

Activity-based and agent-based transport model of Melbourne: an open multi-modal transport simulation model for Greater Melbourne

Afshin Jafari^a, Dharendra Singh^{b,c}, Alan Both^a, Mahsa Abdollahyar^a, Lucy Gunn^a, Steve Pemberton^a, and Billie Giles-Corti^a

^aSchool of Global, Urban and Social Studies, RMIT University, Australia; ^bSchool of Computing Technologies, RMIT University, Australia;

^cData61, Commonwealth Scientific and Industrial Research Organisation, Melbourne, Australia

ABSTRACT

Activity- and agent-based models for simulating transport systems have attracted significant attention in recent years. However, building these types of models at a city-wide level and including motorized (i.e. cars and public transport) and non-motorized (i.e. walk and bicycle) modes of transport is a complicated and involved task. This paper presents an open workflow for creating large-scale multi-modal agent-based transport simulation models. The workflow brings together a number of external tools, for example, an activity-based demand generation tool and a road network generation tool, and a set of tools developed for the agent-based model parameter estimation, calibration, and simulation post-processing. We used this workflow to create an activity- and agent-based model for Melbourne and compared the output of the simulation model with observations from the real world in terms of mode share, road volume, travel time, and travel distance. Through these comparisons, we showed that our model is suitable for studying mode choice and road usage behavior of travelers. The calibrated model could be used to test road network change interventions. In addition, a similar workflow can be applied for building simulation models for other cities or could be expanded to include more complicated travel behaviors.

ARTICLE HISTORY

Received 26 December 2021

Revised 16 May 2023

Accepted 18 June 2023

KEYWORDS

agent-based simulation;
city-scale transport
modeling; MATSim;
multimodal transportation

1. Introduction

Activity-based modeling of transport systems allows us to represent individuals, their trips, activities, and diverse decision-making and behaviors (McNally & Rindt, 2007; Rasouli & Timmermans, 2014). Using a bottom-up and disaggregated approach, activity-based models construct a synthetic population and their corresponding trips and activities. As a result, these models provide an innovative solution for examining individual-level travel behavior and multi-modal trips, including micro-modes of mobility like walking and cycling. In this light, the disaggregated approach of activity-based modeling aligns with agent-based models. These models simulate heterogeneous agents and their interactions within their environment and can be used to experiment with various potential scenarios (Gilbert, 2021). Thus, by combining the two approaches, one could capture both heterogeneous travel plans and complex interactions between travelers (Hørl & Balac, 2020; Tajaddini et al., 2020).

In the activity-based modeling framework, generating transport demand typically involves steps to create a list of agents with their demographics, assign activity patterns (i.e. the activity chain or itinerary), and assign locations to activities (Wang et al., 2021). Over time, a variety of methods have been developed to produce activity-based transport demand. Rule-based activity generation and scheduling are commonly used to create transport demands based on probability distributions from travel survey data. Notable examples of activity-based population generation tools include Travel Activity Scheduler for Household Agents (TASHA) (Roorda et al., 2008), and A Learning-Based Transportation-Oriented Simulation System (ALBATROSS) (Arentze & Timmermans, 2000). Another common approach in activity-based transport demand generation involves using utility maximization theory to build activity sequences and scheduling. A significant example of this approach is Comprehensive Econometric Micro-simulator for Daily Activity-travel

CONTACT Afshin Jafari afshin.jafari@rmit.edu.au RMIT University, GPO Box 2476, Melbourne VIC 3001, Australia

© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Patterns (CEMDAP), which creates a full day sequence of trips and activities for each synthetic traveler using a set of econometric models. More recently, machine learning techniques have been used to improve the accuracy and flexibility of transport demand generation (Allahviranloo et al., 2017; Hadjidimitriou et al., 2022; Hesam Hafezi et al., 2021; Koushik et al., 2020).

Recently, Both et al. (2021) developed an activity-based transport demand generation model for the Greater Melbourne area, combining machine learning, probabilistic, and gravity-based approaches. In their model, Both et al. (2021) used k-means clustering to group demographic profiles (i.e. age and gender cohorts) based on their travel patterns from the travel survey. Then, a daily travel plan comprising a sequence of activities (at specific locations and times of the day) connected by travel legs (using particular modes, for example, driving or Public Transport (PT)) was created, consistent with the travels observed for people of that demographic profile group in the data from the Victorian Integrated Survey for Travel and Activity (VISTA) 2012-18 travel survey (Department of Transport, 2018).

A significant limitation of the activity-based models described above is that they generally do not capture interactions between traveling agents, such as traffic congestion and interactions between different travel modes, such as multi-modal public transport trip scheduling or interactions between motorized vehicles and cyclists. Agent-based transport simulation models like Multi-Agent Transport Simulation (MATSim) (Horni et al., 2016) and SimMobility (Adnan et al., 2016) can address this gap by providing a framework for synthetic travelers (hereafter agents) to interact with each other on a city-wide scale. Notable examples of large-scale MATSim models include those for Switzerland (Bosch et al., 2016), Singapore (Erath et al., 2012), Melbourne (Infrastructure Victoria, 2017), while more recent models include those for Paris (Hörl & Balac, 2021) and Berlin (Ziemke, Kaddoura, et al., 2019). Furthermore, MATSim has been used to model different aspects of the transport system, including PT (Rieser, 2016), cycling (Ziemke, Metzler, et al., 2019), and novel concepts such as shared mobility (Becker et al., 2020), shared autonomous electric vehicles (Müller et al., 2021) and vehicle to vehicle communication (Hu et al., 2022).

There are several key challenges to building agent-based and activity-based models at a large scale, such as for a city. According to Kagho et al. (2020), these challenges include the difficulty in accessing the proper input data for models, models being computationally expensive, transparency of the models and their

validation, and reproducibility of simulation scenarios due to the lack of streamlined model development processes and the lack of standardized presentation of models and procedures. These challenges result in difficulties in gaining trust in the model and its outputs from stakeholders, as well as limited transferability and extensibility of the models developed for one city.

In this paper, we present our streamlined workflow for joining the activity-based model developed by Both et al. (2021) with MATSim and the process of developing and calibrating a *large-scale simulation model of the transport system for Greater Melbourne, Australia*. Our model is based on the MATSim simulation toolkit and is the first multi-modal *calibrated and open*¹ activity-based MATSim model for Melbourne. Active modes of travel, i.e. walking and cycling, are explicitly modeled and calibrated in this paper rather than being considered as auxiliary travel modes as has traditionally been the case in large-scale transport models. Furthermore, the complete workflow to create the model, as well as the tools we developed as part of the process, are open source² with the aim of addressing the need for flexible tools and processes to create large-scale simulation models.

The remainder of this article is as follows. Section 2 describes our workflow and the key tools and methods used to develop the simulation model for Melbourne. The calibration process of the model and the evaluation of the calibrated scenario are discussed in Section 3. Finally, in Section 4, we discuss how this model could be used to help inform decision-making for the transport system in Melbourne and the applicability of the framework to other cases and potential future steps of the model.

2. Model development and calibration workflow

Figure 1 provides an overview of the **Activity-based and agent-based Transport model of Melbourne (AToM)** development workflow. The process started with the construction of the road network (Section 2.1.1). The next step in the workflow was activity-based demand generation (Section 2.1.2). Although the activity-based model that we used to generate transport demand does not rely on the road network as input, we used the network nodes for this step and snapped the activity destinations to the nearest network node to ensure that the destinations are accessible *via* the road network. The process of estimating the model parameters is described in Section 2.1.3. These three components were then used as simulation

Figure 1. The model development workflow overview.

inputs for the agent-based traffic simulation model, as detailed in [Section 2.2](#). The simulation output analysis was the next step, in which the simulated mode share, traffic volume, and travel distance and time were compared to real-world observations. The process of running the simulation model, analyzing and comparing the simulation outputs, and adjusting the model to better fit the observed data, i.e. the calibration loop, is covered in [Section 3](#).

As [Figure 1](#) shows, different waves of the VISTA dataset were used by different components in our workflow. The most recent wave of VISTA, that is, 2016–18, has the most recent sample of travelers in Melbourne. However, destination locations were reported at the level of Local Government Area (LGA) and some LGAs such as the city of Wyndham and the city of Melton have a land area of more than 500 km². In earlier versions of VISTA, specifically for the years 2012 to 2016, destination locations were reported at the Statistical Area level 1 (SA1) according to the Australian Statistical Geography Standard (ASGS), an area with an average population of 400 people.³ In this article, for the estimation of model parameters where greater precision of destination location was desired, we used VISTA data for the years 2012 to 2016. For the calibration of the simulation model where the most recent data were desired, we used VISTA data for 2012–2018.

2.1. Simulation input generation

The three main building blocks of a MATSim model include: (i) a list of agents (synthetic population) and their travel diaries (transport demand); (ii) a representation of the physical environment used by agents

when traveling (transport network); and (iii) a way for agents to interact with each other and their environment (coefficients of utility functions) (Wall, 2016). In this section, we describe our process for creating these three inputs.⁴

2.1.1. Building the transport network (network generation tool)

The network generator tool developed by Jafari et al. (2022) was used to build the transport network input for the model. This tool relies on universal and widely available datasets for cities around the world, including Open Street Map (OSM) for the underlying road network and its attributes, Digital Elevation Model (DEM) for the road elevation, and General Transit Feed Specification (GTFS) for the public transport network and schedules. Moreover, the tool simplifies the road network to make it suitable for large-scale simulation experiments and creates a MATSim readable network that can be used for simulation.

We used the OSM extract for Greater Melbourne from October 2019 and the GTFS data⁵ from October 11 to 17, 2019, as the main datasets to create the transport network for the model. Furthermore, a minimum link length of 20 meters was assumed to simplify the network for run-time efficiency. This means that connected links (i.e. road segments) shorter than 20 meters were merged into a single node, resulting in a simpler representation of complex intersections and roundabouts, significantly reducing the simulation run time without compromising the model's accuracy (Jafari et al., 2022).

The resulting network ([Figure 2\(a\)](#)) is in the form of a set of links representing road segments and nodes representing intersections, roundabouts, or road access

Figure 2. Generated road network and PT network for the study area (base map from # OpenStreetMap).

Figure 3. A Schematic illustration of a car traveler entering traffic from the link's start node in MATSim with and without 500 m access points.

points. In MATSim, vehicles can only enter traffic from the start node of a link. This could cause considerable travel on non-existing roads for long links in MATSim, where agents must walk a significant distance to get to a start node so they can start traveling on the network using their designated mode (Figure 3(a)). Therefore, to minimize this error, we divide any large road links (greater than a threshold length of 500 meters) in areas conducive to active modes (with a speed limit of less than 60 km/h, including a footpath, and permitting both walking and cycling), into several links no longer than the threshold length (Figure 3(b)). In Melbourne, this filtration results in selecting local and residential roads where travelers can enter traffic from their driveways or parking lots, leaving out motorways and major roads where traffic can only enter at designated junctions.

2.1.2. Constructing the activity-based transport demand (demand generation tool)

The activity-based model developed by Both et al. (2021) was used to create a synthetic population of

individuals, representative of 10% of travelers in the Greater Melbourne region, as well as their daily travel diaries. This transport demand generation tool ensures that the locations of the travel destinations, the activity chains and their timing, as well as the demographic profiles of the agents, are representative of the real population at the aggregate level Statistical Area level 3 (SA3) according to the Australian Statistical Geography Standard (ASGS). The distribution of destination type location aggregated at the SA3 level in Greater Melbourne is shown in Figure 4. Interested readers are encouraged to read Both et al. (2021) for a more comprehensive analysis of activity chains and timing.

We considered five types of destinations: home, work, education, commercial, and park. The distribution of destinations in Greater Melbourne, based on the Vicmap Address database of the Victorian government⁶ which contains 2,932,530 addresses and their Mesh Block (MB) land use categories, was used as an input to the transport demand generation tool. The MB is the smallest geographical area defined by

Figure 4. Number of locations of each location type aggregated at SA3 level (base map from # OpenStreetMap).

Australian Bureau of Statistics (ABS), and residential MBs have a dwelling count of approximately 30 to 60 in urban areas.⁷

The generated synthetic population was divided into two sub-population groups, *workers* and *non-workers*, according to whether they had a trip to work or not. These sub-population groups were used during the simulation to implement different behavior change or innovation strategies for each, as explained in [Section 2.2](#).

2.1.3. Choice model estimation

We used data from VISTA 2012–16 as the main input to estimate the parameters of the utility function of the agent-based model. VISTA trip records that began and ended within the Greater Melbourne area and used one of the four modes of travel of driving, PT, walking, and cycling were selected. From the resulting set, commute trips from home to work or education (as primary destinations) were selected for further analysis, giving a sample of 15,038 out of 92,725 total

trips. The selection of mandatory commute trips to primary destinations was intended to minimize the samples affected by factors such as personal goals that are highly relevant for recreational or social trips (Ramezani et al., 2021). VISTA 2012–16 reports the origins and destinations of the trips aggregated at the SA1 level. The latitude and longitude coordinates of the SA1 centroids were considered as the coordinates of each trip origin and destination.

The selected sample was used to estimate the MATSim mode choice parameters for Melbourne. The first step was specifying the utility function for each alternative travel mode based on our model assumptions. We assumed that the effect of distance is fully captured through the travel time and cost components of the utility function, hence the marginal utility of distance was not considered for any of the four mode alternatives. Furthermore, we assumed no monetary cost for walking and cycling trips; therefore, their utility functions could be represented by [Equations 1\(c\)](#) and [1\(d\)](#), respectively.

$$S_{trav, Driving} = \frac{1}{b_m} b_{trav, Driving} t_{trav, Driving} + Dm_{Driving}, \quad (1a)$$

$$S_{trav, PT} = \frac{1}{b_m} b_{trav, PT} t_{trav, PT} + Dm_{PT}, \quad (1b)$$

$$S_{trav, Walking} = \frac{1}{b_m} b_{trav, Walking} t_{trav, Walking} + Dm_{Walking}, \quad (1c)$$

$$S_{trav, Cycling} = \frac{1}{b_m} b_{trav, Cycling} t_{trav, Cycling} + Dm_{Cycling}, \quad (1d)$$

For PT, a constant trip-based fare was used to represent the monetary cost argument Dm_{PT} of Equation 1(b). According to VISTA 2012-16, those who used PT to get to work or education reported on average two PT trips on their survey day. The daily PT pass fare for Melbourne⁸ in 2016 which was \$7.80,⁹ giving an approximated average cost of $7.8/2 = \$3.90$ per trip.

Finally, a distance-based fuel consumption cost function $Dm_{Car} = \frac{1}{b_m} b_{trav, Car} d_{trav, Car}$ was assumed for driving, where $c_{d, Car}$ is the fuel consumption cost per kilometer (km) for an average vehicle and $d_{trav, Car}$ represents the distance traveled by car (Equation 1(a)). According to Australian Transport Assessment and Planning (ATAP) guidelines for road parameter values¹⁰ for a medium car with an average journey speed of 60 km/h, the estimated fuel coefficient was equal to 11.8 lit/100 km. The average annual retail fuel price for 2012–2016 in Victoria was \$1.35 according to the Australian Institute of Petroleum data.¹¹ Therefore, $c_{d, Car}$ was calculated as:

$$c_{d, Car} = \frac{11.8 \text{ lit/100 km} \times 1.35 \text{ $/lit}}{100} = 0.1593 \text{ $/km} \quad (2)$$

The travel time for each alternative transport mode was another key component of the mode choice model to be estimated. Although the self-reported travel time for each trip is recorded in VISTA, since they are self-reported values and not actual, they are often approximations rounded to numbers easier to remember (e.g. quarters, half an hour). Furthermore, VISTA only included information for the mode that the traveler chose to use on the survey day, whereas for building a mode choice model, we needed to have travel time for all four alternative modes of travel (the one that was chosen and those not chosen by the traveler).

We used the Distance Matrix API service¹² from the Google Maps platform to estimate travel routes for the final selected VISTA trips and for all alternative transport modes.^{13,14} The Google Maps platform was selected because it incorporates congestion and

PT schedules. Therefore, it makes it possible to estimate travel times for different modes based on the current or projected road network, traffic congestion, and GTFS schedules.¹⁵

One limitation to be considered in this process is that Google uses recent traffic data to estimate travel times. Given the differences in the transport system at the time of using the Google Maps API compared to the VISTA survey day in terms of road infrastructure and traffic behavior, a deviation was expected from the actual time. Furthermore, the estimates were extracted from Google Maps API for 1 September 2022.

The parameters of the choice model were estimated using Multi-Nomial Logit model (MNL) and based on the maximum log-likelihood estimation (MLE).¹⁶ The estimated parameters for the mode choice model of Equation (1) are presented in Table 1. These parameters were then used to specify the utility function of the simulation model as discussed in the next section.

2.2. Agent-based traffic simulation

The simulation model was based on MATSim version 13.0 and the inputs from the previous steps. MATSim employs an iterative co-evolutionary optimization algorithm to determine how the supply of the road network will be utilized by the demand from the synthetic population (Horni et al., 2016). In MATSim, agents execute their plans simultaneously and travel to their destinations using a queue-based traffic mobility simulator and the road network. The link flow capacity of all network links (Section 2.1.1) was adjusted by a multiplier factor of 0.1 to create a realistic traffic flow compatible with a 10% synthetic population sample size.

Table 1. Estimated mode choice model parameters and their standard error values.

Coefficients	Estimation (s.e.)
b_m	0.19 (0.11)
asc_{PT}	-2.07 (0.61)
$asc_{Walking}$	0.39 (0.17)
$asc_{Cycling}$	-3.23 (0.21)
$b_{trav, Driving}$	-11.48 (1.26)
$b_{trav, PT}$	-12.02 (2.00)
$b_{trav, Walking}$	-11.99 (1.08)
$b_{trav, Cycling}$	-14.75 (2.40)
# estimated parameters	8.00
Number of respondents	15038.00
Number of choice observations	15038.00
McFadden R ²	0.81
AIC	6717.48
AICc	6717.49
BIC	6778.42

$p < .01$;

$p < .05$;

$p < .1$

Driving, PT, cycling, and walking are the four travel modes included in this study. Driving, cycling, and walking were explicitly modeled on the road network, meaning that travelers using these modes navigated the dedicated road network and the traffic dynamics in each road segment (i.e. a network link) was determined by the queue-based road traffic simulator. We used the enhanced First-In-First-Out queue model proposed by Agarwal et al. (2015), where faster vehicles can overtake slower ones. Walking and cycling were configured to not block cars in the queue model. The speed of cars traveling on a link was determined based on the road's speed limit and the maximum speed possible given the traffic flow determined by the queue-based road traffic simulator model. For walking and cycling, as they were not restricted by or impacting the queue model, a constant traveling speed of 1.7 and 5.5 meters per second was set, respectively. PT vehicle movements were simulated using the deterministic Public Transport Simulation (detPTSim) engine proposed by Rieser et al. (2018). In detPTSim, PT vehicles operate according to a strict transit schedule, disregarding the queue network and traffic congestion.

While the use of detPTSim results in a more realistic representation of railway transport (e.g. trains), it may have potential drawbacks for PT vehicles using shared infrastructure with cars (e.g. buses). Walking was the only travel mode considered for access/egress trip legs. The potential start and end stops for each PT trip leg were filtered to those within a certain radius of the trip leg's origin and destination. The initial value of this search radius was set to 1 km. If fewer than two stops were found within this radius, it was increased by another 1 km until at least two stops were found or a maximum radius of 10 km was reached. It should be noted that the search radius of 1 km does not mean that agents travel 1 km to reach their desired PT stop; rather, agents consider all stops within this radius as potential candidates and select the best one based on various factors, including the

amount of walking required and the transit lines servicing each stop.

The estimated mode choice parameters listed in Table 1 were used to construct the utility function, which MATSim used during the simulation to assign a score to all executed travel plans in each simulation iteration. Following Horni et al. (2016), the marginal utility of performing an activity was set equal to the marginal utility of travel time by car, and the marginal utilities of waiting for PT and late arrival were set to be twice and three times this amount, respectively. The marginal utility of travel time by car was set to zero, and the marginal utilities of travel time for other modes were adjusted accordingly. The resulting values are listed in Table 2. It should be noted that the impact of the change in travel time by car is still implicitly included in the simulation model. If an agent is late or is unable to perform an activity, it will receive a negative utility relative to the marginal utility of late arrival and the marginal utility of performing an activity, respectively. These values were used for the initial simulation run, however, as explained later in Section 3, mode specific constant values were further calibrated through a number of experiments to improve how well the simulated mode share matched real-world expected values.

In MATSim, after each iteration, agents compare the score of the executed plan with a limited number of previously executed plans and select the travel plan for the next iteration. We used a memory size of 5 for the highest scored experienced plans. Before the next iteration begins, a given percentage of agents modify their chosen plan following different innovation strategies, such as randomly varying departure times, travel modes, and routes. In this model, we considered only two innovation strategies: *re-routing* and *sub-tour mode choice*. The underlying assumption is that by repeating the process for a sufficient number of iterations, agents will choose routes and travel modes that maximize their score (Equation (1)) while considering the choices made by other agents. Therefore, considering that constant speeds for walking and cycling were applied and that travel time was the only variable included in the

Table 2. Simulation model utility function parameters.

Model parameters	Value			
Generic parameters				
Marginal utility of money (utils/\$)	0.1944			
Marginal utility of performing activity (utils/hour)	11.481			
Marginal utility of late arrival (utils/hour)	−34.441			
Mode specific parameters	Driving	PT	Walking	Cycling
Alternative (mode) specific constant	0.0	−2.067	0.391	−3.234
Marginal Utility of time spent traveling (utils/hour)	0.0	−0.535	−0.506	−3.267
Monetary distance rate (\$/km)	−0.163	−	−	−
Daily monetary cost of using PT (\$/Day)	−	−7.8	−	−
Marginal Utility of waiting at PT station (utils/hour)	−	−22.961	−	−

utility function for these two travel modes in Equation (1), it can be inferred that through the iterative re-routing process, pedestrian and cyclist agents will take the shortest network path.

Both the *workers* and *non-workers* sub-population groups had the MATSim innovation strategy for route choice (*re-routing strategy*) enabled. The *sub-tour mode choice strategy* was enabled only for the workers sub-population, allowing them to change their trip leg modes and to find the one that works best for them. The four main modes (i.e. driving, PT, walking, and cycling) were available for all worker agents to choose from. Cycling and driving were simulated as *tour-modes*, meaning that for an agent to have driving/cycling in one of its trip legs, it must start the trip tour from home with a car/bike and must return to home with a car/bike as well.

No innovation strategy for activity type, location, or timing selection was included as the transport demand generation tool generates trips in a way to match what is expected from real-world data (see (Both et al., 2021) for a discussion on activity type, location, and timing calibration). If an agent did not adopt either of the two innovation strategies, the agent was set to change its plan to another previously experienced plan from its memory with a probability of e^{DScore} , where $DScore$ is the difference in scores between the two plans. This strategy, known as the *ChangeExpBeta* strategy, was selected to encourage agents to seek plans that yield globally optimal scores. A more in-depth discussion of this innovation strategy is presented by Nagel and Flötterød (2016). The weighting of each of the innovation strategies for each subgroup is listed in Table 3.

The simulation model was initially run for 200 iterations, followed by a series of 100 iteration runs for model calibration. Innovation strategies (re-routing and sub-tour mode choice) were disabled for the last 40 iterations (20%) to allow the model to converge to a stable solution (net score).

3. Simulation output analysis

This section analyzes and compares the simulation output with real-world observations to better understand

Table 3. Re-planning innovation strategy weights for different sub-populations.

Strategy	Weight	
	Workers	Non-workers
ChangeExpBeta	0.8	0.9
Re-routing	0.1	0.1
Sub-tour mode choice	0.1	0.0

the accuracy and reliability of our model. Four main measures were analyzed: mode share (Section 3.1), road traffic volume (Section 3.2), public transport usage (Section 3.3), and travel distance and time (Section 3.4).

3.1. Mode share analysis

As explained in Section 2.2, the MATSim sub-tour mode choice strategy was enabled for the *workers* sub-population. To examine and calibrate the mode choice model for this sub-population, we compared simulated trips to work with the VISTA 2016-18 survey data commute to work trips (Department of Transport, 2018). We then followed an iterative process for manual calibration of the mode choice functionality of the model. Initially, we ran the simulation model with the parameters listed in Table 2 for 200 iterations, allowing agents to find their optimal travel mode and route given these estimated parameters. We then compared the mode share of the simulation output to the expected real-world values and adjusted the model's mode-specific constants to achieve a better match. Next, we ran another simulation experiment for 100 iterations with the new adjusted parameters, using the already optimized plans from the previous run as its input. This iterative process of running the simulation, comparing the mode shares from the simulation results, and adjusting the parameters accordingly was repeated until a reasonable fit was achieved. We considered the mode shares of the simulation output for trips to work to be within the $\pm 1\%$ error threshold of the observation data as our calibration target. The adjusted value of the mode-specific constants and the final simulation mode shares for trips to work are listed in Tables 4 and 5, respectively.

Table 4. Adjusted mode-specific constants as a result of the mode share calibration.

	Driving	PT	Walking	Cycling
Adjusted mode specific constants	0.0	4.667	2.891	1.201

Table 5. Mode share comparison between calibrated simulation output and VISTA 2016–18.

	Mode share (%)	
	Simulation	VISTA 2016–18
Trips to work		
Driving	73.6	73.3
PT	15.8	15.6
Walking	8.7	9.2
Cycling	1.9	1.9
Trips to non-work destinations		
Driving	71.5	67.1
PT	10.8	8.1
Walking	16.5	22.5
Cycling	1.2	2.2

We also compared the mode share of non-work trips with real-world data to examine whether enabling the sub-tour mode choice strategy for the workers sub-population was sufficient or if a mode choice strategy for both sub-population groups was needed. To do this, we compared the mode share in all non-work trips from the calibrated simulation output with the share of these travel modes in VISTA 2016–18 non-work trips. Table 5 provides mode shares for the mode choice calibrated simulation model and VISTA travel survey data 2016–18 for work and non-work trips.

3.2. Road traffic volume analysis

The daily volume of driving, cycling, and walking traffic from the calibrated simulation output is illustrated in Figure 5.

Publicly available traffic count data for Melbourne was used to examine the road usage accuracy of the model for driving. We used the Typical Hourly Traffic Volumes (THTV) data from Victoria's open data platform for 2019,¹⁷ which provides typical traffic volumes for major arterial roads across Victoria. THTV was filtered down to the data for school term normal mid-week days. The data was then divided into two categories of roads: those going toward Melbourne CBD (handling most of the AM peak traffic) and those going outward from the CBD (handling PM peak traffic). In each category (that is, AM and PM), the top 10% highest traffic roads were identified, and within each, the road segment with the highest volume was selected for comparison with the simulation output. This resulted in the selection of 87 road segments, 47 for the AM peak hours and 40 for the PM peak hours, for further analysis.

The selected road segments were joined to their equivalent links in the simulation road network. For this purpose, the 'equivalent' link was selected as the link located closest to the midpoint of the road segment that satisfied the conditions of operating in the correct direction to match the road and having a bearing (or azimuth) within 17.5 degrees of the bearing of the road segment.

To compare cycling traffic volume, we used the average daily cycling volume of the automatic cycling volume and speed sensors for Greater Melbourne for weekdays, downloaded from the Victoria Open Data Platform for the period of March 2019.¹⁸ Each sensor was joined to its equivalent link in the simulation road network, selecting the closest link that was a bicycle path or a road with a bicycle lane and operated in the correct direction. In total, 48 counting sensors (some mono-

directional and others bi-directional), corresponding to 70 network links, were selected for further analysis.

For the volume of walking traffic, we used automated pedestrian count data from sensors across Melbourne's central LGA, i.e. City of Melbourne, which encompasses the Central Business District and its surroundings (City of Melbourne, 2021). Data were downloaded for mid-week workdays in March 2019 and joined to the simulation road network by selecting the closest links having similar bearings to the relevant footpaths, in a similar way as described above for driving volumes. Given that the footpaths are bi-directional, the aggregated number of pedestrians passing each sensor, regardless of the walking direction, was used for comparison. Furthermore, for streets with more than one footpath, the aggregated volume from all associated footpaths was used. This resulted in the selection of 48 sensors corresponding to 93 network links for further analysis.

The percentage of daily traffic volume for each hour of the day, h , and for each road segment, r , was calculated for the calibrated simulation output, $s_{r,h}^0$ and the observation data from THTV, $s_{r,h}$, using Equation (3).

$$s_{r,h} = \frac{v_{r,h}}{\sum_{r=1}^N v_{r,h}} \quad (3)$$

Here, N is the total number of road segments analyzed and $v_{r,h}$ is the traffic volume on road r during the hour h . Figure 6(a) depicts the average traffic volume percentage of the daily traffic for every hour of the day on all selected $N=87$ road segments. We then used Weighted Absolute Percentage Error (WAPE) to compare the hourly road traffic volume percentages in the observation data and the simulation results (Equation (4)).

$$WAPE_h = \frac{\sum_{r=1}^N |s_{r,h} - s_{r,h}^0|}{\sum_{r=1}^N s_{r,h}} \quad (4)$$

As shown in Figures 7(a) and 6(a), the simulation model performs well in capturing the volume of car traffic during peak hours in Melbourne, with a WAPE less than 25%. Some potential reasons for the deviations in car traffic volume in the early morning and late evening include a lack of inclusion of freight traffic, travelers from outside the Greater Melbourne area, and airport passengers in the current version of the model. Their absence is likely to be more noticeable during off-peak hours when the roads are not already congested with local commuters. A similar trend for walking was also observed, as shown in Figure 7(c). However, for cycling, the percentage error of traffic volume was high throughout the day and considerably higher during off-peak hours (Figures 6(b) and 7(b)).

Figure 5. Simulation output aggregated daily traffic volume for different modes (base map from # OpenStreetMap).

As discussed in detail in [Section 4](#), this was probably due to the lack of inclusion of the impact of cycling-relevant road infrastructure, such as bikeway type or slope, on cycling route choice behaviors.

3.3. Public transport usage analysis

To validate the use of public transport in the model, we compared the real-world percentage of passenger

flow aggregated at the LGA level with the results of our simulation. To calculate this percentage, we aggregated the passenger flow of 218 train stations across the Greater Melbourne metropolitan area based on the LGA in which they were located. Then, we calculated the aggregated share of the passenger flow of each LGA relative to the total number of passengers in Greater Melbourne for trips by train for further comparison. We used the 2016 station access survey

Figure 6. Aggregated hourly traffic volume percentages in simulation versus observation for different travel modes.

data from Public Transport Victoria for this comparison. This dataset was obtained from Victoria's Department of Transport—Public Transport Victoria. Figure 8 shows the comparison of the real-world observations and our simulation output, indicating that the simulation model was able to capture the PT passenger flow distribution throughout Greater Melbourne.

3.4. Travel distance and time analysis

The mode share and road traffic volume analyses evaluated the model at an aggregated level, that is, aggregated to the travel modes or road segments. We investigated the accuracy of the travel distance and time for a small sample of trips to examine the accuracy of the model at the individual trip level. To do this, we randomly sampled a subset of 1,000 trips stratified by origin SA3 and travel mode from the simulated trips. We extracted the experienced travel distance and time for the sample trips from the simulation output.

The simulated travel time for driving incorporated the impact of road congestion in addition to speed limits and the vehicle's maximum speed. This is due to the MATSim queue model to capture road traffic for the cars. Although walking and cycling trips were based on the road network, they were not set to impact or be affected by road traffic. Therefore, travel

time and distance for walking and cycling reflected the network distance between the origin and destination and their constant speeds, $1.7m/s$ for walking and $5.5m/s$ for cycling. PT travel time was based on the transit schedules extracted from GTFS and the agent's decision on which PT service to use.

We used the Google Distance API to estimate the expected travel distance and time for the sampled trips. For driving and PT, the Google Distance API estimates the travel distance and time based on its historical records, taking into account traffic and network conditions. For cycling and walking, only the fastest route is assumed. Another limitation of using the Google Distance API is that it does not provide estimates for a past trip. Therefore, we estimated the travel distances and time of a sample of trips based on Google data for October 2021.

Figure 9 illustrates the percentage error of the travel distance and travel time for different modes for the 1,000 trips sampled. For example, we calculated the percentage error of travel distance for a sample trip j and mode m , $d_{m,j}^d$, as follows:

$$d_{m,j}^d = 100 \frac{d_{m,j}^s - d_{m,j}^d}{d_{m,j}^d}, \quad (5)$$

where $d_{m,j}^s$ is the experienced travel distance from the simulation for mode m and trip j and $d_{m,j}^d$ is the expected travel distance from Google Distance API.

Figure 7. Weighted absolute percentage error of aggregated hourly traffic volume percentages in simulation versus observation for different travel modes.

4. Discussion and conclusion

In this paper, we presented our model development workflow to build an open multi-modal transport model for the Greater Melbourne area. In doing this, and using Melbourne as our example scenario, we illustrated how a series of open-source tools, including an activity-based transport demand generation tool (Both et al., 2021), a multi-modal network generation tool (Jafari et al., 2022), as well as tools for model output processing and mode choice model parameter estimation, can be used to build a city-wide agent-based and activity-based simulation model. All tools used in our workflow are publicly available in our GitHub repository. Furthermore, these tools were designed to use data sources commonly available for different cities around the world (e.g., travel surveys, traffic counts, OSM, and GTFS). This means that although the data format used in this paper might be specific to Melbourne, the same workflow can be applied to other cities, provided the data structure is compatible with each tool's expected structure. We used two existing standalone tools for demand generation¹⁹ and network generation²⁰ to build the inputs for the agent-based simulation model. Similarly, our algorithm for mode choice model estimation²¹ also is suitable for more general use outside of MATSim.

Over the years, several different Java-based tools have been developed to provide a complete workflow

for transport modeling using MATSim. These include tools to convert raw OSM extracts into MATSim readable transport networks for car traffic (Zilske et al., 2015) and bicycle traffic Ziemke, Metzler, et al. (2019), a tool to add PT routes to the MATSim network (Poletti, 2016) and a set of tools for output analysis included in MATSim's analysis extension.²² In the workflow presented in this paper, MATSim was used as the core agent-based traffic simulator of the model, which is a Java-based program. However, the tasks of input preparation and output processing were performed using tools external to MATSim, written in the R programming language. Given that the R programming language is becoming increasingly more popular among the research community (Bishwal, 2017), a model development workflow based on R provides more flexibility for researchers to use tools that are best suited for their projects and teams when building activity-based and agent-based models. Additionally, an extensive number of powerful spatial and non-spatial data analysis packages exist for R that make it possible for researchers to extend the tools to fit their purposes.

For the Melbourne model, we calibrated the mode choice behavior of work trips for four travel modes of driving, PT, walking, and cycling against the VISTA 2016–18 data (Table 5). To achieve this, we developed a number of open-source tools for simulation output

Figure 8. Passenger flow percentage comparison at the LGA level in real-world and simulation outputs.

analysis, which can be useful to analyze the simulation output of a MATSim model for other cities.²³ Through these analyses, we demonstrated that the simulated mode share percentage for non-work trips closely resembled the figures observed in the travel survey. The mode choice calibrated simulation model could be used as the baseline for examining the potential for mode shift as a result of the built environment, infrastructure and/or monetary interventions that impact travel time and cost, such as constructing new roads, building a new bridge for pedestrians and cyclists to decrease distances, increasing PT services to existing stations or adding new stations, and changing PT fares or motor vehicle fuel prices. For interventions that encourage mode shift for a specific demographic group or are based on built environment attributes that were not included in the mode choice model of the paper, such as the existence of bike lanes for cycling, the mode choice model in Equation (1) needs to be extended accordingly to incorporate these factors.

In addition to mode choice, car traffic volumes and PT passenger flow at the LGA level from the

simulation model output also resembled the volumes observed in the real world (see [Figures 6\(a\)](#) and [8](#), respectively). [Figure 9](#) shows that in addition to the road level and the aggregate level, the results of the model also reflect the expected behavior in terms of travel time and distance at the trip level. The realistic road traffic behavior of the model makes it suitable for examining various traffic management interventions, such as modifying speed limits or blocking certain roads to guide the traffic flow. For example, a snapshot of car traffic on roads within a 10-km radius of Melbourne CBD for 9 AM and 5 PM is illustrated in [Figure 10](#), depicting the heavy congestion on major roads connecting Melbourne CBD to the rest of the metropolitan area. Using agent-based models, it is possible to go beyond high-level snapshots and examine road usage at the individual level. For instance, one of the road segments with heavy congestion both in AM and PM peak hours is the West Gate Bridge, Victoria's most heavily used bridge, which is responsible for connecting the Melbourne CBD to the western suburbs. [Figure 11\(a\)](#) illustrates where vehicles

Figure 9. Percentage error of travel (a) distance and (b) time between simulation output and google distance matrix API estimates for sampled trips.

Figure 10. Snapshots of the simulated car traffic at (a) morning peak (9:00 AM) and (b) evening peak (5:00 PM) for inner melbourne. The colors represent the relative speed with red $\frac{1}{2}$ full stop, yellow $\frac{1}{2}$ traveling speed equal to half of the speed limit, green $\frac{1}{2}$ traveling speed equal to the speed limit.

that used this road segment at 9 AM are coming from and heading to, confirming the critical role that the bridge plays in connecting the western suburbs with the rest of the center and to the east. Furthermore, the travel route of an example agent who used this bridge at 9 AM is also highlighted in [Figure 11\(b\)](#).

All categories of roads accessible to the public were included in the road network of the model, including minor bike paths to local streets and to major arterial roads and highways. Therefore, in addition to common measures such as zone-to-zone movements or traffic on major highways and corridors, our model can be used for exploring local road usage for accessing local destinations. Further calibration is required

to obtain reliable estimates of local road usage for active modes of transport from the model.

In this paper, we considered the shortest network path for cycling and walking routing and only included walking and cycling travel time in the mode choice model of the agent-based model. Distance traveled is one of the key factors important for active travel route choice (Fitch & Handy, 2020); however, numerous studies have shown that cyclists and pedestrians deviate from the shortest route to use routes with better infrastructure (Lu et al., 2018; Lue & Miller, 2019; Rupi et al., 2019). There are several other built-environment factors such as existence of safe infrastructure including bike lanes, footpaths, traffic

Figure 11. West Gate Bridge 9:00 AM snapshot of (a) where the agents using it are coming from and traveling to and (b) simulated travel route of an example agent using the bridge.

signals, and the quietness of the route (Desjardins et al., 2021; Lu et al., 2018; Lue & Miller, 2019), as well as individual factors such as age, gender, and household structure (Aldred et al., 2017; Basu et al., 2021) that are important and should be included in future models to have a more accurate active travel behavior model. Although the impact of these factors on routing was not included in the simulation model described in this article, our transport demand and network generation steps did include several attributes necessary for future modeling of active transport. The input road network includes road categories accessible to pedestrians and cyclists, as well as road slope and bikeway type. Similarly, activity-based transport demand includes individual attributes such as age, sex, occupation, and household structure, which are all important factors for walking and cycling behavior. Including these factors in the input of the model makes it possible for the model to be further extended in the future to better model active travel behavior. The comparatively high inaccuracies observed in the analysis of cycling and walking road use in Figure 7(b) and (c) when compared to the driving analysis are most likely due to the lack of including these factors in the model.

A key consideration for the tools that were used or developed for the workflow of this paper was to rely on the data types and structures commonly found in different cities. Therefore, this makes it possible to extend the workflow and add different impact assessment models as well. For example, the simulation output analysis tools provided in our workflow convert the simulation output into formats (e.g., hourly traffic counts joined to the network, individual travel diary,

minutes spent walking) that are straightforward to join to existing impact assessment tools such as the Transport Health Assessment Tool for Melbourne (THAT-Melbourne) Gunn et al. (2021); Zapata-Diomedes et al. (2021) to estimate non-communicable disease impacts (Veerman et al., 2016; Zapata-Diomedes et al., 2019) and also to include environmental impacts assessment tools such as air quality (Woodcock et al., 2021) arising from change in travel behavior occurring between scenarios. The extended model could be used to examine the health, economic, and environmental impacts of possible changes in travel behavior.

An important limitation of our model is that the mode choice parameters are estimated based on mandatory work and education trips, and mode choice is only permitted for workers. Further research is required to create separate mode choice models based on the purpose of the trip (e.g. mandatory versus discretionary). Although the 2016 Census data were our starting point for creating the synthetic population of the model, data from different years were used to develop and calibrate the model and its inputs. This was to address the challenge of data availability for city-scale agent-based and activity-based models that require large data at the individual level. However, the use of data from different years for model development and calibration makes it not possible to compare the absolute numbers using the simulation model. To address this, we used relative numbers when comparing the output of the model with the real-world observations in Section 3. Consequently, when using the simulation model to test the impact of an intervention on mode share or traffic volume, only relative

numbers must be reported. Additionally, the calibration would have been strengthened if the prediction accuracy of the model had been examined using a historical intervention; and this should be considered for future research if appropriate intervention data can be sourced. Therefore, caution must be exercised when interpreting changes in mode share as a result of interventions and selecting the types of intervention to test using the model.

Finally, PT trips were simulated based on deterministic timings of GTFS and direct dedicated links connecting PT stops (Figure 2(a)). As a result, PT vehicles had no interaction with other modes of transport while traveling and were strictly always on time. Of course, in reality, tram and bus routes typically share the road with cars and are delayed due to traffic congestion or contribute to traffic congestion by occupying a significant amount of mixed-traffic road space. The current state of the model does not capture either of these two scenarios, which again warrants future research.

Notes

1. We note that the first calibrated activity-based MATSim model for Melbourne was MABM (Infrastructure Victoria, 2017). Ours is the first calibrated multi-modal model, which is also *open*.
2. <https://github.com/matsim-melbourne>
3. <https://www.abs.gov.au/>
4. For a more detailed discussion on MATSim and its inputs, see (Horni et al., 2016).
5. GTFS data were downloaded from: <https://transitfeeds.com/p/ptv>
6. <https://discover.data.vic.gov.au/dataset/vicmap-address>
7. <https://www.abs.gov.au/>
8. Public transport fares in Melbourne vary based on the zones the person travels within or between, and whether the traveler has a daily, monthly, or even yearly PT travel pass or pays for each trip individually. For simplicity, we assumed that PT travelers used the standard daily pass (zone1 to zone2).
9. Monetary units in this paper are in Australian Dollars
10. https://www.atap.gov.au/sites/default/files/pv2_road_parameter_values.pdf
11. <https://aip.com.au/aip-annual-retail-price-data>
12. <https://developers.google.com/maps/documentation/distance-matrix/intro>
13. Google Distance Matrix API is a paid service, not an open data source. However, our scripts for preparing the data for the API, sending queries to the API, and processing the results are publicly available
14. URL removed to ensure anonymity. Furthermore, we implemented an alternative option to use OpenRouteService (ORS) to calculate travel time and distance (<https://openrouteservice.org/>). Although ORS is open and free to use, it does not cover PT schedules and congestion.
15. The R package **gmapsdistance (version 3.4)** was used to extract travel times and distances from Google Distance Matrix API.
16. The **mixl** package (version 3.4) in R was used for parameter estimation (Molloy et al., 2019).
17. <https://data.vicroads.vic.gov.au/Metadata/Typical/20Hourly/20Traffic/20Volumes.html>, accessed on 14/05/2021
18. <https://discover.data.vic.gov.au/dataset/bicycle-volume-and-speed>
19. <https://github.com/matsim-melbourne/demand>
20. <https://github.com/matsim-melbourne/network>
21. <https://github.com/matsim-melbourne/choice-model>
22. <https://github.com/matsim-org/matsim-libs/tree/master/contribs/analysis/src/main/java/org/matsim/contrib/analysis>.
23. <https://github.com/matsim-melbourne/useful-scripts>

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

AJ is supported by an Australian Government Research Training Program Scholarship. DS's time on this project is funded by Collaborative Research Project grants from CSIRO's Data61 (2018-19, 2020-21). AB is supported by the NHMRC/UKRI JIBE project (#APP1192788). LG, MA, and SP are supported by the NHMRC-funded Australian Prevention Partnership Centre (#9100001); and BGC is supported by an RMIT VC Professorial Fellowship. The authors acknowledge the use of OpenAI's GPT-4 to assist with proofreading the final version of this manuscript.

ORCID

Afshin Jafari <http://orcid.org/0000-0001-5153-2179>
 Dharendra Singh <http://orcid.org/0000-0001-8369-8878>
 Alan Both <http://orcid.org/0000-0003-4021-9632>
 Mahsa Abdollahyar <http://orcid.org/0009-0002-1222-9833>
 Lucy Gunn <http://orcid.org/0000-0002-9309-3626>
 Steve Pemberton <http://orcid.org/0000-0003-2496-1968>
 Billie Giles-Corti <http://orcid.org/0000-0003-0102-0225>

References

- Adnan, M., Pereira, F. C., Azevedo, C. M. L., Basak, K., Lovric, M., Raveau, S., Zhu, Y., Ferreira, J., Zegras, C., & Ben-Akiva, M. (2016). Simmobility: A multi-scale integrated agent-based simulation platform. In *95th Annual Meeting of the Transportation Research Board Forthcoming in Transportation Research Record*. The National Academies of Sciences, Engineering, and Medicine Washington, DC.
- Agarwal, A., Zilske, M., Rao, K. R., & Nagel, K. (2015). An elegant and computationally efficient approach for heterogeneous traffic modelling using agent based simulation. *Procedia Computer Science*, 52, 962–967. <https://doi.org/10.1016/j.procs.2015.05.173>
- Aldred, R., Elliott, B., Woodcock, J., & Goodman, A. (2017). Cycling provision separated from motor traffic: A systematic review exploring whether stated preferences vary by gender and age. *Transport Reviews*, 37(1), 29–55. <https://doi.org/10.1080/01441647.2016.1200156>
- Allahviranloo, M., Regue, R., & Recker, W. (2017). Modeling the activity profiles of a population. *Transportmetrica B: Transport Dynamics*, 5(4), 426–449. <https://doi.org/10.1080/21680566.2016.1241960>
- Arentze, T., & Timmermans, H. (2000). *Albatross: A learning based transportation oriented simulation system*. Citeseer.
- Basu, N., Haque, M. M., King, M., Kamruzzaman, M., & Oviedo-Trespalacios, O. (2021). A systematic review of the factors associated with pedestrian route choice. *Transport Reviews*, 42(5)pages, 672–694. <https://doi.org/10.1080/01441647.2021.2000064>
- Becker, H., Balac, M., Ciari, F., & Axhausen, K. W. (2020). Assessing the welfare impacts of Shared Mobility and Mobility as a Service (MaaS). *Transportation Research Part A: Policy and Practice*, 131, 228–243. <https://doi.org/10.1016/j.tra.2019.09.027>
- Bishwal, R. (2017). *Potential use of r-statistical programming in the field of geoscience* [Paper presentation]. 2017 2nd International Conference for Convergence in Technology (I2CT), 979–982. IEEE. <https://doi.org/10.1109/I2CT.2017.8226275>
- Bosch, P. M., Müller, K., & Ciari, F. (2016). The ivt 2015 baseline scenario. In *16th Swiss Transport Research Conference (STRC 2016)* 16th Swiss Transport Research Conference (STRC 2016).
- Both, A., Singh, D., Jafari, A., Giles-Corti, B., Gunn, L. (2021). An activity-based model of transport demand for greater melbourne. *arXiv preprint arXiv:2111.10061*. City of Melbourne (2021). *Pedestrian Counting System*. Department of Transport (2018). *Victorian Integrated Survey of Travel and Activity (VISTA)*.
- Desjardins, E., Higgins, C. D., Scott, D. M., Apatu, E., & Paez, A. (2021). Correlates of bicycling trip flows in hamilton, ontario: Fastest, quietest, or balanced routes? *Transportation*, 49(3), 867–895. <https://doi.org/10.1007/s11116-021-10197-1>
- Erath, A., Fourie, P. J., van Eggermond, M. A., Ordonez Medina, S. A., Chakirov, A., & Axhausen, K. W. (2012). Large-scale agent-based transport demand model for singapore. *Arbeitsberichte Verkehrs-und Raumplanung*, 790.
- Fitch, D. T., & Handy, S. L. (2020). Road environments and bicyclist route choice: The cases of davis and san francisco, ca. *Journal of Transport Geography*, 85, 102705. <https://doi.org/10.1016/j.jtrangeo.2020.102705>
- Gilbert, N. (2021). *Agent-based models*. Thousand Oaks.
- Gunn, L., Davern, M., Zapata, D. B., Both, A., Kroen, A., & De, G. C. (2021). Helping planners understand health benefits through the Transport Health Assessment Tool for Melbourne (THAT-Melbourne). *Planning News*, 47(5):23.
- Hadjidimitriou, N. S., Cantelmo, G., & Antoniou, C. (2022). Machine learning for activity pattern detection. *Journal of Intelligent Transportation Systems*, 27(6), 834–848. <https://doi.org/10.1080/15472450.2022.2084336>
- Hesam Hafezi, M., Sultana Daisy, N., Millward, H., & Liu, L. (2021). Framework for development of the scheduler for activities, locations, and travel (salt) model. *Transportmetrica A: Transport Science*, 18(2), 248–280. <https://doi.org/10.1080/23249935.2021.1921879>
- Hörl, S., & Balac, M. (2020). *Open data travel demand synthesis for agent-based transport simulation: A case study of paris and ile-de-france* (pp. 1499). *Arbeitsberichte Verkehrsund Raumplanung*.
- Hörl, S., & Balac, M. (2021). Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. *Transportation Research Part C: Emerging Technologies*, 130, 103291.
- Horni, A., Nagel, K., & Axhausen, K. W. (2016). *The multi-agent transport simulation MATSim*. Ubiquity Press London.
- Hu, W., Winter, S., & Khoshelham, K. (2022). Decentralized spreading of ephemeral road incident information between vehicles. *Journal of Intelligent Transportation Systems*, 28(1), 16–30. <https://doi.org/10.1080/15472450.2022.2095206>
- Infrastructure Victoria (2017). *Model Calibration and Validation Report*. Technical report, Infrastructure Victoria.
- Jafari, A., Both, A., Singh, D., Gunn, L., & Giles-Corti, B. (2022). Building the road network for city-scale active transport simulation models. *Simulation Modelling Practice and Theory*, 114, 102398. <https://doi.org/10.1016/j.simpat.2021.102398>
- Kagho, G. O., Balac, M., & Axhausen, K. W. (2020). Agent-based models in transport planning: Current state, issues, and expectations. *Procedia Computer Science*, 170, 726–732. <https://doi.org/10.1016/j.procs.2020.03.164>
- Koushik, A. N., Manoj, M., & Nezamuddin, N. (2020). Machine learning applications in activity-travel behaviour research: A review. *Transport Reviews*, 40(3), 288–311. <https://doi.org/10.1080/01441647.2019.1704307>
- Lu, W., Scott, D. M., & Dalumpines, R. (2018). Understanding bike share cyclist route choice using gps

- data: Comparing dominant routes and shortest paths. *Journal of Transport Geography*, 71, 172–181. <https://doi.org/10.1016/j.jtrangeo.2018.07.012>
- Lue, G., & Miller, E. J. (2019). Estimating a Toronto pedestrian route choice model using smartphone GPS data. *Travel Behaviour and Society*, 14, 34–42. <https://doi.org/10.1016/j.tbs.2018.09.008>
- McNally, M. G., & Rindt, C. R. (2007). *The activity-based approach*. Emerald Group Publishing Limited.
- Molloy, J., Schmid, B., Becker, F., & Axhausen, K. W. (2019). mixl: An open-source R package for estimating complex choice models on large datasets. *Arbeitsberichte Verkehrs-und Raumplanung*, 1408. IVT, ETH Zurich.
- Müller, J., Straub, M., Naqvi, A., Richter, G., Peer, S., & Rudloff, C. (2021). *Matsim model vienna: Analyzing the socioeconomic impacts for different fleet sizes and pricing schemes of shared autonomous electric vehicles* [Paper presentation]. Transportation Research Board 100th Annual Meeting 2021, In.
- Nagel, K., & Flötterød, G. (2016). Agent-based traffic assignment. In *The multi-agent transport simulation MATSim* (pp. 316–326). Ubiquity Press London.
- Poletti, F. (2016). *Public transit mapping on multi-modal networks in MATSim* [Master thesis]. IVT, ETH Zurich, Zurich. <https://ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/publications/students/501-600/sa530.pdf>
- Ramezani, S., Laatikainen, T., Hasanzadeh, K., & Kytta, M. (2021). Shopping trip mode choice of older adults: An application of activity space and hybrid choice models in understanding the effects of built environment and personal goals. *Transportation*, 48(2), 505–536. <https://doi.org/10.1007/s11116-019-10065-z>
- Rasouli, S., & Timmermans, H. (2014). Activity-based models of travel demand: Promises, progress and prospects. *International Journal of Urban Sciences*, 18(1), 31–60. <https://doi.org/10.1080/12265934.2013.835118>
- Rieser, M. (2016). *Modeling public transport with matsim*. Ubiquity Press.
- Rieser, M., Metrailler, D., & Lieberherr, J. (2018). Adding realism and efficiency to public transportation in MATSim. In *Proceedings of the 18th Swiss Transport Research Conference*, Ascona, Switzerland (pp. 16–18).
- Roorda, M. J., Miller, E. J., & Habib, K. M. N. (2008). Validation of TASHA: A 24-h activity scheduling micro-simulation model. *Transportation Research Part A: Policy and Practice*, 42(2), 360–375. <https://doi.org/10.1016/j.tra.2007.10.004>
- Rupi, F., Poliziani, C., & Schweizer, J. (2019). Data-driven bicycle network analysis based on traditional counting methods and gps traces from smartphone. *ISPRS International Journal of Geo-Information*, 8(8), 322. <https://doi.org/10.3390/ijgi8080322>
- Tajaddini, A., Rose, G., Kockelman, K. M., & Vu, H. L. (2020). Recent progress in activity-based travel demand modeling: Rising data and applicability. In S. D. Luca, R. D. Pace, & C. Fiori (Eds.), *Models and technologies for smart, sustainable and safe transportation systems*. <https://www.intechopen.com/chapters/73240>
- Veerman, J. L., Zapata-Diomed, B., Gunn, L., McCormack, G. R., Cobiac, L. J., Mantilla Herrera, A. M., Giles-Corti, B., & Shiell, A. (2016). Cost-effectiveness of investing in sidewalks as a means of increasing physical activity: A RESIDE modelling study. *BMJ Open*, 6(9), e011617. <https://doi.org/10.1136/bmjopen-2016-011617>
- Wall, F. (2016). Agent-based modeling in managerial science: An illustrative survey and study. *Review of Managerial Science*, 10(1), 135–193. <https://doi.org/10.1007/s11846-014-0139-3>
- Wang, K., Zhang, W., Mortveit, H., & Swarup, S. (2021). Improved travel demand modeling with synthetic populations. In Swarup, S., & Savarimuthu, B. T. R. (Eds.), *Multi-agent-based simulation XXI*. (pp. 94–105). Springer International Publishing.
- Woodcock, J., Aldred, R., Lovelace, R., Strain, T., & Goodman, A. (2021). Health, environmental and distributional impacts of cycling uptake: The model underlying the Propensity to Cycle tool for England and Wales. *Journal of Transport & Health*, 22, 101066. <https://doi.org/10.1016/j.jth.2021.101066>
- Zapata-Diomed, B., Both, A., Kroen, A., De Gruyter, C., Davern, M., & Gunn, L. (2021). *Transport health assessment tool for Melbourne: Transport scenario report*. Technical report. RMIT University, Centre for Urban Research.
- Zapata-Diomed, B., Boulange, C., Giles-Corti, B., Phelan, K., Washington, S., Veerman, J. L., & Gunn, L. D. (2019). Physical activity-related health and economic benefits of building walkable neighbourhoods: A modelled comparison between brownfield and greenfield developments. *The International Journal of Behavioral Nutrition and Physical Activity*, 16(1), 11. <https://doi.org/10.1186/s12966-019-0775-8>
- Ziemke, D., Kaddoura, I., & Nagel, K. (2019). The MATSim Open Berlin Scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. *Procedia Computer Science*, 151, 870–877. <https://doi.org/10.1016/j.procs.2019.04.120>
- Ziemke, D., Metzler, S., & Nagel, K. (2019). Bicycle traffic and its interaction with motorized traffic in an agent-based transport simulation framework. *Future Generation Computer Systems*, 97, 30–40. <https://doi.org/10.1016/j.future.2018.11.005>
- Zilske, M., Neumann, A., & Nagel, K. (2015). *OpenStreetMap for traffic simulation*. Technische Universität Berlin.