Journal of Cycling and Micromobility Research 3 (2025) 100049

Contents lists available at ScienceDirect

Journal of Cycling and Micromobility Research

journal homepage: www.sciencedirect.com/journal/journal-of-cycling-and-micromobility-research

Modelling the impact of road infrastructure on cycling moving speed

Afshin Jafari [a,*], Dhirendra Singh [b,c], Lucy Gunn [a], Alan Both [d], Billie Giles-Corti [a,e]

[a] Centre for Urban Research, RMIT University, Melbourne, Australia
[b] School of Computing Technologies, RMIT University, Melbourne, Australia
[c] Data61, CSIRO, Melbourne, Australia
[d] School of Geospatial science, RMIT University, Melbourne, Australia
[e] The University of Western Australia, Crawley, Western Australia

ARTICLE INFO

Keywords:
Cycling infrastructure optimization
Micromobility solutions
Agent-based simulation
Bicycle speed
Traffic signals

ABSTRACT

Cycling for transport is a sustainable alternative to using motorised vehicles for daily trips and is a key form of micromobility. Travel time is a critical factor influencing cycling route choice behaviour and uptake. Thus, it is important to understand the factors affecting cycling travel time and speed and their impact on cycling behaviour. In this study, an agent-based transport simulation model with heterogeneous cycling speeds was developed and used for Melbourne to study the impact of a hypothetical traffic signal optimisation intervention along six key cycling corridors. Linear regression and random forest models were used to identify factors affecting cycling speed, which informed the parameters of the agent-based model. Simulation outputs showed, on average, an increase of 4.1 % in the number of cyclists on the corridors, as existing cyclists chose to use these corridors, and an average reduction in cyclists' moving travel time of 6.2 % for those using the intervention corridors (excluding time spent waiting at traffic signals). The findings provide insights into the effects of road attributes on cycling speed and behaviour, as well as the effectiveness of interventions aimed at reducing cycling delays. These insights are valuable for developing solutions to optimise urban infrastructure for micromobility, enhancing the efficiency and appeal of cycling as a viable transport option.

1. Introduction

Cycling for transport, that is, cycling to a destination rather than for recreation, is widely recognised as a key solution to a healthier and more sustainable future (Celis-Morales et al., 2017; Handy et al., 2014). Cycling not only provides an active and sustainable mode of transport (Buehler et al., 2020), it is often faster than other modes for short local trips or around Central Business District (CBD) areas or for peak hour commute (Gossling, 2020) offering a competitive alternative to the car for most trips within a cycling distance (Ellison and Greaves, 2011). However, cycling remains one of the least used mode of transport in many cities, particularly those in Australia (Jafari et al., 2024a).

There has been an increasing interest in understanding cyclist behaviour and barriers to cycling as a means of encouraging cycling uptake (Molenberg et al., 2019; 2023). Built-environment factors, particularly safe cycling infrastructure is found to be key barriers for cycling uptake globally, particularly among the demographic groups that are interested to take up cycling and their daily trips can be switched to cycling, for example are within a bikeable distance and or does not require carrying luggage or items, but are not confident to cycle on high stress and unsafe roads.

Another key built-environment element that significantly impacts cycling uptake is intersections. Poorly designed intersections negatively affect cycling in at least two ways. First, they are high-risk and unsafe crossing points, which significantly compromise the overall safety of a cycling journey. Second, they cause delays for cyclists by increasing travel time due to stops and slowdowns at intersections.

Increased travel time is a key factor influencing cycling mode choice and route choice. Borjesson and Eliasson (2012) found that in Stockholm, the value of time on a bicycle—that is, the opportunity cost of time spent travelling on a bicycle—on a street without a bike lane is almost twice that of the time spent on alternative routes with bike lanes. Similarly, in their mode choice model for the UK, Wardman et al. (2007) found that time spent cycling was valued three times higher than the travel time for other modes.

Attempts to reduce cycling travel time typically focus on reducing travel distances for cyclists, either through improving cycling street connectivity and land use change interventions to improve accessibility

* Correspondence to: RMIT University, GPO Box 2476, Melbourne, VIC 3001, Australia
E-mail address: afshin.jafari@rmit.edu.au (A. Jafari).

https://doi.org/10.1016/j.jcmr.2024.100049
Received 24 July 2024; Received in revised form 14 October 2024; Accepted 15 November 2024
Available online 19 November 2024

to destinations (Handy et al., 2014; Ahmad et al., 2020), or by reducing the effect of factors that decrease cycling speed that cause delays throughout the route and journey. There are numerous factors associated with cycling speed, which can be divided into: (i) individual-level factors such as physical ability, trip purpose, clothing, and type of bicycle (Romanillos and Gutierrez, 2020); (ii) social factors such as peer pressure to cycle faster and bicycle congestion (Paulsen and Nagel, 2019); and (iii) environmental factors such as road attributes including slope, traffic signals and intersections, connectivity, and bikeway type (Flugel et al., 2019; Clarry et al., 2019; Boufous et al., 2018; Ellison and Greaves, 2011).

One type of intervention that has gained interest among transport planners in recent years is to optimise traffic signals along the main cycling corridors to minimise disruption for cyclists, creating a green wave of traffic signals for cyclists to bypass intersections. The green wave signal timing for cycling was first introduced in Copenhagen, Denmark, which led to a 17 % travel time reduction for those who cycled at the typical cycling free flow speed of 20 kilometres per hour (km/h) (City of Copenhagen, 2014). Since then, green wave signal timing interventions for cyclists have been implemented in several other cycling corridors in Copenhagen, as well as other cities such as Groningen, The Netherlands (Zhang and Blokpoel, 2018), and in trials in Melbourne,[1] Australia (Bicycle Network, 2019).

Although trial data provides valuable information on the real-world implications of such interventions, a limitation of these trials is that they are often small-scale, typically involving only a few signals, and they lack the ability to compare a number of different alternatives and to identify corridors that could provide maximum gain. City-scale agent-based transport simulation models provide a promising solution to simulate different possible interventions and their potential impact on cyclists and other road users. In recent years, these types of models have become important tools for researchers and planners to examine cycling travel behaviours (Kaziyeva et al., 2021; Jafari et al., 2024a). A common practice in such models has been to assume a constant mode-specific speed of around 15 km/h for cycling (Balac and Horl, 2021), or using queue-based traffic modelling to incorporate bicycle traffic congestion (Paulsen and Nagel, 2019). However, few studies have developed and employed city-scale transport models with heterogeneous cycling speed, and even fewer have included the impact of road attributes such as traffic signals on cycling.

This study aimed to address this gap by investigating the association between road characteristics and cycling speed in Melbourne. The findings were used to inform a city-scale agent-based cycling simulation model to generate more accurate cycling travel speeds. The simulation model was then used to test and understand the potential improvements in cycling travel speed that could be achieved through a series of green wave-like interventions on the main cycling corridors that feed into the Melbourne CBD.

2. Methodology

Fig. 1 provides an overview of the study methods and steps. The first step was creating a road network model for the bicycle infrastructure in Melbourne that included attributes related to the type of cycling infrastructure, slope, and junction. Next, two regression models were developed to estimate the speed of cycling using smartphone application data: a least squares Linear Regression (LR) model and a Random Forest (RF) regression model.

LR and RF models were chosen as each can provide unique insights and information on the association between road attributes and cycling speed. The LR model offers a simple and easy-to-understand approach to understanding the association between road attributes and cycling speed. However, it assumes linearity in the parameters, which limits its predictive accuracy. The RF model does not assume a linear relationship between the predictors and the outcome variables and therefore typically offers better predictive accuracy, however, it is often referred to as a black-box model (Palczewska et al., 2013), which offers little insight into the relationship between variables.

The model with the least prediction error was then used to predict cycling speed for all bicycle accessible roads in Melbourne, which was used to extend an existing agent-based cycling simulation model to consider heterogenous cycling speed during the simulation. We then used this model to predict the impact of improving traffic signals to facilitate cycling speed for the main cycling corridors that feed into Melbourne's CBD. Each of these steps are described in more detail in the remaining of this section.

2.1. Study area

Melbourne is the capital city of the state of Victoria, Australia, with a population of approximately 5 million residents (Australian Bureau of Statistics, 2021), and it covers an area of about 10,000 square kilometres. The population is expected to almost double by 2050, reaching around 9 million people (Department of Environment, Land, Water and Planning, 2017). Accordingly, transport and housing policies seek to meet the needs of this growing population by 2050.

The Australian Bureau of Statistics (ABS) reports census data using a hierarchy of statistical areas, referred to as Statistical Area Levels. For example, Statistical Area Level 2 (SA2) units have populations of around 10,000 people and typically represent suburbs in urban areas, while Statistical Area Level 4 (SA4) units have populations around 300,000 to 500,000 people and represent larger areas, often reflecting labour markets.

Fig. 2 presents a map of Greater Melbourne showing the boundaries of SA4 regions, with population density depicted at the SA2 level in terms of people per square kilometre based on the 2016 Census data. The figure highlights the distribution of population density across Melbourne, illustrating that the highest densities are concentrated in the inner parts of the city, particularly around the CBD and surrounding suburbs, which are mostly within Melbourne – Inner SA4 marked on the map (Fig. 2).

The city is characterised by a monocentric urban structure, with a significant proportion of work trips directed towards the CBD located at the centre of the Melbourne – inner SA4. This centralisation results in high demand for transport infrastructure leading into the CBD, making it a critical area for studies on commuting patterns and transport modes.

Cycling in Melbourne remains a mode with a relatively small share, accounting for about 1.6 % of all commuter trips, while still being higher than the Australian average of about 1 % (Australian Bureau of Statistics, 2022). However, cycling is more prevalent in the inner suburbs, where there is better cycling infrastructure, higher population densities, and a greater mix of land uses (Boulange et al., 2017). Fig. 3 shows the bicycle mode share for trips to work, based on ABS Census 2016 data, at the SA2 level based on place of usual residence.

2.2. Generating the attributed cycling infrastructure road network model for Melbourne

We used the OpenStreetMap extract of the Melbourne region from March 2021 as the main input to create the road network model used in this study. The algorithm developed by Jafari et al. (2022) was used to map the roads from the OpenStreetMap extract and select those that are publicly accessible and can be accessed by bike. This means that roads on privately owned lands that are not accessible, or those where cycling is not permitted, such as footpaths, motorways, and roads explicitly prohibiting cycling, were excluded. Fig. 4 presents the generated road

[1] For simplicity, we will refer to Greater Melbourne simply as Melbourne throughout this paper. Please note that this should not be confused with the City of Melbourne, which is the central Local Government Area within Greater Melbourne.

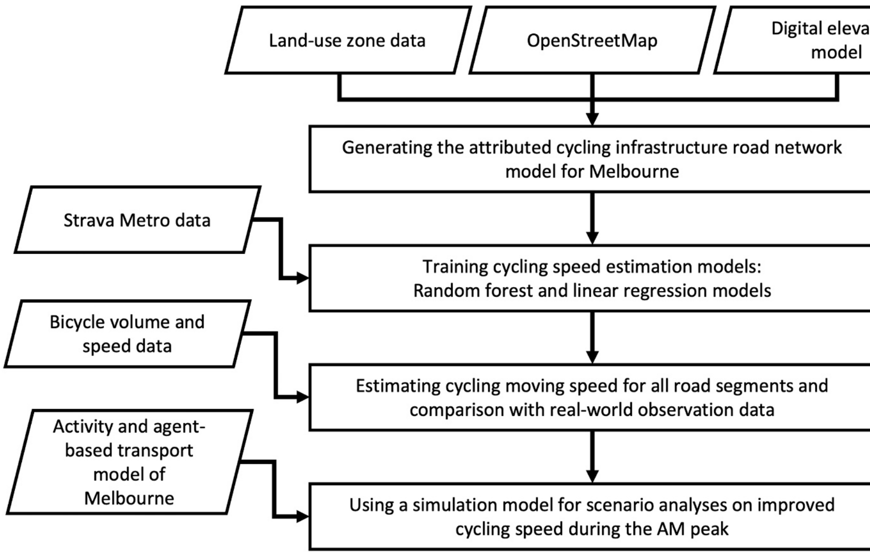

Fig. 1. Overview of the study steps and its methods and data sets.

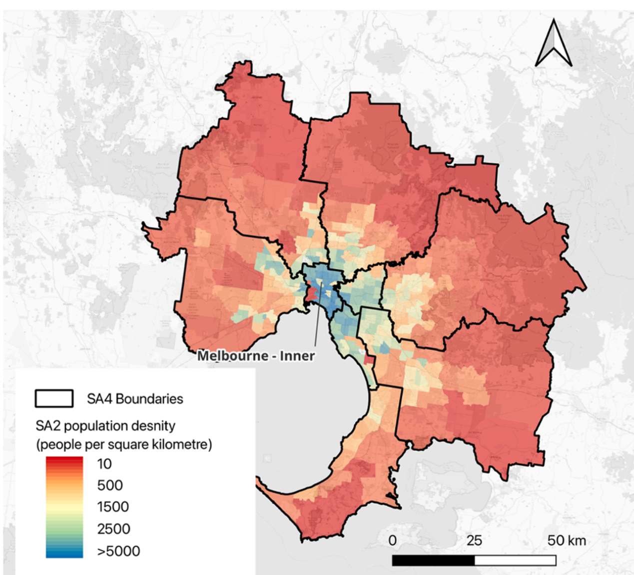

Fig. 2. Map of Melbourne showing the boundaries of Statistical Areas Level 4 (SA4) regions and population density at the SA2 level (people per square kilometre) based on 2016 Census data (base map from OpenStreetMap).

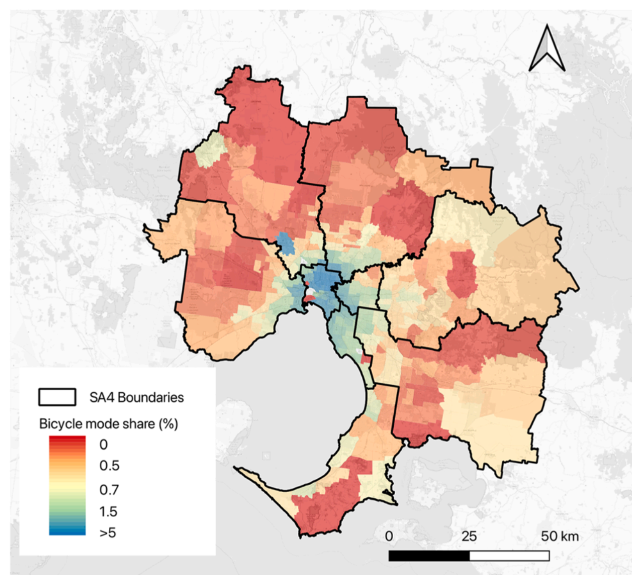

Fig. 3. Map illustrating bicycle mode share for trips to work based on ABS Census 2016 Method of Travel to Work data, shown at the SA2 level (base map from OpenStreetMap).

network for Melbourne. Furthermore, road feature tags from OpenStreetMap were processed to identify the cycling-relevant attributes needed in this study. One of these attributes was bikeway type. Road segments were categorised according to the availability and type of bikeway into five categories: (i) mixed: no bicycle-specific infrastructure; (ii) simple lane: a painted lane on the road without a physical barrier; (iii) separated lane: a bicycle lane on the road but separated from motorised traffic with a physical barrier; (iv) off-road path: an off-road path dedicated to bicycles only; and (v) shared paths which are off-road paths shared between bicycles and pedestrians.

To capture the influence of junctions and traffic signals on cycling speed, we defined two additional road attributes of junction type and signal type based on the OpenStreetMap tags. In this study, we use the term *junction* to refer collectively to both intersections and roundabouts, representing any location where two or more roads meet or cross.

The junction type variable had three possible values: (i) within intersection, that is, a road segment that is physically located within an intersection where two or more roads meet; (ii) within roundabout, that is, a road segment that is within a roundabout; and (iii) mid-block, that is, a road segment that is not within any intersection or roundabout. The junction type variable allows us to assess how different types of junctions affect cycling speed due to factors like navigational complexity, turning movements, and interactions with other road users.

The signal type variable indicates whether a road segment is approaching a signalised junction, with two possible values: (i) approaching a signal, that is, a road segment that is immediately before a signalised junction but is not within the junction itself; and (ii) not approaching a signal, that is, a road segment that is not immediately before a signalised junction. The signal type variable helps capture the impact of approaching traffic signals on cycling speed, as cyclists may adjust their speed in anticipation of stopping or proceeding through the junction.

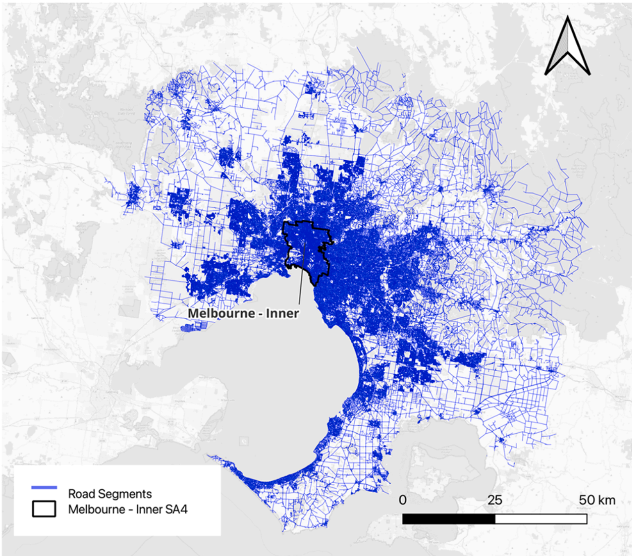

Fig. 4. Road network model of (a) Greater Melbourne, and (b) enlarged map of City of Melbourne with CBD area highlighted (base map from OpenStreetMap).

In addition to the above, we also extracted speed limit, number motor traffic lanes, and road segment length from OpenStreetMap for each of the road segments. Wherever data was not available in OpenStreetMap for speed limit or number of lanes, we used Australian default values based on the hierarchy of the road as indicated by Jafari et al., (2022).

We also used other data sets to add additional attributes to the road segments that are relevant for cycling. A 10 m LiDAR-derived Digital Elevation Model (DEM),[2] which provides high-resolution elevation data of the Earth's surface, was used to determine the elevation at the start and end points of each road segment. By calculating the difference in elevation and dividing it by the length of the road segment, we calculated the slope (%) for each segment.

To account for differences in cycling patterns between the inner, middle, and outer suburbs of Melbourne, we added a variable representing the distance from the CBD. This was calculated based on each road segment's centroid distance to Melbourne's General Post Office in the CBD. Additionally, we used VicMap planning map data from the Victorian Government[3] to add land-use type to the areas surrounding each road, we referred to it as zone type. Land uses were categorised into the following types: activity centres, CBD, commercial, residential, mixed-use, industrial, park and recreation, logistic (areas designated for ports, transit, and freight), roads (areas surrounding arterial roads or higher), and other (for any land uses not classified under the above categories).

These variables were calculated for all road segments in Melbourne. However, for simplicity and due to the large area of Melbourne, we demonstrate these variables in Fig. 5 only for the Melbourne - Inner SA4 area.

2.3. Training cycling speed estimation models

Aggregated and de-identified data from Strava Metro, hereafter the Metro data, for March 2021 was used to capture cycling volume and speed for the study area. The data include 253,424 trips tagged as *commute* undertaken by 31,222 cyclists throughout Melbourne. These data were filtered to weekday records and road segments that carried a minimum of 20 unique cyclists and with an average speed between 5 km/h and 50 km/h. Finally, road segments with an average speed within the two standard deviation ranges of the total average speed, covering approximately 95 % of the road segments, were selected to build the speed estimation model.

To estimate the cycling speed for all road segments, LR and RF regression models were used, with the outcome variable for both models being the average weekday cycling speed extracted from Metro data, denoted as $\bar{v}_{weekday}$, and the physical environment factors shown in Fig. 5 as predictors.

The Metro data was divided into a training set of 28,387 data points and a test set of 9464 data points as follows. Stratified sampling based on the $\bar{v}_{weekday}$ variable with a bin size of four was used to convert the $\bar{v}_{weekday}$ variable which is continuous into a categorical variable for the purpose of dividing the data into a training set and test set. Therefore, random sampling of data points was performed within each stratum, ensuring that both training and test sets had approximately similar proportions of samples within each $\bar{v}_{weekday}$ bin compared with the entire data set.

For both LR and RF models, the continuous $\bar{v}_{weekday}$ (the average weekday cycling speed extracted from Metro data) was used as the outcome variable.

In the LR model, categorical variables were included by coding them as binary indicator variables, with one category designated as the reference category. This method allows categorical predictors to be incorporated into the regression analysis and enables the interpretation of their effects on the outcome variable relative to the reference category.

The RF model does not require the transformation of categorical variables into binary indicators and can handle categorical predictors directly. The RF model is an ensemble learning method that builds multiple decision trees and aggregates their predictions to improve accuracy and control overfitting (Breiman, 2001). In an RF model, a node is a point in a decision tree where the data is split based on a predictor variable and a split refers to the decision rule applied at a node to partition the data into subsets that are more homogeneous with respect to the outcome variable.

The RF model requires tuning of hyperparameters to optimise its performance. Specifically, we tuned the following hyperparameters: (i) *mtry*, which is the number of predictor variables randomly selected as candidates at each split when growing the trees. This parameter controls the diversity of the trees and helps prevent overfitting; and (ii) min_n, which is the minimum number of observations required in a node for it to be split further. This parameter ensures that nodes have a sufficient number of observations before splitting, which can improve the model's generalisability.

We used a grid-based approach with a grid size of 20, which means that 20 parameter sets were created and compared to search for the best parameter values with Root Mean Squared Error (RMSE) and R-squared (R2) as performance metrics. Bootstrap sampling stratified on $\bar{v}_{wd}$ with bin size equal to four and number of bootstrap samples equal to 25 was used for training.

We used the *tune* package in R programming language for hyperparameter tunning and finding the best performing variable values. The hyper-parameter tuning result for the random forest model using a grid of $n=20$ is presented in Fig. 6. Among the 20 parameter pairs analysed, $mtry=3$ and min_n=4 had the best performance with $RMSE=4.41$ and $R2=0.498$. Therefore, these values were used to build the final RF regression model.

The number of trees in the forest, *n_tree,* was another parameter required to build the RF model. This number is not a tuning parameter and only has to be large enough to give every variable the opportunity to be selected (Couronne et al., 2018). Therefore, an *n_tree* of 1000 was selected for the model.

[2] https://discover.data.vic.gov.au/dataset/vicmap-elevation-dem-10m Retrieved May 2021

[3] https://www.land.vic.gov.au/maps-and-spatial/spatial-data/vicmap-catalogue/vicmap-planning Retrieved May 2021

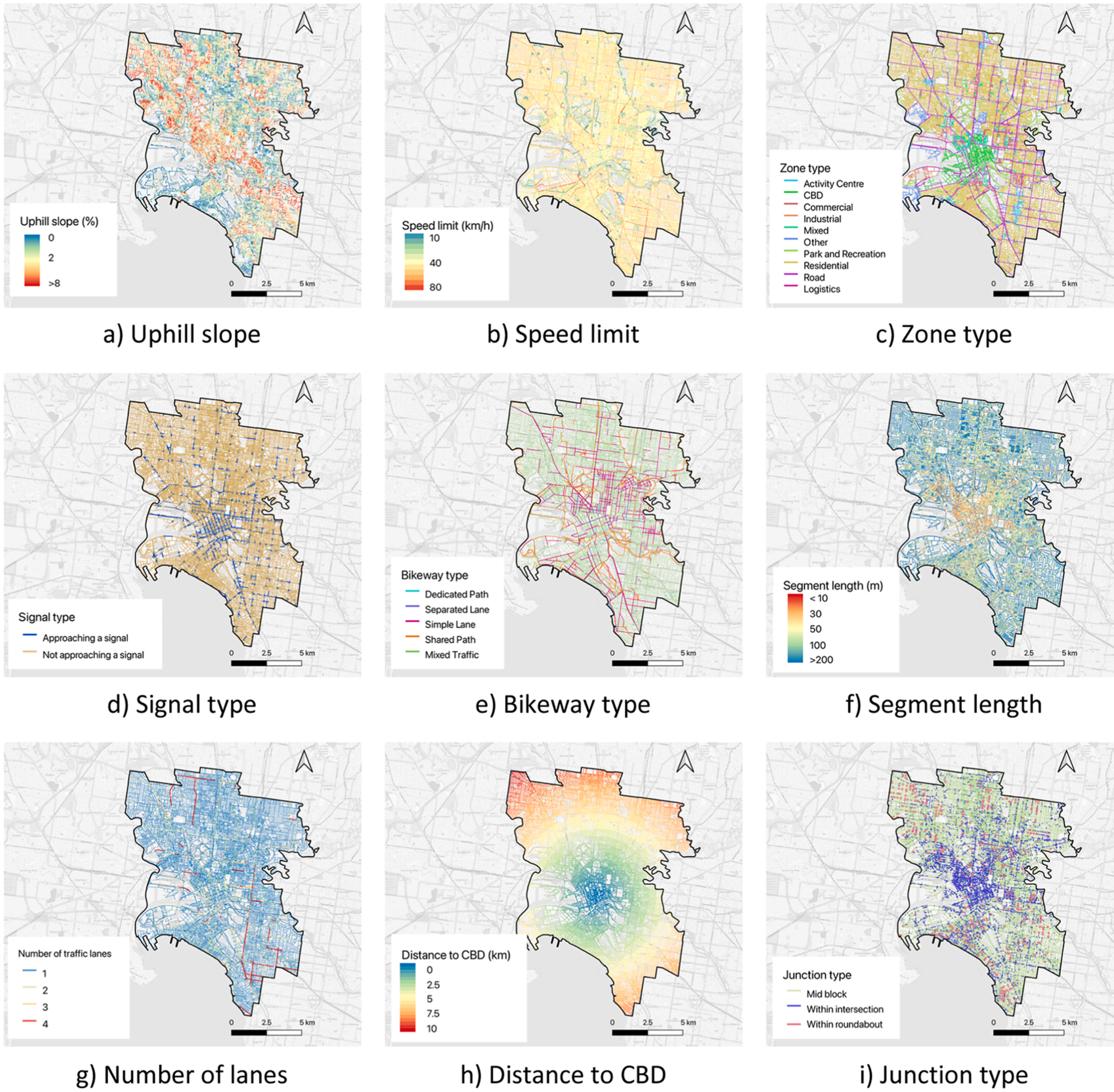

Fig. 5. Attributes added to the selected road segments (base map from OpenStreetMap).

Finally, we use the Permutation-based Variable Importance Measure (P-VIM) to understand the relative importance of the input variables of the model. P-VIM considers a variable to be important if its permutation has a stronger decrease in the accuracy of the model (mean square error in our case) than the permutation of an unimportant variable.

2.4. Estimating cycling moving speed for all road segments and comparison with real-world observation data

The speed estimation models were trained and tested on a subset of all Melbourne road segments with a reliable number of cycling records. However, for a city-wide agent-based simulation model, an associated speed factor was needed for every road segment of the entire network. Therefore, first the physical environment attributes were added to the entire road network following the steps described in Section 2.1. We then used the model with the highest predictive accuracy to assign a cycling speed to each road segment.

To validate the predicted speeds for the full network, we used data from automatic cycling speed sensors across Melbourne. This dataset, hereafter referred to as the sensor data, was downloaded from the Victoria Open Data Platform for the period of March 2021.[4] The sensor network consists of 35 counting stations, each counting station equipped with two sensors, one for each direction, resulting in a total of 70 sensors.

The observed average weekday daily cycling speeds from these sensors varied across different locations. These variations reflect the diversity of cycling conditions in Melbourne, influenced by factors such as infrastructure quality, traffic volumes, and terrain.

[4] https://discover.data.vic.gov.au/dataset/bicycle-volume-and-speed retrieved May 2021

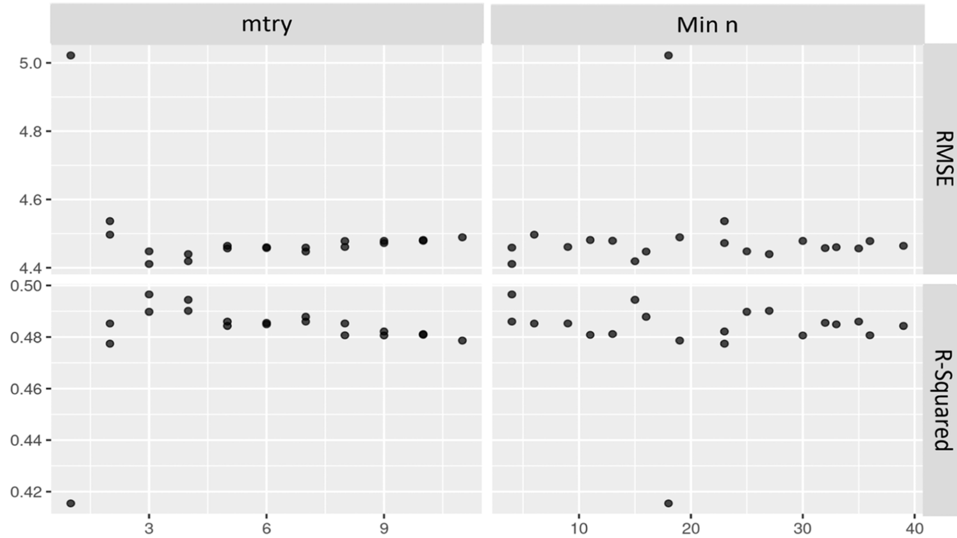

Fig. 6. Random Forest model tuning graph for identifying optimum mtry and min_n.

To analyse the accuracy of speed prediction, for each sensor i, the percentage of speed error denoted as $\delta_{speed,i}$ was calculated based on the predicted speed, $\widehat{v}$, and the average speed observed during the week, $\overline{v}_{sensor,i}$, from sensor data as follows:

$$\delta_{speed,\quad i} = \frac{\widehat{v} - \overline{v}_{sensor,\quad i}}{\overline{v}_{sensor,\quad i}} \times 100 \tag{1}$$

2.5. Using a simulation model for scenario analyses on improved cycling speed during the AM peak

The Activity-based and agent-based Transport model of Melbourne (AToM) developed by Jafari et al. (2024b) was used and extended to incorporate estimated cycling speeds in a cycling traffic simulation model. AToM uses the Multi-Agent Transport Simulation (MATSim) as its core traffic simulator and is calibrated to simulate a typical weekday day of the transport system of Melbourne for a 10 % population sample. The travel demand for the model was created using the activity-based model developed by Both et al. (2021).

We replaced the road network of the AToM model with the road network model with the predicted cycling speed described from the most accurate speed estimation model between RF and LR. All modes except cycling were set to teleport, i.e., travelling in a straight line from the origin to the destination with constant speed and no interaction with the road network or other travellers, with a distance multiplier to bring the simulated distance closer to the actual network distance. These multipliers (Table 1) were estimated using data from the 2014–16 Victorian Integrated Survey for Travel and Activity (VISTA),[5] based on the travellers' recorded trip duration and distance. VISTA is a household travel survey conducted by the Victorian Department of Transport and Planning, in which participants are asked to provide details of their trips and activities on the survey day. These details include the origins and destinations of their trips, main modes of travel, trip purposes, and departure and arrival times.

The cycling utility function in AToM (Jafari et al., 2024b) was set to use the adjusted cycling travel time based on estimated speeds. Furthermore, a margin utility of the bicycle infrastructure was also added based on the parameters and equations proposed by Ziemke et al. (2018). For all other coefficients of the model same values as the baseline AToM was used.

The agent-based simulation model was used to examine the potential impact of predict the impact of improving traffic signals to facilitate cycling speed during AM peak for corridors that feed into the Melbourne CBD. This scenario represents a simplistic approach to model a green wave intervention for cyclists. Six main corridors within a 5 km radius of Melbourne CBD were identified as candidates to test this intervention (Fig. 7). These corridors were selected in consultation with active transport planners from the Victorian Department of Transport and Planning, as well as local government transport planners. The selection was based on a number of reasons. First, the chosen corridors are known to have high cycling traffic, making them significant for commuting cyclists. Enhancements in these areas could benefit a large number of users. Also these corridors are key routes that connect the inner suburbs to the CBD, aligning with the monocentric urban structure of Melbourne where a significant proportion of work trips are directed towards the CBD. And lastly, the traffic signal setup along these corridors were among the potential candidates that a green wave signalling might be implemented for them in future.

To create the simulation scenario, all junctions within the relevant corridors were set to be similar to an un-signalised junction for the AM peak period for cyclists. The cycling speed was then predicted for these corridors using the best performing speed estimation model, and the outcome network was used to test the intervention scenario with the agent-based model. During the simulation, re-routing was the only strategy enabled for the agents. We ran the simulation mode for 200 iterations for baseline and intervention scenarios to allow all agents to find their best route.

Table 1
Speed constant and Euclidean distance multiplier for the simulation teleportation travel modes.

Mode	Speed constant (m/s)	Distance multiplier
Car	7.37	1.30
Public transport	5.44	1.46
Walk	1.34	1.28

3. Results

The first statistical model that we used to predict the speed of cycling based on road attributes was the least squares multiple linear regression model. As shown in Table 2, most predictor variables were found to be

[5] https://www.vic.gov.au/data-and-publications retrieved May 2021

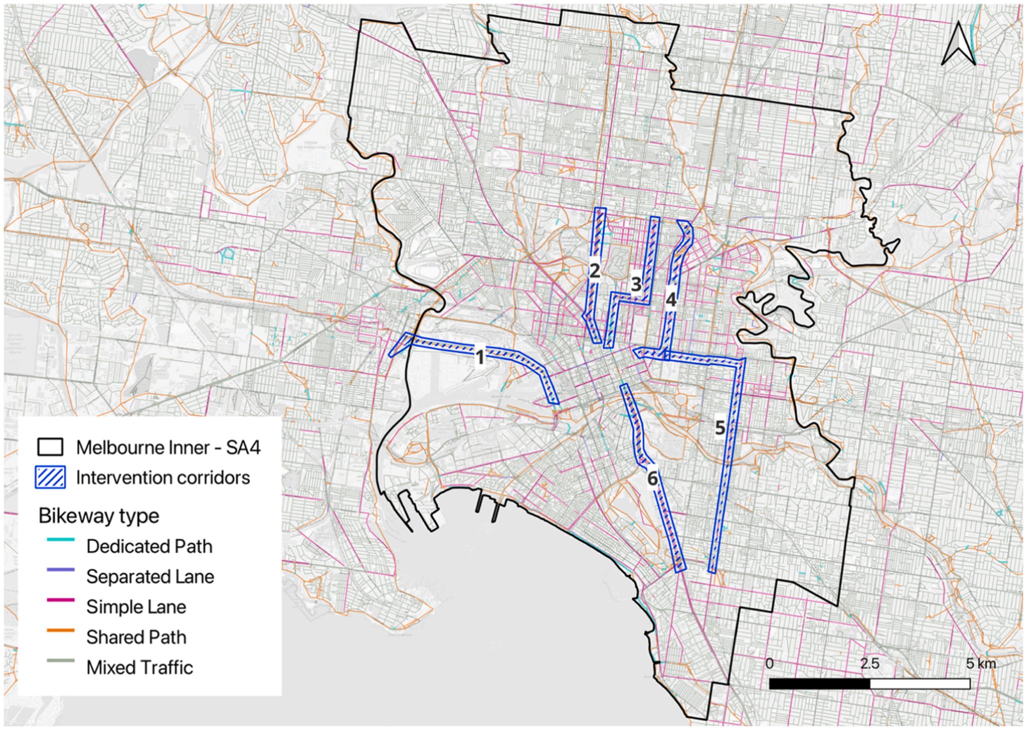

Fig. 7. Selected intervention corridors and bikeway types in Melbourne – Inner SA4 and surrounding (base map from OpenStreetMap).

statistically significant in the LR model ($p < 0.001$), with the exception of Industrial and Other Zones, and Dedicated and Separated Lanes. For bikeway type, cycling on a mixed traffic road showed to be negatively correlated with cycling speed compared to a simple bike lane ($p<0.001$). Interestingly, both going uphill and downhill were found to be negatively correlated with cycling speed. Another notable factor was the negative correlation of cycling speed on a road approaching a signalised junction when compared to unsignalised roads. Both within intersection and within roundabout were found to be associated with higher speeds compared with mid-block road segments, which could be due to cyclists accelerating to exit a junction more quickly. The overall model fit was $R2 = 0.3154$ with RMSE of 4.69.

Fig. 8 shows the results for P-VIM for the predictors of cycling speed based on the tuned RF model. As illustrated in this figure, signal type, distance to CBD, and slope were the three main predictors of cycling speed in Melbourne, followed by road length, zone type, speed limit, number of lanes, bikeway type, and junction type.

Both LR and RF models were used to predict the test dataset to evaluate their predictive accuracy. The prediction graphs in Fig. 9 show that although the prediction values in both models deviated from the actual values, the RF model (Fig. 9b) demonstrates a better fit to the prediction line compared with the LR model (Fig. 9a). The RF model exhibited better prediction accuracy with RMSE = 4.109 compared with RMSE = 4.690 for the LR model. However, both models outperformed the average constant speed of 15 km/h across all the road segments commonly used in simulation models (RMSE = 9.908), as well as the typical assumed cycling free-flow speed of 20 km/h (RMSE = 6.412). Therefore, the RF model was used to predict cycling speed for all road segments across the study area, to be used as input for the agent-based simulation model.

Fig. 10a illustrates the percentage error, δ_{speed}, for the predicted cycling speed from the full network based on the trained RF model versus the observed speed from sensor data at each location of the sensor device. For 40 % of the sensors (28 out of 70), the error percentage was less than 15 %, while three sensors had an error percentage greater than 66 %, mainly in the middle regions of Melbourne. The spread of errors for different types of bikeways is illustrated in Fig. 10b, indicating that the main deviations between the prediction and the actual values are for shared paths, where both pedestrians and cyclists are allowed.

The intervention scenario examined using the simulation model was the implementation of a cycling green wave for each of the corridors shown in Fig. 7. Table 3 shows the moving travel time and the number of cyclists using each corridor before and after the implementation of the intervention scenario. Eliminating the delay caused by traffic signals for cyclists travelling towards Melbourne CBD on the main corridors resulted in an average 6.2 % reduction in cycling moving travel time along these corridors, that is excluding the time stopping behind the traffic signal, and a 4.1 % increase in the number of cyclists using them. This increase is due to the re-routing of existing cyclists who adjust their routes to take advantage of the improved travel times, not due to an overall increase in cycling mode share.

Among the selected corridors, the sixth corridor had the highest decrease in moving travel time (-9.5 %), while the first corridor had the highest increase in cyclist numbers as a result of the intervention (11.1 %). It should be noted that the results in Table 3 only illustrate the impact of the intervention on cycling moving travel time, time spent stopping at traffic signals is not included in the simulation outputs. Therefore, the total reduction in travel time and the potential for attracting more cyclists could be even higher if the impact of traffic signals on stationary components of a trip were also included.

4. Discussion and conclusion

By comparing the LR and RF models in this paper for predicting cycling speed based on Strava Metro data, we found that the RF regression model showed superior accuracy compared with the LR model. This aligns with the findings of Couronne et al. (2018), who compared the predictive accuracy of RF and LR on 243 real datasets. They found that RF performed better in approximately 69 % of the datasets tested. However, both models can provide valuable insights that are discussed below.

The results of the LR model indicated that longer roads (i.e., fewer

Table 2
Summary statistics of the linear regression model for association between road attributes and cycling speed.

Predictors	Estimation	std. error	Statistic	p. value	CI
(Intercept)	21.09	0.34	61.89	<.001	20.42–21.76
Speed limit	0.10	0.02	4.49	<.001	0.06 – 0.15
Road segment length (m)	0.01	0.00	12.76	<.001	0.01 – 0.01
Distance to CBD	0.27	0.01	18.80	<.001	0.24 – 0.29
Uphill Slope (%)	−0.49	0.03	−18.03	<.001	−0.54 – −0.44
Downhill Slope (%)	−0.09	0.03	−3.52	<.001	−0.15 – −0.04
Car lanes (#)	0.35	0.08	4.41	<.001	0.19 – 0.50
Zone type (Ref=Residential)					
Activity Centre	−1.76	0.50	−3.52	<.001	−2.74 – −0.78
CBD	−1.86	0.25	−7.42	<.001	−2.35 – −1.37
Commercial	−1.16	0.28	−4.12	<.001	−1.71 – −0.61
Industrial	−0.63	0.55	−1.15	0.249	−1.71 – 0.45
Mixed	−0.71	0.38	−1.86	0.063	−1.46 – 0.04
Other	0.44	0.32	1.38	0.166	−0.18 – 1.07
Park and Recreation	1.09	0.22	4.98	<.001	0.66 – 1.53
Road	0.40	0.17	2.35	0.018	0.07 – 0.73
Logistics	−2.92	0.41	−7.17	<.001	−3.72 – −2.12
Bikeway type (Ref=Simple lane)					
Dedicated path	−0.40	0.42	−0.96	0.339	−1.22 – 0.42
Mixed traffic	−0.73	0.13	−5.52	<.001	−0.98 – −0.47
Separated lane	0.80	0.57	1.40	0.163	−0.32 – 1.92
Shared path	−0.60	0.26	−2.28	0.023	−1.11 – −0.08
Junction (Ref=Mid-block)					
Within intersection	1.38	0.17	8.12	<.001	1.05 – 1.71
Within roundabout	0.77	0.38	2.05	0.041	0.03 – 1.51
Signal (Ref=Not approaching a signal)					
Approaching a signal	−7.19	0.19	−37.05	<.001	−7.57 – −6.81

N = 7, 547
R2/R2 adjusted = 0.3154/0.3134

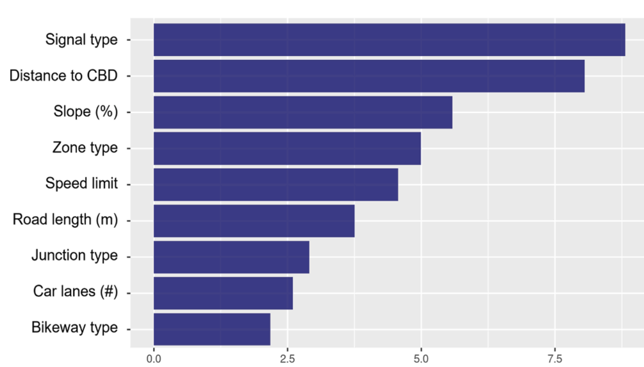

Fig. 8. Variable importance (P-VIM) plot based on the trained Random Forest model.

junctions) and approaching an unsignalised junctions rather than signalised junctions result in faster cycling speeds or, in other words, less delay. This confirms the findings of previous studies on the relationship between cycling speed and bikeways, such as Clarry et al. (2019). The RF model, however, does not provide the direction and expected change in speed per unit change of the predictors, which are common outputs of an LR model. However, the variable importance analysis in Fig. 8 indicated that the most significant road attribute associated with cycling speed was whether the road ends at a signal-controlled junction or not, indicating the importance of a proper signal control strategy to prioritise cyclists. Distance to the CBD was the second most important variable for the model. It represents the accessibility and density of destinations in Melbourne, as the inner-city areas and middle suburbs have greater amenities and typically feature more destinations and higher housing density (Giles-Corti et al., 2022). Furthermore, the results of the LR model demonstrated a positive correlation between the distance from the CBD variable and speed, which means that roads located further away from Melbourne CBD have faster speeds.

Slope was the third most important factor in the RF model for predicting cycling speed. This is consistent with several other studies that found slope to be an important factor (Romanillos and Gutiérrez, 2020), worthy of inclusion in cycling simulations (Ziemke et al., 2018). However, the LR model shows that both uphill and downhill paths are associated with lower cycling speeds. This could be due to cyclists braking more often when going downhill to maintain safe speeds. It may also be the case that simply dividing the slope into uphill and downhill measures is insufficient, and smaller slope intervals, like those considered by Flugel et al. (2019), are needed for the LR model. This is not an issue for the RF model, as it follows a hierarchical classification process and divides variables where the differences are most meaningful.

The RF model predicted cycling speed for all bicycle-accessible road segments in Melbourne. The resulting road network served as input for our city-scale agent-based simulation model, extending the bicycle simulation component of the baseline model proposed in Jafari et al. (2024b). Furthermore, our approach for adding heterogeneous speeds to agent-based models extends the work of Ziemke et al. (2018) by incorporating new factors, such as traffic signals, junction type, road length, and zone type.

The before-and-after comparison of removing the impact of traffic signals on cycling speed for six strategic corridors feeding into Melbourne CBD illustrated the potential time savings that could be achieved by implementing the intervention. However, the results showed that the intervention had varying effects across different corridors. For example, the first corridor saw the highest increase in cycling volume due to the intervention, whereas the highest reduction in travel time was found in corridor six. This indicates that the best corridor for implementing the green wave signal timing depends on whether the goal is to decrease travel time for current users or to encourage those using neighbouring roads to use the desired corridor.

Three north-to-south corridors (corridors two to four) were examined, and the before-and-after comparison illustrated that although the green wave intervention resulted in travel time reductions in all three corridors, cycling volume only increased in one of them. This indicates that when evaluating the success of a series of signal optimisation interventions, it is important to consider the potential impact of competing corridors that are viable alternatives for a large group of cyclists, such as those travelling from the northern suburbs to the CBD. However, even if the cycling volume did not increase in some corridors, the reduction in travel time provides direct benefits to the existing cyclists using those routes. These benefits include shorter commute times, reduced delays at intersections, and an overall more efficient and pleasant cycling experience. Improving travel times can enhance cyclist satisfaction and may contribute to sustained use of these corridors. It is also possible that the existing infrastructure in these corridors already does a fair job of facilitating safe cycling speeds, meaning that the corridors are already close to being *good* and adding a green wave intervention does not result in a significant improvement in attracting more cyclists. For example, Corridors 3 and 4 are favoured by current cyclists due to fewer barriers for cycling and better street design, and therefore, might be performing better than the models estimated. The travel time reductions in these

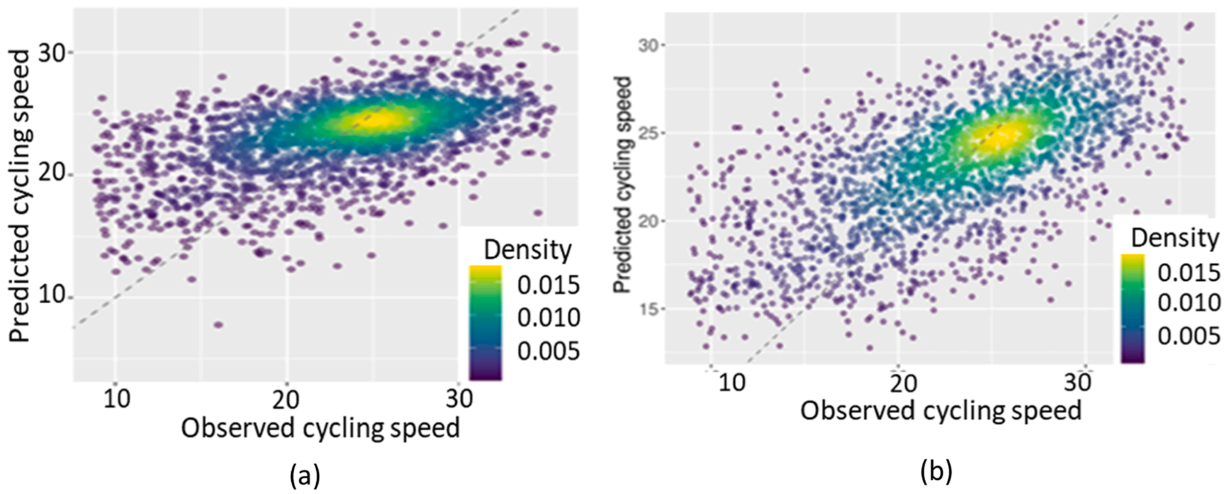

Fig. 9. Prediction plots for the (a) linear regression model and (b) random forest model based on the test data set.

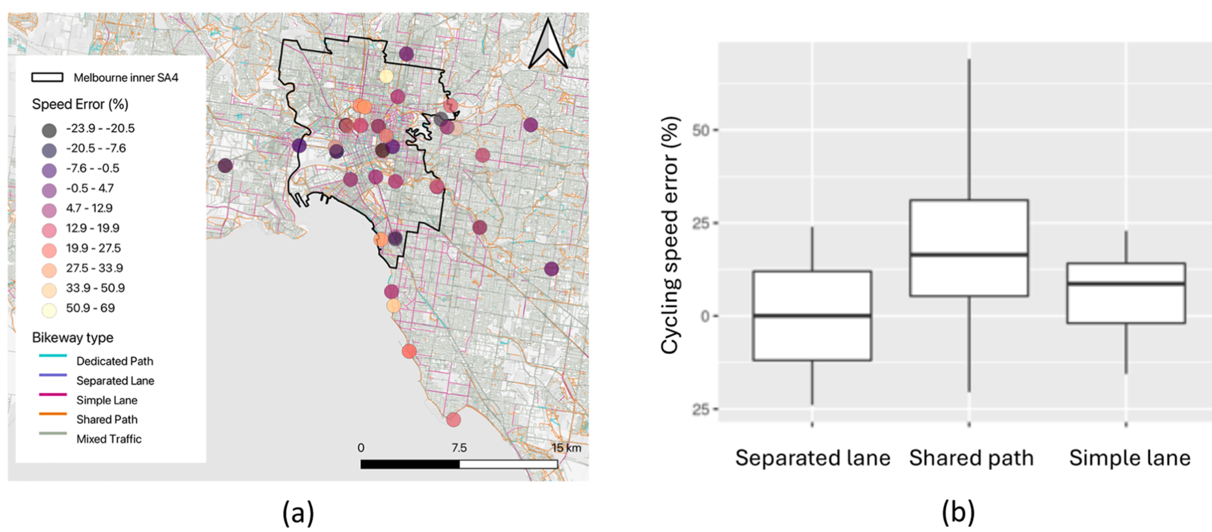

Fig. 10. Cycling speed error comparison between simulation results and sensor data based on (a) sensor locations, and (b) bikeway type (base map from OpenStreetMap).

Table 3
Average AM peak moving travel time of cyclists using the corridors and number of cyclists on the selected corridors before and after the intervention.

Corridor	Travel time			Bicycle traffic volume		
	Before	After	Change (%)	Before	After	Change (%)
1	18.0	16.7	−7.4	292	325	11.1
2	13.6	12.6	−6.7	105	99	−5.5
3	21.6	21.0	−2.7	321	326	1.6
4	20.4	19.7	−3.1	192	191	−0.4
5	22.9	21.2	−7.5	114	125	9.6
6	19.8	17.9	−9.5	124	128	3.3
Average	19.4	18.2	−6.2	191.2	199	4.1

corridors, even without an immediate increase in cycling volume, highlight the value of such interventions in enhancing the quality of cycling for current users.

The results do not suggest that traffic signals and junctions should be avoided when designing a bicycle-friendly neighbourhood. More junctions result in better network connectivity, leading to shorter travel distances (Titze et al., 2010), which favours using bicycles as a mode of transport. The quandary of uninterrupted roads to reduce cycling delay versus more junctions for better connectivity to reduce distance, and more signalised junctions for better control of traffic flow and safety, could perhaps best be analysed based on the purpose of the road used by cyclists. For example, it is reasonable to lean towards a green wave intervention for corridors where most cyclists use these routes to travel to Melbourne's CBD. In this case, the focus is on improving cyclist flow and reducing travel time delays. However, where local accessibility to destinations is a priority for cyclists, connectivity and safety are more pressing issues. This aligns with the Movement and Place framework recently adopted by the Victorian Government (Department of Transport, 2019), which classifies streets based on their level of locality and function into different categories, from Connectors (roads with high movement significance and low place significance) to City Places (roads and streets with high demand for pedestrian activities and lower vehicle movement) to City Hubs (dense and vibrant places that also have a high demand for movement). Following this framework, an intervention to minimise cycling delays could be a desirable approach for Connector roads, while for City Places, higher connectivity might be of greater importance. More research is needed to examine the interdependencies and implications of each approach for different situations and road types, such as the different categories of the Movement and Place framework.

Moreover, it is important to recognise that travel time is not the sole factor influencing cyclists' route choices. Studies have shown that cyclists are often willing to accept longer travel times to avoid uncomfortable roads, preferring routes that lead through more attractive areas such as parks or quieter streets, which may also feature fewer traffic lights (Broach et al., 2012). These preferences highlight the significance of route attractiveness, safety, and comfort in addition to travel time. Therefore, interventions focused solely on reducing travel time, such as optimising traffic signals, may not be sufficient to significantly increase cycling volumes if they do not also address these other important factors.

Furthermore, our simulation model indicates that improvements in travel time can lead to a redistribution of existing cyclists within the network, with more cyclists choosing the improved corridors. However, the model does not account for potential mode shifts from car or public transport to cycling. Assessing the impact of reduced bicycle travel time on encouraging non-cyclists to switch modes is an important area for future research. Incorporating mode choice into the simulation could provide insights into how infrastructure improvements might influence overall cycling participation.

This study does not account for individual cyclist characteristics and attributes that could be important correlates of cycling speed. The absence of these factors in our models likely contributes to the relatively high prediction errors we observed. These attributes include the cyclist's level of physical fitness, experience and confidence, bicycle type (e.g., fixed-gear bike, road bike, or e-bike), and whether the person is wearing business attire or not. Furthermore, the use of Strava data may introduce a bias towards more competitive or active cyclists, potentially overestimating the average cycling speeds of the general population (Boss et al., 2018). This bias could explain some of the deviations observed in our models, despite cross-validation with sensor data. Therefore, when adjusting the road network to attract new cyclists, it is important to consider the speed and needs of less experienced or casual cyclists, who may travel at lower speeds. Modelling these differences is challenging with the available data, but acknowledging this limitation is crucial for interpreting our findings and for future research efforts aimed at promoting cycling among a broader demographic. Incorporating these attributes into the simulation model is possible given that heterogeneous agents can be modelled in the agent-based model. However, there is a dearth of data at this level of detail, and primary data collection might be needed.

Additionally, emerging micromobility technologies, including e-bikes and e-scooters, which also make use of cycling routes and infrastructure, and the impact of the built environment on their moving speed, were not considered in this study due to limitations in the Metro data shared with us. While the data provides counts of rides per road segment for bicycles and e-bikes separately, we were unable to differentiate speeds between cyclists and e-bikes. However, the counts of e-bike rides were negligible compared to traditional bicycles, and therefore, we expect that they have had limited impact on the average speeds per road segment calculated in our study. E-bikes, with their electric assistance, can reach higher speeds and accelerate more quickly, creating different behaviour compared to regular bicycles. The integration of e-bikes into existing cycling infrastructure can impact overall flow and safety, and their increasing popularity necessitates a broader understanding of how they interact with conventional cyclists. Another interesting avenue for future research is to explore how to best optimize traffic signals to accommodate these different behaviours.

Data used in this study was for the period that the transport system in Melbourne was affected by the travel restrictions due to the COVID-19 pandemic period. Although this likely led to fewer bicycle trips for travel to work and more for other purposes, given the focus of this research on built-environment attributes and their association with cycling speed, we expect this to have limited implications for the usability of our findings.

Lastly, another limitation of our approach was that traffic signal timing for different phases, as well as time spent waiting at traffic signals, were not considered. Although it is possible to incorporate traffic signals into a MATSim model using the extension by Kuhnel et al. (2018), it requires details of the timing and locations of all signals, which could be difficult to obtain and process for city-scale models. We expect that incorporating traffic signal timing, whether exact or heuristic, would show that building a green wave for cyclists has a more significant impact in terms of travel time savings and increased cyclist numbers.

Funding

Afshin Jafari is supported by an Australian Government Research Training Program Scholarship. Dhirendra Singh's time on this project is funded by Collaborative Research Project grants from CSIRO's Data61 (2018–19, 2020–21). Alan Both is supported by the NHMRC/UKRI JIBE project (#APP1192788). Lucy Gunn is supported by the NHMRC funded Australian Prevention Partnership Centre administered by the Sax Institute (#9100001); and Billie Giles-Corti is supported by an RMIT VC Professorial Fellowship.

CRediT authorship contribution statement

Afshin Jafari: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Dhirendra Singh:** Writing – review & editing, Supervision, Conceptualization. **Lucy Gunn:** Writing – review & editing, Funding acquisition, Conceptualization. **Alan Both:** Writing – review & editing, Conceptualization. **Billie Giles-Corti:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study used aggregated and deidentified data from Strava Metro. The authors would like to acknowledge the contribution of Mr Nick Collins from the Victorian Department of Transport and Planning for his valuable input in conceptualising this study.

Data availability

The authors do not have permission to share data.

References

Ahmad, S., Goodman, A., Creutzig, F., Woodcock, J., Tainio, M., 2020. A comparison of the health and environmental impacts of increasing urban density against increasing propensity to walk and cycle in Nashville. Usa. *Cities Health* 4 (1), 55–65.

Australian Bureau of Statistics, 2021. Population: Census. ABS. ⟨https://www.abs.gov.au/statistics/people/population/population-census/2021⟩.

Australian Bureau of Statistics, 2022. Australia's journey to work. November 3. ABS. ⟨https://www.abs.gov.au/articles/australias-journey-work⟩. November 3.

Balac, M., Horl, S., 2021. Simulation of intermodal shared mobility in the San Francisco bay area using MATSim. *In 2021*. IEEE International Intelligent Transportation Systems Conference (ITSC). IEEE, pp. 3278–3283.

Bicycle Network (2019). Green wave a go on Albert Street. ⟨https://bicyclenetwork.com.au/newsroom/2019/09/16/green-wave-a-go-on-albert-street/⟩.

Biernat, E., Buchholtz, S., Bartkiewicz, P., 2018. Motivations and barriers to bicycle commuting: Lessons from Poland. Transp. Res. Part F: Traffic Psychol. Behav. 55, 492–502.

Borjesson, M., Eliasson, J., 2012. The value of time and external benefits in bicycle appraisal. Transp. Res. Part A: Policy Pract. 46 (4), 673–683.

Boss, D., Nelson, T., Winters, M., Ferster, C.J., 2018. Using crowdsourced data to monitor change in spatial patterns of bicycle ridership. J. Transp. Health 9, 226–233.

Both, A., Singh, D., Jafari, A., Giles-Corti, B., and Gunn, L. (2021). An activity-based model of transport demand for greater Melbourne. *arXiv*. ⟨https://arxiv.org/pdf/2111.10061⟩.

Boufous, S., Hatfield, J., Grzebieta, R., 2018. The impact of environmental factors on cycling speed on shared paths. Accid. Anal. Prev. 110, 171–176.
Boulange, C., Gunn, L., Giles-Corti, B., Mavoa, S., Pettit, C., Badland, H., 2017. Examining associations between urban design attributes and transport mode choice for walking, cycling, public transport and private motor vehicle trips. J. Transp. Health 6, 155–166.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Broach, J., Dill, J., Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transp. Res. Part A: Policy Pract. 46 (10), 1730–1740.
Buehler, R., Pucher, J., Bauman, A., 2020. Physical activity from walking and cycling for daily travel in the United States, 2001–2017: Demographic, socioeconomic, and geographic variation. J. Transp. Health 16, 100811.
Celis-Morales, C.A., Lyall, D.M., Welsh, P., Anderson, J., Steell, L., Guo, Y., Maldonado, R., Mackay, D.F., Pell, J.P., Sattar, N., et al., 2017. Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study. bmj 357, j1456.
City of Copenhagen (2014). Better mobility in Copenhagen: ITS Action Plan 2015-2016.
City of New York (2021). Green Wave Progress Report.
Clarry, A., Faghih Imani, A., Miller, E.J., 2019. Where we ride faster? Examining cycling speed using smartphone GPS data. Sustain. Cities Soc. 49, 101594.
Couronne, R., Probst, P., Boulesteix, A.-L., 2018. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinforma. 19 (1), 270.
Department of Environment, Land, Water and Planning, 2017. Plan Melbourne 2017–2050. Victorian Government. ⟨https://www.planning.vic.gov.au/guides-and-resources/strategies-and-initiatives/plan-melbourne/the-plan⟩.
Department of Transport (2019). Movement and Place in Victoria. ⟨https://www.vic.gov.au/movement-and-place-victoria⟩.
Ellison, R.B., Greaves, S., 2011. Travel time competitiveness of cycling in Sydney, Australia. Transp. Res. Rec. 2247 (1), 99–108.
Flugel, S., Hulleberg, N., Fyhri, A., Weber, C., Ævarsson, G., 2019. Empirical speed models for cycling in the Oslo road network. Transportation 46 (4), 1395–1419.
Giles-Corti, B., Saghapour, T., Turrell, G., Gunn, L., Both, A., Lowe, M., Rozek, J., Roberts, R., Hooper, P., Butt, A., et al., 2022. Spatial and socioeconomic inequities in liveability in Australia's 21 largest cities: Does city size matter? Health Place 78, 102899.
Gossling, S., 2020. Why cities need to take road space from cars-and how this could be done. J. Urban Des. 25 (4), 443–448.
Handy, S., van Wee, B., Kroesen, M., 2014. Promoting cycling for transport: Research needs and challenges. Transp. Rev. 34 (1), 4–24.
Jafari, A., Both, A., Singh, D., Gunn, L., Giles-Corti, B., 2022. Building the road network for city-scale active transport simulation models. Simul. Model. Pract. Theory 114, 102398.
Jafari, A., Singh, D., Both, A., Abdollahyar, M., Gunn, L., Pemberton, S., Giles-Corti, B., 2024b. Activity-based and agent-based transport model of Melbourne: an open multi-modal transport simulation model for Greater Melbourne. J. Intell. Transp. Syst. 1–18.
Jafari, A., Pemberton, S., Singh, D., Saghapour, T., Both, A., Gunn, L., Giles-Corti, B., 2024a. Understanding the Impact of City-wide Cycling Corridors on Cycling Mode Share Among Different Demographic Clusters in Greater Melbourne. Research Square, Australia. https://doi.org/10.21203/rs.3.rs-4538181/v1.
Kaziyeva, D., Loidl, M., Wallentin, G., 2021. Simulating spatio-temporal patterns of bicycle flows with an agent-based model. ISPRS Int. J. Geo-Inf. 10 (2), 88.
Kuhnel, N., Thunig, T., Nagel, K., 2018. Implementing an adaptive traffic signal control algorithm in an agent-based transport simulation. Procedia Comput. Sci. 130, 894–899.
Molenberg, F.J., Panter, J., Burdorf, A., van Lenthe, F.J., 2019. A systematic review of the effect of infrastructural interventions to promote cycling: strengthening causal inference from observational data. Int. J. Behav. Nutr. Phys. Act. 16 (1), 1–31.
Palczewska, A., Palczewski, J., Robinson, R.M., Neagu, D., 2013. Interpreting random forest models using a feature contribution method. 2013 IEEE 14th International Conference on Information Reuse & Integration (IRI). IEEE, pp. 112–119.
Paulsen, M., Nagel, K., 2019. Large-scale assignment of congested bicycle traffic using speed heterogeneous agents. Procedia Comput. Sci. 151, 820–825.
Pearson, L., Berkovic, D., Reeder, S., Gabbe, B., Beck, B., 2023. Adults' self-reported barriers and enablers to riding a bike for transport: A systematic review. Transp. Rev. 43 (3), 356–384.
Romanillos, G., Gutierrez, J., 2020. Cyclists do better. analyzing urban cycling operating speeds and accessibility. Int. J. Sustain. Transp. 14 (6), 448–464.
Titze, S., Giles-Corti, B., Knuiman, M.W., Pikora, T.J., Timperio, A., Bull, F.C., van Niel, K., 2010. Associations between intrapersonal and neighborhood environmental characteristics and cycling for transport and recreation in adults: Baseline results from the RESIDE Study. J. Phys. Act. Health 7 (4), 423–431.
Wardman, M., Tight, M., Page, M., 2007. Factors influencing the propensity to cycle to work. Transp. Res. Part A: Policy Pract. 41 (4), 339–350.
Zhang, X., Blokpoel, R., 2018. A scale-up network level study of green wave with speed advice for cycling. ITS World Congr.
Ziemke, D., Metzler, S., Nagel, K., 2018. Bicycle traffic and its interaction with motorized traffic in an agent-based transport simulation framework. Future Gener. Comput. Syst.