adjacent to one another.

2. The *intermediate activity* models govern preferences for performing activities on the way to a mandatory activity (referred to as *outbound* activities) as well as on the way back home (*inbound* activities). For non-mandatory out-of-home tours, the number of intermediate activities equal the number of activities within the whole tour (excluding the home activities).
3. For work and school tours, additional *subtour activity* models introduce the opportunity to leave the workplace for an intermediate activity before returning back to the workplace.
4. The last choice step governs the *activity type* of all intermediate activities. This includes the non-mandatory activity types consisting of leisure, shopping, accompanying someone, and business (only relevant during work tours).

Figure 23: Sub-model steps within activity generation



For each of the mentioned sub-model steps, individual logit models were estimated. As an illustrative example, Table 16 presents the estimated coefficients for the activity type model within non-mandatory tours.

Table 16: Preferences for performing specific activities within non-mandatory tours

Variable	Shopping	Leisure	Accompanying
ASC	- (ref.)	0	0.7156
Age 18-25 years	-	0.0720	-2.2477
Age 26-45 years	-	-0.1380	-2.0355
Age 46-65 years	-	-0.1568	-2.4087
Age > 65 years	-	-0.3655	-2.7240
Gender male	-	0.1574	-0.1373
Part-time worker	-	-0.1404	0.0483
Full-time worker	-	-0.1497	-0.1263
Primary/secondary student	-	0.1598	-1.0220
Part-time tertiary student	-	0.1228	0.1041
Full-time tertiary student	-	-0.0572	-0.2383
Child present in household	-	-0.1770	0.7315
Car available in household	-	0.2279	0.5438
Medium log. accessibility (7-8)	-	0.0302	0.1883
High log. accessibility (>8)	-	0.1439	-0.0808

All values are statistically significant at the 99% confidence level.

The coefficients indicate that:

- The likelihood of performing leisure activities decreases with age compared to shopping activities.
- Being a worker has a significant negative impact on the number of leisure activities generated compared to being a non-worker.

- Being located in areas of high multi-modal accessibility (expressed as a logarithmic summation over all destinations j for each SA2 zones based on Equation 3) is associated with higher preferences towards leisure activities.
- The preference for performing accompanying activities is positively influenced by the presence of a dependent child in the household.
- Part-time workers have a higher preference for accompanying activities compared to full-time workers.

The accompanying activity (also referred to as ‘escort’ activities) usually involves multiple agents from the same household (e.g., picking a child up from school) and is therefore somewhat different from other activities. Monty only models this behaviour implicitly as the presence of a child in a household significantly increase the likelihood of performing such an accompanying activity. However, explicit joint household interactions are not modelled within either the Monty AcBM or network simulation. This means that the current Monty implementation does not constrain individuals to do accompanying tasks together. The increased probability can be seen as a soft constraint allowing for some, but limited, behavioural realism when it comes to modelling this type of behaviour.

4.2.3.2 Non-Mandatory Locations

As mentioned previously, non-mandatory activities comprise of activity types such as leisure, shopping, accompanying someone, and business. The locations of these activities are determined individually for each instance. In contrast to a person’s place of work or study (which is usually stable across a long-term period), a person may go shopping at several different places on the same day for example.

In many cases, decisions surrounding the placement of non-mandatory locations are constrained by the activity chain itself. For example, a shopping activity might be situated in an activity chain that consists of home to shopping to work back to home. This means that the location of the shopping activity is constrained by not only the trip origin (in this case the home), but also by the work location. In such an example, it would be relatively unlikely for a person to choose a shopping location that is both far away from home and in the opposite direction to the workplace. This means that they must consider not only whether it is convenient to go from the home location to the shopping location, but also whether the workplace is accessible from the shopping location in order to arrive at work on time.

Methodologically, the decision regarding non-mandatory locations follows similar principles as those outlined for the long-term location choice models (see Subsection 4.2.2). To consider the constraints of the activity chain, Equation 3 was extended in order to estimate the preferences for choosing a non-mandatory activity location j given trip origin i and mandatory location k :

Equation 6

$$V_{j|i,k} = \theta \cdot \log A_j + \omega \cdot V_{ij} + \phi \cdot V_{kj}$$

with A being the attraction of a specific destination zone j and θ the logsize of the attraction, ω being the weight of the utility for travelling from trip origin i to j and ϕ the weight of travelling from the mandatory activity location k to j . This methodology is commonly referred to as *rubber-banding* around two anchor activities since the weights ω and ϕ increase the likelihood of choosing a location that is accessible from both activities.

Figure 24: Intermediate activity location choice

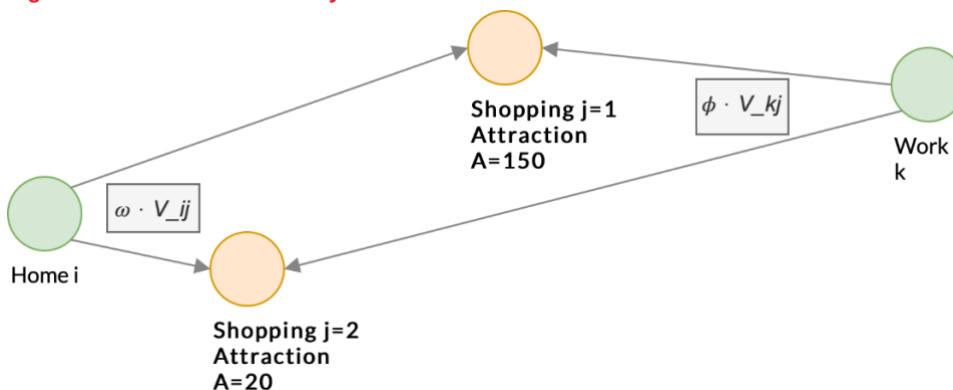


Figure 24 illustratively presents Equation 6. It shows the decision process made between two possible shopping locations with different attraction values (in this case an indicator of the size of the shop). The agent can choose between a small shop that is close to the household residence and a large shopping mall that is in the middle of the home and workplace. This means that the choice based on Equation 6 considers both the attraction of possible locations as well as the overall connectivity between all activities.

Compared to Equation 3, the utility in the non-mandatory location model is specified as a sequential travel mode and location choice model. This means that the intermediate location choice model treats the expected maximal utility (*EMU*, or also referred to as the ‘*logsum*’) from a mode choice model as a static input. Mathematically, the utility formation for both V_{ij} and V_{kj} is given by:

Equation 7

$$V_{ij} = \beta_{dist} \cdot tdist_{ij} + EMU_{ij}$$

$$V_{kj} = \beta_{dist} \cdot tdist_{kj} + EMU_{kj}$$

with *tdist* representing the routed car network distance in uncongested traffic conditions. The expected maximal utility *EMU* is calculated based on utility terms from the trip-based mode choice models:

Equation 8

$$EMU_{ij} = \log \sum_m e^{ASC_m + \sum_k^{\mathcal{K}} \beta_{k,mode} \cdot k_{ijm}}$$

Variable definitions in Equation 8 are the same as those specified for the long-term location choice model equations (see Subsection 4.2.2). A two-step process is applied in order to calibrate such non-mandatory location choice models within the Monty AcBM:

5. *Travel mode choice coefficients.* A trip-based mode choice model estimates the alternative-specific constants and the mode-specific coefficients in Equation 8.

Location choice coefficients. The resulting expected maximum utilities from the mode choice model are fed into the location choice model. The coefficients for the location choice models in Equation 6 and Equation 7 are then estimated jointly.

To calibrate this model, HTS data from 2017 to the first quarter of 2020 (three years’ worth cumulatively) was used, containing a sample of over 120,000 reported trips. The mode choice model included five mode alternatives: car driver, car passenger, public transport, cycling and walking. The estimated coefficients are shown in Table 17. Compared to the mode-specific coefficients in the workplace model, regional and behavioural constants were additionally included in this specification. The constants express preferences towards certain mode alternatives. They are fully exclusive, meaning one person cannot be in multiple *ASC*-rows, receiving exactly one constant per mode alternative.

The resulting values can be interpreted in relation to each other. For example, workers (both part-time and full-time) living in Wellington have a stronger preference for choosing public transport compared to workers living in Auckland. This is expressed by the less negative constant in Wellington compared to Auckland. Other notable examples are the relatively stronger preferences of people living in Wellington to use active modes compared to other regions. Additionally, people who are not working nor studying show a tendency to use the car passenger mode more often.

Additional supply indicators allow for consideration of the impact of the transport system on individual preferences. For example, the coefficients for travel time (*tt*) penalise being a car driver more strongly than being a passenger. This is expected, as time can be used by a car passenger for other simultaneous activities such as reading. Another notable influence is the average waiting time for public transport. The dis-utility of waiting times is substantially higher compared to actual in-vehicle travel times⁷. This indicates that people have a strong preference for avoiding waiting times, meaning that public transport service frequencies can have a high impact on the overall utility of using public transport.

⁷ In contrast to the workplace location model coefficients shown in Table 13, the public transport in-vehicle time and access/egress times are combined into one coefficient. This is because their separation in this instance led to insignificant model coefficients.

Table 17: Mode choice model coefficients

Variable	Car Driver	Car Passenger	Public Transport	Cycle	Walk
ASC primary/secondary student	-(ref.)	2.7809	1.0136	2.3721	8.4741
ASC student	-	-2.0936	-2.0936	-1.2582	5.1248
ASC Auckland worker	-	-2.8049	-3.2203	-2.3840	3.8889
ASC Wellington worker	-	-2.2885	-2.1695	-1.6132	5.4486
ASC Canterbury worker	-	-2.8379	-4.5228	-1.0405	4.1780
ASC Other regions worker	-	-2.7384	-5.3810	-1.9725	3.8722
ASC Worker without car	-	-1.1312	0.0388	0.5914	6.8933
ASC Auckland non-worker/student	-	-1.0953	-2.8836	-1.2833	4.7601
ASC Wellington non-worker/student	-	-1.3237	-2.7779	-2.2291	5.5629
ASC Canterbury non-worker/student	-	-1.6148	-6.2649	-1.5756	4.6840
ASC Other regions non-worker/student	-	-1.2917	-3.9626	-1.4205	4.7168
ASC Non-worker/student without car	-	0.4601	0.3026	1.8678	7.6940
β_{tc} (log.)	-0.3880	-0.3880	-0.3880		
$\beta_{tt,car}$ (linear)	-0.0255	-0.0205			
$\beta_{pc,car}$ (box cox*)	-0.2332	-0.2332			
$\lambda_{pc,car}$	0.5648	0.5648			
$\beta_{ivt\&acceg,pt}$ (linear)			-0.0039		
$\beta_{waiting,pt}$ (box cox*)			-0.0742		
$\lambda_{waiting,pt}$			0.7173		
$\beta_{tt,active}$ (log.)				-1.0838	-1.8437

*($x^\lambda - 1$)/ λ

All values are statistically significant at the 99% confidence level.

The coefficients from Table 17 result in an estimated VTT of \$11.20/h for car drivers, \$6.70/h for car passengers, and \$3.40/h for public transport excluding waiting times. The VTT for public transport is on the lower side, explained by the non-negligible impact of waiting times. Excluding this from the utility specification results in a VTT for public transport between \$6.00/h and \$7.00/h. It was thus a design decision to keep the waiting time in the specification, which results in a low VTT for public transport.

Using these coefficients, the individual expected maximal utilities EMU_{ij} were then calculated and fed into Equation 7. Having EMU_{ij} in Equation 7 and feeding the resulting V_{ij} back into Equation 6, the remaining coefficients include the distance-based coefficients β_{dist} , the ω - and ϕ -weights and the logsize term of the attraction θ . All these remaining coefficients were estimated as a logit model using maximum likelihood estimation. The alternatives of this model were all SA2 zones, meaning that over 2,000 alternatives were considered.

Table 18 contains all the estimated coefficients for the intermediate activity location choice model. The coefficients can be analysed across three distinct clusters:

6. *θ -coefficients.* Four different types of attractions were included in the model:

- Population counts by zone.
- Total jobs across all industries by zone.
- Total jobs in the retail/commercial industry by zone.
- Leisure counts, derived from the most frequently visited zones in the HTS data.

For most activity types, the total number of jobs was the most significant attraction variable. One exception was shopping activities, which only considered destination zones containing retail/commercial jobs. The other exception was leisure activities which considered a combination of total job counts as well as the number of leisure visits reported in the HTS as potential zonal attractions. Additionally, a small weight for population attraction was manually added to certain activity types. These were not statistically significant but were nonetheless critical for avoiding edge cases where persons did not live near zones containing jobs.

ω - and ϕ -coefficients. These represent the weight of incoming and outgoing trips. For non-mandatory tours, the first trip was handled separately because in this case both the incoming and outgoing trip are the same except for the direction of travel. Outside of this edge case, the weight of the incoming trip ω was higher than the weight of the outgoing trip ϕ . The higher the ω , the less a person is likely to consider a trip at all. For leisure activities in particular, the return trip had a comparably smaller weight. Some of these coefficients were manually adjusted higher to improve model calibration compared to their statistically estimated values.

β_{dist} -coefficients. Distance was transformed into a piece-wise linear variable. Short-distance trips have relatively high dis-utilities, flattening out with increasing trip length.

Table 18: Intermediate activity location coefficients

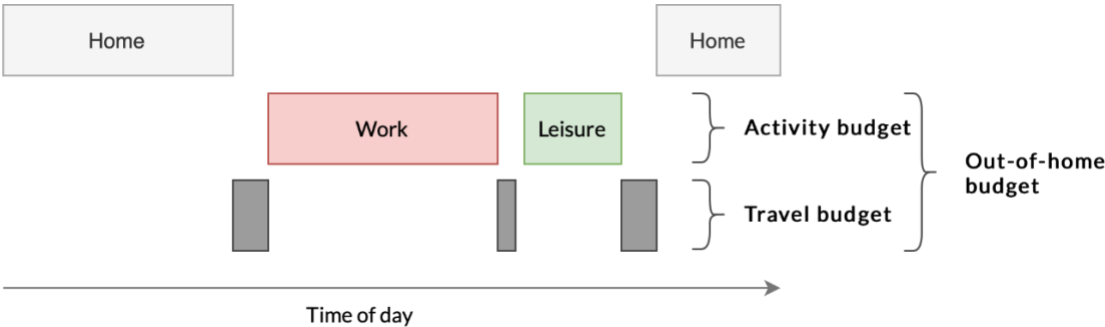
Variable	Non-Mandatory Tour			Work Tour				School Tour
	Leisure	Shopping	Accomp.	Business	Leisure	Shopping	Accomp.	All
$\theta_{total\ population}$	0		0.0001*	0	0.0001*	0.0001*	0.0001*	0.132
$\theta_{total\ jobs}$	0.256		0.725	0.729	0.670		0.578	0.513
$\theta_{retail\ jobs}$	1.000	1.000	1.000			0.82		
$\theta_{leisure}$	0.643							
$\omega_{first\ trip}$	1.000	0.960	1.240					
ω	0.890	0.680	0.820	0.800*	1.380	0.800*	0.790	0.600*
ϕ	0.360	0.510	0.640	0.600*	0.600*	0.500*	0.300	0.500*
$\beta_{dist\ 0-5km}$	-0.491	-0.366	-0.353	-0.441	-0.192	-0.484	-0.417	-0.931
$\beta_{dist\ 5-10km}$	-0.225	-0.350*	-0.228	-0.154	-0.060	-0.266	-0.215	-0.323
$\beta_{dist\ 10-40km}$	-0.071	-0.065*	-0.067	-0.049	-0.033	-0.098	-0.094	-0.149
$\beta_{dist>40km}$	=0.011	-0.025	-0.017	-0.019	-0.004	-0.022	-0.053	-0.064

*coefficients manually set for calibration purposes.

4.2.3.3 Activity Timings

With activity chains and activity locations determined for each agent, activity timings are then assigned to each daily activity schedule. The primary goal of the adopted procedure is to fit all the previous decisions into a consistent daily activity pattern. This can be a challenging task if a person wishes to perform many activities with long-distance trips between them. In such a case, the person must prioritise some activities whilst also shortening some activity durations in order to successfully complete everything within a single day. The methodology applied within Monty is based on the procedure described in (Sherr, Manser, Joshi, Frischknecht, & Métrailler, 2020).

Figure 25: Daily time budgets



To ensure integrity across both time and space during the course of the simulated day, the algorithm for choosing activity timings builds on time-budget principles. The time budgets are presented conceptually in Figure 25. The algorithm considers the total time a person needs for travelling and the time that is left for performing activities. The time needed for travelling during the entire is referred to as the *travel budget* while the total time available for performing activities is referred to as the *activity budget*. The sum of both budgets equals the total *out-of-home-budget*.

In Monty's current implementation, a global value has been assigned to each type of time budget, derived based on available HTS data. This means that the same time budgets are applied to every agent within the population. In reality, different types of agents would likely have different time budgets. Using this starting point, the activity timing algorithm executes the following steps:

1. Check if the total daily time spent travelling is feasible (in Monty's case, less than a travel budget of six hours). If this condition cannot be met, re-evaluate the choice of intermediate locations and try to find locations that lead to less travelling. Note that the long-term locations are not changed during this process.
2. Check if the total daily non-home activity duration is smaller than the activity budget of 14 hours. If the daily activity duration is too high for a given agent, repeat the choice of the activity durations by taking another Monte Carlo draw from the duration distribution.
3. Check if the combination of durations and travel times together is feasible (i.e., below the out-of-home budget of 16 hours). If not, try another random combination of both the re-evaluated intermediate location choice and activity duration choice until this condition is met.
4. With a combination of activity durations that satisfies all time budget criterion, find activity start times for each activity. The initial foundation of activity start time choice is based on observed start time distributions as observed in the HTS. An iterative process is then applied to assess whether an agent is satisfied with their selected start times throughout the simulated day by prioritising certain activities over others. For example, a person might want to shop at 8am initially, leading to a late arrival at their workplace. By prioritising the work activity over the shopping activity, the agent moves the timing of the shopping activity to ensure they arrive at work on time.

The following example is used to illustrate the practical implementation of this process. An agent living in Hamilton has been assigned a workplace in the Auckland CBD as part of previous steps within the activity synthesis process. After work, they wish to perform a leisure activity in Silverdale before turning home to Hamilton. The daily activity chain is thus home (Hamilton) to work (Auckland CBD) to leisure (West Auckland) back to home (Hamilton).

This example requires the agent two hours to commute in the morning. The trip from the workplace to the leisure activity may take another 1.5 hours as it involves crossing the Auckland Harbour Bridge during the evening peak period. The return trip home then takes three hours. This results in a total travel time of 6.5 hours during the day which violates the constraint surrounding the daily travel budget. The agent must re-evaluate the leisure activity since it would be too much in one day to need to travel to Silverdale and back to Hamilton. A generated alternative results in a leisure activity in the Auckland CBD instead, reducing the total amount of travel required to 4.5 hours rather than 6.5 hours.

With an out-of-home-budget of 16 hours, 11.5 hours remain for performing non-home activities. The person may decide to not spend the entire out-of-home budget, working for eight hours then completing a leisure activity for only two additional hours. These decisions are fully valid as all time-budget constraints are met.

Finally, the start time algorithm attempts to iteratively fit the activities into the day. Taking the example at hand, the agent may choose to start working in the Auckland CBD at 7am in the first iteration, which then implies the start time of the leisure activity at 3:30pm (taking into consideration travel time between the work and leisure activity in the afternoon). The iterative process evaluates if all timings are convenient. In this case, 5am may be deemed too early to leave home. After a few iterations, the algorithm may land on a later work start time that is more personally satisfying for this agent (e.g., 8:30am). The result is a complete daily activity schedule that is consistent in time and space.

4.2.4 Facility Sampling

The previous subsections described the process used to generate fully consistent daily activity schedules across time and space. In this process, long-term locations as well as non-mandatory locations were chosen at the zonal level (SA2 zones). The ambition of Monty is to simulate locations as precise point-coordinates for the purposes of the network simulation, which are referred to as *facilities*.

In order to assign a facility to each chosen activity within an SA2 zone, Monty applied a simplistic procedure. The underlying dataset containing facilities is provided by Land Information New Zealand (Land Information New Zealand, 2019). This is a geographic dataset containing building outlines without any further contextual information attached (e.g., whether a structure is a shopping mall or residential house). The currently facility-sampling implementation in Monty assigns a facility to each activity given an SA2 zone based on a purely random draw.

In the case of long-term locations (i.e., household residence, workplaces and schools/universities), the random draw is cached for each agent. If the same activity type occurs multiple times during the day, the location is taken from the agent's cache. This assures that those types of activities take place at the same building throughout the simulated day.

This procedure will be re-addressed in future work to integrate building-specific information into the activity-facility joining process, as well as to potentially incorporate information about proximity to public transport stops.

4.2.5 Simulation Procedure

The simulation of the Monty AcBM computes a daily activity schedule for each agent. It is implemented within a Python module (heavily inspired by (Swiss Federal Railways, 2023)) using the following inputs:

- Synthetic population.
- Travel skim matrices (origin-destination matrices of travel times and other simulation statistics, described in Chapter 5) including parking cost and cost for travelling.
- Behavioural coefficients as introduced throughout this chapter.
- Zonal attraction values (e.g., the number of jobs per zone).

The run times of the AcBM are mainly driven by the number of agents within the population as well as the number of zones considered across the various location choice models. In Monty, the synthetic population for the 2018 base scenario includes approximately 4.7 million agents. Excluding agents aged 4 years and younger from actively making activity decisions, 4.4 million agents remain in the simulation. The SA2 geographic unit contains approximately 2,200 boundaries.

Given these inputs, it takes approximately two hours to compute the activity schedules and associated locations for the entire 2018 base population using a laptop⁸. The computational complex location choice models apply highly performant parallel-computing libraries. The result of the simulation is a large file containing a list of activities for each agent in a temporally sequential order with accompanying information about activity type, start- and end-times as well as geographic coordinates.

4.2.6 Validation

The ability of the AcBM to accurately replicate observed activity decisions was assessed across several dimensions:

- *Activity generation.* Quantification of whether a realistic number of daily trips were being produced for each agent across demographic groups and across each activity type.

⁸ MacBook Pro: Apple M1 Pro, 8 cores, 15 GB of RAM.

- *Location choice.* For those trips being produced, were their length distributions accurate compared to available observed data? Where more granular observed data was available, evaluation of whether the specific distribution of destinations relative to origins was also undertaken.
- *Activity timings.* When considering each person's entire activity schedule, were activities starting at realistic times and were they assigned a correct distribution of durations?

The subsequent subsections summarise the AcBM's validation performance across these dimensions respectively. The HTS was used as the primary source of observed data driving model performance benchmarking. 2018 census outcomes were also utilised for specific tests surrounding workplace location choice. Mode choice outputs from the AcBM were not directly validated at this stage as these results are only used as an initial seed for the mode choice process that occurs during the agent-based simulation. Mode choice accuracy is discussed in Chapter 5 after the introduction of the agent-based traffic assignment process.

4.2.6.1 Activity Generation

The quality of activity generation within the AcBM is characterised by its ability to answer the following questions:

- Are agents within the simulation being assigned a realistic number of trips overall and within subsets across different demographic groups?
- Commuting trips represent a critical subset of travel. Are agents within the simulation being assigned the right number of commuting trips, and, by association, a realistic number of non-work trips within their schedules?
- Is the distribution of trips across activity types realistic when considering all travel that is planned to occur?

Figure 26 shows the mean number of trips generated for agents by the AcBM compared against available valid data for respondents from the HTS. This measure demonstrates a mean absolute error (MAE) of 0.1% using the denoted age groups, with very slight overestimation of daily trips made by each person.

Figure 26: Mean daily trips by age group

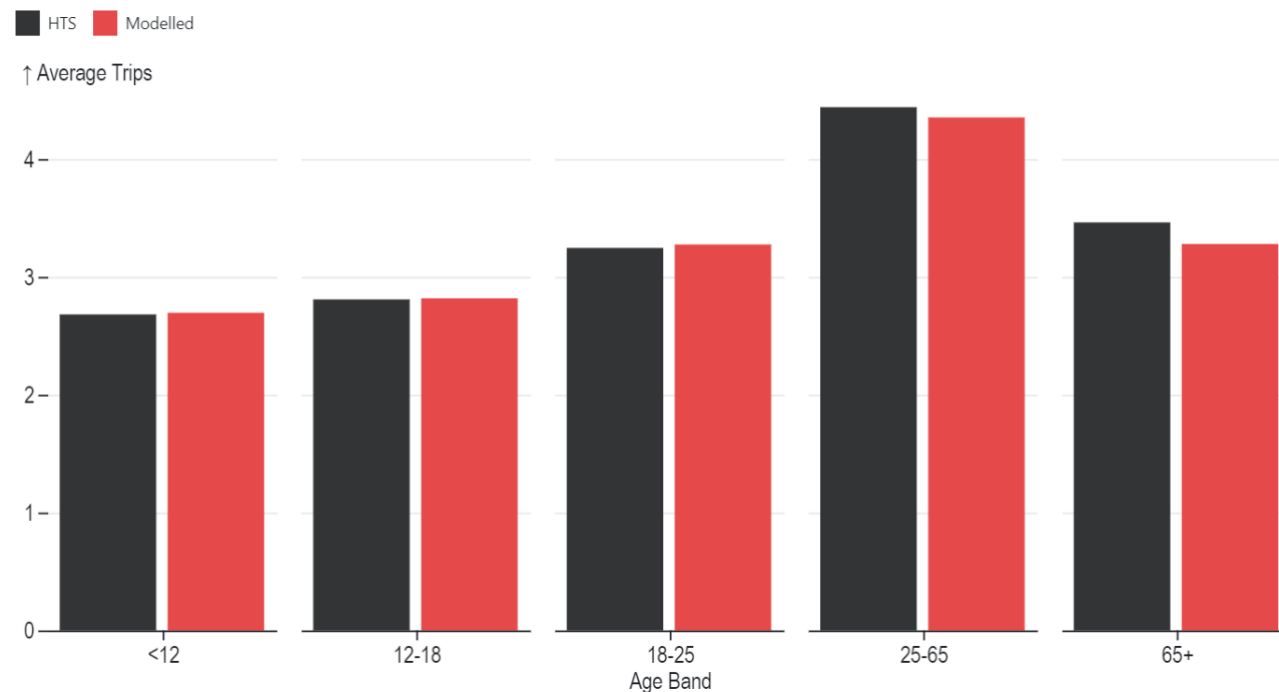


Figure 27 and Figure 28 similarly show the proportion of agents making specific numbers of daily trips segmented by either their labour force status or studying status respectively. All of these distribution show excellent agreement between the patterns observed across HTS data and the schedules generated by the AcBM.

Of particular note is the successful capturing of the bimodal distribution underlying trip counts for those working or studying full time.

Figure 27: Trip count distribution by labour force status

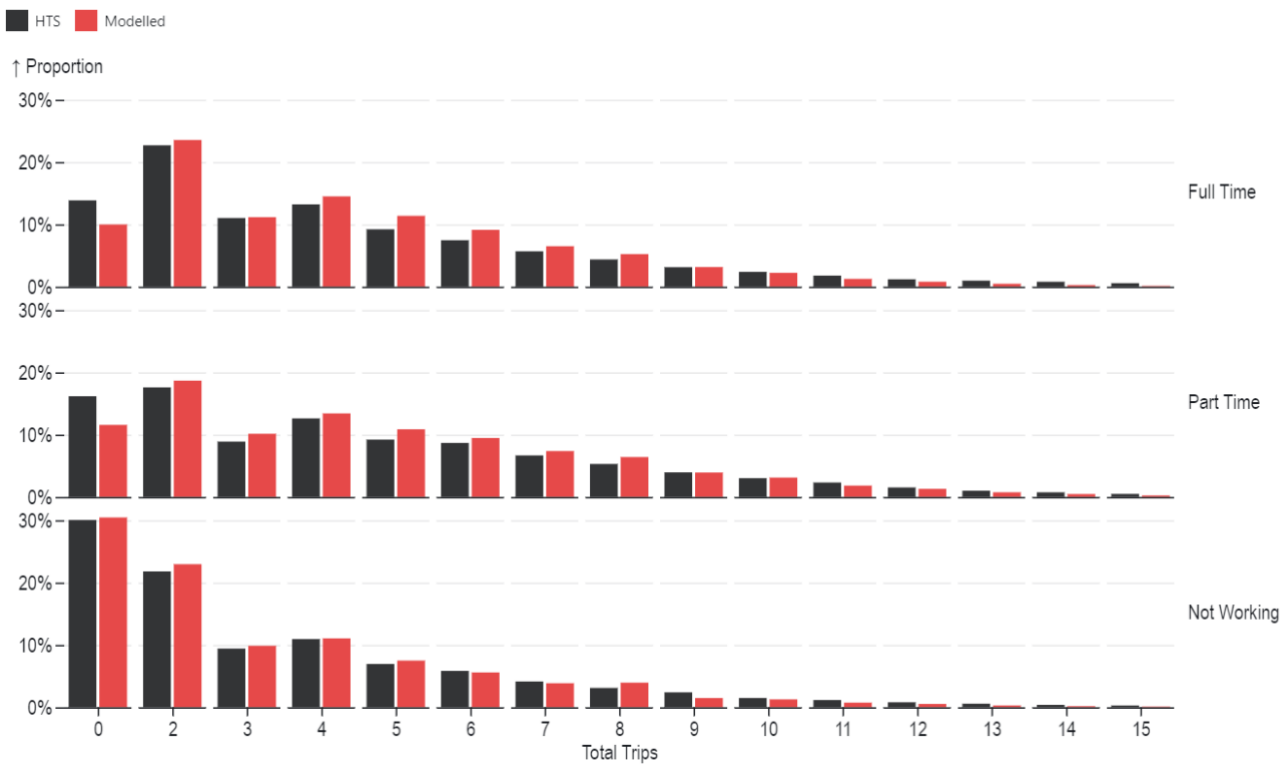
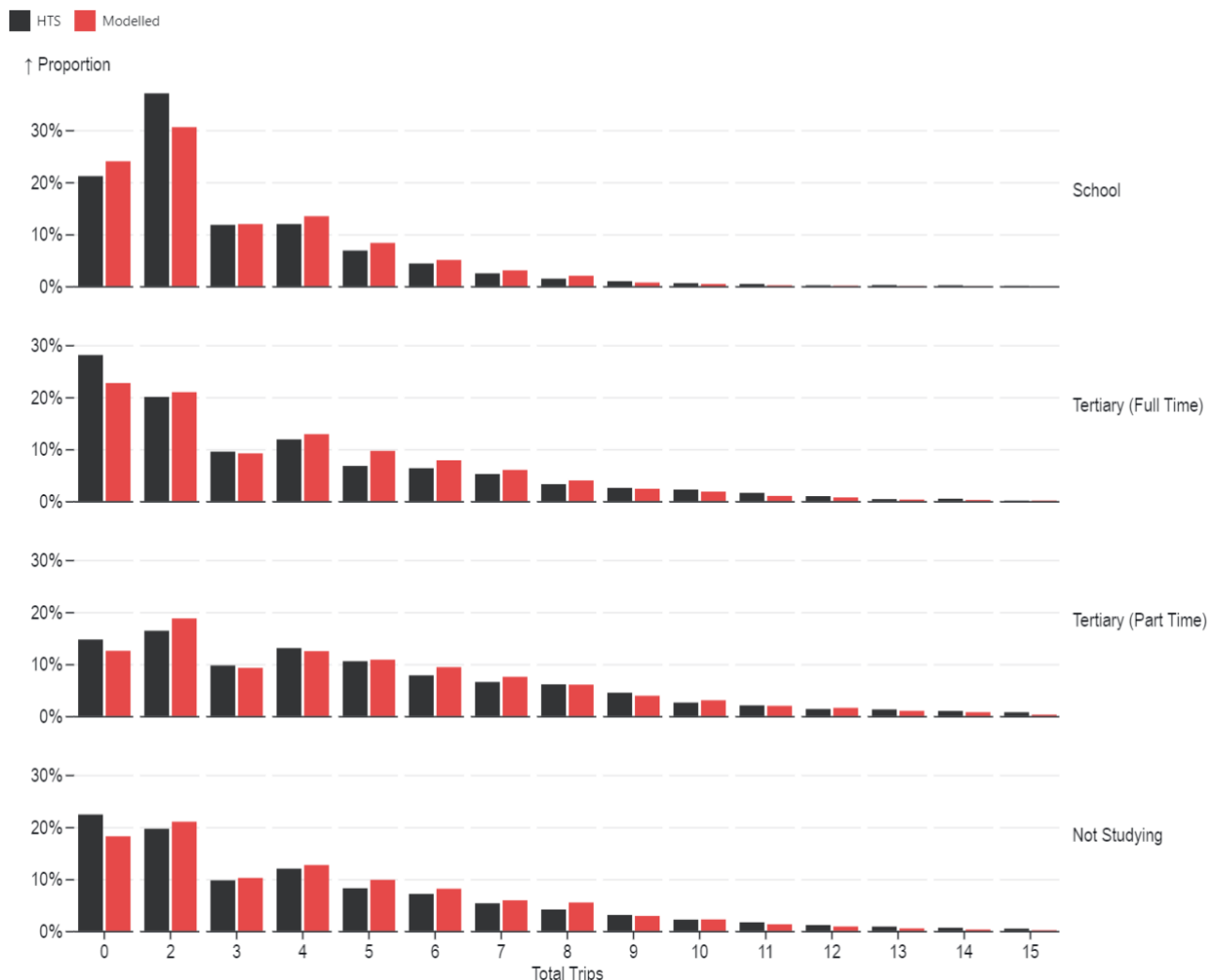


Figure 28: Trip count distribution by studying status



Beyond itemising activity schedules by individual trips, activity sequences can also be decomposed into *tours*. A tour is defined as a sequence of activities that starts and ends at an agent's home. For example, a home to work back to home activity sequence could be described as a work tour with a single work activity. Figure 29 shows the distribution of work tour counts undertaken by each person across HTS and AcBM generated activity schedules. Those working full-time or part-time are assigned a realistic number of work tours, including the circumstance where an agent may be working full-time but decides not to travel to work during the simulated day. This case is slightly under-represented within Monty, meaning that the average number of work tours shows a tendency to be higher compared to the HTS.

Figure 29: Work tour count distribution by labour force status

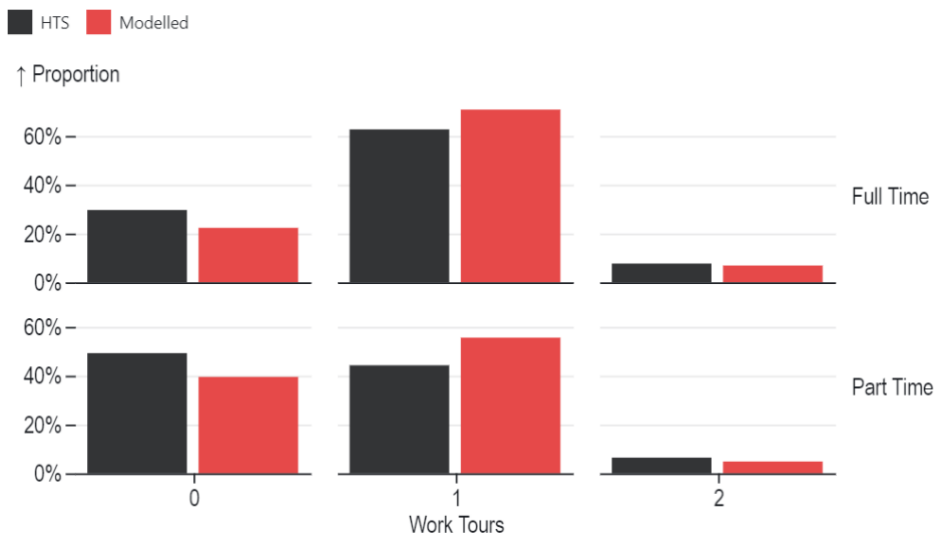
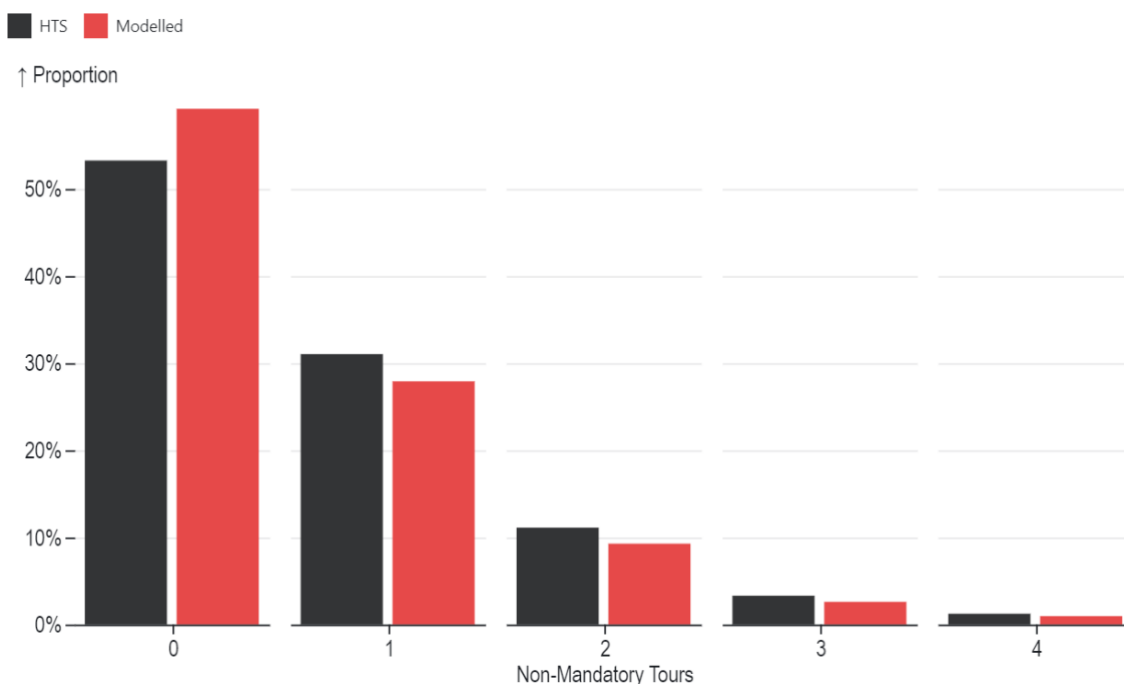


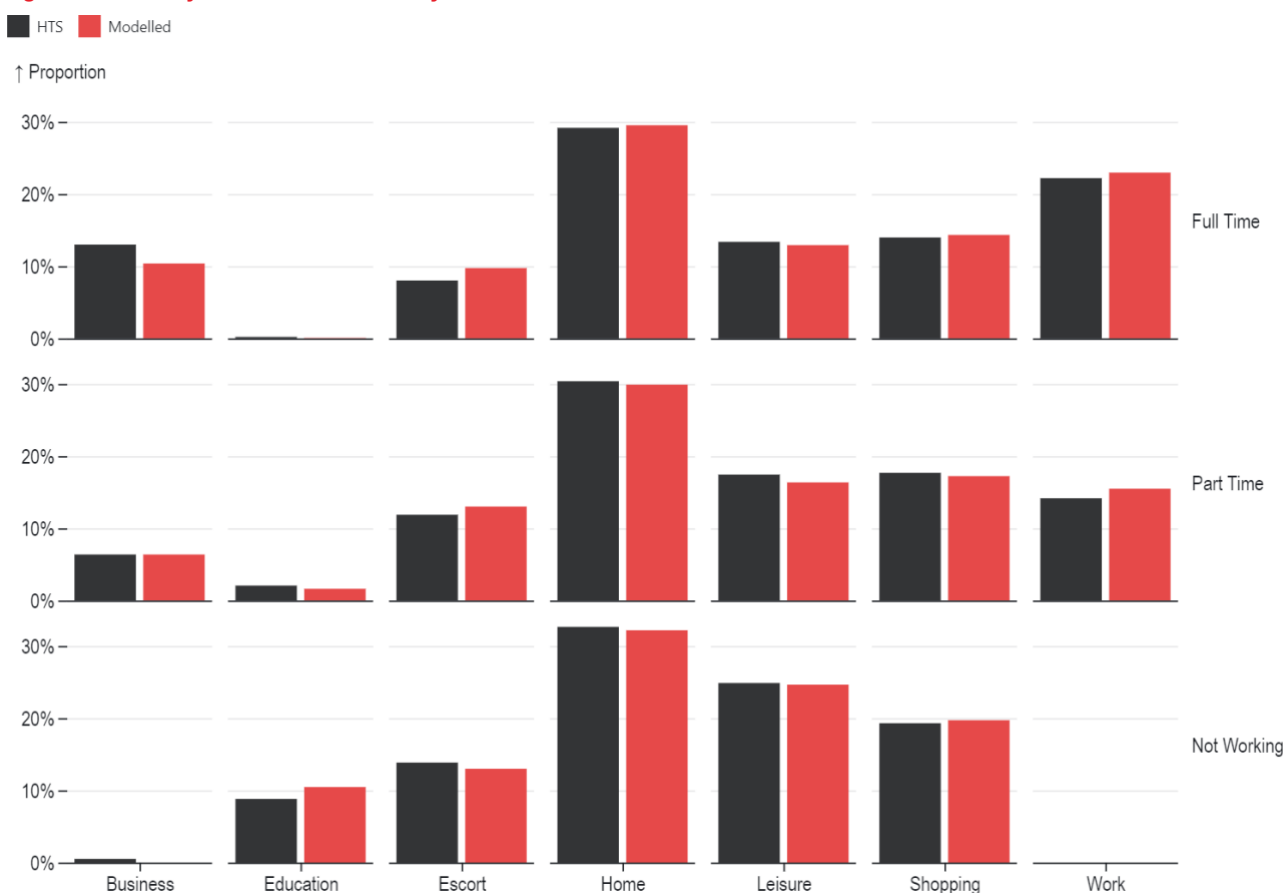
Figure 30 shows the corresponding number of non-mandatory tours undertaken by each agent. This refers to home to activity to home sequences where the intermediary activity is not one of either working or studying. There is good agreement between the patterns produced by the AcBM and respondent activity schedules seen within the HTS.

Figure 30: Non-mandatory tour count distribution



The final comparison presented of this type is activity distributions based on labour force status as shown in Figure 31. It can be seen that an attribute like labour force status is associated with a reasonable distribution of activity types. Correlations with other attributes like age, studying status and having a child show similarly acceptable outcomes.

Figure 31: Activity count distributions by labour status



4.2.6.2 Location Choice

Location choice quality within the AcBM can be assessed through the following questions:

- Are trips of a specific purpose and origination from specific regions demonstrating the correct trip distance distributions?
- Where detailed origin-destination data is available, are people travelling to the right share of locations for specific purposes?

Figure 32 shows the length distribution of all trips generated by the AcBM compared to trips observed within the HTS. Figure 33 similarly shows the individual length distributions for trips by each regional council of origin. There is a strong match in distribution shape between the HTC and AcBM-generated data, with the largest amount of variation seen for the shortest trip lengths. This level of variation is expected due to the following factors:

- Short trips can often be underestimated within household travel survey-style data as respondents neglect to record minor journeys that take a very short amount of time.
- The trip distances for intra-zonal journeys (i.e., those that stay within the confines of a given SA2) are only accurate to a level of detail that corresponds with the size of that SA2. More granular location choice within an SA2 was not undertaken beyond facility selection.

The regional council-specific analysis in Figure 33 additionally provides insights into the non-homogeneous behaviour across the country. Large cities like Auckland demonstrate more distributed patterns of trip length, associated with the greater variety of journey types associated with these locations. In contrast, less dense regions like Otago demonstrate more concentrated short trip behaviour as would be expected. Figure 34 shows trip length distribution by purpose. Again, other than a small over-estimation of short trips across some purposes, individual length distributions show strong agreement with HTS-derived values.

Figure 32: Trip length distribution

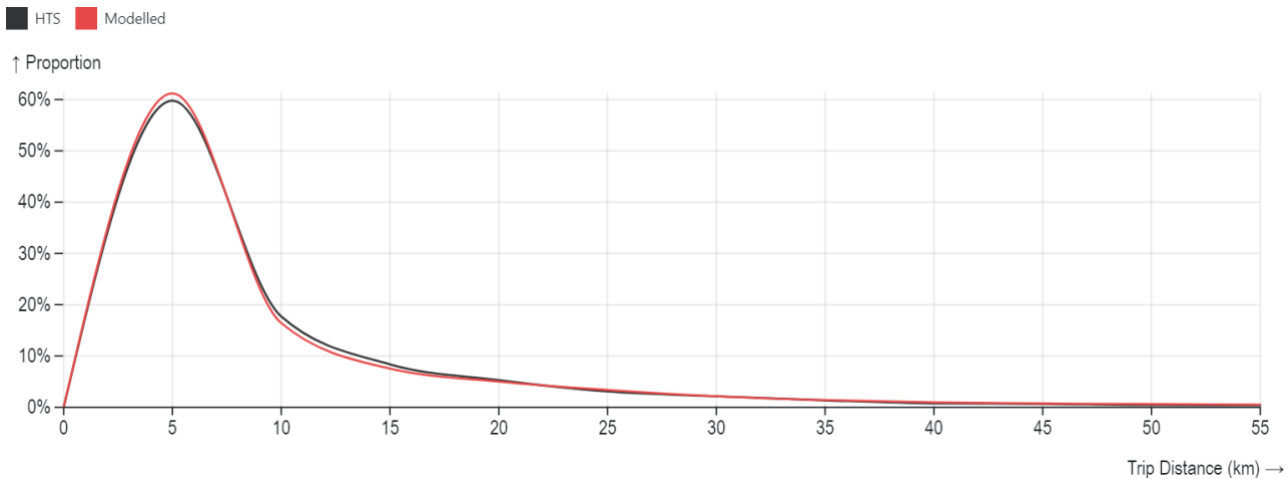


Figure 33: Trip length distribution by regional council

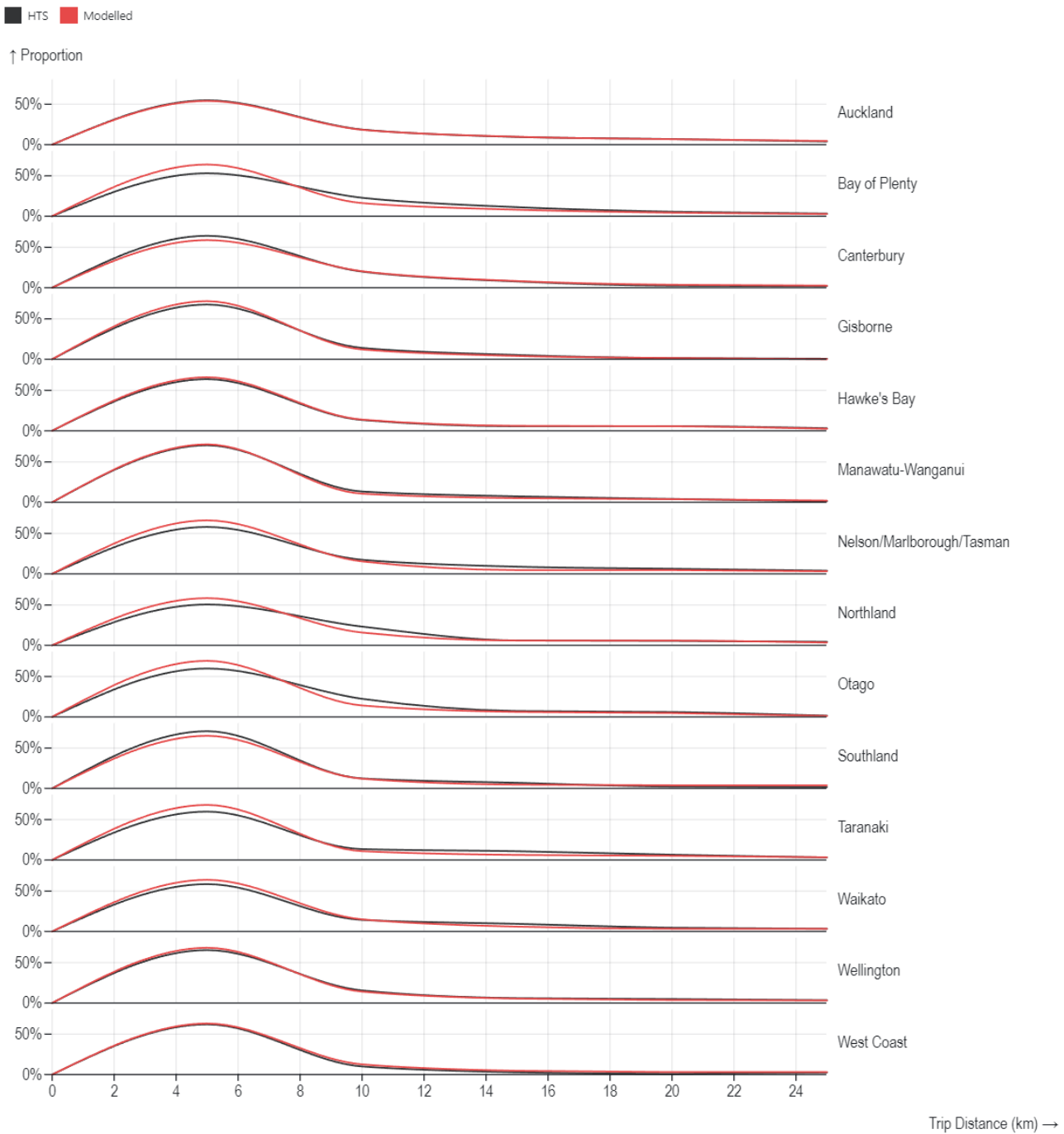
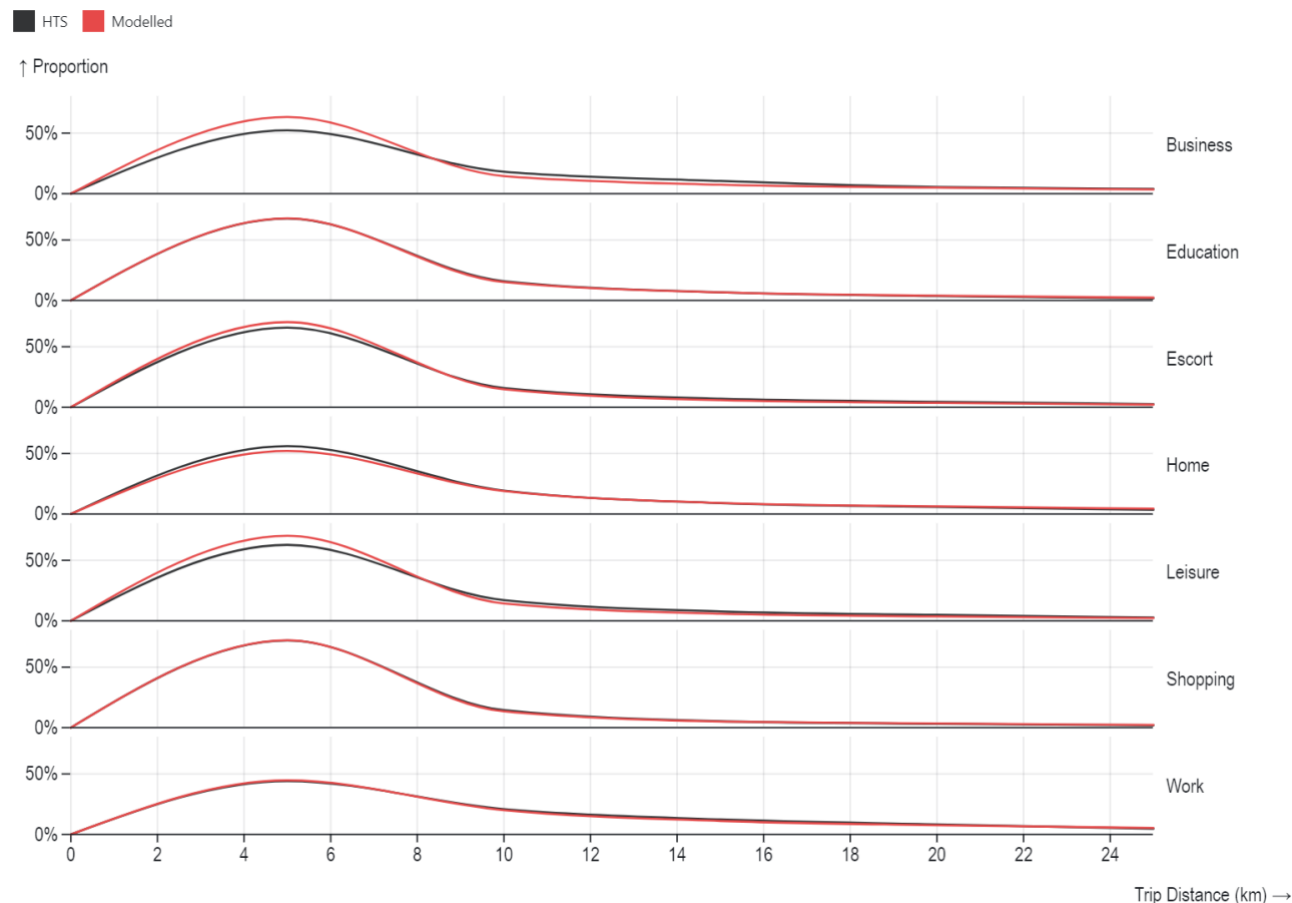
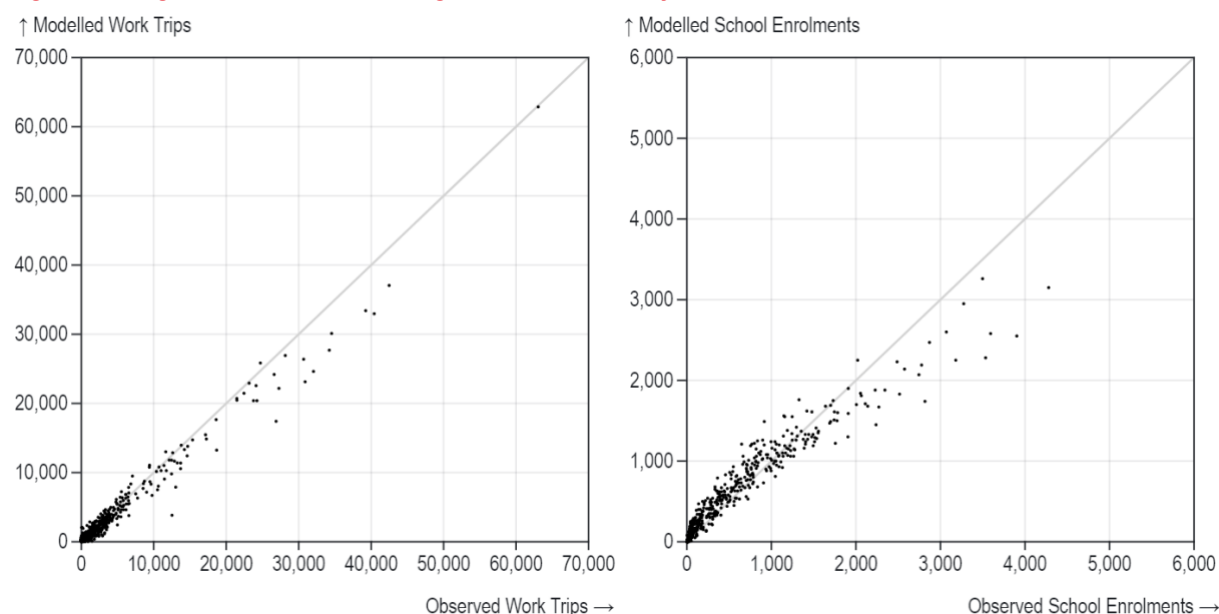


Figure 34: Trip length distribution by purpose



2018 census data provides a tallying of all commuting trips that occurred on census day, including information surrounding origin and destination SA2. Figure 35 compares the number of people commuting from zone to another zone between the AcBM generated outcomes and this census commuting data. Correlation between the datasets is very high, exhibiting an R^2 of 0.98. Additional data is made available by Stats NZ summarising the number of enrolments present across all private and public schools in 2018. Aggregating this dataset to an SA2 level, this figure also shows the number of primary and secondary enrolments generated by the AcBM against this dataset. Secondary school enrolments demonstrate very strong correlation between the AcBM-generated locations and the observed enrolment data with an R^2 of 0.92. Primary school enrolments display more variance, exhibiting an R^2 of 0.88.

Figure 35: Origin-destination commuting/enrolment values by zone



4.2.6.3 Scheduling

Scheduling within the AcBM is concerned with the correct sequencing of activities, expressed through the establishment of activity start times and durations within each agent's activity schedule. Quality of the scheduling can be characterised through the following questions:

- Are trips of specific types starting at realistic times of the day?
- Are trips of specific types demonstrating the correct distribution of durations?
- Does the overall share of activities taking place throughout the modelled day match those seen in observed data sources?

Figure 36 shows the distribution of all activity start times across all types of trip purpose, comparing AcBM generated outcomes to patterns seen in the HTS. Accuracy of this general profile is very high, with a RMSE of 0.0047% across hourly groups. The distinct morning and evening peaks are captured across this profile, alongside the correct magnitude of inter-peak trips as well as the long evening tail.

Figure 36: Activity start time distribution

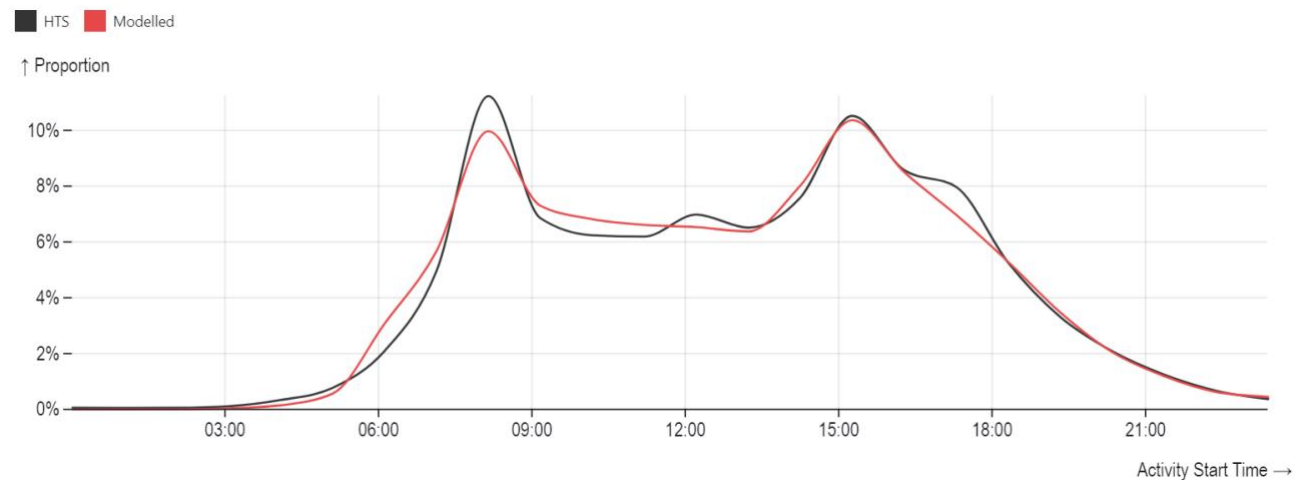


Figure 37 shows the same data segmented by activity type. Variation between HTS data and AcBM-generated values is higher at this level of granularity, but outcomes are still acceptable especially given the AcBM's ability to capture the individual peaks of each purpose's distribution throughout the modelled day.

Figure 37: Activity start time distribution by activity type

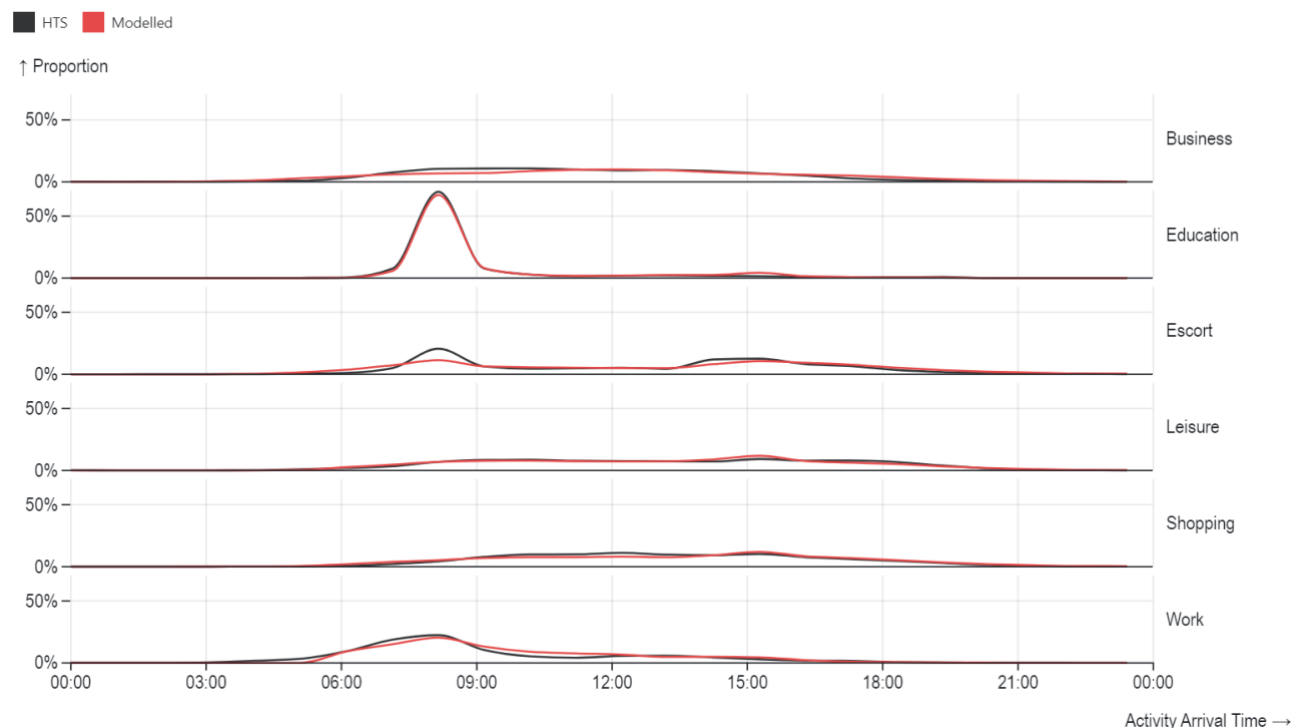
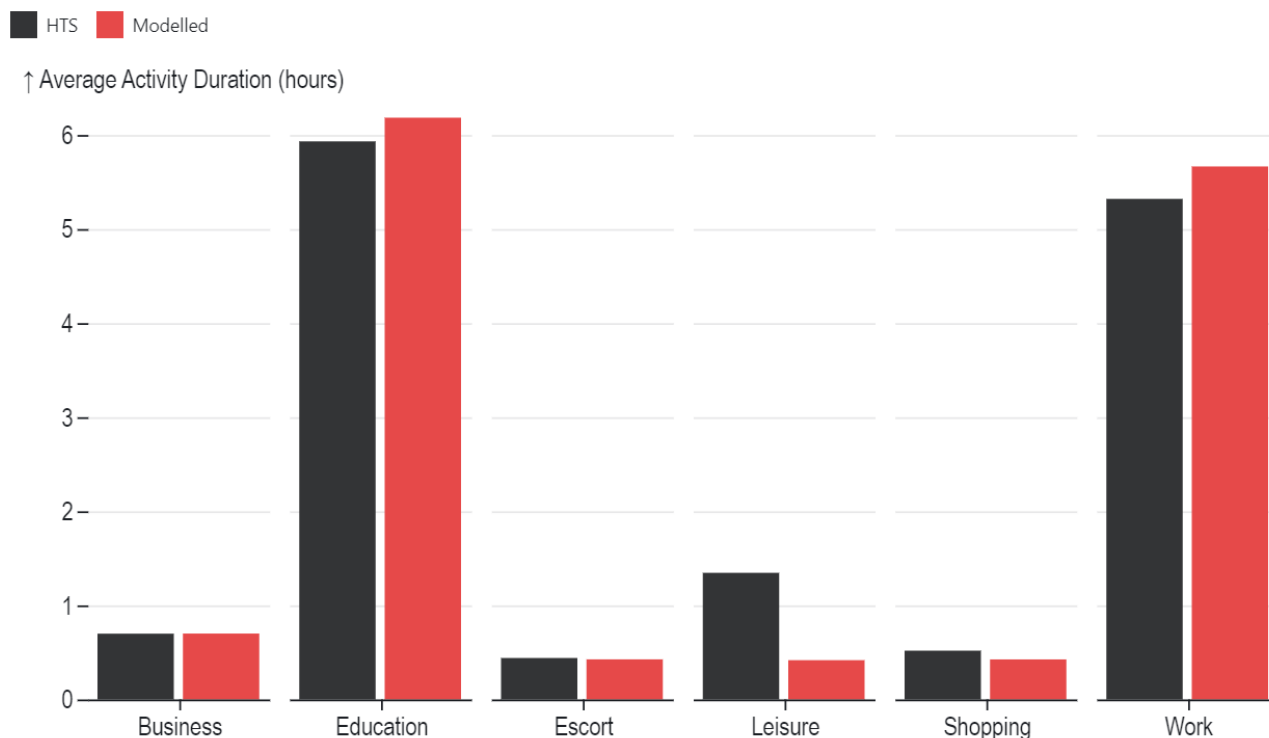


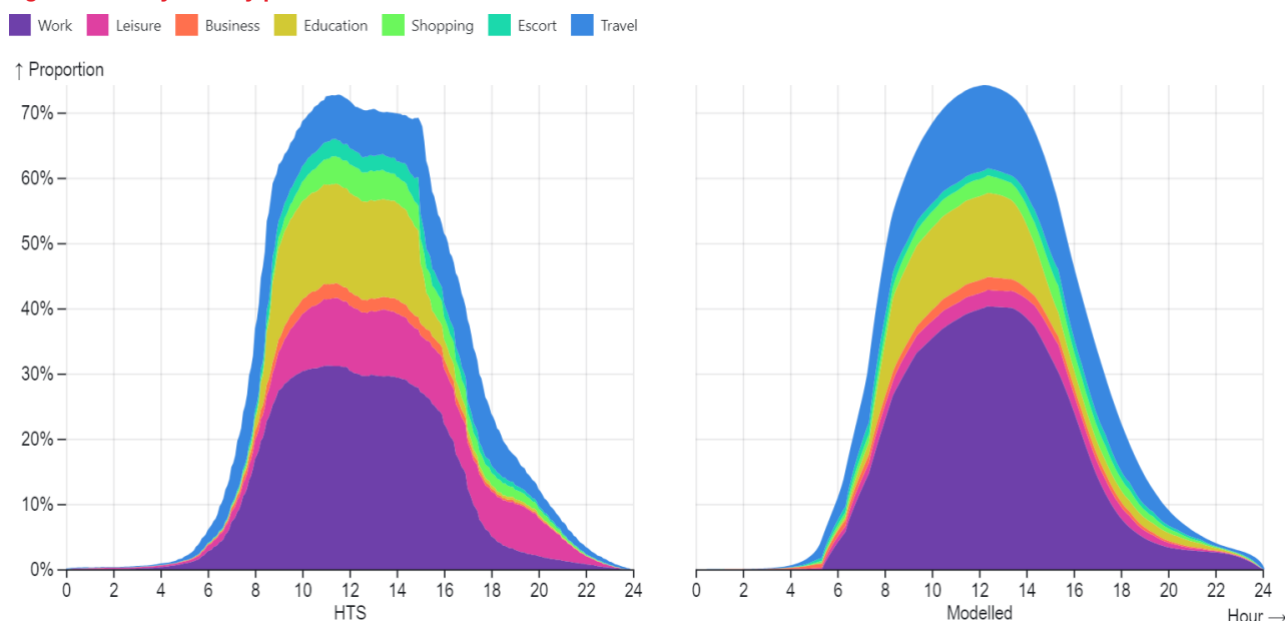
Figure 38 shows the mean activity duration times generated by the AcBM compared to activity durations seen across the HTS data available. All purposes demonstrated realistic levels of duration, with the largest deviation seen across the leisure purpose – the AcBM tends to under-estimate the duration of leisure activities by approximately one hour compared to examples of leisure activities seen within the HTS.

Figure 38: Mean activity durations by trip purpose



A final aspect of scheduling undertaken by the AcBM is the placement of activities throughout the day. Figure 39 shows a side-by-side comparison of the share of particular activity types being undertaken throughout the day as generated by the AcBM compared to the HTS. As expected, work and education constitute a majority of time being spent between the hours of 6am and 7pm, with other activities and travel making up the remaining balance. The activity profiles confirm the observations made in the previous Figure 38, with the duration of leisure activities tending to be under-estimated by the model. In contrast, the blue area representing travel is slightly larger in Monty compared to the HTS. Conservative travel times have been assigned by the AcBM on purpose to ensure that agents are provided enough time to travel. This will be updated with more accurate information with the network simulation stage.

Figure 39: Daily activity profiles



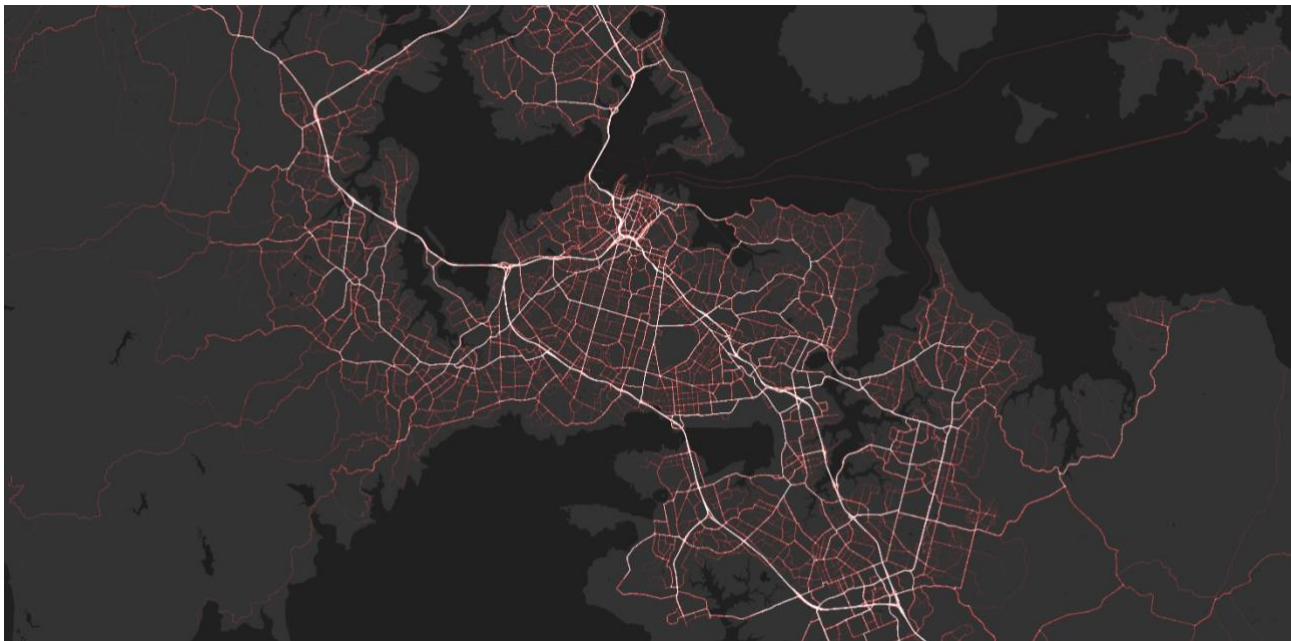
4.2.7 Discussion

The demand model has been demonstrated to be capable of accurately reproducing observed behaviours seen in survey data for the 2018 base year scenario. This includes choice dimensions such as number of activities per type, their locations as well as timings. Monty's AcBM is primarily based on behavioural models that express individual preferences. This enables the capability to not only to produce a base year scenario, but also facilitates predictions about potential future changes such as demographic shifts, changes in land use, or induced travel due to infrastructure investments.

The following limitations of Monty's current AcBM implementation have been identified as most critical to be addressed in future development efforts:

- *Sensitivity testing.* Whilst all coefficients in the choice models are statistically significant and have been validated with the means at hand (such as through calculation of VTT savings), elasticities are also an important quality target as well. Systematic testing should be conducted to calculate the model's response to changes such as impacts to travel demand due to travel time improvements. These elasticity values can be validated against historic data and figures available in the literature.
- *Behavioural variables.* The behavioural models used for the generation of activities capture a rich selection of variables. This includes socio-demographic variables such as age, household-specific attributes (such as the presence of a child), or network measures such as accessibility to jobs. Currently not implemented are factors that influence activity behaviour such as income or available private vehicle type. Once the mentioned variables are included as part of the population synthesis process, the utility specifications in activity synthesis can be extended to measure their potential impact on decision-making behaviour.
- *Intra-household dependencies.* Household composition can have a significant impact on people's behaviour since this can often manifest as hard constraints during the day. For example, the need to drop children off at school often represents a relatively inflexible activity households with school-age children must contend with. Currently, Monty treats household composition as a behavioural factor, i.e., the presence of children increases the likelihood of an adult agent needing to perform a pick-up or drop-off activity. However, these activities are not directly coupled with the activity schedule of the child. This means that the child may not go to school at the same time as the parent does the corresponding drop-off activity. Future extensions of the AcBM will aim to more accurately capture this type of intra-household dependency.
- *Land-use data.* The location choice models within the AcBM rely on metrics that measure the attractiveness of a zone for performing certain activities. Whilst high-quality data is openly available for the number of jobs as well as the number of school enrolments, the attraction measurements for shopping and leisure activities are less direct. In the current implementation, these are captured through a combination of industry-specific job counts and reported trip destinations observed within the HTS. This approach comes with limitations, namely that sampling issues can arise where zones are erroneously not considered as shopping or leisure locations because they neither appear in the HTS or industry-specific job counts. In the future, additional data sources will be leveraged to better represent shopping and leisure attractiveness.
- *Facility sampling.* The current approach to dis-aggregate chosen locations from SA2 zones to specific coordinates is based on simplistic random principles. It does not consider either the size of a building as a weight or the connectivity of a building to the network. Future work aims to address this limitation through the application of data containing specific building characteristics. This dataset can then introduce weights into the sampling, ensuring that larger buildings attract more people to perform specific types of activities for instance.

5. Network Simulation



Following population synthesis and activity synthesis (see Sections 4.1 and 4.2), each agent needs to travel through the network from one activity location to the next while interacting with other agents. This chapter elaborates on this step, describing the process employed to leverage outputs of the Activity-Based Model (AcBM) within an Agent-Based Model (ABM) for the purposes of scenario testing.

Conceptually, each agent within the ABM aims to find optimised travel modes, routes, and departure times to undertake their assigned activity schedule. The ABM itself can be characterised as an iterative learning framework whereby these agents interact within a transport network (the structure of which was described in Chapter 3) and learn based on each other's decisions. Such interactions may happen on the road network (leading to forms of congestion) or within the public transport system (leading to crowding).

Consider a situation whereby an agent desires to travel from their household residence to their workplace by car. Within the simulation, this agent may face a situation where many other agents are using the same roads as them. The resulting congestion on those specific roads ends up causing substantial delays to that agent's journey. Given this outcome, the agent may explore alternative options such as travelling via rail. These newly explored options may end up being faster, resulting in the agent changing its method of commuting. This example presents a typical decision that is made within the ABM's iterative learning algorithm.

By iteratively testing new options and learning from the outcomes of previous decisions, the simulation aims to reach an equilibrium state. New options in this case can refer to testing alternative routes, modifying trip departure times or switching transport modes. Once the simulation has reached an equilibrium, each agent theoretically cannot improve their individual daily utility score (a measure of satisfaction in a sense) any more given the decisions of all other agents.

The end state of the ABM simulation can provide valuable insights across all interactions between agents and the transport system, both at an aggregated level and at the level of individual agents in specific circumstances. For example, it is possible to test which demographic groups may benefit most from a piece of infrastructure depending on the attributes made available during population and activity synthesis. Aggregated statistics can also help understand broader impacts of policies such as those aimed at meeting emissions reduction targets.

This chapter begins with a broad overview of the ABM methodology in Section 5.1. Following this, the associated iterative learning algorithm is described in Section 5.2. Behavioural coefficients and other ABM configuration parameters are discussed in Section 5.3, whilst special representation of freight demand is treated in Section 5.4. Skim matrix generation – an important output of transport models – is covered in Section 5.5. Computational resource considerations are discussed in Section 5.6. The chapter concludes with ABM validation (Section 5.7) and a discussion of next steps in the development process (Section 5.8).

5.1 Agent-Based Model Overview

The agent-based simulation relies upon three major inputs as illustrated in Figure 40:

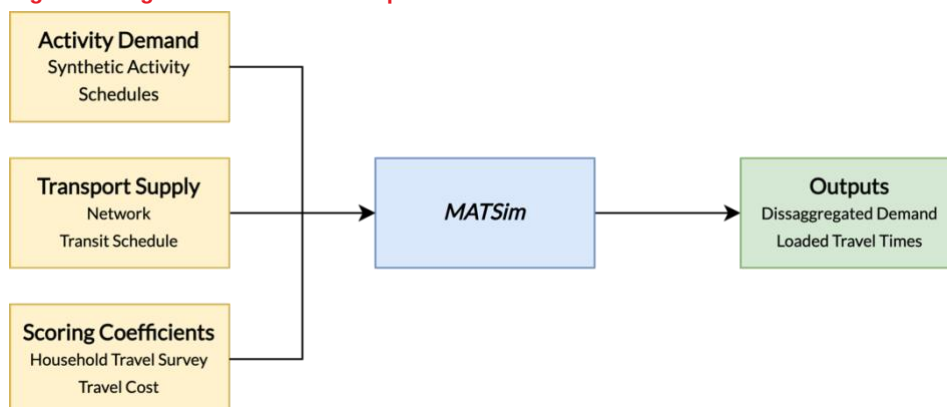
- The first is the synthetic activity schedules as generated by the AcBM (see Section 4.2). These contain a planned activity sequence for each agent within the synthetic population, each of which consists of activity types, corresponding timings as well as spatial information regarding activity locations. The agent-based simulation does not further modify the sequence of activity types and locations, but it does optimise activity timings and the travel episodes between activities.
- The second input is the transport network for New Zealand (see Section 3). Within the transport network, the representation of public transport is of particular importance – the ABM simulates such services as individual vehicles over the course of the modelled day according to the specified timetables.
- The third input is utility scoring coefficients that express preferences for how agents weigh certain decision outcomes against others (see following Section 5.3.5).

These inputs are then processed using the open-source software package MATSim (Horni, Nagel, & Axhausen, 2016). MATSim stands for *Multi-Agent Transport Simulation* and has its roots in academia, with core development originating from ETH Zurich and TU Berlin. It has a steadily growing community and a well-established Java codebase (MATSim, Multi-Agent Transport Simulation, 2023). Over the past decade, an increasing number of practitioners and organisations from broader industry have leveraged MATSim to facilitate agent-based simulation work in the transport modelling space.

Based on the input demand, network and scoring coefficients, MATSim applies an iterative learning algorithm as previously mentioned and described in more detailed in the following sections. During this entire process, MATSim treats agents as individuals with time and space fully disaggregated (no zones or time periods). That means that the resulting output is a complete list of synthetic persons, each of which with their travel behaviour over the day. This goes down to travelled routes as a sequence of link, accompanied by by-the-second time stamps of when the agent entered and left a certain road segment or a public transport vehicle.

Those high-resolution outputs generated by MATSim can then be used for a wide range of analysis, ranging from aggregated forecasts of travel volumes on roads to the routes of individual agent trips.

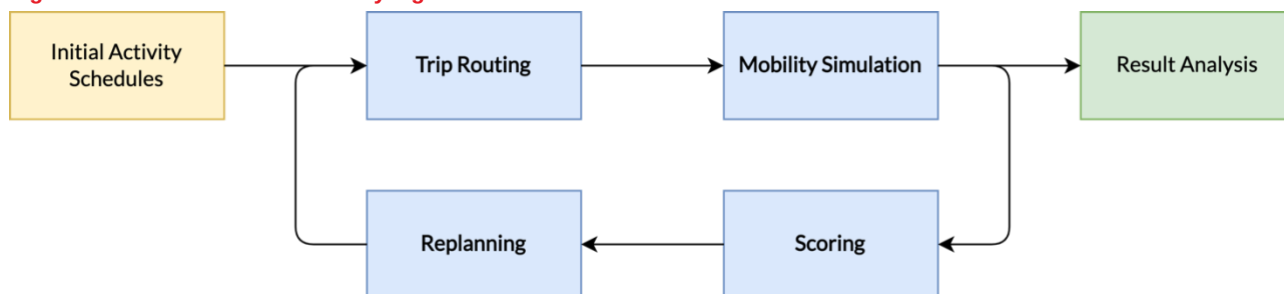
Figure 40: Agent-based simulation process



5.2 Co-Evolutionary Learning Algorithm

The iterative learning process as implemented within MATSim builds on co-evolutionary principles. Figure 41 depicts how MATSim structures this process within the simulation system. A scenario is tested across a series of iterations (often numbering in the hundreds). Across these iterations, agents continuously modify their travel behaviour, testing them in the simulation whilst interacting with other agents. Schedule modifications that result in higher utility scores are more likely to be kept by each agent whilst low-scoring alternatives are rejected. As this process is repeated across iterations, every agent in the population slowly maximises their ultimate daily utility score. This selection process corresponds to principles of ‘survival of the fittest’, in which random mutations (in this case schedule modification) evolve to an adapted overall population.

Figure 41: Iterative co-evolutionary algorithm



Individual stages of the MATSim core algorithm are described as follows:

- *Initial activity schedules.* From the activity synthesis, each agent is assigned an initial activity schedule that forms a skeleton of their simulated day. This schedule contains an activity sequence, with each activity associated with a type, start time and location. At this point, an initial transport mode for each trip between activities is also generated as an initial estimate.
- *Trip routing.* MATSim will consider the activity locations in a given agent’s schedule and generate an appropriate route using the indicated trip modes. The path search takes into consideration network travel speeds as informed by the previous MATSim iteration (if there was one), thus accounting for congestion and any associated public transport delays as relevant for each agent. For each trip, the routing algorithm returns a precise sequence of network links for the agent to use during the upcoming simulation stage.
- *Mobility simulation.* Following trip routing, MATSim will then simultaneously simulate all of the generated agent movements on the transport network. As described previously, if many agents choose to travel on the same network links or public transport vehicles, congestion and crowding will result. The mobility simulation will then reflect this in the network conditions experienced by agents (e.g., by manifesting congestion delays). Such effects will be considered in the trip routing of the subsequent simulation iteration, ultimately affecting agent decision-making through the subsequent scoring and replanning stages.

From a technical perspective, the precise mobility simulation adopted within MATSim is a form of *mesoscopic* traffic flow simulation. Each vehicle is simulated through the network individually but turns and intersections are handled based on simplistic first-in-first-out principles.

- *Scoring.* Following each iteration of mobility simulation, the scoring stage is used to calculate the utility (or satisfaction) of each agent based on how ‘well’ they were able to carry out their activity schedule. The higher the utility score, the more optimal the adopted solution for a given agent. This utility score is comprised of positive utilities received for performing activities and (typically) negative utilities (also referred to as *generalised cost*) for travelling between activities (additionally capturing the need to pay tolls and fares, as well as more intangible forms of travel friction such as public transport transfers).

Given this construction, agents are incentivised to minimise the generalised cost associated with their travel whilst maximising the time used for performing activities. While discarding plans with low scores and keeping highly scoring alternatives, the population converges towards a maximum in possible aggregate utility.

- *Replanning.* Once all agents have a total utility score calculated for them, the replanning stage is then used to generate (if necessary) any modifications that each agent will adopt in the subsequent simulation

iteration. This can be a route, departure time or transport mode modification. The aim of these modifications is to find a solution for each agent that has the potential of scoring higher than their previous attempts. Better-scoring plans replace worse-scoring plans within the ‘memory’ of each agent, eventually trending the entire simulation towards equilibrium given enough iterations.

In any given iteration, it is worth noting that only a specific randomly selected share of agents will be allowed to randomly modify their schedules and explore opportunities. If all agents were able to modify their plans in every iteration, the simulation would devolve into a random exploration of the plan space with no chance of converging. By only allowing certain portions of agents to modify their plans in each iteration, stability is induced providing a greater chance of shifting towards equilibrium.

- *Results analysis.* This is the final stage after the iterative simulation process has converged towards a maximum in average agent scores. The raw output consists of an event log that details every single action that each agent undertook during the final simulated iteration. This can then be used to generate aggregated statistics such as link volumes and public transport patronage, matrix outputs as well as more specialised analyses focusing on individual agent movements.

5.3 Model Configuration and Parameters

The 2018 base year scenario of Monty uses a standard release version of MATSim (version 15) with additional software add-ons tailored to New Zealand-specific requirements. The following subsections describe the MATSim settings as applied within Monty. Subsection 5.3.1 explains the different transport modes used within the simulation. The configuration options adopted for the simulation are described in Subsection 5.3.2. Subsection 5.3.3 gives insights into the applied methodology to consider parking cost. Following this, Subsection 5.3.4 provides an overview of how Monty groups agents into behaviourally similar clusters for both scoring and replanning. For these groups, the utility specification and associated behavioural coefficients are outlined in Subsection 5.3.5 and replanning strategies in Subsection 5.3.7 respectively, with Subsection 5.3.6 detailing the methodology behind the time-of-day decisions if the agents in Monty.

5.3.1 Transport Modes

The agent-based transport simulation provided by MATSim allows for the representation of a range of different transport modes. Monty specifically differentiates between the modes of travel shown in Table 19.

Table 19: Simulated transport modes in MATSim

Type	Mode	Routing Mechanism	Simulation Representation
Private Vehicles	Driver	Loaded Network	In Traffic
	Passenger		Teleportation
Public Transport	Rail	Timetable	Timetable
	Bus		
	Ferry (non-car)		
	Cable Car (Wellington)		
Active Modes	Cycle	Average Speed	Teleportation
	Walk		

- For the private vehicle modes, of particular note is the differentiation between car drivers and car passengers. Both car driver and passenger trips are routed on the network, meaning that travel times consider congestion effects. However, these modes differ in that car driving journeys are directly represented within the mobility simulation and thus cause congestion, whilst passengers are teleported using simulated travel times from the previous simulation iteration. The passenger mode is thus modelled simplistically. The agent does not actually traverse the network in a modelled vehicle with a driver that drives to the same destination at the same time.

Regardless of whether a car driver or passenger trip is being considered, the network-based router applied for these journeys is called *speedyALT*, an efficient implementation of the A-star landmark algorithm. Given the trip start and end activity locations, the router first looks for the closest link around those activity coordinates. Certain links such as motorways are access restricted, meaning that they are not allowed as

trip start or end links. Based on the found trip start and end links, the router then calculates the least-cost link sequence in the car network (excluding for example dedicated bus lanes).

- Public transport comprises of rail, bus and ferry (with the exception of the Cook Straight ferry) journeys, all of which are handled in the same fundamental way. Multiple modes can be used during the same trip, for example bus-to-rail interchanges. Both routing and simulation of these modes are timetable-based, meaning that public transport vehicles strictly follow the planned timetable. The routing algorithm implementation used for the timetable-based public transport modes is called *SwissRailRaptor*. The least-cost path search considers all elements of a public transport trip from access to a stop, in-vehicle times, number of interchanges, to egress from the last stop. It lists all available alternatives and chooses the route that is associated with the minimal generalised cost for a given agent.
- Active modes include cycling and walking. For both, the routing in MATSim is based on a straight-line distance calculated from the trip origin to trip destination multiplied by a detour factor of 1.3. The travel time is computed based on average speeds, which are set at 15 km/h for cycling and 3 km/hr for walking. In the simulation, active modes do not interact with other modes in the network. They use *teleportation* to travel from trip origin to destination and use the estimated travel time as described.

5.3.2 Mobility Simulation

The mobility simulation is the step within MATSim in which agents execute their travel decisions and interact with other agents on the transport network. MATSim has two main implementations for this step, named *QSim* and *Hermes*. QSim offers more functionality such as dynamic speed limit changes over day, but Hermes comes with faster overall simulation run times (around a factor of 2 in the case of Monty). By default, Monty uses the Hermes implementation of mobility simulation within MATSim (also for presented validation in this report), but also allows for dynamically changing to QSim. Both QSim and Hermes send vehicles through the network based on the free-flow-speed and capacity of network links. If a link is at capacity, first-in-first-out (FIFO) principles are used to decide which vehicle can enter the link once it has free capacity. The other vehicles must wait, and thus experience delays in their trips relative to the free-flow speeds available. This assignment process means that any queues in the network may propagate to impact upstream links and will vary dynamically across simulation steps, features that differ from strategic models reliant on individual link-based volume delay functions and static demand periods.

In MATSim, it is common practice to down-sample the population to save on computation time. This is done by, for example, simulating only a randomly selected portion of households in the population. This relies on the assumption that realistic travel times can be approximated through such down-sampled scenarios. To facilitate this style of testing, the mobility simulation requires two scaling parameters to be defined:

- The flow capacity factor is responsible for scaling the number of vehicles that can cross a link in a certain timeframe, i.e. vehicles per hour.
- The storage capacity factor defines how many vehicles a link can hold based on its length and the number of lanes.

The factors shown in Table 20 are those adopted by Monty, empirically determined to minimise the differences in travel times between the different sample sizes and best match observed volume and travel time data. To illustrate the impact of these factors, take the storage capacity factor of 0.40 for a 25% population simulation as an example. Naively, one might assume that under such a scenario the storage capacity of each link should be 0.25 – i.e., 25% of that assumed for a 100% sample population. However, due to the dynamics of the network simulation this can overrepresent congestion to unrealistic levels. Increasing this to 0.40 absolves these issues, resulting in more realistic flow dynamics for small sample sizes. Calibration of such factors is an important step in model development as it is often not feasible to run 100% sample simulation with available resources and time constraints.

Table 20: Simulation scaling parameters

Sample Size	Flow Capacity	Storage Capacity
10% population	0.10	0.17
25% population	0.25	0.40

5.3.3 Parking

The cost for parking a vehicle plays a crucial role in the process of deciding about the travel mode. Typically, it requires time and effort to find parking and may be costly in specific areas at certain times. Monty estimates a parking cost model based on urban density as introduced in Section 3.1.2. This parking model outputs an hourly parking cost for each SA2 zone. In MATSim, those hourly parking cost are assigned to each network link based on the SA2 in which the link is located.

During the mobility simulation in MATSim, Monty implements a vehicle tracker that registers a vehicle whenever an agent parks it at a link with parking rates larger than zero. The agent then performs the scheduled activity (for example running errands). When returning to the vehicle, the agent must pay the parking cost based on the duration that the vehicle has been parked at the link and the link's parking rate. The activity type 'home' is excluded from paying for parking, but all other activity types require paying the amount assigned to the link in combination with the parking duration.

This means that Monty includes a parking component in the scoring of a trip (see Subsection 5.3.5). As a result, the model is sensitive to changes in parking cost, as the agents would try to explore travelling with alternative modes or spending less time at certain activity locations with increasing parking cost.

5.3.4 Subpopulations and Scoring Groups

MATSim allows for differentiating between different *subpopulations* – subpopulations are groups of agents with similar behaviour in terms their replanning strategy. For example, this allows for constraining certain agents to not modify their travel modes between iterations if necessary. Each subpopulation must also be given a set of scoring parameters. Monty adopts a differentiation between two subpopulations: *regular agents* and *freight demand*. Regular or 'normal' agents include all residents within New Zealand with their activity behaviour generated by the activity synthesis. Regular agents test all replanning strategies and optimise travel mode, route, or departure times during the MATSim simulation. Freight agents are only allowed to explore different network routes for their planned journeys.

Additionally, Monty introduces the custom concept of *scoring groups*. This is motivated by significant differences in observed behaviour when comparing the certain groups of agents. For example, full-time workers in the HTS report to travel by car more often than other groups of agents. Methodologically, scoring groups will experience a nuanced delta or shift in some of their utility coefficients. This shift is applied to the original scoring coefficients of the agent's subpopulation. In the example of full-time workers, Monty would give them a positive delta in their preference of driving a car (constant).

Additional scoring groups are used for income-dependent money scoring to differentiate the willingness to pay for a reduction in travel time depending on the agent's income. All scoring groups and their corresponding utility shifts are presented in Table 21.

5.3.5 Scoring Parameters

The scoring coefficients within MATSim are used to help measure the satisfaction of agents as they are testing plan modifications. For example, an agent may test a later departure time for their commuting trip to work in order to avoid congestion. This may indeed lead to a shorter travel time, but also consequently result in a late arrival at their workplace. Scoring coefficients are used to determine whether such a decision ends up being a net benefit or loss to the agent's overall daily utility score, biasing behavioural preferences towards certain outcomes.

In the mentioned example, the total daily utility for that agent decreases if the negative weight for being late to work was greater than the positive weight gained by saving travel time. In that case, the agent would be better off not changing their trip departure time and contending directly with congestion to avoid being late to work. To capture such interwoven dependencies between activity and travel decisions, the scoring formulation of each individual activity schedule contains two components as seen in the following equation:

Equation 9

$$U_{daily} = \sum_{activity}^A U_{activity} + \sum_{trip}^{A-1} U_{trip,mode}$$

where A is the number of activities that are to be performed during the course of a day by an individual agent. Each *activity* contains a performing period, which is defined by a start time and a duration. A *trip* follows each activity (except for the last activity, typically being at home). For each given time stamp in the simulation, an agent can either be performing an activity *or* travelling which means there will never be an overlap between the two utility components.

$U_{activity}$ in Equation 9 represents the total utility for spending time at each individual activity *act*:

Equation 10

$$U_{activity} = U_{duration,act} + U_{early\ start,act} + U_{late\ start,act}$$

where $U_{duration,act}$ is the reward for performing a certain activity (this is typically positively weighted). $U_{early\ start,act}$ and $U_{late\ start,act}$ penalise arriving early or late to an activity. This means that agents aim to maximise the time spent performing activities whilst simultaneously being incentivised not to start too early or too late. More detailed information about each utility component can be found in (MATSim, Multi-Agent Transport Simulation, 2023).

The second component in Equation 9 is used to score trips, including mode-specific travel times and costs:

Equation 11

$$U_{trip,mode} = asc_{mode} + \sum_k^K \beta_{k,mode} \cdot x_{trip} - \beta_{money} \cdot (\gamma_{cost,mode} \cdot dist_{trip} + \gamma_{parking} \cdot dur_{act, follow})$$

where asc_{mode} is the mode-specific constant. K is the set of explanatory trip variables, typically containing at least travel time and distance components. In the case of a public transport trip, K additionally includes the waiting time and number of transfers. $\beta_{k,mode}$ is the corresponding mode-specific coefficient or weight of each trip variable k . To include monetary travel cost, a distance-based cost is used for each mode. For this purpose, the term β_{money} translates monetary cost $\gamma_{cost,mode}$ into generalised utility units. The last term is also money-related and refers to the parking cost. The monetary cost is a result of the hourly parking rate as looked up from the parked link multiplied with the duration of the following activity. To convert the monetary parking cost into negative utility of parking, it is also multiplied with β_{money} .

Table 21 shows the specific scoring coefficients adopted by Monty, informed through both (manual) calibration and existing literature (Waka Kotahi NZ Transport Agency, 2020). In some cases, nuanced scoring shifts were applied for certain groups of the population as highlighted in cursive letters (as mentioned in Subsection 5.3.3). The shifts are added as utility deltas to the original parameter depending on the agent's attributes. For example, the original coefficient for converting monetary cost into utility (β_{money}) is 1.3. If the agent is a high-income worker, a negative utility shift of -0.2 is applied to this value, totalling in 1.1. This means money has less weight for those agents and makes them less sensitive to changes in pricing.

In general, the utility for performing an activity is highly positive (7.4). Being early results in an indirect penalty of -5.0 as the agent does not gain any performing utility during the waiting period ($\beta_{early\ arrival} - \beta_{performing}$). Similarly, being late is penalised with -5.0 in Monty. This decision was informed through calibration outcomes.

Table 21: MATSim scoring coefficients

Variable	General	Car Driver	Car Pass.	Bus	Rail/Ferry	Cycle	Walk
β_{money} [util/\$]	1.3						
β_{money} <i>shift high income</i>	-0.2						
β_{money} <i>shift low income</i>	+0.2						
$\beta_{performing}$ [util/h]	7.4						
$\beta_{early\ arrival}$ [util/h]	2.4						
$\beta_{late\ arrival}$ [util/h]	-5.0						

Variable	General	Car Driver	Car Pass.	Bus	Rail/Ferry	Cycle	Walk
<i>asc</i>		0.0	-2.5	-2.0	-2.0	-4.5	0.0
<i>asc shift high income</i>				-3.0	-3.0		
<i>asc shift low income</i>		-3.0					
<i>asc shift full-time worker</i>		+4.0	-2.0				
<i>asc shift Auckland residents</i>		+6.0	-2.0	-2.0	-4.0		
<i>asc shift Wellington residents</i>		-4.0		-1.5	+6.0		
$\beta_{travel\ distance}$ [util/km]		0.8	-0.2	0	0	-1.6	-1.0
$\beta_{travel\ time}$ [util/h]		-8.5	-8.0	-4.0	2.0	-0.5	-0.5
γ_{cost} [\$/km]		0.29	0.29	0.35	0.35	0	0
$\gamma_{parking}$ [\$/h]		link-based	link-based	0	0	0	0
$\beta_{waiting\ pt}$ [util/h]				3.0	3.0		
$\beta_{transfer\ pt}$ [util]				-2.0	-2.0		

A central concept in MATSim is representation of the opportunity cost of travel (i.e., $\beta_{travel\ time} - \beta_{performing}$). This basically means that whilst an agent is travelling, they are not solely penalised by the coefficients related to travel time – the agent is also simultaneously missing out on the positive utility for performing an activity as it is not possible to perform an activity and travel at the same time.

Take travelling via bus as an example – the $\beta_{travel\ time}$ equals -4.0. The opportunity cost (or negative utility) for travelling by bus can thus be characterised roughly as 11.4 utility cost units per hour ($-4.0 - 7.4$). This opportunity cost of travel can then be used to translate utility units into hours. For example, each transfer from one bus line to another public transport service is penalised with two utility units ($\beta_{transfer\ pt} = -2.0$). This means that each transfer is penalised with the utility equivalent of $\frac{-2.0}{-11.4} = 5.7$ minutes.

Additionally, opportunity cost of travel (see explained above, $\beta_{travel\ time} - \beta_{performing}$) can be used to calculate VTT savings. The VTT can provide a sensible measure to validate each mode-specific coefficient (as has already been applied in Subsection 4.2.3 regarding activity generation) depending on the marginal utility of money β_{money} . As the term β_{money} in turn is depending on an agent's income, the VTT is also different for certain income groups. It can be expressed as the opportunity cost of travel divided by β_{money} ⁹.

Doing so, the VTT for medium-income car drivers is \$12.20/hour and for car passengers it is \$11.80/hour. For public transport, the VTT for bus use of medium-income agents is \$8.80/hour and for rail \$4.20/hour. This means that people are willing to spend more money to save travel time in private vehicles compared to public transport. Framed differently, people are more likely to accept higher travel times on public transport compared to when using private vehicles. This can be explained by the fact that time in public transport can often be used to do other things, such as answering work emails or watching/reading content. For high-income car drivers, the VTT totals in \$14.50/hour. This means that high-income agents are willing to pay more to save travel time. This makes them less sensitive to price changes such as peak-hour tolls.

5.3.6 Time-of-Day Decisions

Making a choice between different transport modes is relatively trivial. It is beneficial for an agent to change their transport mode if it results in a net utility gain based on trip scoring parameters. As there are a limited number of modes to choose from, the choice set is clear and can be exhaustively enumerated. In contrast, MATSim treats time as a continuous variable, meaning an agent can choose from an almost infinite number

⁹ See chapter “A closer look at scoring” in (MATSim, Multi-Agent Transport Simulation, 2023).

of activity timings (by the second). This is based on a fundamentally novel methodology as Monty addresses the question of how flexible people are in shifting each of their planned activity start times and durations.

For example, a peak-hour toll might encourage certain agents to travel earlier or later than they otherwise would depending on how flexible their activities are. The key to determining such outcomes lies in the activity scoring as introduced in Equation 10. This equation consists of two main components:

- *Activity start time*: The start time of an activity may be relevant if the agent is not flexible in arriving earlier or later. For example, dropping off children at school often has a non-negotiable timing. This means that the agents must arrive at a given time. This behaviour can be modelled in MATSim through defining the ‘opening time’ and ‘latest start time’ parameters. Arriving earlier than the opening time results in an early arrival or waiting penalty in combination with not gaining utility whilst spending time at the activity. Arriving later than the latest start results in a penalty for the late arrival as shown in Table 21.
- *Activity duration*: The duration utility is influenced by the time spent at an activity location. Agents can maximise their utility by setting all their activity durations as close to the specified ‘typical’ duration of each activity type as possible. This typical duration can also be seen as a ‘desired’ duration, meaning how long the agent prefers to stay somewhere. For example, the typical duration for a work activity is usually around 9 hours in the case of a full-time worker. Not conforming with the desired duration for a certain activity does not result in a direct utility penalty. However, the agent misses out on gaining additional utility which is an indirect utility penalty (also referred to as opportunity cost).

The combination of desired activity timings and durations allows modelling of the flexibility of agents in regard to undertaking certain activity types. In MATSim, the aforementioned penalties for being early or late are not dependent on the activity type, meaning that the penalty or the flexibility for being late at work or school is the same. Instead, MATSim introduces one set of typical durations and desired start timing for each activity type. For example, a ‘work’ activity can be defined through the typical duration 9 hours, with an opening time of 7am and the latest start time of 9am. This means that the agents prefer to work according to this pattern. The ‘school’ activity may start at 9am and have a typical duration of 6 hours, typically ending at 3pm.

However, specific groups of persons may have more nuanced desired start times and durations. For example, a completely flexible part-time worker may have a unique schedule and their activity start time might not be constraint at all. To capture such behaviour, Monty introduces additional sub-activity types based on the initial activity schedules generated by the activity synthesis:

- *Typical duration*: The typical or desired activity duration for each activity type is defined based on the initial activity duration. For example, consider an agent that decides to perform a two-hour leisure task. This information is transferred into MATSim by assigning a typical duration of two hours to this leisure task (sub-activity type ‘leisure_120min’). This means that, under ideal conditions, the agent would perform this leisure task for two hours. However, if the agent can avoid severe congestion or a time-dependent toll, it might be beneficial to spend some time longer or shorter at the leisure facility.
- *Start time ‘education’*: Consistent school starts at around 9am can lead to sharp morning traffic peaks. Whilst it is certainly not good to be late at school, people also prefer to not arrive early at school. For this reason, Monty introduces a sub-activity type ‘education_9am’ for primary and secondary pupils, ensuring that they will be at school on time and not waiting for too long.
- *Start time ‘work’*: Some people are more flexible with starting their work activity. However, if an agent decides to travel during morning peak in their initial schedule, Monty assumes that this agent must be at work at a given time with introducing work sub-activities. For example, if an agent’s initial schedule contains a work activity that starts at 8am, this agent is likely not flexible with moving that activity. For this reason, we convert this initial ‘work’ activity into a ‘work_8am’ activity to model the penalties for arriving early or late. It means that the agent can still arrive at work later than 8am, but only if it pays off between gaining utility for avoiding traffic versus losing utility for being late.
- *Start time ‘home’*: Literature has shown that people prefer arriving home in the evening at a certain time, for example to spend time with their family. This means that they are generally not flexible with the start time of this home activity. For this reason, Monty assigns a given desired start time to all home activities initially starting between 4pm and 7pm.

This requires the assumption that the activity sequences as chosen in the activity synthesis are generally good and desired. The agents will only move away from those timings in the case of an external disruption, such as more traffic leading to delays or a certain pricing scheme. Allowing the agents to modify and optimise their activity timings gives Monty the ability to forecast complex behaviour changes such as peak hour spreading.

5.3.7 Replanning Annealing

Replanning strategies within MATSim are used to define both the:

- Percentage of agents that are allowed to modify their daily travel decisions during each iteration of the simulation (referred to as the global innovation rate).
- How those agents are allowed to modify their decisions. Modification types include transport mode, route and departure time.

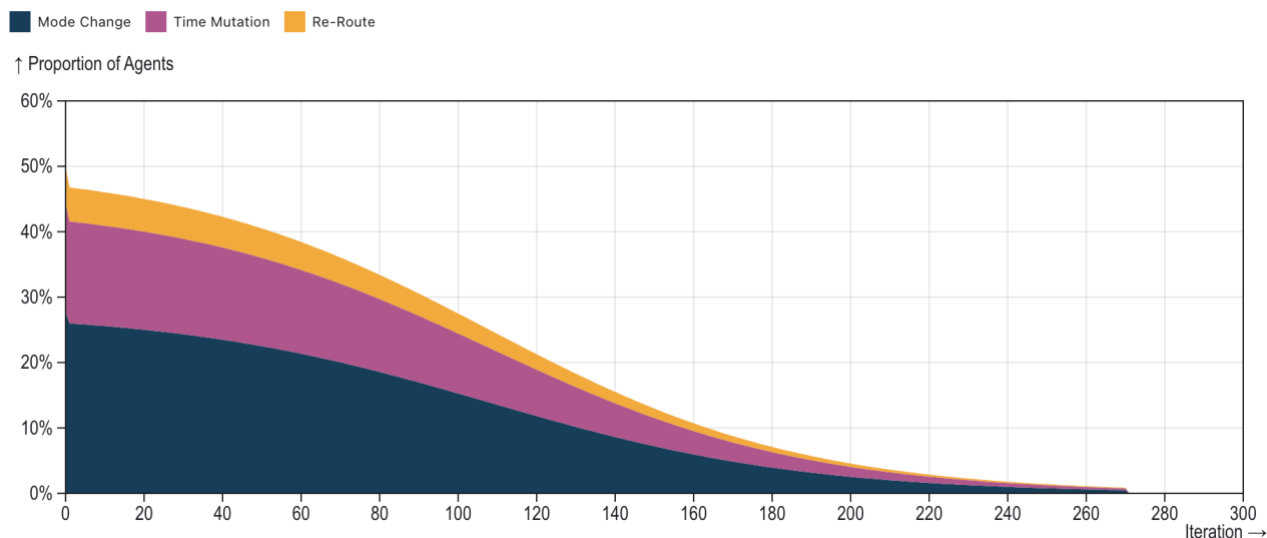
In MATSim, it is common practice to introduce a constant overall replanning rate of 30% for a specific proportion of simulation iterations (e.g., 90%). In the case of a 300-iteration simulation for example, this would be 270 of the total iterations, referred to as the *innovation phase*.

The 30% replanning rate may be divided into 15% mode, 5% route, and 10% departure time modifications. During each of the innovation phase iterations, 30% of the agents are randomly selected, for which each would be assigned the ability to modify their schedule according to the replanning strategy proportions of mode, route and departure time modifications.

After the innovation phase, the replanning rate is switched to 0%, meaning that the agents can no longer introduce new modifications. They can only change their daily decisions based on their existing memory of explored daily activity schedules seen during the innovation phase. This would typically be comprised of the best 3-5 schedule alternatives seen within the simulation to date. The remaining iterations of the simulation (30, in the case of a 300-iteration simulation with a 90% innovation phase) would be undertaken with these constraints.

The sudden lack of innovation can manifest some disadvantages in terms of the simulation's convergence behaviour since network loading can differ substantially from the innovation phase to the final memory-based selection of agent schedules. To overcome this issue, Monty applies an *annealing* replanning curve – a feature that has been recently introduced within MATSim. The curve follows a gradual, sigmoid-shaped form as shown in Figure 42. It starts with an overall replanning rate of 50% and then slowly reduces this rate until it reaches 0%. This allows the agents to gradually adapt to new travel conditions without a stepwise disruption.

Figure 42: MATSim replanning annealing strategy example



5.4 Freight

Freight demand within Monty is treated as a static input, meaning that freight trip origins and destinations are not responsive to new infrastructure such as delivery hubs or terminals. In its current state, the main purpose

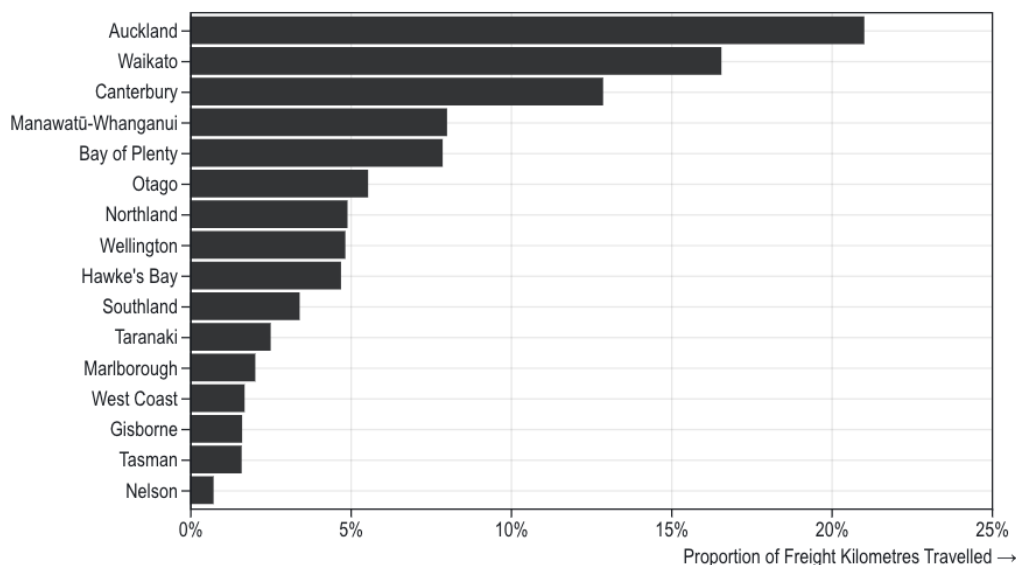
of this static freight demand input is to provide a realistic picture of travel conditions on roads, where a non-trivial portion of private vehicle travel is composed of freight-related trips. Freight agents are thus limited in their ability to optimise their plans in that they are only able to modify their route choice. Under the current methodology, it does not make sense for a freight agent to be presented with the options of switching their mode of travel as rail freight is not currently represented.

Monty's freight demand is derived from external data provided by EROAD – a supplier of freight telematic systems within New Zealand. The data contains GPS-tracked freight movements on a spatially anonymised level. Over the day, a freight vehicle may conduct multiple trips, resulting in a sequence of 'delivery' activities. The dataset includes data from many major freight operators and covers most of the country spatially. Overall, it claims around a 35% market share of large freight operators with some region-dependent differences in the market share.

A procedure was developed by the model development team to convert this information into a format compatible with Monty's network simulation, including a spatial disaggregation step used to transform the anonymised, region-based origin and destination data into realistic trips (a similar disaggregation procedure was applied for activities generated within the AcBM as described in Subsection 4.2.4).

The final transformed freight demand dataset contains approximately 0.48 million daily heavy vehicle trips, involving aggregated travel of more than 11.9 million freight kilometres travelled during the simulated day. In average, freight vehicles conduct 6.5 trips per day. Figure 43 shows a breakdown of the 2018 freight kilometres travelled in specific regional council areas. Auckland, Waikato and Canterbury attract most freight traffic volumes, together accounting for 50% of the freight kilometres travelled in New Zealand. The freight distances observed within other regions is significantly lower. This largely follows distributions of population density in combination with the presence of major ports such as the Port of Tauranga in the Bay of Plenty region. A more detailed heavy vehicle count validation is shown in Subsection 5.7.5.

Figure 43: Freight kilometres travelled by trip destination regional council



5.5 Skim Matrix Generation

Skim matrices refer to origin-destination zone matrices of general network statistics (such as average travel times during a given time period). These represent key outputs of the network simulation process and are commonly produced by strategic transport models to inform internal feedback mechanisms as well as scenario analysis. One such example of a skim matrix is average private vehicle travel times observed between every pair of SA2 zones in a given scenario for a specific part of the day such as the morning peak. Such an output is directly used within components such as the AcBM to help inform the accessibility of moving between zones and thus decisions regarding location choice and initial mode choice.

As described, the ABM represents travel discretely at the individual agent level. Thus, average network statistics such as travel times, travel distances, public transport waiting times etc. must be aggregated both

temporally and spatially to fit a skim matrix structure. This section outlines the methodology employed by Monty to generate such skim matrices.

In Monty, two different types of skim matrices are calculated:

- *Network-based skims* are used to represent road traffic metrics, such as those related to car driving journeys. They are a function of travel time and distance, taking into consideration modelled congestion levels.
- *Schedule-based skims* are used to represent public transport-related metrics such as bus travel. These are more complex than their network-based counterparts, as public transport services are constrained by the associated timetable.

The following procedure is applied to generate and aggregate both network-based and schedule-based skim matrices:

1. Five coordinates are selected within each SA2 zone. The selection process currently applies a random sampling of road network nodes present within the zone to produce these five coordinates. This is motivated by the fact that such a mechanism increases the likelihood of selecting coordinates within areas of high node densities, which is assumed to be correlated with the number of trips that originate and terminate in proximity to those locations.
2. For each given time period (see Table 2), multiple discrete points in time are selected for the network-based skim matrices. For example, 7:30am, 8:00am and 8:30am may be chosen for the morning peak period. For the resulting travel time, the average over those timings is taken. For the schedule-based skims, a time interval is defined per time period. For example, an interval from 7:30am to 8:30am may be selected. This is necessary to capture the high variance of travel times that public transit schedules can produce – for example, the fastest public transport route between two destinations may only occur every 45 minutes.
3. For every possible combination of coordinates and timings selected in steps 1-2, journeys are routed within the transport network to calculate network metrics such as average travel speed using the outcomes of the final simulation iteration. For example, with 1,000 zones and five coordinates per zone, 25 million different routes would be calculated to facilitate this. The calculation process is repeated for each discrete time selected for the network-based matrices. If three times were selected, 75 million routes would be assessed. For the schedule-based matrices, routes are calculated every two minutes within the given time interval, resulting in a large number of route searches. For this purpose, the least-cost-path tree algorithms implemented within MATSim (MATSim, MATSim-Extensions by SBB, 2023) were applied to optimise this process. Through use of this, the problem space can be reduced to approximately 5,000 tree calculations per timing interval.
4. For each zonal pair, the average value network metrics for all coordinate combinations and discrete timings is calculated for the network-based skim matrices. The schedule-based skim matrices use the weighted average of multiple route calculations within the specified time period, weighted by the travel time of the route (referred to as roof-top method). Slow routes are assigned lower weights and are only marginally considered within the final value if there is a fast alternative serving the route at the same time.

The results of this procedure are network-based as well as schedule-based skim matrices for multiple time periods. Monty uses network skims for all four time periods, but only schedule-based skims for the morning peak when feeding simulation results back into the AcBM where necessary. This represents a pragmatic trade-off balancing computation time and accuracy since the schedule-based matrices require significant computational resources.

Monty currently produces the following skim matrices:

- *Network-based*: average travel times, average travel distances.
- *Schedule-based*: average in-vehicle times, average in-vehicle distances, average access/egress time, average number of transfers, average stop waiting time.

5.6 Computational Resources

Monty's network simulation processes extremely large amounts of data as it seeks to represent individual vehicles within a detailed network representation. MATSim provides efficient parallel-computing algorithms

to attack this problem, but computational challenges related to resources still exist when attempting to run simulations as large as those required by Monty. Unlike the population synthesis and activity synthesis steps, the network simulation cannot be run on normal laptops or desktop-class machines. Instead, Monty leverages elastic compute infrastructure provided by Amazon Web Services (AWS).

As mentioned in Subsection 5.3.2, it is common practice to simulate a certain fraction of the population in order to make scenarios more feasible to execute. A population sample of 25% (around one million agents for Monty) is typical, requiring approximately 120 GB of RAM per scenario and 42 CPU cores. Under such a configuration, a 25% sample simulation of 300 iterations takes approximately 28 hours to complete. This computation time scales fairly linearly with sample size, with a 10% size simulation taking approximately 10 hours.

5.7 Validation

The performance of the ABM can be verified against observed volumes and travel times for accuracy. This is an important aspect of model development, essential in building confidence that the simulation is able to capture correct insights into agent travel behaviour covering route choice, mode choice and other journey characteristics. This section summarises the validation process and outcomes for this component of Monty at the time of writing, noting that ongoing model development progress can significantly alter such outcomes going forward as new functionality is developed. Subsection 4.2.6 provided an overview of an equivalent exercise as carried out for the AcBM.

Assignment validation performance has been characterised across the following measures:

- Is the number of iterations adopted across the ABM simulation sufficient? Has the simulation converged to a state in which the agents' behaviour and associated utility scores do not change significantly anymore?
- Are the network conditions in the skim matrices stable? If we re-calculate the skim matrices based on new simulation outputs, do zonal accessibilities still vary?
- Is the share of use across transport modes in alignment with what observed data tells us?
- Is the simulated routing of journeys of sufficient accuracy that private vehicle and public transport passenger flows align with observed survey counts where data is available?
- Do vehicle and passenger flows follow a realistic temporal distribution throughout the modelled day? Are the shapes of the morning and evening peaks congruent with reality?
- Are the right numbers of trips originating and terminating from each region within the model compared to available observed data?
- Are the travel times as experienced in the simulation realistic, taking into consideration road congestion and public transport delays?

To answer these questions, several observed data sources were used including the HTS, road volume counts, and public transport passenger count data. The validation is undertaken for the 2018 base representation in Monty. With every new base year model feature (might be as small as changing capacities on a specific link), all the validation statistics are to be re-generated and re-assessed. This means that validation in general is seen as an ongoing and repeating process. This process aims to constantly improve the model outcomes and build additional confidence.

Additionally, it is worth highlighting that Monty simulates all choice dimensions such as routes, travel modes, and departure times simultaneously on an agent level. This means that all outcomes of Monty are a result of behavioural utility maximisation of individuals, without any corrections on an aggregated matrix level. This has the clear advantage of preserving the unique relationship between specific pieces of infrastructure and individual agents. However, it presents a fundamentally different way of modelling travel compared to historic approaches. As a result, guidelines tailored to historic approaches may expect high accuracy for certain metrics (for example traffic volumes). But they do not include a suggestion for other metrics that are important for Monty (for example mode shares by different socio-demographics of agents). As mentioned above, a high dimensionality model such Monty simulates multiple choices simultaneously. During the calibration of the model, a perfect fit for one dimension such as mode share targets for specific types of trips often led to under-

or over-representation of other metrics such as traffic volumes once assignment took place. This validation presents a holistic overview of an ABM validation with reference to historic guidelines and makes pragmatic and sensible trade-offs between high accuracy and the high dimensionality of Monty.

The following sub-sections provide a comparison of simulated outputs against a range of different statistics to validate the quality of Monty's current scenario representation.

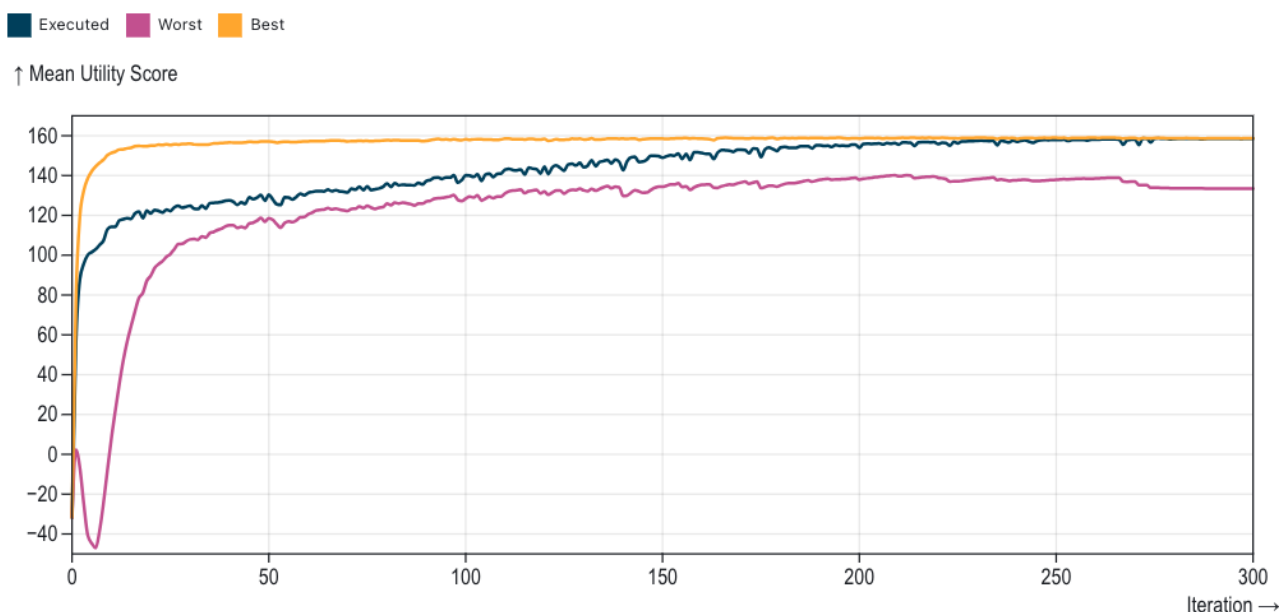
5.7.1 Simulation Convergence

As explained in Section 5.2, Monty applies an iterative framework to simulate agent decision-making and consequent network use. In each iteration, random processes are applied to explore new travel options. This means that it is important to demonstrate that each simulation run has reached a converged state after the applied number of simulation iterations. If a simulation has not converged towards equilibrium, network volumes and public transport patronage may vary heavily from one simulation to the next, based on nothing but stochastic noise. This is an undesired state as it can reduce confidence when comparing different scenario tests, making it difficult to analyse the actual impact of a given intervention. This is why stable network flows are important, particularly as the simulation gets closer to its final iteration.

Reporting on simulation convergence is not only important for a base year representation, but for *all* simulation runs including future year scenarios which may contain higher levels of congestion. As a result, the following statistics are recommended to be part of routine checks for major scenario changes such as significant population growth. They should be reported with relevance to their particular use case.

Figure 44 shows different variants of average agent utility score, including average of the executed score of all agents within the simulation across iterations.

Figure 44: Score convergence



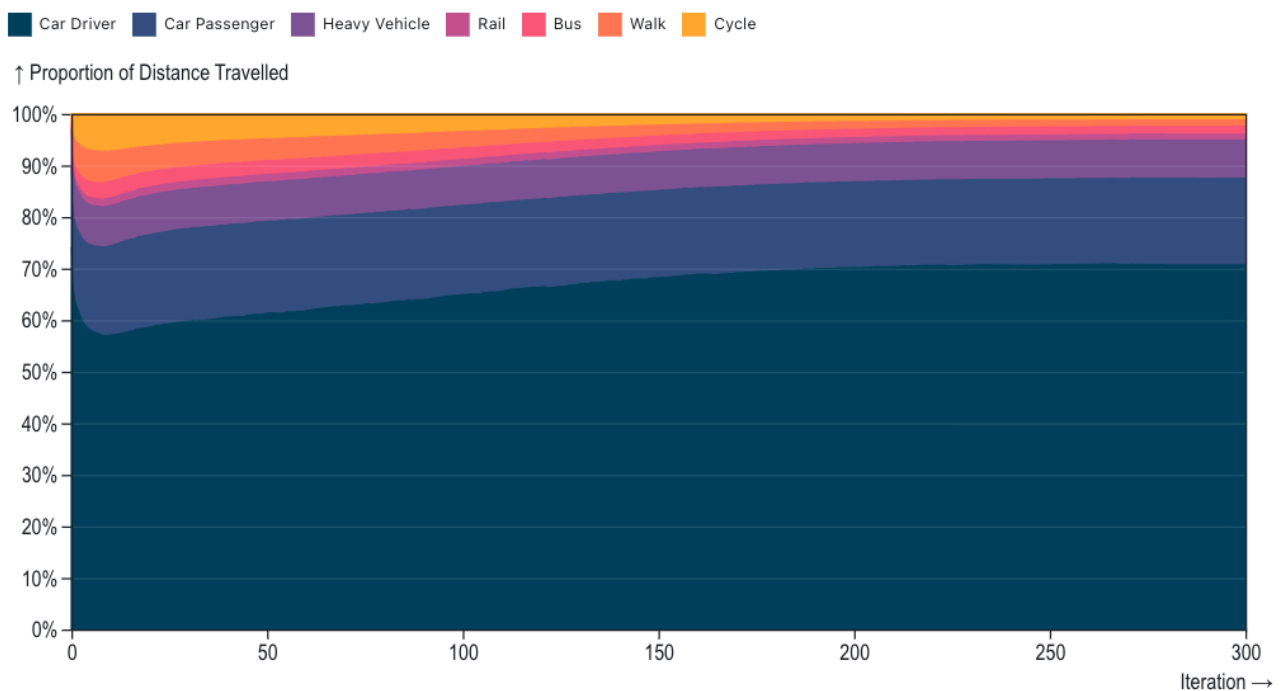
Several indicators are used to track convergence rate within MATSim for a typical Monty scenario run. As defined in Subsection 5.3.5, an agent's utility score measures how satisfied they were with their simulated travel decisions and schedule. Intuitively, a stable score over many iterations means that a given agent cannot significantly improve their travel behaviour through changing their route, mode or trip departure times. It also means that all other agent scores must be stable as well, as the performance of surrounding agents can greatly affect each individual agent.

As seen in Figure 44, typical Monty scenarios converge around an average score of 160 utility units between iteration 200 and 300. As the agents keep multiple plans in their memory, it is also possible to analyse the average of the best and worst plan scores over all agents. The average of the best plan scores converges extremely fast, whilst the worst score stabilises at a lower utility level. The executed score is first closer to the curve of the worst score during the innovation phase in the early iterations with high replanning rates (as described in Subsection 5.3.7). This means that newly explored plan modifications are often non-optimal and will be rejected. With increasing iterations and lower replanning rates, the executed score converges towards

the best score of the agent as they become less innovative. There is high confidence that the Monty simulation converges satisfactorily from an agent scoring perspective, typically the project team have run simulations for up to 500 iterations to confirm these observations.

Figure 45 depicts the average travelled distance by transport mode across each of the simulation iterations. In a sense, this is a form of distance-based mode share. From a convergence perspective, it is desirable for the simulation to rapidly stabilise its mode share as this indicates that agents are successfully finding optimal means of achieving their activity schedules directly through mode choice and indirectly through route choice. Similar to the scores of the agents, Monty's mode share evolution across iterations shows a clear convergence trend. The distance share of car drivers converges at around 70% during the simulation end phase, which is a valid indicator for a converged simulation.

Figure 45: Mode share convergence

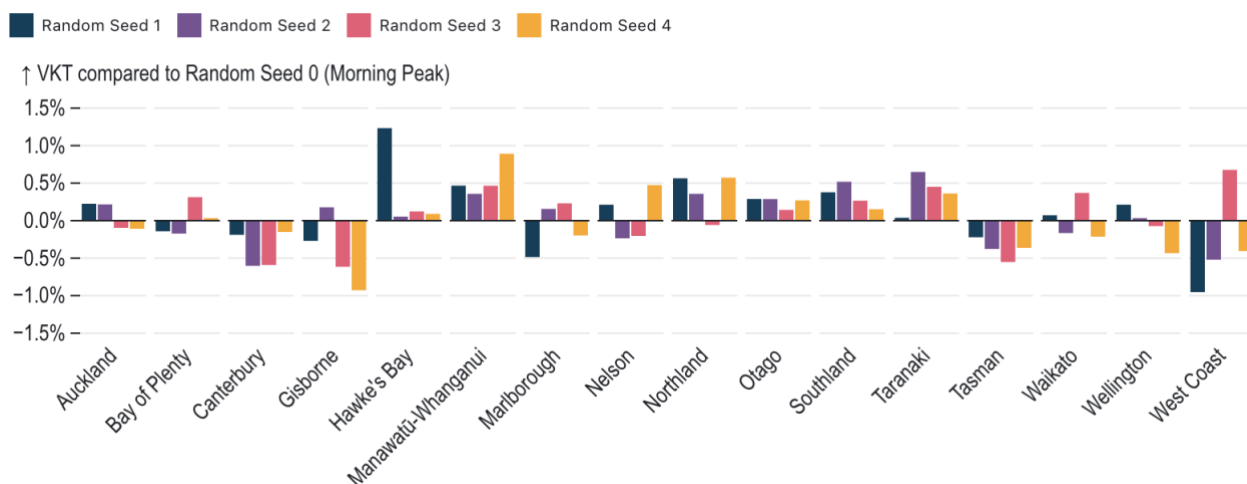


Both agent scores and mode shares show a satisfying convergence behaviour as they stabilise towards the final iterations. However, transport simulation frameworks such as Monty are usually applied to compare traffic volumes across multiple scenarios, down to selected links and time intervals. As mentioned above, the traffic volumes outcomes in Monty are the result of an iterative process with random plan modifications. The randomness can be controlled through an initial random seed value. Modifying the random seed parameter consequently impacts the traffic outcomes, even if no other parameter is changed. This means that the random seed can be directly used to analyse the magnitude of the so-called ‘*random simulation noise*’.

For this purpose, an experiment is introduced to explore the magnitude of random simulation noise in Monty on a link level. The goal is to quantify a confidence interval for the random noise depending on the traffic volume level on a link. When comparing two scenarios (for example with and without new road infrastructure), the confidence intervals can be used to make statements such as ‘we can say with 90% confidence that the volume differences on this link are due to simulation randomness only and not due to the new road infrastructure’.

To measure the level of confidence, multiple simulations have been run, only varying the random seed while leaving all other simulation parameters unchanged. Figure 46 shows some first insights on an aggregated level. It contains simulation outputs of five runs, each with a different random seed. For each of those outputs, vehicle kilometres travelled (VKT) during the morning peak are calculated and aggregated by regional council. The figure shows the relative differences in VKT of random seed 1-4 compared to random seed 0. This gives an initial feeling how the simulation results vary depending on the seed value. Large councils such as Auckland or Wellington have a high stability in their numbers, as the VKT differences are in the range of +/- 0.5%. Only for one council and one random seed, the relative VKT difference exceeds the range of 1.0%. Figure 46 gives high confidence in the stability of simulation results on an aggregated scale.

Figure 46: Aggregated VKT statistics depending on random seed



Aggregated numbers can help to get a first feeling of the magnitude of the simulation noise. For more in-depth measures, the distributions of the relative link volume differences within a certain volume category are analysed. As an example, a volume category might be 500-1,000 cars on a link during the morning peak. For every individual link falling into this category, the relative volume difference between two simulation runs is calculated. A certain link might end up with a relative volume difference of -4%, and another with +9%. As mentioned above, it is ensured that the differences come from simulation randomness only. All relative link differences in this category together are assumed to follow a normal distribution which allows for calculating a certain level of confidence.

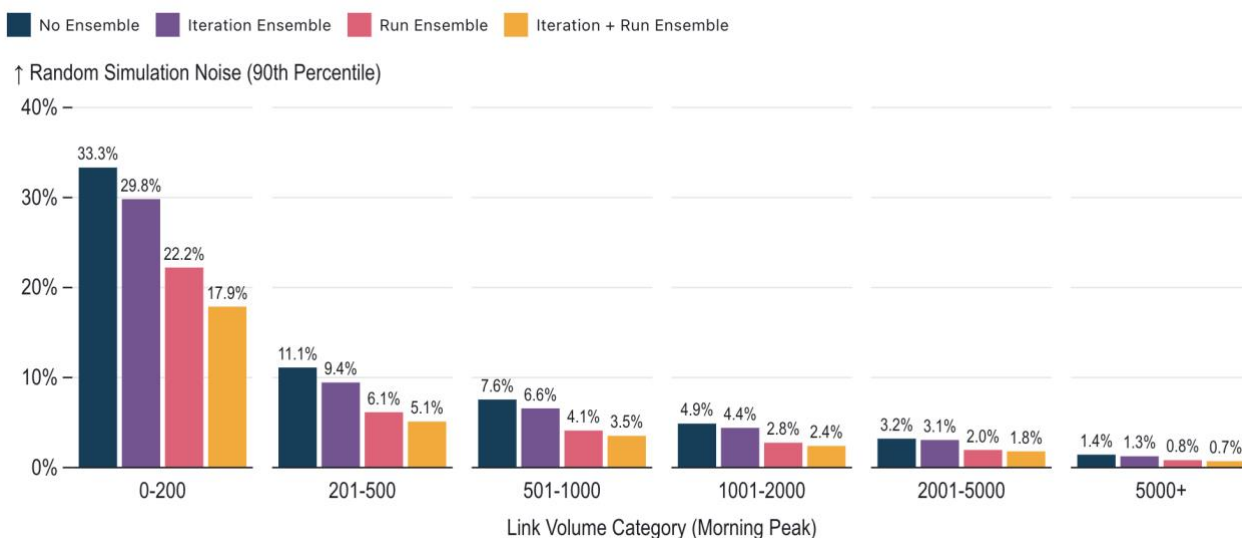
The literature reports on several approaches to reduce the random noise. The most intuitive approach is averaging the results over multiple simulation runs which is commonly referred to as '*Ensemble Runs*'. The confidence intervals of random simulation noise in Monty are quantified for four different strategies:

- *No ensemble*: Compares traffic volumes on a link level during the final iteration between two simulation runs. Note that the population sample in the simulation runs contains 25% of New Zealand's residents.
- *Iteration ensemble*: Averages traffic volumes on a link level over the last ten iterations of one simulation run and compares the results to the average volumes over the last ten iterations of another simulation run.
- *Run ensemble*: Averages traffic volumes of the final iteration over three simulation runs and compares the results to the average volumes of three other simulation runs.
- *Iteration + Run Ensemble*: Combination of averaging over the last ten iterations and three simulation runs. Results in the average over 30 traffic volume values for each link and ensemble.

Figure 47 shows the 90th percentile of the random simulation noise depending on the ensemble approach and the link volume category. The relative differences are higher for low link volumes. This is not surprising as small numbers of agents have a higher relative impact on low-volume roads. Expectedly, the ensemble strategies help to reduce the random noise. For medium-frequented roads (around 500-1,000 cars from 7am to 9am) and no ensemble averaging, the random noise is quantified as 7.6%. This means that 90 out of 100 links have a relative volume difference (or random noise) of 7.6% or less over the morning peak. Averaging over the last ten iterations helps to reduce the noise but is not as effective as averaging over multiple simulation runs. This confirms that the simulations generally converge well but might find different local optima. Taking the average over three simulation runs almost halves the random noise and could be even further reduced with additional simulation runs.

This analysis confirms the importance of assessing the random simulation noise based on the particular use case. Expectedly, small volumes such as traffic flows on individual links come with higher random simulation noise. The random noise can be minimised and controlled through different run ensemble strategies.

Figure 47: Random simulation noise



5.7.2 Skim Matrix Convergence

To confirm that the Monty outcomes are robust and stable, it is also important to demonstrate that the resulting skim matrices converge. The average traffic conditions should not be subject to significant variance when updating the skim matrices without changing any other parameter in the Monty ecosystem. The iterative process has been introduced and illustrated in Figure 4. It feeds the skim matrices (which are an output of the network simulation) back into the demand generation module. Monty’s activity-based demand model component is then responsive to changes in traffic conditions as found by the agents during the network simulation (Section 4.2). For example, a heavily congested road segment leads to less agents choosing activity locations that rely on passing this segment.

Zonal accessibilities can be used as a metric to measure the convergence behaviour of skim matrix updates. For demonstration purposes, the workplace model coefficients (see Table 12) are taken to calculate zonal accessibilities as a function of the skim matrices (multi-modal accessibilities). The accessibility numbers are expressed as a logarithmic summation over all destinations j for each SA2 zones based on Equation 3.

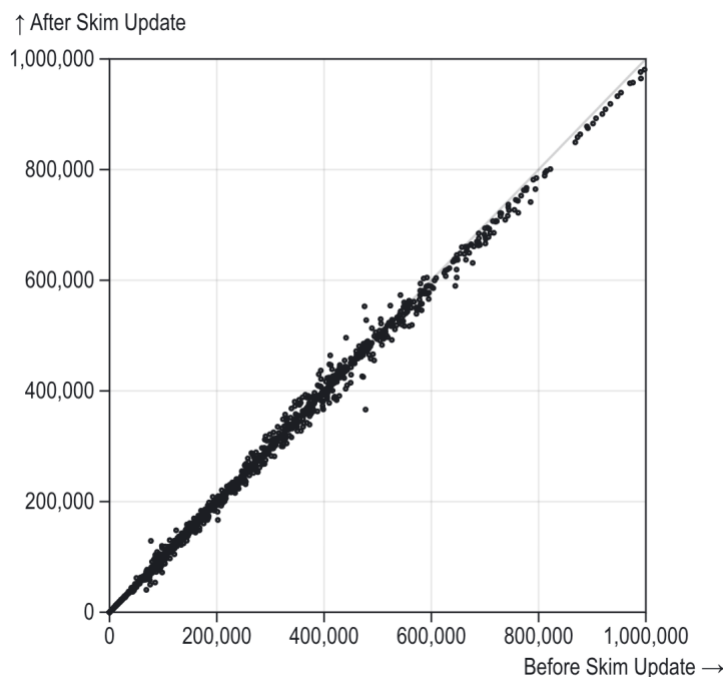
Table 22 presents some statistical metrics to demonstrate the stability of the zonal accessibilities between skim matrix updates. Each iteration shows the linear regression metrics between the zonal accessibilities before and after updating the skim matrices (a full loop of skim calculation – demand generation – network simulation – skim calculation – demand generation). The table shows that the error metrics are overall satisfying with high R^2 and low %RMSE while the slope stays close to 1.0. With increasing iterations, the metrics further improve and show a tendency to converge.

Table 22: Zonal accessibility variance between skim matrix updates

Iteration	R^2	Slope	%RMSE
1	0.99603	1.00	6.23%
2	0.99684	1.00	5.54%
3	0.99773	0.99	4.78%
4	0.99788	0.99	4.73%

Figure 48 shows a scatter plot of the workplace accessibilities before and after the last skim update in the fourth iteration. It confirms the overall conclusions made from Table 22 as it demonstrates strong correlation between the accessibility values.

Figure 48: Zonal workplace accessibilities before and after fourth iteration



5.7.3 Mode Shares

As outlined in Table 19, Monty differentiates between several mode types including private vehicles (driver and passenger variants), public transport (bus, train, ferry) as well as active transport (walking and cycling). During the simulation, agents attempt to optimise their chosen transport modes based on utility scores derived from network conditions, associated costs and requisite travel times. If a trip with a newly tested mode leads to a higher utility score for that agent, they are likely to keep that mode for future simulation iterations. The coefficients used to calculate an agent's utility score are outlined in Table 21.

Once the simulation has reached a converged state, the resulting transport mode shares can be compared to observed data as reported in the HTS. Figure 49 shows trip-based mode shares for the entire country, calculated from the HTS and compared against Monty's 2018 base scenario outcomes. Trips are clearly dominated by private vehicles (drivers and passengers combined), which have a share of more than 80% overall. This effect is captured correctly by the model, albeit with a slight over-prediction of car driver mode use.

Figure 49: Trip-based mode share for New Zealand

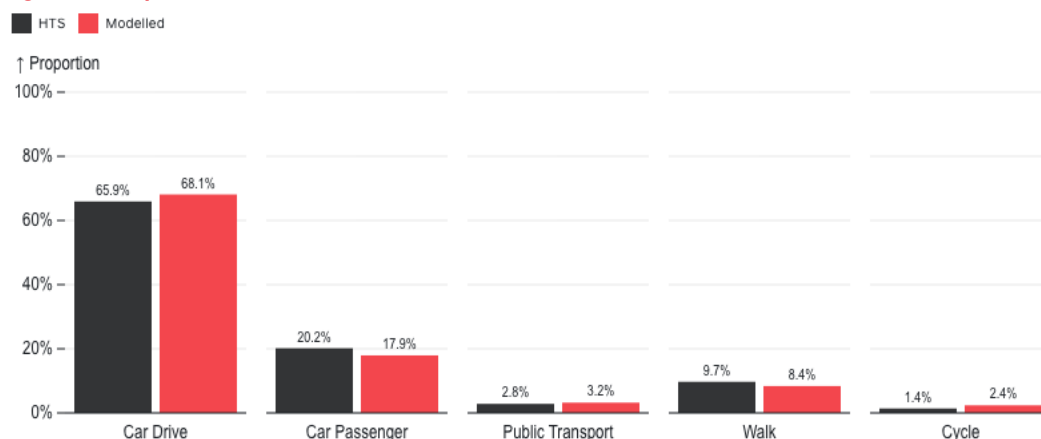


Figure 50 disaggregates mode shares by trip distance in kilometres. Strong agreement is seen between the Monty 2018 base scenario and patterns of behaviour seen within the HTS, particularly in replicating patterns of use for private vehicles including both car drivers and passengers. Public transport usage shows a peak at around 35 km in the HTS (presumably rail legs), which is under-represented in Monty. Other distances show a good agreement for public transport usage between Monty and the HTS. Active modes (both walking and cycling) are well represented over the distance classes, with a small tendency to under-predict mid-range walking trips (around 2-3 km). Overall, the distance-based mode shares in this figure confirm the car-orientated

travel behaviour seen generally across New Zealand, particularly for trips longer than 5 km. Above a trip distance of 40 km, the share of private vehicle travel reaches nearly 100% in both Monty and as observed from the HTS.

Figure 50: Distance-based mode shares for New Zealand

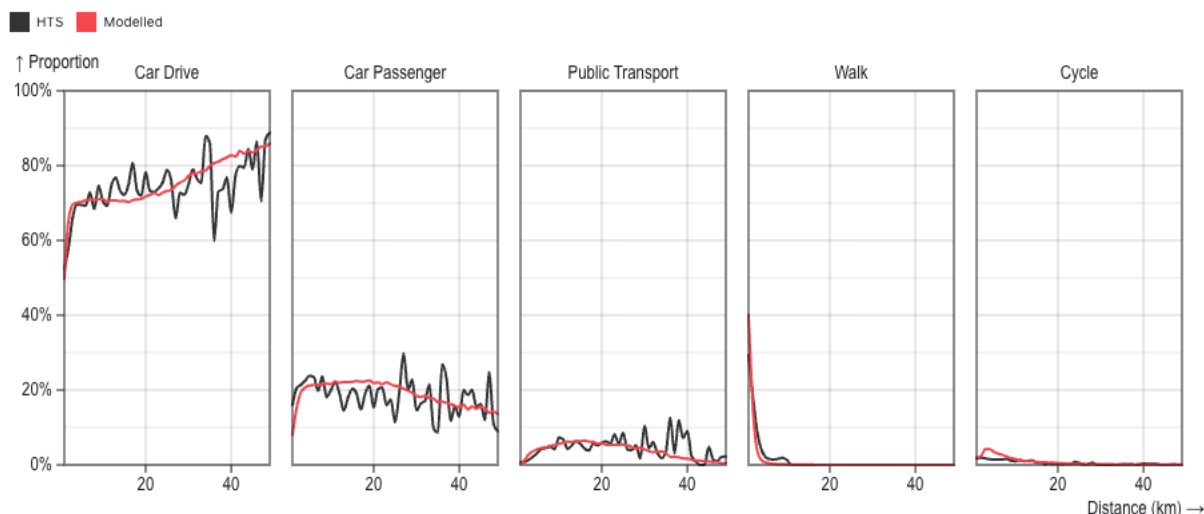


Figure 51 shows trip-based mode shares for specific types of agents. As an explanatory analysis, this figure shows the relationship between different income groups and trip-based mode shares in New Zealand. Medium- and high-income agents show a similar mode choice behaviour, with high-income agents tending to use active modes more often. Low-income agents show higher shares of public transport usage and sharing private vehicles (car passengers) which comes with a reduced car driver share compared to the other income groups. This behaviour is well captured by the Monty utility specification.

Figure 51: Trip-based mode shares by income group

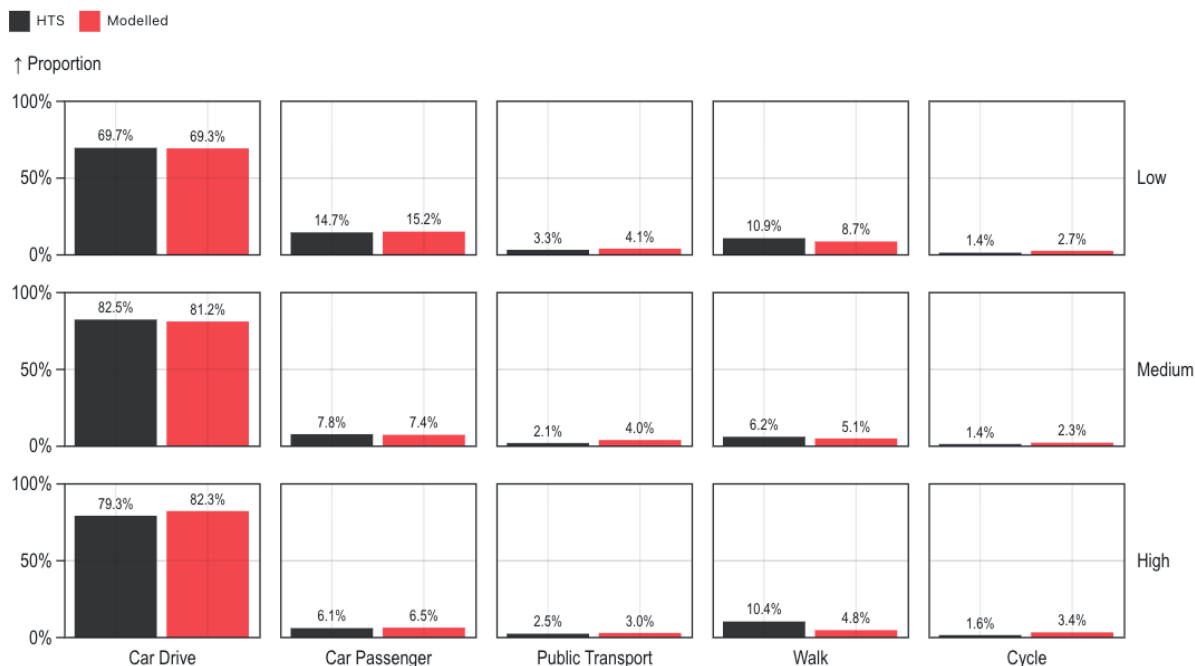


Figure 52 shows trip-based mode shares across some of the largest regional council areas in New Zealand. This shows notable region-dependent differences in mode use. For example, the share of private vehicle traffic in Wellington is significantly lower than in Auckland and other regions. Instead, Wellington has a particularly high share of short-distance walking trips in comparison. This confirms observations made during the location-choice step of the AcBM regarding shorter trip lengths in Wellington of this type.

Monty captures regional differences well and is aligned with observed regional council trip-based mode shares observed from the HTS data. It is worth noting the slight over estimation of car mode share in some regions

like the Bay of Plenty region, and in the previous analyses, against HTS is artificially introduced by the modelling team to better align the modelled road volumes with observed traffic counts (further explored in Subsection 5.7.4).

Figure 52: Trip-based mode shares for select regional councils



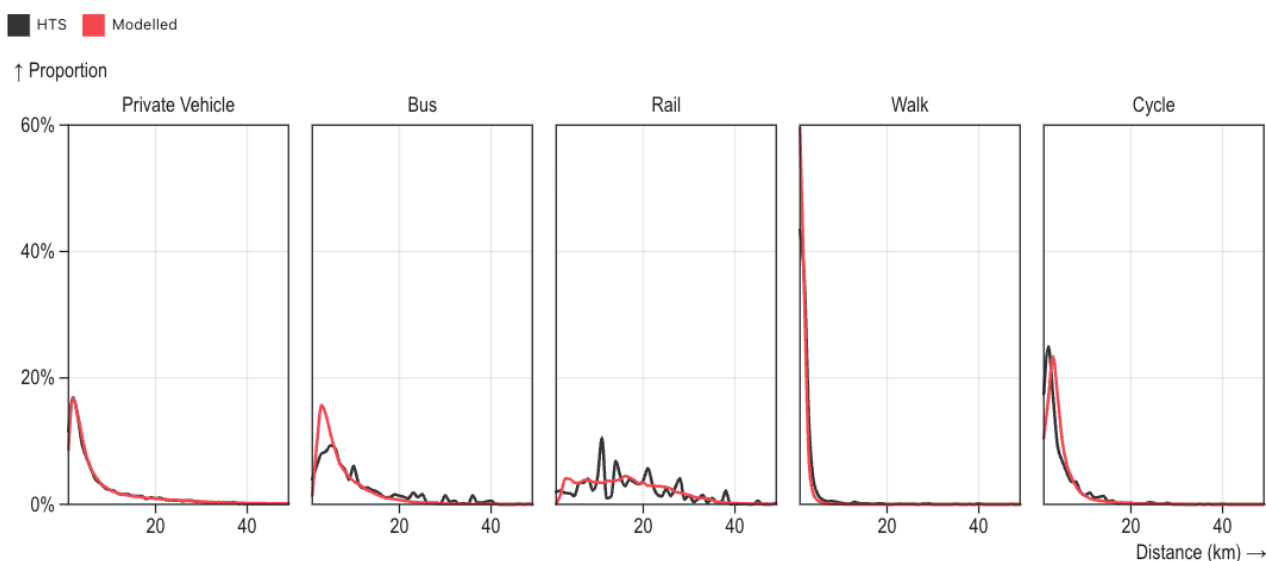
Transport mode shares also differ between trip purposes servicing different activity types as seen in Figure 53. For example, education trips expectedly show a high share of active mode use, particularly walking. In contrast, work commuting trips and shopping are associated with a very high private vehicle mode share. Such mode share patterns are captured well by Monty, even without specifying dedicated scoring coefficients for attributes like student status within MATSim. School trips in Monty show a lower share of public transport trips when comparing to the HTS. This can be explained by the fact that Monty does only limitedly represent school buses. The over-prediction of driving a car to leisure activities could be improved by lowering the convenience of such journey types. In the same way, the high shares of car drivers for both education and accompany activities could be corrected. This is the subject of ongoing investigation and implementation by the model development team.

Figure 53: Trip-based mode shares for select trip purposes



Figure 54 provides additional comparisons of trip length distributions depending on mode. As expected, the walking and cycling distributions are dominated by very short trips. Rail trips also expectedly show a higher share of long trips than bus trips, peaking at 20 km whilst bus trips peak between 5-10 km. Overall, Monty shows good agreement across all modes, particularly for active modes and cars. The model slightly underestimates the average trip lengths for bus trips compared to patterns of behaviour seen in the HTS.

Figure 54: Distance distribution by mode



5.7.4 Road Traffic Volumes

The Monty ABM provides a high-resolution representation of vehicle traffic, simulating individual vehicles as they move through the network during the simulated day. For the purposes of road network validation, traffic counter data provided by Waka Kotahi NZ Transport Agency (Waka Kotahi NZ Transport Agency, 2020) was used as a source of observed movements on an aggregated level. The traffic counter data (also called ‘TMS Sites’) offers directional traffic volume counts on a 15-minute aggregation level, differentiated by light and heavy vehicles. Across the entire country, around 1,930 count sites were ultimately processed and matched using an automated algorithm to the equivalent road segments within the strategic transport network to assist with this for Monty.

Figure 55 shows daily and hourly modelled traffic counts compared to observed directional traffic counts for each count site across New Zealand. The traffic counts include the sum over both light and heavy vehicles and whenever possible, it is differentiated by traffic direction. A linear regression line is presented on each chart with its intercept forced through zero, as is typical for comparisons of this nature when validating strategic transport models. The associated R^2 and slope statistics for this regression line are also presented on each chart.

The daily traffic volumes range up to 90,000 vehicles per traffic direction on large motorways mostly around State Highway 1 in Auckland. Monty slightly under-predicts traffic on some of the busier roads but does generally reproduce the traffic counts well in terms of R^2 and slope statistics. On an hourly traffic volume level, the R^2 statistic for the regression line decreases, but still remains higher than 0.9.

Figure 55: Directional traffic volume comparisons for New Zealand (left: daily volumes; right: hourly volumes)

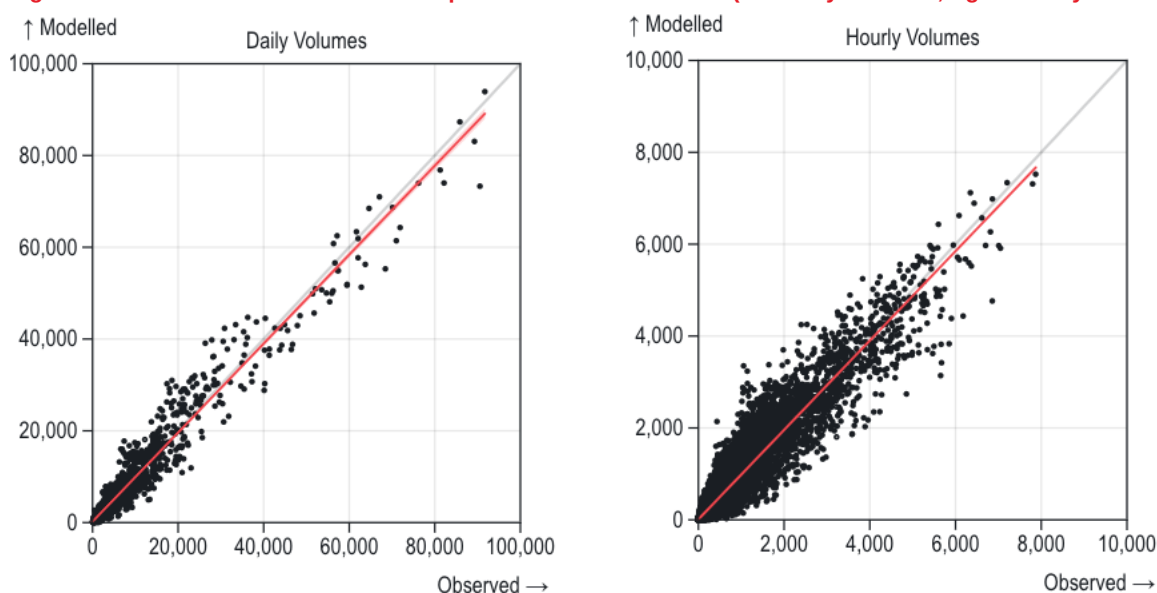
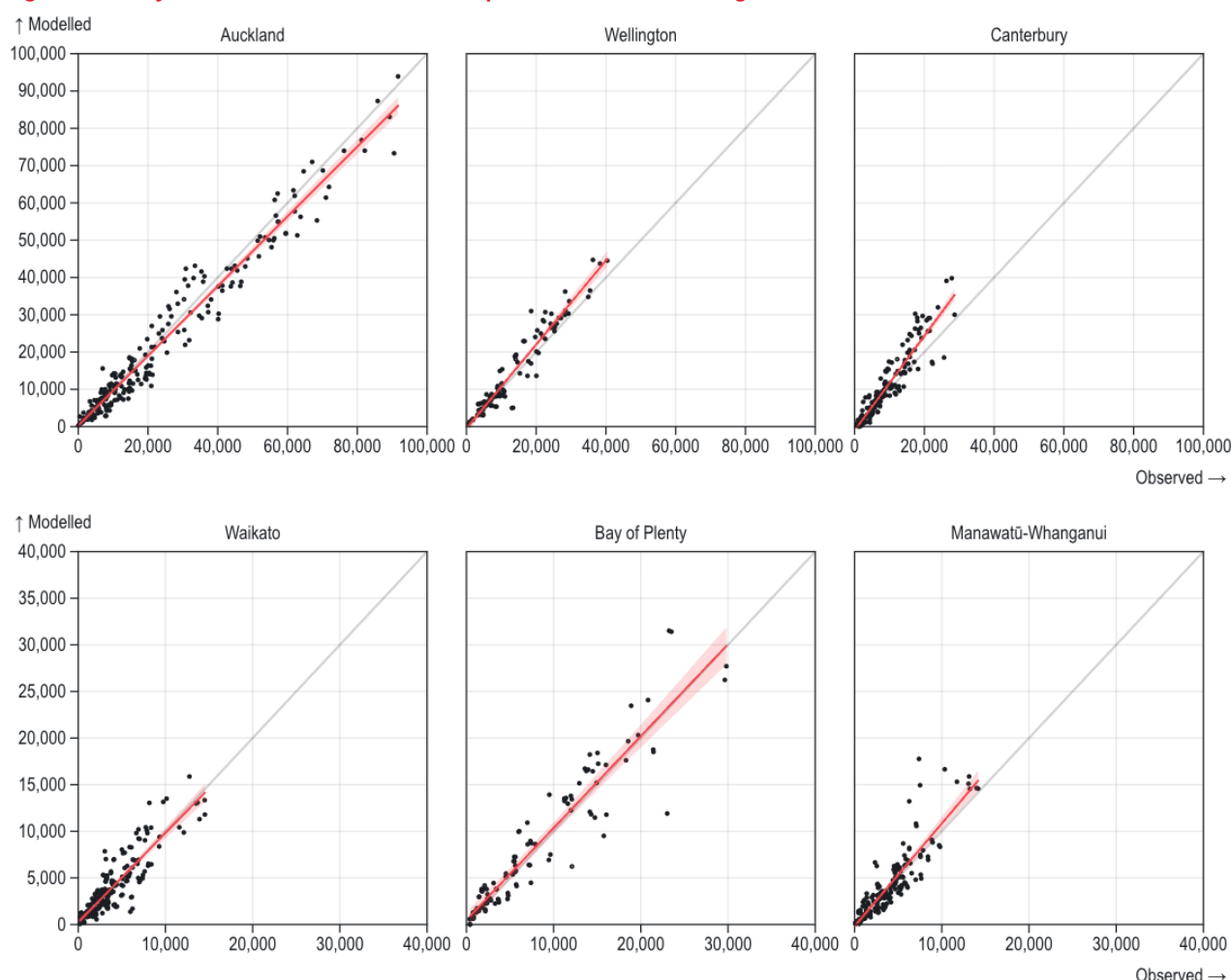


Figure 56 shows daily modelled traffic counts compared to specific large regional councils including Auckland, Canterbury, and Wellington. Calibration of the model for this development round focused primarily on the regions of Auckland and Wellington. Christchurch in Canterbury is subject to currently ongoing calibration efforts, as traffic on major roads is over-predicted in Monty.

It is also worth noting that the algorithm matching the count sites to the road segments in the Monty network is fully automated. The task of matching count sites can be non-trivial and challenging, particularly in urban settings with high road densities. The development team has conducted in-depth cross-checking of the automated matching, but some outliers might still occur simply because of a mismatch between count site and the road segment in Monty’s strategic network representation.

Figure 56: Daily directional traffic volume comparisons for selected regional councils



When considering these daily traffic volumes in both Figure 55 and Figure 56, the R^2 over all roads with a matched count site in New Zealand is above 0.9. Precise values for R^2 and slope coefficients are presented below in Table 23. The R^2 of the Auckland region is particularly high, landing above 0.95 whilst the volume accuracy for Wellington is slightly lower at 0.94. The slope coefficient is close to 1x for all most regions, indicating a lack of overarching bias that would result in systematic underestimation or overestimation of volumes. As mentioned above, Canterbury needs further investigations from the model development team.

Of note are a few outliers within the Wellington scatter plot (bottom right) which were determined to be along routes heavily interacting with Wellington airport. This also occurs in Auckland but to a lesser extent. Since Monty currently does not simulate domestic or international air travel, this can be explained through missing visitor demand arriving and departing from the airport. This is an area of active investigation by the model development team. It is worth noting that demand related to employment opportunities and other types of activities captured by the HTS are still generated within airport precincts within Monty through the AcBM, hence why there is some simulated travel demand rather than nothing at all.

Figure 57 shows the same statistics for New Zealand but limited to specific time periods rather than the entire day. In general, the R^2 are expectedly slightly lower than the equivalent daily outcomes. Overall, at a nationwide level R^2 values range between 0.9 and 0.95 for all time periods. This is considered an acceptable fit for a regional model according to the *New Zealand Transport Model Development Guidelines* (Waka Kotahi NZ Transport Agency, 2019). The line of best fit slopes range between 0.9-1.1, also conforming to the guidelines. Traffic during the evening peak shows a slight under-representation bias. In contrast, Monty over-predicts the number of vehicles around the intermediate peak.

Figure 57: Time-period specific traffic volume comparisons for New Zealand

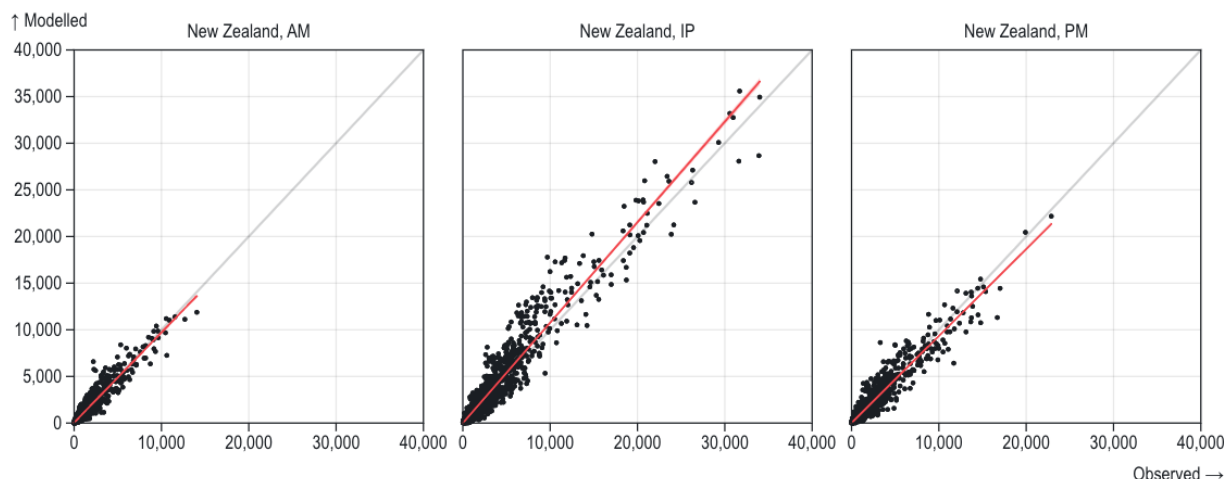


Figure 58 provides time-specific volume comparisons for Auckland. Agreement between observed traffic volumes and Monty is high, with an R^2 of approximately 0.91 in the morning and evening peak periods. This improves to almost 0.95 in the middle of the day. Auckland traffic volumes are well represented in Monty, with some outliers seen in the evening peak requiring further investigation by the model development team.

Figure 58: Time-period specific traffic volume comparisons for Auckland

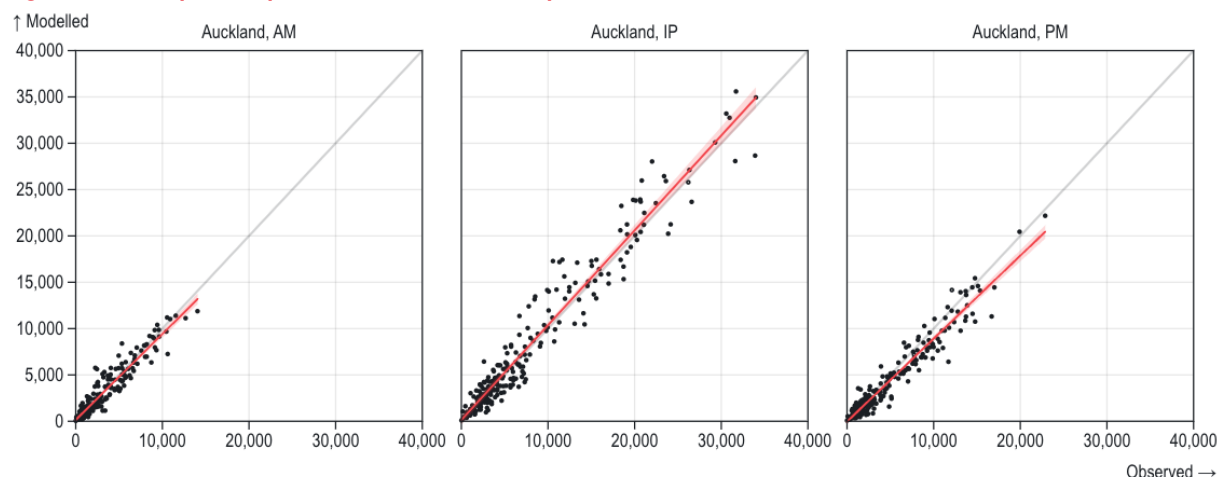


Figure 59 shows the volume scatter graphs for the Wellington region by time period. The fit is lower here with the R^2 ranging between 0.85 and 0.90. In contrast to Auckland, traffic tends to be higher in relation to observed count values, particularly in the intermediate peak. This is confirmed below by Figure 60, showing more concentrated traffic peaks in Wellington in relation to the intermediate peak.

Figure 59: Time-period specific traffic volume comparisons for Wellington

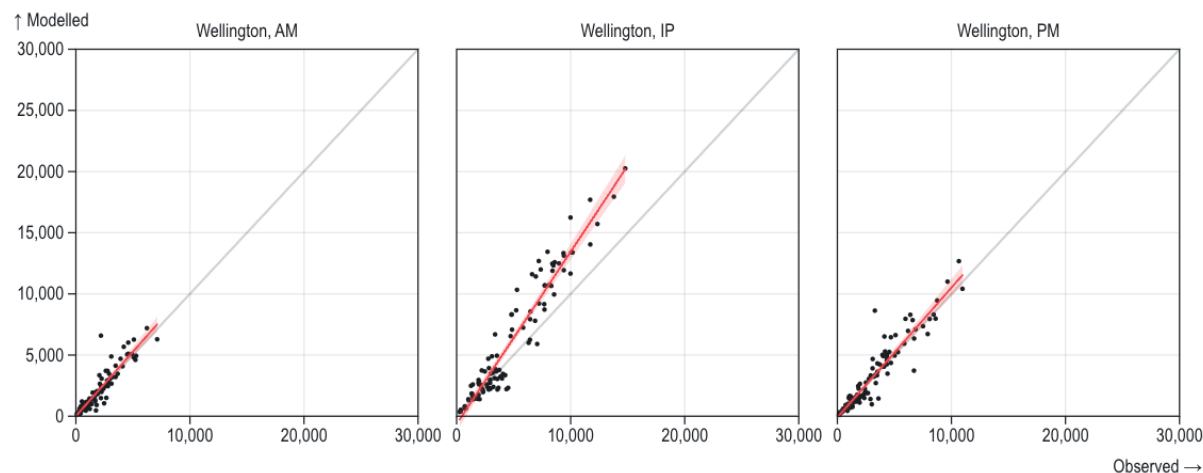


Table 23 summarises all the previous road network validation findings and also provides additional error measures for each region of interest. As demonstrated in the figures above, the regression statistics (R^2 and slope) are in line with the expectation stated in the *Transport Model Development Guidelines* (Waka Kotahi NZ Transport Agency, 2019). Also, the percentage of traffic counts below an hourly GEH of 10 is above 80%, indicating a sufficient fit between Monty and traffic counts. In contrast, the %RMSE ranges between 30% to 45% across sites in New Zealand, which is classified as ‘requires classification’. As highlighted previously, a slight over-prediction of vehicle traffic has been tolerated because this needs to be balanced against mode share comparisons outcomes with the HTS. Given inconsistencies between available data sources, the model development team could not simultaneously satisfy a more accurate trip-based mode share measure whilst maintaining good road network volume accuracy (especially when simultaneously considering excellent representation of trip length distributions). The calibration outcome in this instance thus represents a pragmatic trade-off between attempting to represent the behaviour described within the HTS (suggesting less car use) and TMS (suggesting more car use) data sources. The Scalable Quality Value (SQV)¹⁰ provides another way of reporting the model performance and the values are in a comparable range as reported by other activity- and agent-based models of a similar implementation structure (Sherr, Manser, Joshi, Frischknecht, & Métrailler, 2020).

At this stage of the model development process, main calibration efforts have focused predominantly on the Auckland and the Wellington region, which is well reflected in Table 23. Nearly all values in the Auckland region robustly meet the requirements for a strategic network model according to the modelling guidelines. For the Wellington region, values are just within an acceptable range considering the above-mentioned trade-off made between mode share and traffic count accuracies. One outlier worth discussing is the intermediate peak volumes in the Wellington, during which the model over-predicts traffic. In comparison to other regions like Auckland, the time-of-day profile in Wellington is more concentrated around the peaks. This could be actively corrected in the model by specifying region specific time parameters and is part of ongoing work.

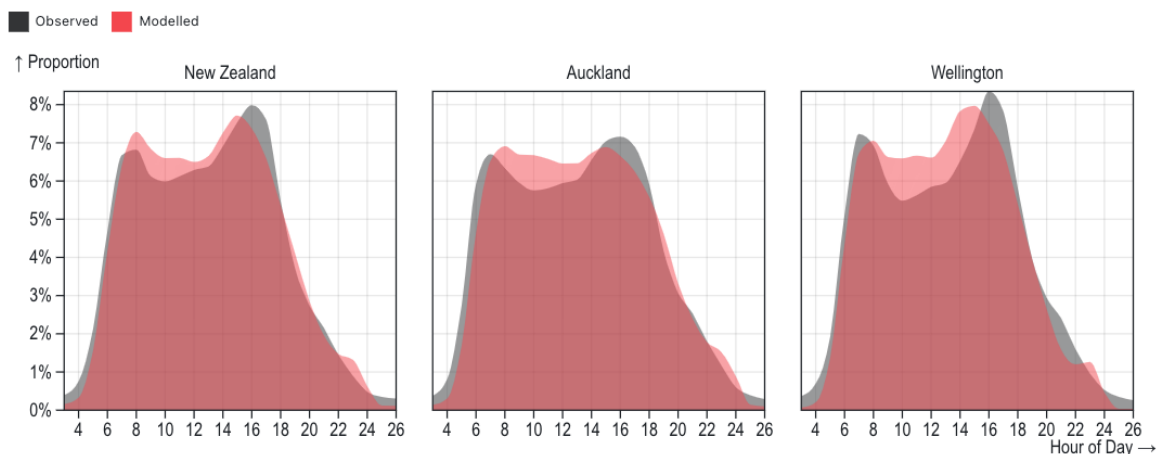
Table 23: Traffic count performance measures

Region	Period	Sites	R^2	Slope	%RMSE	%GEH<10	%SQV>0.8 (Friedrich, Pestel, & Simon, 2019)
New Zealand	Daily	1'931	0.95	0.98	34.2%	88.2%	76.8%
	AM		0.91	0.98	46.3%	84.2%	83.4%
	IP		0.93	1.08	45.0%	84.2%	64.4%
	PM		0.92	0.93	40.4%	82.1%	77.8%
Auckland	Daily	231	0.96	0.94	19.2%	73.1%	56.3%
	AM		0.91	0.95	29.5%	66.7%	65.4%
	IP		0.95	1.03	23.5%	75.2%	50.6%
	PM		0.94	0.89	25.6%	69.8%	61.5%
Wellington	Daily	313	0.94	1.10	24.9%	76.4%	72.2%
	AM		0.85	1.06	35.1%	81.0%	81.5%
	IP		0.92	1.32	49.0%	64.8%	41.7%
	PM		0.88	1.04	30.2%	77.8%	73.1%

Figure 60 compares profiles of hourly traffic between Monty’s 2018 base scenario and observed survey counts where hourly data was available. Previous observations are reinforced with this figure, with Monty following a similar distribution to the observed counts. Peak volumes are generally well aligned while the intermediate peak has higher volumes in Monty, particularly in Wellington. Traffic is seen to ramp up and down rapidly in the morning, with a longer peak observed across the evening as expected. The actual evening is slightly early in Monty, but ramps down following the same distribution as observed in the TMS dataset.

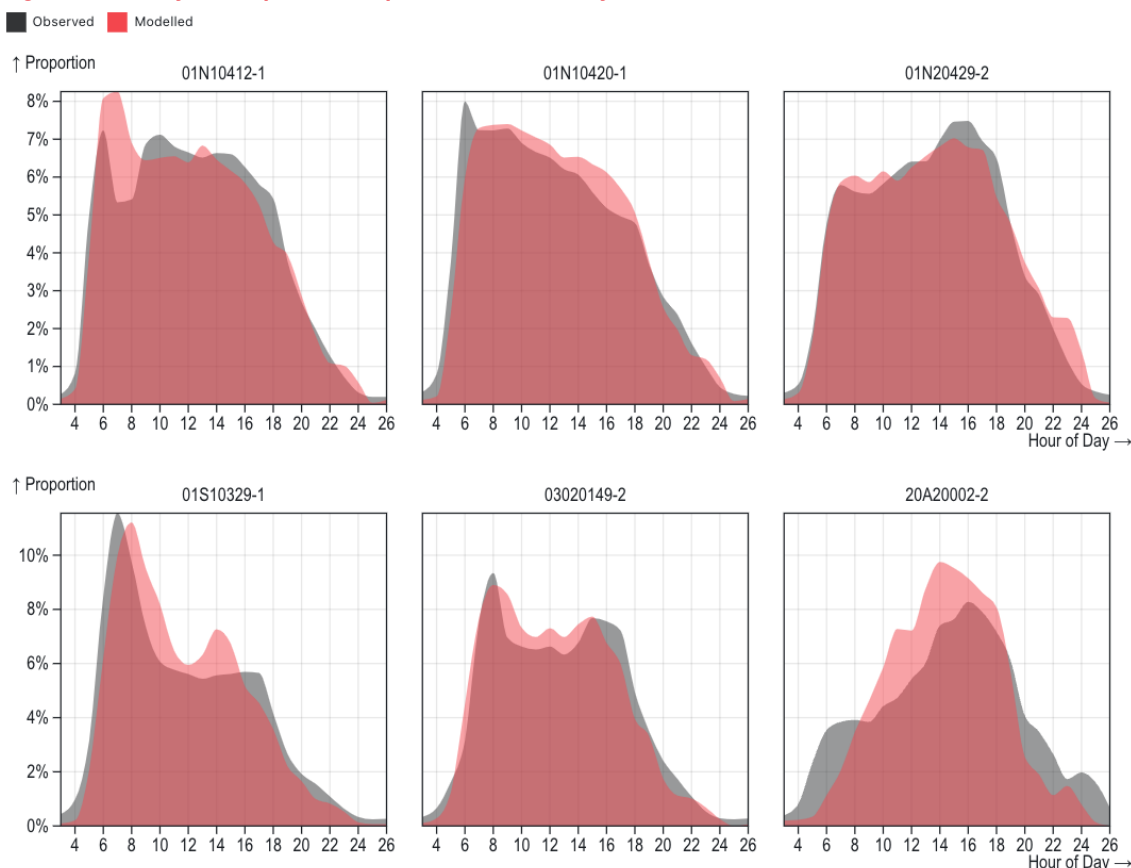
¹⁰ Scalable GEH: A Quality Measure for Comparing Observed and Modelled Single Values in a Travel Demand Model Validation (Friedrich, Pestel, & Simon, 2019).

Figure 60: Hourly traffic proportion comparison



Finally, Figure 61 selects six random major count sites and shows the comparison between the time-of-day distribution in Monty and the count site. On the level of specific count sites, Monty reproduces general traffic levels and captures time-of-day and directional trends well.

Figure 61: Hourly traffic profile comparison for randomly selected count sites



5.7.5 Heavy Vehicle Traffic

In addition to comparing total traffic volumes, the ‘TMS Sites’ from the Waka Kotahi NZ Transport Agency (Waka Kotahi NZ Transport Agency, 2020) also offer counts for heavy vehicles (above 3.5 tonnes) specifically. Monty uses those counts to verify the simulated freight demand as introduced in Section 5.4. The freight demand generation samples trips from an external origin-destination matrix based on local market share assumptions. Those local market shares are in turn derived from the heavy vehicle traffic counts.

As a result, Figure 62 presents regression statistics with a Slope being tailored close to 1.0 as the Slope was used to derive the regional market share factors. However, the R^2 value validates the original freight matrix as

the origin-destination pairs were not modified. An R^2 value of almost 0.85 is considered a good fit given that it is validating unmodified GPS-tracked data.

Figure 62: Daily directional traffic volume comparisons for heavy vehicles New Zealand

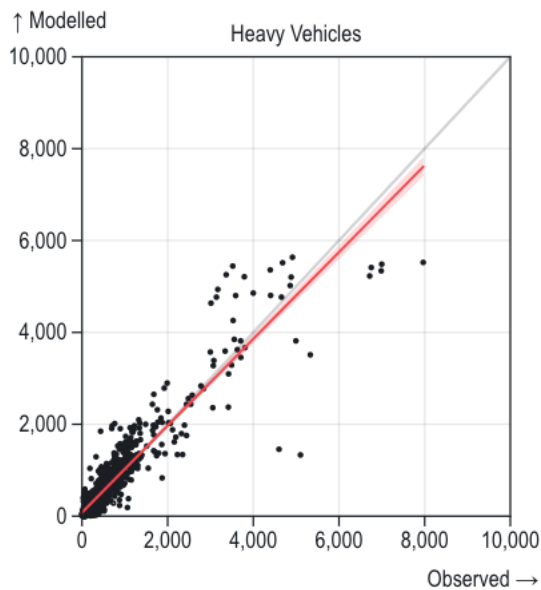
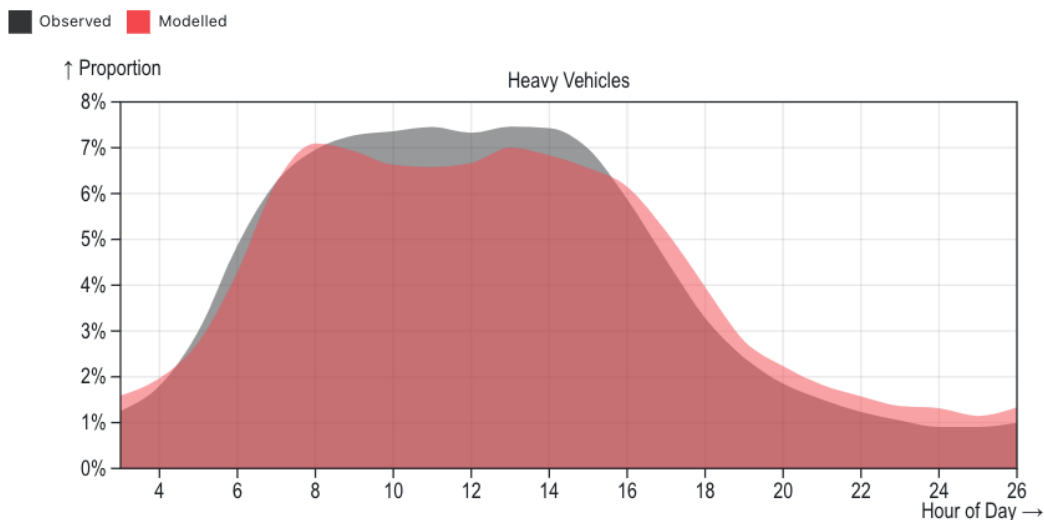


Figure 63 demonstrates how freight movements are distributed over the day. In contrast to the total vehicle traffic, heavy vehicles have a significant proportion of over-night traffic. Over the day, heavy vehicle is more evenly distributed and less concentrated around peak hours. This is well captured by the freight demand simulated in Monty.

Figure 63: Hourly heavy vehicle traffic proportion comparison



5.7.6 Public Transport Patronage

In addition to the performance validation of road volume counts, patronage on the public transport system must also be compared to available observed data. For the calibration cycle described in this report, the model development team was provided access to specific public transport ticket counts from service providers in both Auckland and Wellington:

- For Auckland, counts were provided by Auckland Transport (AT) and averaged over a period of 3 working days in March 2018 between 13 March 2018 to 15 March 2018.
- For Wellington, rail ticketing data collected over a one-month period was collected in November 2018 as a source of boardings information.

Figure 64 shows the Auckland-wide total daily bus tag-on and tag-off events aggregated by Ward region compared to Monty bus boarding and alighting events in the 2018 base scenario. In general, Monty and the AT bus counts are similar overall. When looking at specific Ward regions, Monty results in a significant under-usage of buses in the central Waitemātā region. Current investigations are ongoing by the development team, specifically focusing on an opportunity to improve route choice between bus and rail. Other regions worth noting are Manukau and Waitākere in which Monty tends to over-predict bus usage. This can be interpreted by region-specific behavioural preferences towards certain mode alternatives that must be accounted for.

Figure 64: Auckland bus passenger boardings and alightings comparison

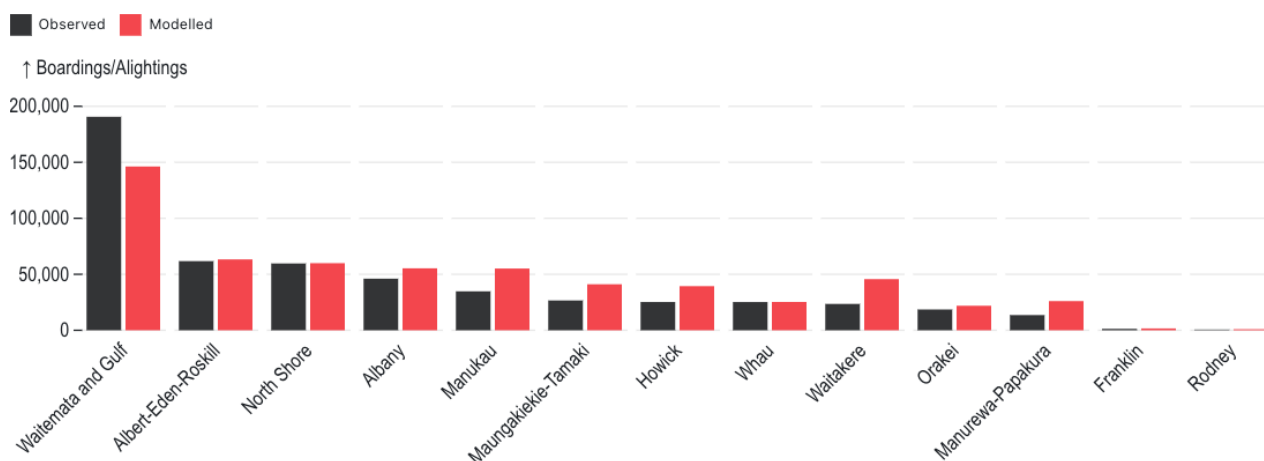


Figure 65 shows the same statistics for daily bus usage across Ward regions in Wellington. Monty agrees well with empirical counts collected from the ticketing data and there is less regional variety compared to Auckland.

Figure 65: Wellington bus passenger boardings and alightings comparison

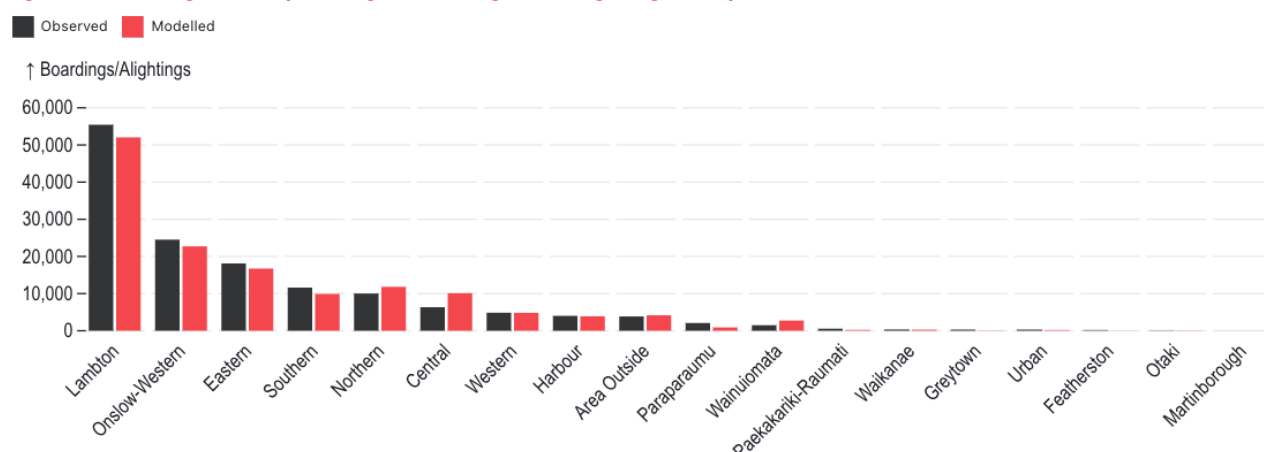


Figure 66 provides insights in terms of distribution of bus usage over the day for both Auckland and Wellington. It shows the aggregated number of bus tag-on and tag-off events in hourly bins over the course of a day. In general, Monty and the ticketing data agree in terms of both profile shape and magnitude. When looking at the distribution over a day, Monty provides a reduced morning peak and under-predicts bus patronage in the hours between 7-9 am. This is less significant in Auckland compared to Wellington, which shows a massive peak in bus usage in the morning. The evening peak follows a very similar shape, with the ticketing data indicating a slightly reduced number of bus passengers in the evening than the values produced by the model.

Figure 66: Hourly bus passenger profile comparison

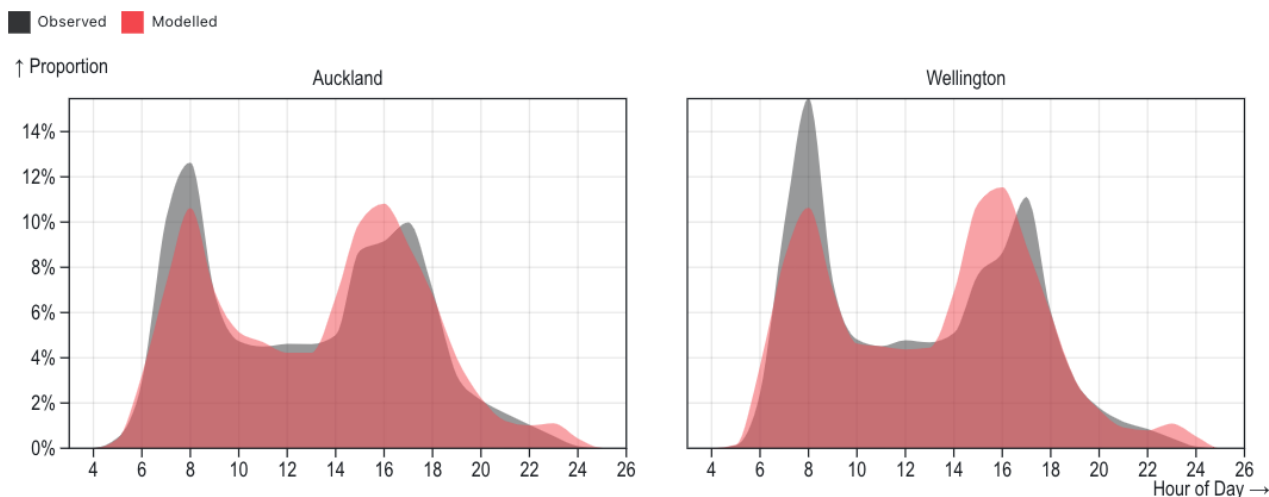
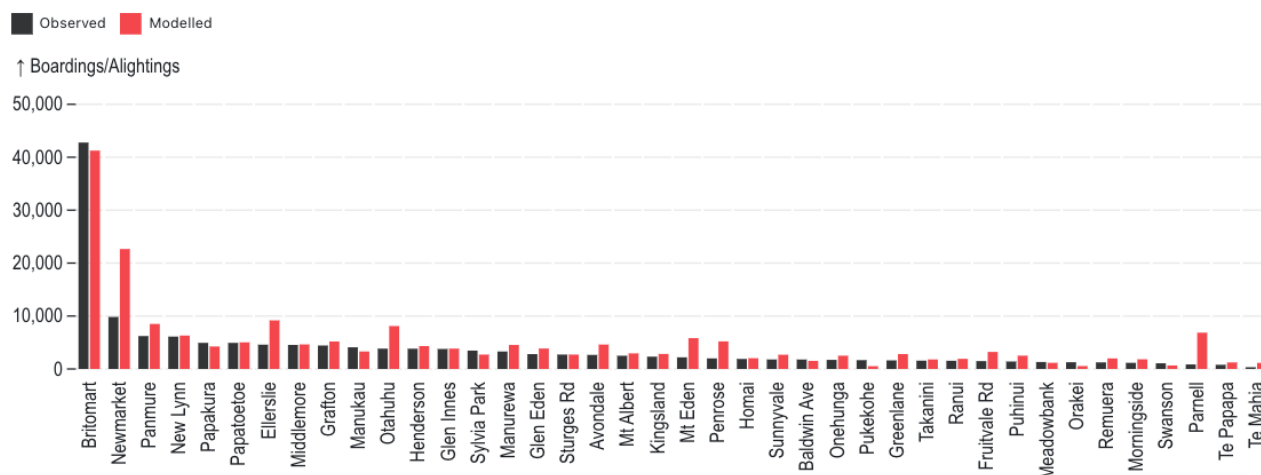


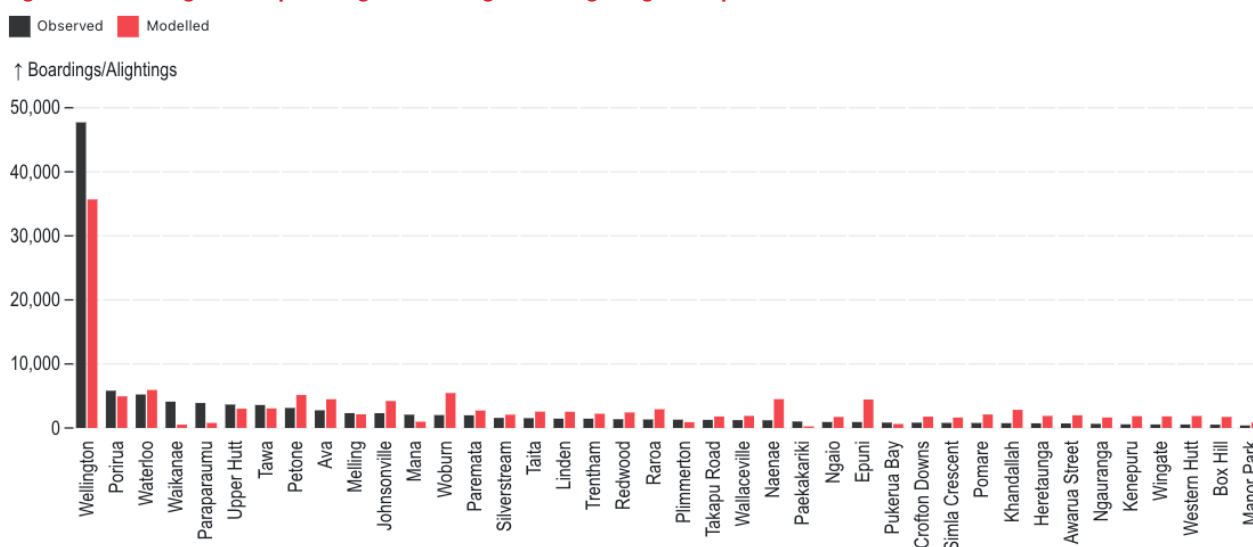
Figure 67 shows the total daily boardings and alightings at rail stations within Auckland. Similar to patterns seen for the bus mode, general passenger counts are very comparable between observed and modelled statistics. However, there is a notable deviation for train stations around the Auckland CBD (Newmarket and Parnell), where the model is significantly over-predicting station use compared to observed behaviour. As mentioned in the bus count validation, route-choice issues causing this discrepancy are currently being investigated by the model development team as part of the next calibration round. For remaining medium and small train stations, Monty generally matches observed daily passenger volumes closely.

Figure 67: Auckland rail passenger boardings and alightings comparison



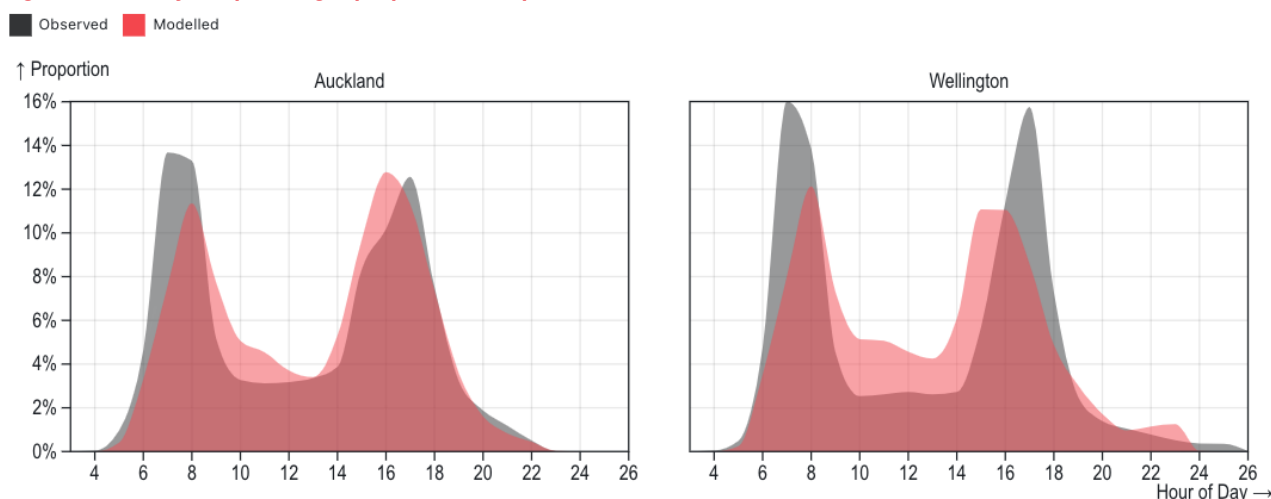
Rail passenger counts for the Wellington region are shown in Figure 68. It shows a slightly different distribution as seen within Figure 67 for Auckland as the large central Wellington Station is substantially under-used by the agents in Monty. This is a particularly extreme case, with trips starting or ending at this main station accounting for almost 90% of all train trips. The number of passengers modelled for small and medium stations are typically well-represented in Monty. Some exceptions are Waikanae and Paraparaumu not attracting enough rail passengers in Monty. These are very long-distance trips to get into the central area of Wellington and might be indicating a location choice issue. As was the case for similar discrepancies seen for Auckland, this deviation is currently being addressed part of the next model calibration round.

Figure 68: Wellington rail passenger boardings and alightings comparison



The hourly distribution profile of rail passengers in Figure 69 confirms the observations made in the distribution of bus trips over the day. The rail patronage in Auckland follows a similar pattern between Monty the ticketing data. In contrast, ticketing data reveals that rail patronage in Wellington is concentrated around sharp peaks in the morning and the evening. Those peaks are currently under-predicted in Monty.

Figure 69: Hourly rail passenger proportion comparison



5.7.7 Journey Times and Speeds

The travel times and the corresponding journey speeds produced by Monty's ABM are fed back into the activity synthesis step as part of a feedback loop involving the generation of skim matrices (see Section 5.5). Based on the travel conditions that the agents experienced in the simulation, they re-schedule their activity patterns. Hence, it is of importance that Monty models an accurate representation of travel times and average journey speeds, including variations between different time periods during the simulated day. As pointed out in Subsection 5.3.1, the public transport vehicles do not experience congestion and rely on reasonable timetabled journey times. For this reason, this section focuses on road journeys only.

The Google Directions API was used to query observed average travel times and distances as collected by Google along routes within the road network. These observed journeys were then compared to modelled travel routes within Monty as a basis for the validation comparison. Unfortunately, Google does not allow for the querying of routes from the past. Thus, the adopted travel times used for validation represent typical weekday conditions in 2023. Origins and destinations forming journey routes were sampled from the reported HTS trips across New Zealand weighted by population density. Overall, 2,301 routes were generated for this purpose, facilitating travel time validation across the morning peak, inter-peak and evening peak time periods.

Table 24 shows some high-level statistics about the routed travel time comparisons. Throughout the entire day, 69% of all modelled journey times are within a 15%- or 1-minute error range when comparing to the observed travel times. 85% of all journeys fall into a 25%- or 1.5-minute error range. The journey durations show a good agreement during free-flow conditions (off peak), with a drop in accuracy during the day.

Table 24: Travel time statistics

Metric	Daily	OP	AM	IP	PM
Journey times within 15% or 1 minute	69.0%	80.4%	66.8%	67.2%	61.7%
Journey times within 25% or 1.5 minutes	85.1%	92.5%	82.8%	83.5%	81.5%

Figure 70 confirms those observations, presenting the travel time comparison as scatter plots per time period. As mentioned above, the observed travel times are based on traffic conditions in the year 2023, while the modelled network state targets to represent 2018. The model development team has assessed major outliers in Figure 70, confirming that a large proportion of the outliers are due to additional capacity in road infrastructure between 2018 and 2023. Also, there is a significant number of routes that under-predict travel times during the evening peak. This means that traffic delays due to congestions are under-represented in the evening. Below Table 25 is able to specifically narrow down that issue to the Auckland region.

Figure 70: Journey durations per time period

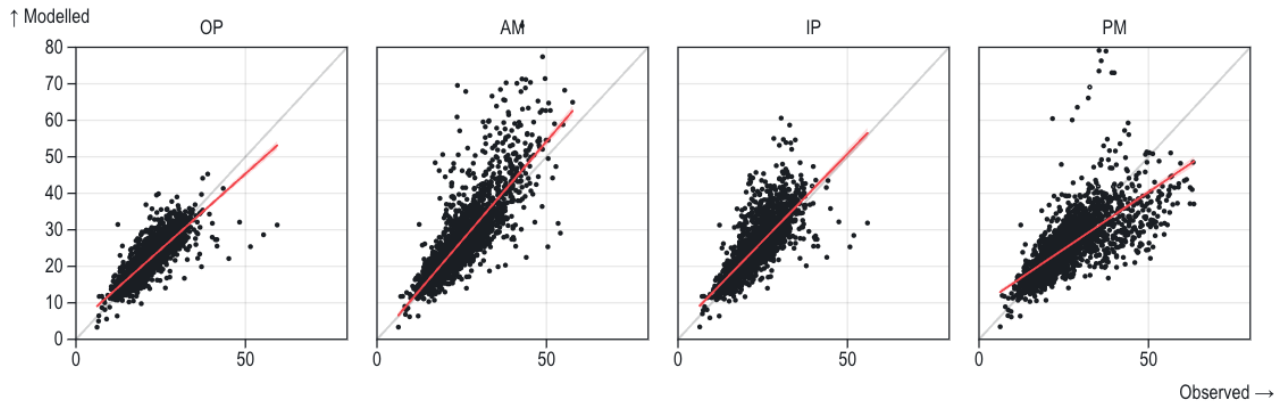


Table 25 presents a comparison between observed average journey speeds and those generated by Monty. Across New Zealand, Monty captures well the differences between free-flow conditions and peak-hour traffic. This is shown in a drop from an average speed of 62km/h in free-flow conditions to around 54km/h in the morning and evening peak. The intermediate peak is subject to ongoing work in order to reduce traffic and congestion levels between the two peak hours.

When looking at specific regions, Table 25 demonstrates an overall good agreement between traffic conditions in Monty and the queried Google journeys. The most notable exception is the average speed during the evening-peak period in Auckland. This can be actively resolved by checking the road capacities in Auckland, particularly around key routes such as motorways.

Table 25: Average road journey speeds for selected regional councils

Region	Period	Journeys	Modelled [km/h]	Observed[km/h]
New Zealand	Free Flow	2,301	62.2	62.1
	AM		53.2	54.4
	IP		55.4	58.4
	PM		55.4	54.0
	Free Flow		59.8	60.0
Auckland	AM	505	38.5	42.4
	IP		43.4	52.6
	PM		46.0	37.6
	Free Flow		59.8	60.0
	Free Flow		59.8	60.0

Wellington	Free Flow	248	63.9	61.3
	AM		47.5	50.0
	IP		54.8	58.3
	PM		49.3	49.9
Canterbury	Free Flow	252	58.7	62.5
	AM		53.4	56.0
	IP		54.5	58.1
	PM		54.0	56.1
Otago	Free Flow	123	62.0	60.4
	AM		58.9	58.2
	IP		58.9	58.5
	PM		59.2	58.8

5.8 Discussion

The current implementation of the ABM within Monty uses the open-source software framework MATSim. It simulates traffic decisions and network conditions including route, travel mode and departure time choice on an individual agent level, manifesting as a detailed representation of traffic and passenger flows. This approach preserves the unique relationship between individuals and infrastructure, allowing for detailed analysis of how specific pieces of infrastructure are used by different types of agents. The adopted scoring coefficients are within a reasonable range when comparing them to known econometric values of travel time available in literature.

Certain additional functionality is desired to enhance the current traffic simulation capabilities of Monty in the future:

- *Bus and private vehicle traffic interaction.* One of the primary sources of journey delays for buses is that they can get stuck in surrounding road congestion. New Zealand only incorporates partial coverage of dedicated bus lanes, making this a key aspect of accurately representing bus journey characteristics. The current traffic simulation as provided by the Hermes assignment mechanisms does not consider such traffic interactions, meaning that journey delays need to be explicitly accounted for exogenous to the model within the represented timetable. Means of introducing this level of interaction directly within the simulation are currently being investigated.
- *Dynamic lane simulation.* New Zealand has a dynamic road network, with many lanes being only open to traffic during specific times of the day. This includes parking lanes, transit lanes and even tidal flow lanes in specific locations such as along the Auckland Harbour Bridge. MATSim generally supports the simulation of a dynamic road network. However, the functionality requires further in-depth testing and validation. Some version of this functionality is planned for a future version of Monty.
- *Active mode representation.* Walking and cycling types are currently simulated as point-to-point teleportation modes without considering network characteristics like gradients or person characteristics such as age. Future feature developments are aimed at implementing a more realistic active mode representation, including a routing engine and consideration of more factors influencing travel speed and perceptions of safety.

Model validation of the ABM outcomes showcases a wide range of statistic comparisons, including mode share, road volumes, public transport patronage and travel times. With minor exceptions, most metrics fell within expected performance benchmarks for strategic models of this scale, especially when considering that the existing focus of development has concentrated on representing areas with the highest population densities across the country. Development of Monty is actively ongoing, and is likely to consider the following tasks as part of future calibration efforts:

- *Calibration of more regions.* The overall ambition of Monty is to provide a high-quality transport representation for the entire country of New Zealand. This requires additional ongoing calibration efforts

in regions outside of Auckland and Wellington to improve validation performance in these areas. Work is currently ongoing to acquire more data to help facilitate such efforts. Traffic demand in some regions has high shares of visitors and tourists. Incorporating such demand plays a crucial role as highlighted in below bullet point.

- *Public transport patronage.* Existing public transport validation performance has revealed issues in the representation of relatively sharp passenger peaks within the city centres of both Auckland and Wellington. Work is currently being planned concerning a more sophisticated representation of public transport travel that will aim to account for these local effects more accurately.
- *Travel times.* Analysis of journey durations and average speeds showed issues for some regions during specific time periods. The reason for this can be very isolated, the capacity information of one specific link may have a massive impact on traffic conditions in Monty. Ongoing work addresses this issue, aiming to find ways to efficiently identify such capacity bottlenecks.
- *Visitor demand.* Some specific road segments in Wellington showed significantly lower volumes in the model compared to observed count data. Further investigation revealed that this was due to missing air traffic visits, i.e., people who travel to the airport and then leave Wellington by plane or vice versa. A future work package will likely focus on adding this type of demand directly to the model in areas with large airports to address instances like this. A longer-term ambition is to also introduce domestic air travel mode representation within the model as this can be directly sourced from the HTS. International demand adds further complexity as it is not covered by the survey.

6. Outlook

Monty applies a combined activity- and agent-based modelling methodology based on the application of behavioural principles. This is a cutting-edge approach, with Monty representing one of the first models of this kind globally. All components are fully synthetic, meaning that agents adapt their behaviour in response to changes in their environment as expressed through model inputs. For example, a new shopping centre will likely attract additional travel to that area since people desire to undertake shopping activities there. Similarly, an ageing population would likely lead to changes in overall patterns of activity, causing changes in travel demand patterns. The model has currently demonstrated its ability to provide a validated 2018 base year scenario using the implemented framework as described within this report.

Monty's capability of predicting the impact of these types of behavioural changes unlocks the opportunity to conduct a wide range of case studies. Section 6.1 briefly describes how future year scenarios have already been developed for Monty based on official population projections provided by Stats NZ (Statistics New Zealand, 2022). Other early applications of Monty are described in Section 6.2, including road pricing studies and testing of planned future infrastructure projects. More case studies are being undertaken as part of ongoing work of the model development team.

As has been mentioned on numerous occasions, Monty is still under continuous development. New and enhanced functionality is intended to further improve its predictive power and widen the potential range of future application opportunities. Beyond the specific examples already covered across previous chapters, Section 6.3 seeks to provide a high-level insight into planned feature developments ordered by priority.

6.1 Future Year Scenarios

A 2038 future year has been developed for Monty to assist with sensitivity testing of the model and application in case study analyses. The ambition was to create a future year baseline scenario, upon which additional changes could be implemented to cover scenarios such as increased working from home behaviour, new infrastructure provision, changes in land use, and combinations of various interventions. To facilitate this, the following inputs were collected and processed:

- Population projections (median case) from Stats NZ (Statistics New Zealand, 2022) were used as control totals to create a synthetic population for the year 2038. The projections included forecasts at an SA2 level for both gender and age distributions. Additionally, labour force participation forecasts were available at a national level. For some control totals, no official projection data was available. In these cases, simplified scaling assumptions were used to transform control total distributions as seen in 2018 to match the patterns specified in 2038.
- A new road network representation was created incorporating funded, committed and planned major highway projects in the Auckland and Wellington regions, including the new State Highway route from Pūhoi to Warkworth (opened in July 2023).
- A new public transport schedule representation for Auckland was developed based on outcomes of the Auckland 2031 Regional Land Transport Plan (Auckland Transport, 2021). For Wellington, the public transport schedule created for a future scenario with the existing city-scale strategic transport model was adapted for Monty.
- Simplified assumptions were made regarding land use statistics used to drive job opportunities and school enrolments, as well as the freight demand representation. Scaling factors were adopted to increase opportunities proportionally with local population growth.
- The monetary cost of travelling (e.g., vehicle operating costs and public transport fares) were not modified compared to the 2018 base scenario.

One of the key trends made clear from the population projections is an average shift upwards in age across the population. Whilst the absolute number of teenagers is forecast to stay relatively constant across the next 15 years, the number of persons above the age of 60 are expected to increase by over 60%.

In terms of mobility behaviour, vehicle-kilometres travelled correspondingly increases proportional to population growth with some regional differences. Particularly in Auckland, public transport patronage is predicted to experience a disproportional level of growth. For some regions including the city centre, the number of public transport trips almost doubles according to early scenario testing. However, even with these shifts private vehicle usage continues to dominate overall trip mode shares.

6.2 Early Applications

Early applications of Monty have been conducted with the main purpose of demonstrating the model's possible use cases. Each application also represents a stress test for both the model and the model development team, enabling learning from real scenario testing situations. The following list describes some of the case studies that have already been conducted by the team:

- *Road pricing.* A wide range of different road pricing schemes have been evaluated across both Auckland and Wellington with both the 2018 base and future year scenarios. This has included implementation of complex area, corridor and cordon-based schemes with a temporal ramp up and ramp down of prices across time periods. Implementations have also included daily toll price caps.
- *Flood road closure response.* Testing has been conducted investigating the short and medium-term road closure impacts that resulted from the 2023 North Island flooding events. The conducted analysis focused on changes to freight movements and the demographic groups that would experience the largest journey delays as a result of road network changes. The application focused primarily on the Hawke's Bay region, demonstrating the advantage of a national scale model.
- *Auckland infrastructure testing.* Combined assessments of city-shaping infrastructure projects and accompanying road pricing mechanisms are being conducted to support ongoing policy planning. This has included the implementation of various road and public transport projects in the Auckland context, with analysis focusing on demand response, mode share shift and equity outcomes.
- *Electric Vehicle Charging.* The team have also started to experiment with the model's ability to optimise. For example, using Monty to identify electric vehicle charging infrastructure locations on the State Highway network.

The team are also looking to improve the country-wide coverage of future infrastructure plans implemented in the 2038 future scenario, to ensure that improvements to the network outside of the main centres are appropriately captured.

6.3 Project Roadmap

Alongside the implementation of case study analyses, the model development team is working on continuously improving model functionality. As relatively novel transport modelling methodologies and technologies have been used, Monty is confronted with an almost limitless list of possible improvements. Each chapter of this report has provided a detailed discussion of possible and ongoing future work packages aimed at enhancing the model's performance in some manner. Further, the AcBM component has been peer-reviewed by Prof. Eric Miller – the outcomes of which are reproduced in Appendix A. This review provided additional suggestions and directions for the future development of the project.

Table 26 summarises the planned work packages and assigns a tentative priority to each task. Items not shown in this table are currently part of the feature backlog, representing a longer-term collection of efforts. The priority order in combination with the backlog are regularly reviewed by the model development team and modified in reaction to emerging requirements.

Currently, revision of the network generation pipeline is of the highest priority. The ambition (as for every component of Monty) is to create an intuitive and reproducible generation procedure for the transport network. In the same work package, the calibration of traffic flow speeds will be addressed. Better income modelling follows closely after this, as this will unlock more sensitive choice modelling across numerous components of the modelling process. The third component with a high development priority is a better approach to the mutation of departure timings in the network simulation. This will improve the model's ability to capture effects from interventions such as peak-hour tolls.

The team also continue to work on and develop the IT infrastructure required to support the ambitious scope of the model, including the ability to provide effective output analysis, visualisation and to work on meta-analyses including equity considerations.

Table 26: Upcoming work packages

Component	Task	Priority
Network Generation (Chapter 3)	Pipeline Revision	★★★
Population Synthesis (Section 4.1)	Income Model	★★★
	Vehicle Ownership Model	★★
Activity Synthesis (Section 4.2)	Income-Dependent Behaviour	★
	Better Workplace Assignment	★
Network Simulation (Chapter 5)	Time Mutation	★★★
	Parking Cost Incorporation	★★
Validation (Section 5.7)	Traffic Flow Speed Calibration	★★★
	Rail Patronage Accuracy	★
	Sensitivity Testing	★★

References

- ActivitySim. (2022). *ActivitySim 1.1.3 Documentation*. (GitHub) Retrieved 12 8, 2022, from <https://activitysim.github.io/activitysim/v1.1.3/>
- Auckland Transport. (2021). (New Zealand Government) Retrieved 8 30, 2023, from <https://at.govt.nz/media/1986141/final-regional-land-transport-plan-2021-2031-web-version.pdf>
- Bierlaire, M. (2020). A short introduction to PandasBiogeme. Lausanne, Switzerland: TRANSP-OR 200605.
- Bowman, J. L., & Ben-Akiva, E. M. (2001). Activity-based dis-aggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice*, 35(1), 1-28.
- Castiglione, J., Bradley, M., & Gliebe, J. (2015). Activity-based travel demand models: A primer. In *SHRP 2 Report S2-C46-RR-1*. Washington DC, USA: The National Academies Press.
- Friedrich, M., Pestel, E., & Simon, R. (2019). Scalable GEH: A Quality Measure for Comparing Observed and Modeled Single Values in a Travel Demand Model Validation. *Transportation Research Record*, 2673(4), 722-732.
- Hess, S., & Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, 32, 100170.
- Horni, A., Nagel, K., & Axhausen, K. (2016). *The Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press.
- Land Information New Zealand. (2019). *NZ Building Outlines*. (New Zealand Government) Retrieved 10 26, 2022, from <https://data.linz.govt.nz/layer/101290-nz-building-outlines/>
- Mackay, J. (2023). *The Establishment of a Nationwide Agent-Based Transport Model for Aotearoa New Zealand: Architecture and Technology*. European Transport Conference.
- MATSim. (2023). *MATSim-Extensions by SBB*. (GitHub) Retrieved 8 30, 2023, from <https://github.com/matsim-org/matsim-libs/tree/master/contribs/sbb-extensions>
- MATSim. (2023). *Multi-Agent Transport Simulation*. (GitHub) Retrieved 8 30, 2023, from <https://github.com/matsim-org>
- Miller, E. J., & Roorda, M. J. (2003). Prototype model of household activity-travel scheduling. *Transportation Research Record*, 1831(1), 114-121.
- Ministry of Education. (2018). *Student Rolls by School 2010-2022*. (New Zealand Government) Retrieved 4 26, 2022, from <https://www.educationcounts.govt.nz/statistics?>
- Ministry of Transport. (2020). *New Zealand Household Travel Survey*. Retrieved 5 26, 2022, from <https://www.transport.govt.nz/area-of-interest/public-transport/new-zealand-household-travel-survey/>
- OpenStreetMap contributors. (2022). Planet dump retrieved from <https://planet.osm.org>.
- Sherr, W., Manser, P., Joshi, C., Frischknecht, N., & Métrailler, D. (2020). Towards agent-based travel demand simulation across all mobility choices – the role of balancing preferences and constraints. *European Journal of Transport and Infrastructure Research*, 20(4), 152-172.
- Statistics New Zealand. (2018). *Regional Council 2018 (generalised)*. (New Zealand Government) Retrieved 8 28, 2023, from <https://datafinder.stats.govt.nz/layer/92204-regional-council-2018-generalised/>
- Statistics New Zealand. (2018). *Statistical Area 2 2018 (generalised)*. (New Zealand Government) Retrieved 8 28, 2023, from <https://datafinder.stats.govt.nz/layer/92212-statistical-area-2-2018-generalised/>
- Statistics New Zealand. (2020). *Statistical area 1 dataset for 2018 census - updated March 2020*. (New Zealand Government) Retrieved 5 26, 2022, from <https://www.stats.govt.nz/information-releases/statistical-area-1-dataset-for-2018-census-updated-march-2020>

- Statistics New Zealand. (2022). *Subnational population projections*. (New Zealand Government) Retrieved 16, 2023
- Swiss Federal Railways. (2023). *Activity-Based Modelling in PTV Visum*. (GitHub) Retrieved 8 29, 2023, from <https://github.com/SchweizerischeBundesbahnen/abm-in-visum>
- Waka Kotahi NZ Transport Agency. (2019). *Transport model development guidelines*. (New Zealand Government) Retrieved 08 29, 2023, from <https://www.nzta.govt.nz/resources/transport-model-development-guidelines/>
- Waka Kotahi NZ Transport Agency. (2020). *Monetised benefits and costs manual*. (New Zealand Government) Retrieved 9 20, 2022, from <https://www.nzta.govt.nz/assets/resources/monetised-benefits-and-costs-manual/Monetised-benefits-and-costs-manual.pdf>
- Waka Kotahi NZ Transport Agency. (2020). *TMS daily traffic counts CSV*. (New Zealand Government) Retrieved 5 26, 2022, from <https://opendata-nzta.opendata.arcgis.com/datasets/tms-daily-traffic-counts-csv/>
- Waka Kotahi NZ Transport Agency. (2023). *Monetised benefits and costs manual*. Wellington: New Zealand Government. Retrieved from <https://www.nzta.govt.nz/resources/monetised-benefits-and-costs-manual>

Appendix A

Demand Generation Peer Review

The AcBM component of Monty was peer-reviewed by Prof. Eric Miller from the University of Toronto. This examined at the processed employed to generate both the synthetic population of agents and the activity schedules assigned to them. The transport supply and network simulation components were not within the scope of this peer review. Eric’s entire peer review has been reproduced in this report, followed by a response from the model development team.

AcBM Peer Review

Prof. Eric Miller’s response has been reproduced exactly, with minor edits provided by the model development team to update references to sections and figures that had since moved in the original interim model development report.

REVIEW OF DEMAND GENERATION COMPONENTS OF “MONTY”, THE NEW ZEALAND ACTIVITY/AGENT-BASED TRAVEL DEMAND MODEL

By:

Prof. Eric J. Miller, PhD

University of Toronto

For:

Te Manatū Waka – Ministry of Transport, New Zealand & Arup

July 14, 2023

EXECUTIVE SUMMARY

The New Zealand Ministry of Transport (MoT) initiated the development of a nationwide, multimodal transport model, known as “*Monty*”. The model aims at supporting emerging questions in New Zealand related to social equity, emissions or future pricing interventions. MoT partnered with Arup to build and apply this model based on the latest state-of-the-art in transport planning involving the application of an agent- and activity-based microsimulation approach to modelling travel demand and transport system performance.

At a very high level, the model system consists of three major components:

1. A population synthesis procedure for generating a synthetic nationwide population of individuals and their salient attributes (age, etc.) to provide a representative set of trip-makers.
2. An agent- and activity-based travel demand model that generates the out-of-home activities and associated trips made by this synthetic population on a typical weekday.
3. An agent-based model that simulates path choice by travel mode for the predicted origin-destination trips and the associated road and transit network performance levels.

This report presents a peer review of the first two components: population synthesis and activity-based demand generation. This review includes an assessment of both the modelling methods used and the performance of these methods in capturing base year travel behaviour.

Overall, the current *Monty* activity- and agent-based travel demand forecasting system is a well-designed and well-implemented model system that conforms well to current industry standards with respect to both design and performance. It represents a robust balance between theoretical considerations and practical issues of data availability, computing requirements, etc. It is well-grounded in best-practice experience with such models in both North America and Europe.

The model system is still under development and so, inevitably, possesses areas for future improvement. This report discusses a number of these. Key findings about both “minor” (typically short-run) and “major” (typically long-run) recommendations for model improvement are summarized in Section 4 of this report. In both cases, these are categorized by the three major model system components reviewed in this report: population synthesis; long-term decisions; and daily activity scheduling.

1. INTRODUCTION

The New Zealand Ministry of Transport (MoT) initiated the development of a nationwide, multimodal transport model. The model aims at supporting emerging questions in New Zealand related to social equity, emissions or future pricing interventions. MoT partnered with Arup to build and apply this model based on the latest state-of-the-art in transport planning involving the application of an agent- and activity-based microsimulation approach to modelling travel demand and transport system performance.

At a very high level, the model system consists of three major components:

1. A population synthesis procedure for generating a synthetic nationwide population of individuals and their salient attributes (age, etc.) to provide a representative set of trip-makers.
2. An agent- and activity-based travel demand model that generates the out-of-home activities and associated trips made by this synthetic population on a typical weekday.
3. An agent-based model that simulates path choice by travel mode for the predicted origin-destination trips and the associated road and transit network performance levels.

This report presents a peer review of the first two components: population synthesis and activity-based demand generation. This review includes an assessment of both the modelling methods used and the performance of these methods in capturing base year travel behaviour. The report is organized as follows:

- Section 2: Review of the Monty population synthesis procedure.
- Section 3: Review of the Monty activity-based travel demand.
- Section 4: Summary of review findings and recommendations.

Unless otherwise indicated, all discussion of the model system is based on the report *New Zealand Agent-Based Transport Model: The Monty Book* (May 1, 2023). For brevity, the model system is generally referred to as *Monty* and references to the document use the format (*Monty*, pp) where “pp” is the page number in the report.

2. POPULATION SYNTHESIS

2.1 Introduction

Agent-based microsimulation models such as *Monty* predict the behaviour of a set of individual persons, generally referred to as *agents*. An agent is an entity that perceives its environment and acts within this environment in order to fulfil its needs and wants. In this case, individual agents (people) participate in daily out-of-activities and associated travel in order to meet their daily needs (engage in work/school, shop, etc.). The agent-based model requires a list of the agents, and their salient attributes (such as age, income, employment status, etc.) upon which to operate.

A population synthesis procedure is a statistical process that generates this list of individual agents and their attributes from more aggregate data, typically derived from sources such as a national census. For example, if an aggregate two-dimensional table of persons by employment status and income category is available, a synthesis procedure might generate a list of individual persons, each with an employment status and an income level. More specifically, all population synthesizers require:

- A “seed sample” of disaggregate observations of agents and their individual characteristics that provide information concerning how these characteristics are correlated within the population. This seed sample is often obtained from a survey, such as household travel survey.
- “Control totals” that describe the aggregate distributions of key attributes within the population being simulated. These control totals are often obtained from a national census.

The population synthesis procedure utilizes these two data to create a synthetic list of all agents in the population, each with an individual set of attributes.

Numerous population synthesizers have been developed over the past several decades. Criteria for selecting and assessing the performance of a population synthesizer are discussed in Section 2.2. Section 2.3 then assesses the *Monty* population synthesizer with respect to both its methodology and its performance in replicating the base year New Zealand population.

2.2 Criteria for Assessing Population Synthesis Procedures

Methodologically, a population synthesis procedure must satisfy several criteria. The procedure must:

1. Utilize best available datasets for both the seed sample and control totals.
2. Generate both persons and the households within which the persons are contained. Household constraints and interactions are important in determining a person's activity and travel behaviour.
3. Generate all person and household attributes required by the activity-based travel demand model.
4. The individual person and household attributes need to be logically consistent and representative of the population being modelled.
5. Be able to generate future year forecasts, not just reproduce the base year distribution.
6. Be statistically valid.
7. Be computationally efficient.¹¹

In terms of performance, the synthesis procedure can only be judged in terms of how well the synthesized population reproduces known aggregate control totals and distributions. The assumption is that if the procedure reproduces the base case well, then future year populations will also be well synthesized, providing that the future year control totals are forecasted reasonably well. Standard criteria defining acceptable “goodness of fit” are not generally available. It is expected, however, that the procedure should be able to reproduce relevant observed distributions “well”, with, at most, minor deviations from control totals reflecting only minor random simulation errors.

2.3 Assessment

2.3.1 Methods

Monty uses a well-known open-source population synthesis procedure, *PopulationSim*, which is part of the US-based open-source *ActivitySim* software system.¹² *PopulationSim* is, in turn, an upgraded version of *PopSynIII*, another common synthesizer.¹³ The *Monty* implementation is assessed with respect to the seven evaluation criteria defined in Section 2.2.

1.1 Input Data, Seed Sample: The person and household seed samples were obtained from New Zealand Household Travel Survey (HTS) data for the financial years 2003/04 to 2017/18, inclusive. That is, the annual survey samples were combined into a single dataset in order to create a larger sample to use in the population synthesis process. This approach assumes that household structures within New Zealand remained relatively unchanged over this time period. This is a relatively mild assumption that is justified by the increased much larger total sample that it generates, relative to using the base year 2017/18 alone. This enables a more statistically reliable estimate of the national distribution of households and persons to be used in the synthesis.

After cleaning, 33,400 household are in the household seed sample, representing approximately 2.0% of total households in 2018. This is a quite typical sample size for large-scale population synthesis applications, many of which use samples of 1% or less.

1.2 Control Totals: All control total data were obtained from the 2018 national census, which is the standard approach for base year population synthesis applications.

2. Generate both households & persons: The procedure simultaneously creates a synthetic population of households and associated persons by attaching weights to each household sample record so that the weighted totals for a set of household and person attributes sum as closely as possible to the population control totals. This is an industry standard approach. One possible limitation of this approach is that, given the computational burden of the optimization problem being solved, the number person and household attributes that can be

¹¹ While computational efficiency is always a desirable property, it is of somewhat secondary importance for population synthesis. Population synthesis usually is only run when a new forecast year “land use” scenario is to be generated, with many demand model runs typically using the same synthesized population. Thus, some inefficiency in the population synthesis algorithm can sometimes be tolerated.

¹² *PopulationSim* is documented in detail in <https://activitysim.github.io/populationsim/>.

¹³ Paul, et al. (2018).

included in the optimization procedure is somewhat limited. This is not a significant problem in the current application given the relatively small number of attributes being synthesized, but may become an issue if additional attributes (notably occupation, income and auto ownership levels) are added to future versions of the model.

3. *Generate required attributes:* All household and person attributes required by the demand model are generated. If this set of attributes were to be extended in future versions of the model, the population synthesis procedure would need to be extended accordingly. Discussion of the “completeness” of this attribute set for activity/travel demand modelling is taken up in Section 3.

4. *Appropriate distribution of attributes:* Use of the high-quality HTS and census datasets ensures that the attributes assigned to the synthesized persons and households are logical and representative. In particular, since all households in the synthesized population are generated directly from NHS households, their attributes correspond to households observed in the survey.

5. *Synthesizing future year populations:* The current *Monty* application only deals with synthesizing a base year 2018 population. Future work will be required to demonstrate the procedure’s application to future year forecasts. Principally, this will require developing appropriate methods for robustly generating the future year control totals for input into the procedure.

6. *Statistical validity:* The synthesis procedure is an industry standard that is known to generate a synthesized population which is the statistically most likely population given the control totals and the seed sample.

7. *Computational efficiency:* The original *PopSynIII* code was quite inefficient. While the reviewer has not tested the *Monty PopulationSim* run time, it is understood that it runs very quickly on a modestly-sized computer (e.g., a laptop), and is well within acceptable limits for efficient use.

2.3.2 Performance

The accuracy of the population synthesis procedure in generating a valid base year (2018) disaggregate population can be assessed at multiple spatial levels. At the overall national level, the totals by individual attribute are essentially perfect (*Monty*, Table 9, 29). This is to be expected of any valid synthesizer and is essentially a check on the internal workings of the algorithm.

A more important test is the population generation at an appropriate level of spatial disaggregation, in this case the SA2 zone level. The generation of total households and persons by SA2 (Figure 15) and counts of persons and households by individual attribute (Table 10) are also excellent (*Monty*, 29). Again, this is an expected result for a well-founded synthesizer.

More interesting validation tests involve attributes that emerge out of the synthesis that are not directly controlled for within the procedure, as well as correlations among attributes that, again, are emergent out of the process. The current documentation provides two examples of this:

1. Comparison of the generated distribution of the number of dependent children within households vs. the census distribution. The fit is excellent, indeed, near-perfect (*Monty*, Figure 16, 30).
2. For families with dependent children, the age distribution of the youngest child in the household (*Monty*, Figure 17, 31). This is a more challenging attribute to generate, and so a bit more scatter is observed, but the fit is well within normal expected tolerances.

In terms of additional validation, it would be useful to see a few comparisons between synthesized and observed cross-tabulations of key variables that are likely to have significant impact on activity and travel, notably:

- Age by employment status.
- Age by student status.
- Number of household workers by car availability.

3. ACTIVITY-BASED DEMAND MODEL

3.1 Introduction

Travel is derived demand; i.e., people generally do not travel for the sake of the travel *per se*, but rather in order to participate in out-of-home activities. Thus, it has been long recognized that travel can be best modelled as an *activity-based* process in which people choose the activities in which they need engage during a given day, with trip-making emerging from the need to travel to these activity locations. Further, activity and travel participation varies across individuals, given their various needs, preferences and capabilities. Thus, a *disaggregate* modelling approach is required in which the behaviour of individual agents is *microsimulated* in order to account for this heterogeneity within the population.

Given these observations, activity-based travel demand models have been under development for a number of decades, with an increasing number and variety of models entering operational practice. *Monty* is derived from a Swiss national model used by Swiss Federal Railways, with this model, in turn, being consistent with operational practice in many US urban applications (*Monty*, 33).

Criteria for selecting and assessing the performance of activity-based travel demand models are presented in Section 3.2. Section 3.3 then assesses the *Monty* long-term decision model, while Section 3.4 assesses the daily tour/activity model. In both cases the assessment deals with both methodological and performance criteria.

3.2 Criteria for Assessing Activity-Based Travel Demand Models

Methodologically, there is no universally adopted approach to activity-based travel demand modelling. Common features of all such models, including *Monty*, however, are:

- They are disaggregate, agent-based models, in which the travel behaviour of individual persons is simulated.
- They typically model a 24-hour weekday.
- They model all trip purposes (activity types) and relevant travel modes.
- They model trip mode choice as a tour-based process.
- They typically interface with a network modelling software package for road and transit assignments; in the case of *Monty*, this is *MATSim*. I.e., the activity/travel model *per se* can be thought of as replacing the first three stages of the four-step procedure (generation, distribution, mode choice).
- They typically can be divided into two main components (in addition to the population synthesis procedure discussed in Section 2): (1) “mid/long-term” decisions concerning work/school locations for workers/students, household auto ownership, driver’s licence possession, etc.; and (2) the simulation of daily activity and travel given these longer-term decisions.

Beyond these basic, common features, operational models can be categorized along two important dimensions. The first is the treatment of household interactions in determining individual trip-maker behaviour, which can take two general forms:

1. *Household-based models* in which household interactions (joint activities, sharing of household resources such as cars, etc.) are explicitly modelled and condition/constrain individual level activities and trip-making.
2. *Individual trip-maker models*, in which some household characteristics (e.g., number of household autos, household income) may influence individuals’ activity/travel decision-making, but the behaviour of the household *per se* is not directly modelled.

The household-based approach is arguably the more “behaviourally sound” since it provides a more complete context for modelling activity/travel, but a majority of operational models are individual-based, including *Monty*.

The second major categorization is between (1) *tour-based* models, in which the focus of the model is to predict tours directly and (2) *activity-scheduling* models, in which activity participation is directly modelled, with tours emerging from the resulting activity patterns. Tour-based models are the most common operational type, but activity-scheduling models are also in operational use. *Monty* is something of a hybrid formulation in that it is primarily tour-based but with activity-scheduling elements (as discussed further in Section 3.4).

Regardless of the overall model system structure, methodological criteria for assessing the model system include:

1. Completeness: All activity/trip purposes and travel modes are modelled over an entire 24-hour day.
2. Consistency: All calculations are internally consistent (trips sum to known totals; autos cannot be used by two drivers at the same time; etc.).
3. Behavioural soundness:
 - a. An appropriate set of explanatory variables expected to influence activity/travel behaviour are included.
 - b. Model mathematical formulations correspond to best practice.
 - c. The model embeds best theoretical understanding of activity/travel behaviour.
4. Policy sensitivity. The model system is able to forecast future travel behaviour in a robust and credible fashion as a function of a variety of policy-relevant factors (as per *Monty*, 34):
 - a. Population and land use changes.
 - b. Behavioural changes in trip-making.
 - c. Road and transit infrastructure and service operation changes.
 - d. External factors (changes in fuel prices, transit fare policies, etc.)
5. Feasibility for generating forecasts (or scenarios) of future year explanatory variable for input into the model system.
6. Computational efficiency.

Criteria for assessing the performance of the model system include:

1. Individual models need to satisfy standard statistical evaluation criteria:
 - a. Parameters need to be statistically significant and of correct sign.
 - b. The overall goodness-of-fit of the model needs to meet industry standards.
 - c. The model must display appropriate elasticities / sensitivity to changes in explanatory variables.
2. As discussed in detail in NZ Transport Agency (2019), the model system must be able to reproduce observed base year trip-making to acceptable levels of fit to observed data with respect to:
 - a. Original-destination (O-D) trip matrices.
 - b. Trip mode shares.
 - c. Screenline counts.
3. Going beyond trips per se, an activity-based model must also be able to reproduce the daily *activity patterns* that generate the travel:
 - a. Activity episodes by type, location and start time.
 - b. Activity episode durations.
 - c. Activity patterns/sequences.

3.3 Long-Term Decision Model Assessment

3.3.1 Methods

Monty assigns a work and school location to each worker and student in the synthesized population. A joint trip mode and location choice model is used, which takes the form of a “doubly-constrained” gravity/logit model. The “mode choice” component of the model is only used to define the relative accessibilities of different work/school locations for workers and students given their residential locations; actual trip mode choices are determined later in the model system. This is a quite standard approach to determining work and school locations. Strengths of this approach include:

- The model has strong theoretical foundations both behaviourally (random utility theory) and statistically (information theory).
- The doubly-constrained formulation:
 - Accounts for workers and students competing for the finite numbers of jobs and school capacities within each analysis area.
 - Ensures internal consistency in that each worker is assigned exactly one job and vice versa (and the same for student location assignments being consistent with school enrolment).

- The mode sensitivity ensures that work/school commuting patterns will change as the accessibility provided by the transportation system changes due to infrastructure investments and other policies.
- This approach separates the long-term choice of work/school location from the daily decision of whether the person will travel to work/school on the given day being simulated. This is a very important improvement over conventional trip-based models which typically include a work/school “distribution” (“destination choice”) model within the daily trip model system.

Two concerns, however, exist with the current work location choice model:

1. It does not distinguish between full-time and part-time employment locations. The population synthesis procedure assigns full- and part-time status to the synthesized workers, but this categorization is not carried forward into the work location model. It is to be expected that these two types of workers may have different perceptions of modal travel times. Perhaps more importantly, they should be matched with the same type of employment opportunities. Otherwise, it may well be the case that part-time workers will be assigned to full-time jobs and *vice versa*, with associated possible distortions in O-D work flows (which can then ripple through the modelling of non-mandatory activities). Hence, in future versions of the model it should ideally be segmented into two models: one for full-time workers and one for part-time workers.
2. Similarly, worker occupation is not represented in the model system at all. As with employment status, however, work locations and activity participation vary by occupation, and workers in each occupation group should only be assigned to jobs of this occupation type. This has always been an important criterion for modelling work commuting, but it has been even more important in this post-COVID period of increased working from home (WfH), the rates of which vary considerably across occupation groups.

3.3.2 Performance

The work location choice model parameters are all statistically significant and of correct sign (*Monty*, Table 12, 39), and the model’s overall goodness-of-fit (adjusted R^2 of 0.530) is quite good for this type of model, especially one being applied at the national scale. The model specification is well-founded (the caveats raised above concerning lack of employment status and occupation aside).

Similar comments hold for the school and university models (*Monty*, Table 13, 40 and Table 14, 40, respectively), although the school model adjusted R^2 is only 0.326 and the university model adjusted R^2 value is not reported. Predicting school locations is a much more challenging task than work locations, since school location choice can be subject to a wide variety of factors that are beyond the modeller’s ability to observe. The performance of the *Monty* models is quite good relative to current industry standards.

In terms of validating the work and school location choice models’ performance in reproducing observed base year data it is observed that:

- The modelled trip length frequency distributions for both work and education achieve excellent fits to the observed distributions (*Monty*, Figure 34, 55).
- The fit between modelled and observed O-D linkages at the SA2 level are also very good, with an R^2 of 0.92 for work and 0.88 for education (*Monty*, Figure 35, 55).

3.4 Daily Activity Schedule Model Assessment

Monty builds a daily activity schedule for a given individual through the sequential application of four models:

1. Generation of mandatory (work and school) and non-mandatory (no work/school activities) tours.
2. Location choice for non-mandatory activity episodes.
3. Determination of start times and durations for all generated activities.
4. Assignment of initial travel modes to each trip in each tour. These mode choices will subsequently be updated in MATSim (which is not under review within this report).

The tour-based nature of the model is evident in this cursory description of the model structure. Three tour types are modelled: work tours, school tours and “non-mandatory” tours, where the work and school tours can include non-mandatory activities as well. All tours begin and end at home. Non-mandatory activities are further categorized into three types: leisure, shopping, accompany someone and business (for work tours; i.e., work-

related business activities outside the home and the normal workplace). Each of the four main steps in the process are discussed in more detail below.

3.4.1 Daily Tour & Activity Generation

Multinomial logit (MNL) models are used to determine the number and type of home-based tours in which an individual will engage. Usage of MNL models in such applications is standard practice. Each tour generated is specified in three primary steps:

1. The type of tour (work, school, mandatory) is chosen.
2. “Intermediate activities” are chosen for this tour.
 - a. For work and school trips, any combination of three types of intermediate activities can be selected:
 - i. Inserting an activity on the way from home to work/school.
 - ii. Inserting an activity on the way from work/school to home.
 - iii. Inserting a work/school-based subtour in which the person travels from work to another activity and then returns to work.
 - b. For non-mandatory tours, this model selects the number of non-mandatory activities in which the trip-maker engage on this tour.
3. For each non-mandatory activity, the activity type is selected.

While this is a typical example of current tour-based modelling practice, a possible limitation of this model structure is that the generation of participation in an activity prior to knowing its purpose is not particularly behaviourally sound. I.e., the model system first asks the question “do you want to engage in an activity” and then subsequently asks “what type of activity”? A more explicitly activity-based approach would be to start with the question “in what activity types do you want to engage?”

Full specifications and parameter estimation results for these models are not provided in the documentation. Based on the available documentation, it is clear that the models are largely driven by trip-maker and household attributes (which is appropriate and expected), but that they also include accessibility terms that make them sensitive to transportation system service levels. This is a very desirable feature, since it means that if the transportation system is improved in some way (and/or land uses change to increase the attractiveness of activity locations) that activity participation will increase (and *vice versa* if the transportation system deteriorates). This “elasticity” of activity participation to the transportation system is a very desirable property for policy analysis, and can be contrasted with the “inelasticity” traditionally found in conventional trip-based models.

3.4.2 Non-Mandatory Activity Location Choice

A nested mode-location choice model is used to assign locations for all non-mandatory activities. As with the work/school location choice models, the mode component is used to define the accessibility of competing locations, not to assign actual modes to the trips to/from the activity location (which are determined later in the model system). A “time-space prism”, “rubber-banding” (*Monty*, Figure 24, 43) is used to constrain the feasible activity location choice set, given the known locations of prior and posterior activities within the tour. This modelling approach is very typical of current industry best practice. The use of region-specific parameters is also welcome, since it is very reasonable to expect these to vary across the nation.

The mode choice model specification and parameter estimates (*Monty*, Table 17, 45) and location choice model specification and parameter estimates (*Monty*, Table 18, 46) are both sensible. The model developers note the relatively low values of time in this model (*Monty*, 45), but this is not particularly of concern in this application, especially given that actual travel mode choices are made later in the model system.

The one minor technical concern with this model is that the two logit models are estimated sequentially. That is, the “lower-level” mode choice model is estimated first. “Logsum terms” are then computed based on the estimated mode choice model parameters and entered into the “upper-level” location choice model as an explanatory variable. This will result in the location choice parameter standard errors being biased downwards (since the modelling error in computing the logsum variables is not accounted for), and so the reported t-statistics will be biased upwards (McFadden, 1981). While not a major problem given the good performance of the model (and the parameter estimates are not biased by this procedure, which is the critical issue), in this

age when full-information joint estimation methods are available for nested logit models, this concern might well be rectified in future versions of the model.

3.4.3 Activity Start Times & Durations

This procedure assigns start times and durations to each non-home activity in each tour. This introduces something of an “activity scheduling” element to the model system, which is welcome and somewhat advanced relative to many US-based tour-based models. The procedure is based on the Swiss Rail model as documented Scherr, et al. (2020). It makes use of a combination of survey-based probability distributions for activity start times and durations and feasibility rules concerning logical daily time budget constraints. Although the overall model structure is fairly different, this approach is similar in concept to that used in the TASHA activity-scheduling model that is in operational use in Canada to generate activity start times and durations (Miller and Roorda, 2003; Miller and Duggal, 2023).

3.4.4 Trip Mode Choice

Once all details of all tours for each individual have been determined, a provisional travel mode is assigned to each trip in each tour for each person. No documentation is provided in the Monty report concerning this model. These mode choices, however, simply represent an “initialization” of the trip/tour records that are provided to MATSim for assignment to the road and transit networks. MATSim may subsequently update these choices, as the agents “learn” about the detailed modal service levels experienced in the MATSim network simulations. Evaluation of the MATSim component of the *Monty* system is beyond the scope of this review.

3.4.5 Missing Variables & Other Issues

While the current activity generation model appears to be performing well in terms of base year calibration, four areas for possible future improvement can be identified.

1. Household income is absent from the model. Income is a major explanatory variable for auto ownership, activity participation, activity location choice, and travel mode choice, among possibly other elements of the model system. Best-practice models should include income. This is a gap that should be rectified in future versions of the model. This will require extending the population synthesis procedure to include income as another household attribute to generated.
2. Similarly, the current mode is overly simplistic in its treatment of auto ownership, another key variable in explaining activity location choice and mode choice. The current model only has a binary “0-1” variable indicating the absence or presence of one or more vehicles in the household, but not the actual number of vehicles. While this binary indicator is quite useful (going from no cars to at least one car has a dramatic impact on travel behaviour, the ability to differentiate auto usage between one-car and multi-car households is also extremely important. In particular, it is not clear from the documentation that two persons in a one-car household could not be using that car simultaneously since there is no information concerning the number of cars and, hence, constraint on their usage.
3. This issue of car allocation is one of several examples where the lack of explicit modelling of household activities and interactions may limit the model’s performance. Joint trips, ridesharing, allocation of “household activities” (shopping, household maintenance, childcare duties, etc.) all are difficult or impossible to model within an individual-centric model structure.
4. Related to this last point, the “escort” non-mandatory activity is not well defined in the documentation. This may simply represent the need for more explicit discussion of this activity. But at the moment it is not clear whether this includes “serve passenger” activities (dropping one’s child at school on the way to work), joint activities (going out to dinner together), and/or accompanying others on a trip (a child going with a parent on a shopping trip; a parent taking a child to a soccer practice). These are quite different activities. They also all derive from the interaction of two or more household members schedules (the child’s school or soccer team participation) and/or joint decision-making (deciding to go to dinner together).

3.4.6 Performance Assessment

Based on the documentation presented in the *Monty* report, comments concerning the performance of the model system in replicating observed base year activity and travel participation are presented below, by model system component.

3.4.6.1 Activity Generation

Observations concerning the performance of the activity generation model include the following:

- The model is doing a very good job in reproducing the national mean number of daily trips by age group (*Monty*, Figure 26, 49). The mean trips per day are slightly low for the 25-65 and 65+ age groups, which should be investigated in further model development given the importance of these age groups in the population; but nevertheless, the model performance is very good. Note that these results are the aggregate, emergent result of both tour generation and the insertion of activities within tours and so are not a “directly calibrated” outcome.
- Similarly, the overall distribution of the number of trips per day by employment status (Figure 27) and studying status (Figure 28) are also very good (*Monty*, 50). Slight deviations of possible interest include:
 - Trips by workers appear to be slightly over-estimated, except in the right-hand tail of very large daily trip counts. This may be consistent with the comments above about the model structure limiting very large trip chains on work trips, but the errors are very small in absolute terms.
 - A similar observation can be made about the Tertiary full-time distribution.
 - Simple 2-trip school tours (Home-School-Home) are rather significantly under-predicted, with a compensating over-prediction of complex school daily patterns of 4 or more trips.
- Consistent with the above observations, the number of daily work tours per worker are slightly over-predicted (Figure 29) and the number of non-mandatory tours per person (Figure 30) are somewhat under-predicted (*Monty*, 52). While the results are consistent with industry experience, some additional calibration of the models might well improve these results.
- The distribution of counts by activity type (*Monty*, Figure 31, 53) by employment status is excellent, with very little error across the board. Two points to note are:
 - The excellent fit to “home” (i.e., return home trips) is particularly noteworthy, since this is the emergent outcome of both the tour and activity generation processes and is not directly modelled.
 - The excellent fit to the non-mandatory trip counts indicates that the non-mandatory activity choice model is working well.

3.4.6.2 Non-Mandatory Location Choice

The choice of non-mandatory locations affects not just trips lengths and O-D patterns for non-mandatory activities but also for mandatory activities, since non-mandatory activities and their locations “intervene” within what would otherwise be “direct” trips between home and work/school locations. Thus, for example, an overall (O-D) trip length frequency distribution is a complex outcome of work/school location choices, non-mandatory location choices and the Observations concerning the performance of the non-mandatory activity choice model include the following:

- The overall (all-trips) trip length frequency distribution (Figure 32) and the trip length frequency distributions by both region (Figure 33) and trip/activity purpose (Figure 34) look very good (*Monty*, 54-55).
- As the report authors note, *Monty* tends to overpredict short trips somewhat, across most trip purposes. But, as the authors also note, travel surveys almost always under-report short trips due to several systematic limitations in survey methodology.

3.4.6.3 Activity Start Times & Durations

The model is reproducing the overall daily pattern of activity start times well (*Monty*, Figure 36, 56), although AM peak period starts are somewhat low and mid-day starts are a bit high. It is very difficult to reproduce start times perfectly, and the performance of the model is certainly well within current industry standards. It is also important to note that the model is doing an excellent job in getting the work and education activity start times

right (*Monty*, Figure 37, 56). It is clear that the overall discrepancies in start times are with business-related travel and “escort”, which are difficult activity types with which to deal.

A very important feature of activity-based models is the prediction of activity durations, since these determine the start times for non-home-origin trips. As indicated in Figure 38, *Monty* is currently over-predicting work and school durations somewhat (*Monty*, 57). Correspondingly, leisure is significantly under-predicted and shopping somewhat. This is an area for future investigation to see if results could be improved somewhat.

Figure 39 (*Monty*, 57) presents a consistent message with work being over-estimated in the overall activity profiles and leisure being notably under-estimated. The allocation to travel is over-estimated. This might indicate an over-estimation of trip distances (which is not observed in the trip length frequency distributions), but is more likely due to the use of “conservative” travel times in the activity generation model (*Monty*, 57). This use of conservative travel times is a reasonable assumption to maximize the likelihood of the generated activity patterns being feasible as the mode assignments and resulting travel times are updated in MATSim. But it does possibly skew the activity scheduling results somewhat since the amount of time of activity participation is being under-estimated somewhat. It is likely, however, that this is largely corrected in MATSim, which is able to update activity start times and durations as improved modal travel times are generated.

3.4.6.4 Sensitivity Testing

Sensitivity testing is an extremely important component of model validation. Such tests typically take such forms as:

- Across the board changes in a key explanatory variable (e.g., an x% change in auto travel times) to explore the aggregate elasticity of the model’s response to such a change.
- Changing socio-economic inputs (x% growth in population and employment; x% change in average real income; etc.).
- Implementing a major new transport investment (significant improvement in transit services in the major cities).

Such tests take the model beyond calibration by “stressing” it in a variety of ways through the generation of “out of sample” predictions. In so doing, they provide strong indications of how it will perform in application. To have confidence in the use of the model in operational planning practice it is important that the model displays sensitivity to system changes, etc. that fall within expected ranges based on international experience. Such sensitivity testing does not appear have yet been done (*Monty*, 58) but should be high priority among “next steps”, since it may well expose elements of the current system that would most benefit from improvements of one form or another.

4. SUMMARY OF FINDINGS & RECOMMENDATIONS

Overall, the current *Monty* activity- and agent-based travel demand forecasting system is a well-designed and well-implemented model system that conforms well to current industry standards with respect to both design and performance. It represents a robust balance between theoretical considerations and practical issues of data availability, computing requirements, etc. It is well-grounded in best-practice experience with such models in both North America and Europe.

The model system is still under development and so, inevitably, possesses areas for future improvement. The previous sections of this report have discussed a number of these. Key findings about both “minor” (typically short-run) and “major” (typically long-run) recommendations for model improvement are summarized in the following two sections. In both cases, these are categorized by the three major model system components reviewed in this report: population synthesis; long-term decisions; and daily activity scheduling.

4.1 *Minor Issues & Recommendations*

“Minor” recommendations deal with short-term improvements within the current model system, additional validation and/or improvements in documentation.

4.1.1 *Population Synthesis*

1. Additional validation of the population synthesis procedure should involve few comparisons between synthesized and observed cross-tabulations of key variables that are likely to have significant impact on activity and travel, notably:
 - a. Age by employment status.
 - b. Age by student status.
 - c. Number of household workers by car availability.

4.1.2 *Long-Term Decisions*

No minor issues identified.

4.1.3 *Daily Activity Scheduling*

1. The non-mandatory activity location choice nested logit model should be re-estimated using current full-information likelihood methods available in a number of standard software packages.
2. Better documentation of the activity timing and final trip mode choice models could be provided to provide enhanced understanding of the methods currently being used and their base year calibration performance.
3. Better description of the “escort” activity is required to understand what it consists of and how it is being currently modelled.
4. While the model appears to be reproducing base year activity and travel well within typical industry standards, there are a number of small discrepancies that might be reduced with some additional tweaking of model parameters (see Section 3.4.6.1 for detailed examples).
5. Extensive sensitivity testing of the current model should be undertaken, both the further enhance confidence in the model’s performance, as well as to identify possible areas for further model improvement.

4.2 *Major Issues & Recommendations*

“Major” recommendations consist of recommended extensions to the current model system to incorporate an enhanced set of explanatory variables and improvements to the representation of decision processes.

4.2.1 *Population Synthesis*

1. In the longer term the population synthesis procedure will need to be extended to generate additional personal and household attributes required by an extended model system. These should include:
 - a. Household income.
 - b. Replacement of the 0-1 auto availability variable with the number of autos per household
 - c. Worker occupation class.

4.2.2 *Long-Term Decisions*

1. Future versions of the work location choice model should incorporate separate models for:
 - a. Full-time and part-time workers.
 - b. For both full- and part-time workers, be divided into major occupation categories.
2. An endogenous household auto ownership model should be incorporated into the long-term decision component that assigns an explicit number of vehicles to each household as a function of household attributes and transportation accessibility measures (transit availability, etc.).

4.2.3 *Daily Activity Scheduling*

1. Household income should be included in these models.
2. Explicit representation of household auto ownership should be included in (at a minimum) the mode choice models, and consistency checks should be implemented to ensure that a car cannot be driven by two household members at the same time.
3. The model developers should consider possible extensions to:
 - a. More explicitly deal with intra-household interactions and constraints.
 - b. Deal with “escort” / “joint” activity participation and travel.

REFERENCES

- McFadden, D.F. (1981). “Econometric models of probabilistic choice”, in C.F. Manski and D.F. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: MIT Press.
- Miller, E.J. and M. Duggal (2023). “Agent-Based Microsimulation of Activity & Travel: Canadian Operational Implementation Experience”, presented at the 58th Annual Conference of the Canadian Transportation Research Forum, Toronto, May 7-10.
- Miller, E.J. and M.J. Roorda (2003). “A Prototype Model of Household Activity/Travel Scheduling”, *Transportation Research Record, Journal of the Transportation Research Board*, 1831, 14-121.
- NZ Transport Agency (2019). *Transport Model Development Guidelines*.
- New Zealand Agent-Based Transport Model: The Monty Book* (May 1, 2023).
- Paul, et al. (2018). “Multi-level Population Synthesis Using Entropy Maximization Based Simultaneous List Balancing”, *97th Annual Meeting of the Transportation Research Board*, January 2018.
- Scherr, et al. (2020). “Towards agent-based travel demand simulation across all mobility choices – the role of balancing preferences and constraints”, *EJTIR*, 20(4) 152-172. <https://doi.org/10.18757/ejtir.2020.20.4.4463>

Model Development Team's Response to Peer Review

The peer review classifies issues and recommendations into *minor* and *major* categories. The following list provides a response for each identified issue.

Minor Issues (or short-term improvements)

- **Population Synthesis**

1. Cross-tabulations of key variables such as age, labour force status and student have since been added to the report (see Figure 18 and Figure 19). They are validated against the demographic compositions as reported in the HTS.

- **Daily Activity Scheduling**

1. The re-estimation of the non-mandatory location choice model (see Subsection 4.2.3.2) would have clear advantages in terms of theoretical correctness of the model. It has been put into the development backlog. The main concern of the peer review was biased choice parameter standard errors. Since all parameters used across the specified models were highly statistically significant, the risk of the mentioned bias can be minimised. The practical benefits of the re-estimation are expected to be small in comparison to its contributions to the theoretical correctness of the model.
2. Descriptions of the activity timing and mode choice models has now been added to Section 4.2 were not available at the time of Prof. Eric Miller's review. The mode choice model is specifically described in Subsection 4.2.3.2. The model for choosing final activity timings is part of a future work package with a high priority (see *time mutation* in Section 5.8).
3. Better descriptions of the accompany (or 'escort') activity is now provided in Subsection 4.2.3.1. Basically, it refers to an activity that people typically must perform together such as bringing a child to school. Monty does model this behaviour implicitly by introducing increased probabilities for doing accompany tasks when a child is present in the household. However, household members in Monty are not constrained to do accompany tasks together at the same time.
4. Tweaking of model parameters to reduce small discrepancies has been assigned as a medium priority item within the backlog. Such adjustments usually come with the price that they are based on manual adjustments rather than being entirely transparently driven by modelled processes. The current calibration of Monty values automated and reproducible estimation routines more than manual changes where pragmatic. The inclusion of other new model capabilities and refinements (as set out Section 6.3) will require re-estimation of many model parameters and will provide the opportunity to address any remaining discrepancies of this type.
5. Testing the sensitivity of the current model structure is a work package with a high priority. It has been partially done as part of the analysis of future year scenarios (see Section 6.1). However, more systematic tests are currently in planning to analyse the response of Monty to standardised changes in the transport network as well as across population attributes.

Major Issues (or long-term developments)

- **Population Synthesis**

1. Additional personal and household attributes are prioritised based on the specific attribute. A more sophisticated modelling of income has the highest priority currently and will be included in the next calibration phase. The car ownership model (e.g., how many or what type of cars a household possesses) is being considered after this. Integrating worked occupation class is now discussed in Subsection 4.1.6.

- **Long-Term Decisions**

1. Differentiating between part-time and full-time workers is a desired future enhancement of the location choice models. This is of importance since the perception of mode-dependent travel times are different for these groups as mentioned by Eric Miller. The HTS confirms this assertion by showing a significantly lower average commuting distance for part-time workers.
2. Explicitly assigning the number of vehicles to each household is part of a future work package with priority just after modelling income explicitly. The aim is to include a range of decision factors into such a model, such as accessibility and household attributes.

- **Daily Activity Scheduling**

3. Household income as a decision factor for daily (short-term) activity scheduling is planned to be integrated once the attribute has been assigned as part of the population synthesis process. Preliminary analyses have revealed that the frequency of work and particularly business activities increase with a higher income level.
4. Consistency of car ownership in the mode choice model is a desired capability within the network simulation. Given the complexity of this task, it is currently considered low priority but on the backlog.
5. Intra-household interactions such as joint activity participation or sharing a car across multiple household members is a complex modelling task. This is a backlog item with a low priority at the current time.

The model development team thanks Eric Miller for this comprehensive and positive peer-review report.