Transportation Research Part A 201 (2025) 104679

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Cycling route choice preferences: A taste heterogeneity and exogenous segmentation analysis based on age, gender, Geller typology, and e-bike use

Tanapon Lilasathapornkit[a], Debjit Bhowmick[b], Ben Beck[b], Hao Wu[c], Christopher Pettit[c], Kerry Nice[d], Sachith Seneviratne[d], Mohit Gupta[b], Hai L. Vu[e], Trisalyn Nelson[f], Meead Saberi[a,*]

[a] *Research Centre for Integrated Transport Innovation (rCITI), School of Civil and Environmental Engineering, University of New South Wales, Sydney, 2052, Australia*
[b] *School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia*
[c] *City Futures Research Centre, School of Built Environment, University of New South Wales, Sydney, Australia*
[d] *Transport, Health and Urban Systems Research Lab, Melbourne School of Design, The University of Melbourne, Melbourne, Australia*
[e] *Institute of Transport Studies, Monash University, Melbourne, VIC, Australia*
[f] *Department of Geography, University of California, Santa Barbara, USA*

ARTICLE INFO

Keywords:
Bicycle route choice
Taste heterogeneity
Market segmentation
E-bike
Geller typology

ABSTRACT

A range of factors influences cyclists' route choices, yet infrastructure design often fails to account for the diverse preferences and needs of different groups. This study examines cycling route choice preferences using revealed preference GPS data from Melbourne, Australia. Path Size Logit (PSL) and Mixed Path Size Logit models are estimated to capture path correlation due to overlapping routes and taste heterogeneity in route choice preferences among cyclist groups, segmented by age, gender, e-bike use, and Geller typology. Using a hybrid generalized Breadth-First Search on Link Elimination (BFS-LE) approach, the study enhances the quality and diversity of the generated choice set. Results indicate significant taste heterogeneity in route choices, with distinct preferences across cyclist segments. Risk-averse cyclists, particularly women and the "interested but concerned" group, showed a strong preference for protected bike lanes and off-road paths. In contrast, more confident cyclists, such as "enthused and confident," exhibited greater flexibility and were less sensitive to infrastructure types, slopes, and turns. Traditional bike riders were found to be more sensitive to infrastructure variability compared to e-bike users. Findings also revealed that cyclists, on average, perceived a 1 % increase in the proportion of a route on an off-road bike path as equivalent to a reduction of 80 m in trip length, though this effect varied across individuals. Similarly, a 1 % increase in the proportion of a route on a protected bike lane was, on average, equivalent to a reduction of 61 m, while each additional turn was perceived, on average, as adding 121 m, highlighting the variability in how route complexity influences cyclists' choices. Overall, the study offers valuable insights for urban planners and policymakers, emphasizing the need for inclusive cycling infrastructure that accommodates the diverse preferences of different cyclist groups to encourage broader participation.

* Corresponding author.
E-mail address: meead.saberi@unsw.edu.au (M. Saberi).

https://doi.org/10.1016/j.tra.2025.104679
Received 22 January 2025; Received in revised form 5 September 2025; Accepted 11 September 2025

1. Introduction

Cycling is widely recognized as an essential mode of transportation for building sustainable cities, offering significant environmental, health, and social benefits (Garrard et al., 2012; Götschi et al., 2016; Del Rosario et al., 2024). It reduces traffic congestion, lowers air pollution and carbon emissions, and promotes physical well-being among individuals. As cities around the world increase their investments in cycling infrastructure, understanding cyclists' route choices has become crucial for planning and designing safe, equitable, and efficient active transportation networks. While GPS-based cycling route choice modeling has advanced significantly (Łukawska, 2024), variations in findings across regions-due to differences in infrastructure, traffic conditions, culture, and safety perceptions-highlight the need for more localized studies. Additionally, there remains a limited understanding of taste heterogeneity, referring to variations in preferences, across different cyclist groups segmented by age, gender, cycling experience, and e-bike use. Furthermore, cycling route choice models are valuable for estimating link-level cycling volumes when integrated with cycling demand models. These models effectively capture the complexities associated with cycling route choices (Ryu et al., 2019; Bhowmick et al., 2022; Meister et al., 2024), unlike oversimplified, deterministic, and non-representative route assignment strategies (e.g. shortest distance path) that are prevalent in the literature (Wallentin and Loidl, 2015). Addressing these gaps is crucial for developing more targeted urban planning strategies that cater to the diverse needs of cyclists.

Recent studies leveraging GPS-based revealed preference (RP) data have offered valuable insights into cyclist behavior across different contexts (Łukawska et al., 2023; Łukawska, 2024; Cubells et al., 2023b; Sobhani et al., 2019; Dane et al., 2020). More specifically, Meister et al. (2023) showed that route preferences were shaped by factors such as bicycle infrastructure quality, rider demographic characteristics, and travel purpose. However, these studies often face limitations, including narrow geographical scope, small sample sizes, sampling bias that over-represents frequent and confident cyclists, and difficulties in generalizing findings to the wider population. Additionally, challenges in generating high-quality choice sets and segmenting users by demographic and behavioral factors have persisted, limiting the accuracy and transferability of results. Prato et al. (2018) developed a bicycle route choice model based on GPS-tracked trips in Copenhagen, aiming to integrate cycling considerations into transportation planning. The study highlighted cyclists' diverse preferences for various route features, underscoring the value of detailed cyclist behavior data for sustainable transport planning.

This study addresses these gaps by presenting one of the most extensive datasets on cycling route choice to date, based on 19,782 journeys recorded from 673 cyclists in Greater Melbourne, Australia (Bhowmick et al., 2025). The dataset, comprising 35.6 million GPS data points collected over seven months, offers a unique opportunity to explore route choice behavior at a granular level. In addition to examining traditional factors such as cycling infrastructure and traffic speed, we incorporate behavioral and demographic variables, including gender, age, the Geller typology (a classification system based on individuals' comfort and willingness to cycle under different conditions), and e-bike usage, as well as a traffic stress indicator that integrates bicycle infrastructure type, road hierarchy, traffic volume, and speed limit into a single Level of Traffic Stress (LTS) metric. This enables a deeper understanding of variations and complexities in route choice preferences.

A key component of this study is the exploration of taste heterogeneity in cyclists' route choice preferences, which refers to variations in individual preferences that influence route selection behavior. Using the well-established mixed logit modeling framework with a path size component, we measure the presence of unobserved heterogeneity, capturing differences in individual preferences that are not fully explained by observable characteristics alone. This approach provides a more accurate representation of cyclists' route selection behavior. Additionally, the exogenous segmentation analysis in the study builds upon the mixed logit findings by examining variation in route choice preferences across predefined cyclist segments, offering a more detailed understanding of how demographic and behavioral factors influence choices.

In cycling route choice modeling, generating choice sets is a critical step to ensure that the modeled alternatives reflect realistic and diverse route options available to cyclists (Bovy, 2009; Ben-Akiva et al., 1984; Broach et al., 2010; Rieser-Schüssler et al., 2013; Hess et al., 2015; Tahlyan and Pinjari, 2020). Cycling is strongly determined by infrastructure and individual preferences, necessitating a more sophisticated approach to choice set generation. Previous studies have utilized various methods to generate choice sets, such as labeling approaches (Ben-Akiva et al., 1984; Broach et al., 2010) and Breadth-First Search on Link Elimination (BFS-LE) (Rieser-Schüssler et al., 2013; Hess et al., 2015; Tahlyan and Pinjari, 2020), which is often computed based solely on distance. However, these standard approaches often fail to adequately capture the complexity and diversity of cyclists' preferences, particularly in terms of infrastructure use and spatial variation.

A data-driven path identification approach, known as DDPI, was introduced by Ton et al. (2018) that defines choice sets using observed GPS routes to capture more realistic choices for model estimation. They found that while DDPI effectively represents actual behavior and provides valuable insights without additional network data, it is less effective for predictive purposes due to limitations with out-of-sample data. The BFS-LE approach, on the other hand, produces a choice set with less spatial diversity than labeling approaches but results in the highest percentage of separate bike paths and the fewest intersections in the choice set. This study advances the BFS-LE method by introducing a generalized cost function that accounts for bicycle infrastructure, slope, and distance to better capture cyclists' perceived costs, thereby improving the diversity and realism of generated choice sets.

One of the contributions of the study is the development of a hybrid choice set generation process that combines BFS-LE with a generalized cost function and the Google Directions API. This generalized hybrid approach improves the quality and diversity of the generated choice sets, addressing limitations in previous studies that relied on a single method for generating alternatives. By incorporating the proposed approach, we ensure a comprehensive set of realistic route alternatives, reflecting diverse infrastructure conditions and route attributes, while minimizing false positives and negatives in the choice set.

A core feature of this study is the segmentation of the population into distinct subgroups based on demographic and behavioral factors, such as age, gender, Geller typology (which categorizes cyclists into four types: strong and fearless, enthused and confident, interested but concerned, and no way no how), and whether the cyclist is riding an e-bike or a traditional bike. This distinction between e-bike users and traditional bike users is particularly important due to the unique characteristics and advantages of e-bikes. For instance, e-bikes have the potential to overcome topographical challenges, making routes with steeper slopes more accessible, and they may also appeal to a broader range of users, including those who might otherwise be deterred by physical exertion. This exogenous segmentation enables the estimation of separate route choice models for each population segment, providing an in-depth understanding of how different groups interact with the built environment. We then conduct a comparative analysis that highlights variations in route choice preferences across these groups.

This approach is particularly significant because it moves beyond the limitations of homogenous choice models that assume homogeneity across all cyclists. Homogenous choice models often mask important behavioral differences, which can lead to inaccurate or generalized conclusions about cycling behavior. By segmenting the population, we identify latent behavioral patterns specific to each group, revealing, for example, how factors such as traffic stress, cycling infrastructure, road type, or environmental features impact women differently than men, or how older cyclists' route preferences diverge from younger cyclists. The analysis also enables us to explore how e-bike users, who may experience cycling differently due to factors like energy consumption and effort, make route choices compared to non-e-bike users. For example, modeling results suggest that younger cyclists prioritize flatter and more direct routes, while older cyclists may prefer routes with lower traffic stress. Similarly, e-bike users are found to be less averse to steep gradients than traditional cyclists, given the assistive nature of their bikes.

As the first large-scale study of this kind in Australia, our research provides new empirical evidence on cycling route choice behavior. The findings offer actionable insights for policymakers, urban planners, and transport authorities, helping them design more inclusive cycling infrastructure. This research also contributes to the global discourse on sustainable mobility by addressing key methodological challenges in GPS-based route choice modeling. The study makes four main contributions: (1) introducing a hybrid choice set generation approach that leverages BFS-LE; (2) enhancing the BFS-LE cost calculation by incorporating trip distance, slope, and bicycle infrastructure; (3) estimating path size logit and mixed logit models with a path size component that use LTS as a proxy for traffic characteristics; and (4) capturing variability in cyclists' route preferences and choices through taste heterogeneity analysis and exogenous segmentation.

2. Literature review

The literature on bicycle route choice preferences has expanded rapidly over the past decade, with several studies utilizing GPS-based revealed preference (RP) data to analyze cycling behavior. However, differences in study scope, dataset size, and the inclusion of key demographic and behavioral factors have led to varied findings across regions. In Table 1, we summarize several key studies in this domain, comparing them based on dataset size, the number of participants, and the inclusion of socio-demographic factors such as gender, age, and the use of e-bikes. Many earlier studies, such as Broach et al. (2012) in Portland, Oregon and Hood et al. (2011) in San Francisco, California relied on relatively small sample sizes and did not include critical demographic variables like age or e-bike usage. More recent studies, such as Meister et al. (2023) in Zurich and Dane et al. (2020) in the Noord-Brabant region of The Netherlands, have begun to include gender and e-bike segmentation, but still, the scope remains limited.

Building on this, Łukawska (2024) categorized the factors influencing bicycle route choices into three main groups: network, contextual, and individual factors. Network factors-such as trip length, slope, built environment, motorized traffic volume, speed

Table 1
Summary of GPS-based cycling route choice studies including dataset size, number of participants, and considered cyclist characteristics.

Study	Location	Number of trips collected	Number of trips for modelling	Number of cyclists	Gender	Age	Geller Type	E-bike
Broach et al. (2012)	Portland, Oregon		1449	164	✗	✗	✗	✗
Ghanayim and Bekhor (2018)	Tel Aviv, Israel	618	545	221	✗	✗	✗	✗
Hood et al. (2011)	San Francisco, California, USA	7096	2777	366	✗	✗	✗	✗
Ton et al. (2018)	Amsterdam, The Netherlands	3045	2819	-	✗	✗	✗	✗
Dane et al. (2020)	Noord-Brabant, The Netherlands	17,626	-	742	✓	✓	✗	✓
Sobhani et al. (2019)	Toronto, Canada	5123	500	-	✗	✗	✗	✗
Cubells et al. (2023b)	Barcelona, Spain	-	911	89	✓	✓	✗	✗
Łukawska et al. (2023)	Copenhagen, Denmark	365,813	159,451	8555	✗	✗	✗	✗
Meister et al. (2023)	Zurich, Switzerland	5000	4432	100	✓	✓	✗	✓
This study	Melbourne, Australia	19,782	12,224	646	✓	✓	✓	✓

limits, turns, and land use-have been the focus of many recent studies due to their measurability through GPS data. However, some studies extend beyond these network factors by incorporating contextual factors, including trip purpose, time of day, and weather conditions, which add further complexity to route choices (Hood et al., 2011; Prato et al., 2018; Sobhani et al., 2019). For instance, utilitarian trips that prioritize minimizing time and distance exhibit distinct patterns compared to recreational trips, which often favor more scenic, longer routes (Dill and Gliebe, 2008; Bernardi et al., 2018). Likewise, commute trips tend to be more sensitive to distance while being less influenced by infrastructure improvements (Broach et al., 2012).

In addition to route length, two other key factors influencing cyclists' decisions are bicycle infrastructure (Fitch and Handy, 2020; Fosgerau et al., 2023; Łukawska et al., 2023; Pettit et al., 2024) and slope (Broach et al., 2012; Lin and Wei, 2018; Ryu et al., 2021; Meister et al., 2023). Cyclists generally avoid hilly routes, particularly those with steep slopes, and are often willing to accept detours if the alternative routes offer better cycling facilities. The presence and quality of bicycle infrastructure significantly impact ridership, with better infrastructure encouraging more frequent cycling (Pucher and Buehler, 2017; Arellana et al., 2020; Fosgerau et al., 2023).

Several studies have incorporated cyclists' socio-demographic characteristics, such as gender, age, and income, into analyses (Monsere et al., 2014; Misra and Watkins, 2018; Fitch and Handy, 2020; Meister et al., 2023). Cyclists using e-bikes tend to be less sensitive to slopes and often undertake longer trips compared to those using conventional bikes (Dane et al., 2020; Meister et al., 2023). In contrast, e-scooter riders are more likely to take longer detours than cyclists (Cubells et al., 2023a).

Regular commuters may be less inclined to use bike lanes or paths compared to recreational cyclists, as suggested by Arellana et al. (2020). For utilitarian trips, some studies have found that cyclists may be less willing to deviate from their route to access off-street paths, instead preferring on-street facilities (Larsen and El-Geneidy, 2011). Time constraints, particularly during commuting to work, appear to play a critical role in shaping route choices. During peak hours, Łukawska et al. (2023) observed that commuters are less likely to deviate from the shortest path compared to other utilitarian trips or off-peak travel. Additionally, the routine nature of commuting might heighten cyclists' awareness of time and distance variations between routes. Familiarity with regular routes could also allow commuters to better manage delays and address safety concerns as they arise. Interestingly, Akar and Clifton (2009) and Misra and Watkins (2018) found that experienced cyclists may prefer on-street infrastructure over off-street paths, suggesting a potential divergence in preferences based on experience and cycling purpose.

This study significantly expands upon these previous works by integrating a more comprehensive demographic segmentation approach, capturing cyclists' behavior across characteristics such as age, gender, Geller typology, and e-bike use. This approach allows for a detailed examination of how different population subgroups interact with various built environment factors, such as traffic stress, infrastructure quality, and road characteristics. By using one of the largest datasets to date, collected in Melbourne, Australia, this research provides more robust and in-depth insights into how these factors influence cycling route choices considering taste heterogeneity.

3. Data

We collected GPS data from a diverse group of 673 cyclists across Greater Melbourne, documenting 19,782 cycling journeys, translating to 35.6 million GPS data points (Bhowmick et al., 2025). The data collection took place over seven months, from January 2022 to August 2022. During this period, participants recorded smartphone GPS data, including location coordinates, timestamps, and speeds, for approximately two months. Prior to collecting GPS data, we conducted a survey to gather socio-demographic information such as age, gender, income, occupation, employment status, primary language, bike ownership, and bike type. Additionally, participants were classified into one of the four Geller typologies based on their stated comfort levels while riding in specific types of infrastructure scenarios. These scenario questions were derived from the 'Four Types of Cyclists' tool (Pearson et al., 2022; Dill and McNeil, 2013) with appropriate alterations made for the Australian context. Our data collection approach paired a smartphone app with a Bluetooth beacon attached to participants' bikes, ensuring accurate recording of all biking trips while reducing self-reporting bias and privacy concerns, and minimizing participant dropouts. We developed a sampling framework to ensure collecting data from a representative sample of the underlying adult bike-riding population of Greater Melbourne, enabling an accurate assessment of representativeness. Our study sample closely mirrored the spatial and demographic distribution of the overall bike-riding population. For further details, please refer to Bhowmick et al. (2025). The observed raw cyclists' trajectories are illustrated in Fig. 1(a).

Links or street segments of the underlying bikeable network were classified into different infrastructure categories. Initially, the raw network from OpenStreetMap was spatially merged with publicly available and proprietary datasets, such as motor vehicle volume, slope, and posted speed limits. The network was then classified based on (a) the type of bicycle infrastructure installed on the link, (b) road hierarchy, (c) Bicycling Level of Traffic Stress (LTS)-which combines information on bike infrastructure type, road hierarchy, motor vehicle volume, and posted speed limit-and (d) the combination of bike infrastructure and road class.

We obtained Light Detection and Ranging (LiDAR)-derived Digital Elevation Model (DEM) data at 10 m resolution from the VicMap Elevation LiDAR DEMs Collection (Water and Planning, 2021). This dataset covers approximately 60 % of the state's land area and provides over 99 % coverage of populated areas, with data collected between 2007 and 2023 within the state of Victoria (Vicmap Spatial Services Branch, 2022). Researchers use various approaches to incorporate slope into bicycle route choice models, including classifying slopes into distinct categories such as flat and steep (Broach et al., 2012; Łukawska et al., 2023; Meister et al., 2023), focusing solely on uphill gradients (Zimmermann et al., 2017; Cubells et al., 2023a), or using average slope (Hood et al., 2011; Cho and Shin, 2022). A road segment with a very steep uphill gradient is a strong deterrent, often leading cyclists to detour around

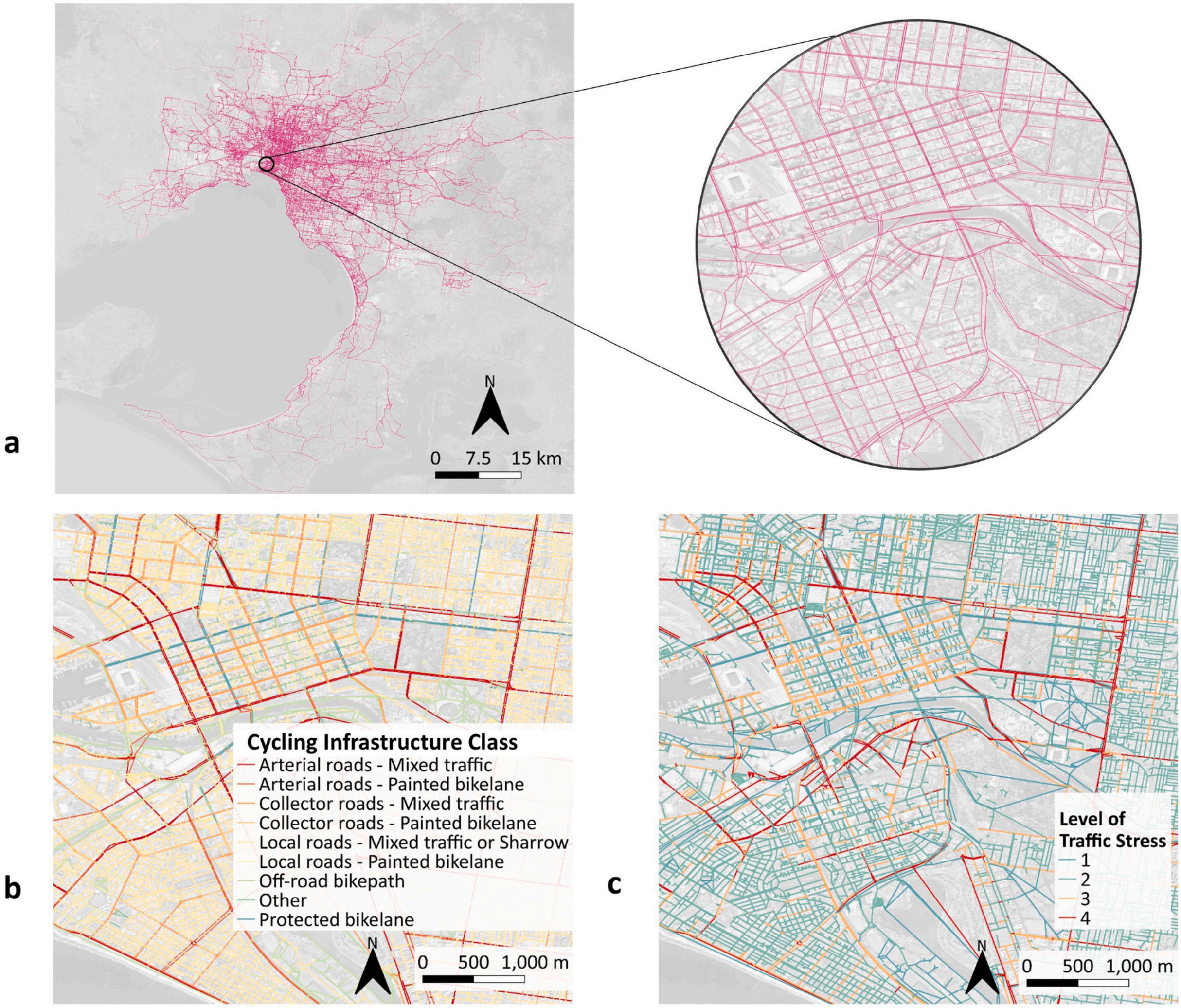

Fig. 1. An overview of the input data used in this study: (a) observed cycling trajectories with a zoomed-in view, (b) cycling infrastructure classes at the link level, and (c) the level of traffic stress (LTS) at the link level, ranging from LTS 1 (low stress) to LTS 4 (high stress), used to represent the comfort and safety of cycling conditions across the network..

the hill (Menghini et al., 2010; Broach et al., 2012; Meister et al., 2023). Since cyclists are particularly sensitive to extreme slopes, we adopt the maximum slope value, similar to Menghini et al. (2010).

We also obtained Point of Interest (POI) data from OpenStreetMap (OSM) using the Pandana Python library (Foti et al., 2012). A POI denotes a location tagged with descriptive features, which can refer to a specific point or area. POIs are categorized into various groups, such as residential, education, healthcare, market, transportation, and financial services (Wibowo et al., 2021), or into four categories: economic, educational, government, and health (Zhuwaki and Coetzee, 2021). Each POI encountered along a path represents an amenity, such as a café, restaurant, post office, or store, typically located at ground level, which tends to attract more pedestrian traffic and potentially cause crowdedness for cyclists. While pedestrians often prefer routes with numerous ground-floor amenities like shops, parks, cafés, and restaurants (Sevtsuk et al., 2021), some past studies suggest that cyclists generally avoid densely populated areas with a high concentration of shops and eateries (Park and Akar, 2019; Desjardins et al., 2022; Cubells et al., 2023a). The OSM dataset includes various nodes, such as road intersections, junctions, and traffic lights, which may not be pertinent to our study. To extract relevant POIs from the OSM, we define each POI as a single coordinate tagged with its amenity type and name. All types of POIs are treated equally within the scope of this research. See Table 2 for the description of the route attributes considered in this study.

Table 2
Attribute description.

Attribute	Description
Length	Route distance or length in meter
Turns	The number of turns along the route
Maximum gradient	The maximum change in elevation over distance among multiple segments on the route
Average gradient	The weighted average change in elevation over distance
Prop. of Arterial roads – Painted bike lane	Proportion of route with Arterial roads – Painted bike lane
Prop. of Arterial roads – Mixed traffic	Proportion of route with Arterial roads – Mixed traffic
Prop. of Local roads – Mixed traffic or sharrow	Proportion of route with Local roads – Mixed traffic or sharrow
Prop. of Collector roads – Mixed traffic	Proportion of route with Collector roads – Mixed traffic
Prop. of Protected bike lane	Proportion of route with Protected bike lane
Prop. of Off-road bike path	Proportion of route with Off-road bike path
Length of LTS 1	Route distance or length in meter with the Level of Traffic Stress Class 1
Length of LTS 2	Route distance or length in meter with the Level of Traffic Stress Class 2
Length of LTS 3	Route distance or length in meter with the Level of Traffic Stress Class 3
Length of LTS 4	Route distance or length in meter with the Level of Traffic Stress Class 4
POIs	The number of points of interest from OSM along the route. Each POI along a path represents one amenity (e.g. a cafe, restaurant, post office, or storefront) on the ground floor that cyclists might find attractive.
Path size	The degree of similarity between alternatives within the choice set. Refer to Equation 8.

Note: [1] The calculation is a weighted average using link lengths as weights across multiple segments on the route.

4. Methodology

4.1. Data processing

Because raw GPS traces often deviate from actual streets due to signal interference from buildings, trees, or urban canyons, we applied a map-matching procedure to align the trajectories with the underlying bicycle network. Using the OSMnx Python package (Boeing, 2017), we extracted the bicycling network from OpenStreetMap as of April 30th, 2022, including all streets and paths accessible to cyclists while excluding freeways and pedestrian-only footpaths. The resulting network contained approximately 785,648 directed links and 339,277 nodes. We then used the `Leuven.MapMatching` Python package (Newson and Krumm, 2009), which implements a probabilistic algorithm to associate each GPS trace with the most likely route on the network. To improve computational efficiency without compromising accuracy, the GPS data were downsampled from 1-second to 15-second intervals, following the method detailed in Bhowmick et al. (2025). A maximum allowable distance of 500 metres between a GPS point and a road segment was applied to improve completeness for trips with erratic traces. For the final matched dataset, the mean distance between GPS points and matched links was 18 metres (minimum: 0 m; maximum: 500 m; standard deviation: 45 m). In total, 19,782 trips were collected, of which over 98 % were successfully matched to the network; unmatched or poorly matched trips were excluded from further analysis. To further assess map-matching quality, we computed the maximum distance between each GPS point and the matched network link per trip. Most trips had small deviations, with about 19,309 trips (97.6 %) exceeding 10 m, 12,520 trips (63.3 %) exceeding 50 m, and 4267 trips (21.6 %) exceeding 200 m. Fig. 2 provides an overview of the data processing workflow. We generated choice sets and calculated route attributes, resulting in an additional 11 % data loss. To ensure that only utilitarian and well-represented trips were used in the modeling process, we applied three criteria, leading to a 30 % data reduction, including the observed distance $L_n < 30$ km; the detour factor $DF_n < 3$; and the ratio of the average distance in the choice set to the observed distance $DD_n < 1.3$ for every trip n in the data. See Appendix A for the sensitivity analysis of the distance difference threshold and its effect on the model outcomes. Note that DF_n was computed using the beeline distance as a proxy for the shortest path. While using actual network distance could improve precision in areas with physical barriers (e.g., lakes), it is unlikely to significantly affect trip classification outcomes. More details are discussed later in Section 4.3. Finally, we randomly split the trips into a 90 % training set and a 10 % testing set.

4.2. Choice set generation

Route choice estimation using revealed preference (RP) data involves identifying the route actually taken by the cyclist (the observed or chosen route) along with a set of plausible alternative routes that may have been considered but ultimately not selected. RP refers to real-world choices observed from GPS data, as opposed to stated preference (SP), which captures hypothetical choices made in response to survey scenarios. The presence of alternatives allows a choice model to evaluate whether and which particular route attributes are systematically related to a higher likelihood of preference. The set of alternatives plays a key role in determining the estimated coefficients. Generating this set of alternative routes presents several challenges. The size and complexity of the choice set can significantly increase the computational task. Additionally, the lack of existing observed bicycle routes necessitates the development of new choice set generation techniques. Common algorithms based on travel time and street hierarchy are not directly applicable, as bicycle travel times are not affected by road speed limits, congestion, and functional class in the same ways as auto travel times (Broach et al., 2012). Furthermore, cycling route choice is influenced by multiple factors, which are not limited to travel

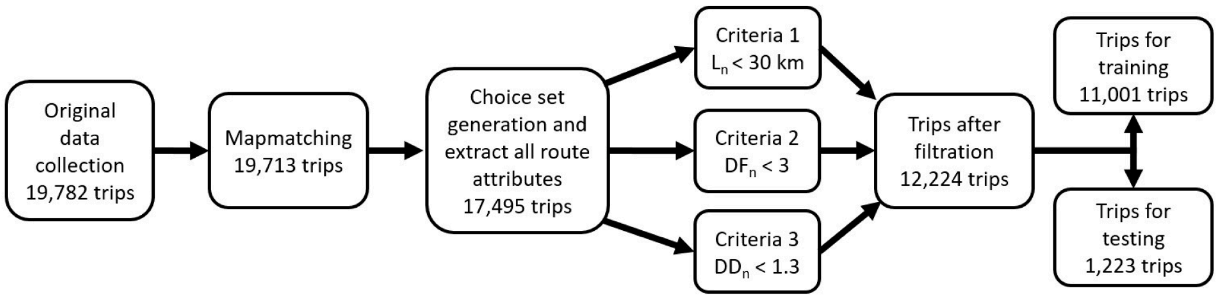

Fig. 2. Data processing flow of the original cycling data including map matching, choice set generation, exclusion of non-utilitarian trips and dividing the data into training vs testing sets.

time, travel distance, or road hierarchy alone, and therefore needs careful consideration when generating appropriate choice sets. Many of the alternatives could overlap with the actual route or each other, thus not constituting independent choices, which would violate the independence of irrelevant alternatives (IIA) property in decision theory. However, in practice, a relatively small number of alternatives (e.g., 1-5) is typically used in route choice models, with alternatives chosen to offer variation in estimated route attributes. Several approaches have been developed for choice set generation, including k-shortest path (Bovy, 2009), doubly stochastic shortest path (Bovy and Fiorenzo-Catalano, 2007; Hood et al., 2011), Double Stochastic Generation Function (DSGF) (Halldórsdóttir et al., 2014; Koch et al., 2019), labeling (Ben-Akiva et al., 1984; Broach et al., 2010), BFS-LE (Rieser-Schüssler et al., 2013; Hess et al., 2015; Tahlyan and Pinjari, 2020), and data-driven approaches (Ton et al., 2017; Lu et al., 2018).

Generating a choice set based on the k-shortest path method is easy to implement but is unlikely to capture all observed routes (Wang et al., 2023). The BFS-LE approach provides a set of shortest routes with less diversity, while the labeling approach offers more diverse routes that better mimic observed behavior in terms of spatial and behavioral patterns, although BFS-LE produces a larger number of alternatives (Ton et al., 2018). Koch et al. (2019) introduced the concept of route complexity to create more diverse choice sets by counting the number of locations where cyclists deviate from the shortest paths. They found that route complexity generated by BFS-LE is significantly affected by network density, while the DSGF method is less influenced by this factor (Koch et al., 2019). DSGF draws random costs and parameters from probability distributions, which adds heterogeneity to the network. However, BFS-LE requires significantly less computation time compared to DSGF (Halldórsdóttir et al., 2014; Koch et al., 2019).

The choice set algorithm inherently produces two types of errors: false negatives, when the algorithm fails to reproduce the chosen alternatives, and false positives, when alternatives that were not considered by the individual are included. In practice, when building a route choice model, a false negative can be entirely mitigated by adding the observed route as one of the alternatives in the choice set. The consequence of a false positive error is not limited to increased computational burden; including extremely poor alternatives may also affect model estimation by distorting the likelihood function value, potentially leading to convergence on suboptimal parameter estimates (Ton et al., 2018). A larger number of alternatives reduces false negatives due to the higher likelihood of matching the observed route but increases false positives because irrelevant alternatives are more likely to be included in the choice set. This study addresses false negatives by including the observed route when absent from the generated choice set, while also reducing false positives by incorporating a generalized cost function that accounts for bicycle infrastructure, slope, and distance as follows:

$$c_a = x_a^D(1 - \gamma_{sep}x_a^{sep} - \gamma_{pnt}x_a^{pnt} + \gamma_{slope}x_a^{slope}) \tag{1}$$

In this formulation, c_a denotes the perceived cost of link a, and x_a^D is the actual distance of the link. The variables x_a^{sep} and x_a^{pnt} represent the proportion of the link with physically separated and painted bike lanes, respectively (both ranging from 0 to 1). The variable x_a^{slope} denotes the average slope of the link, calculated as the rise-over-run ratio (i.e., elevation gain divided by horizontal distance), and is a continuous variable typically ranging from 0 to 0.4 in our dataset, based on urban terrain. The slope term enters the equation with a positive sign to reflect the assumption that steeper gradients increase perceived cycling distance and, consequently, perceived cost. This is consistent with the idea that cyclists are more likely to avoid steep routes, all else being equal. The calibrated parameters obtained through grid search are $\gamma_{sep} = 0.785$, $\gamma_{pnt} = 0.860$, and $\gamma_{slope} = 0.067$. This formulation assumes that both the presence of cycling infrastructure and topographic slope influence how cyclists perceive distance, which in turn affects route choice. The grid search procedure optimized these parameters to maximize alignment between modeled and observed routes, effectively estimating how each feature distorts perceived distance relative to actual distance.

We utilize two methods for generating our choice set: Breadth-First Search on Link Elimination (BFS-LE) and the Google Directions API. BFS-LE involves iterating through the minimum cost path (MCP) search by gradually eliminating links from the MCP, starting at the origin and progressing toward the destination. This method ensures route uniqueness, defined as a lower degree of overlap among alternatives rather than complete dissimilarity, by applying a similarity constraint (Cascetta et al., 1996). Two routes are considered unique if they meet a minimum threshold of dissimilarity. We generate up to five alternatives based on trip length only and five additional alternatives based on trip length, slope, and cycling infrastructure using the cost function described earlier. We find that the cost function, which incorporates multiple route attributes, significantly impacts the diversity of the generated choice sets. Conversely, the Google Directions API provides alternative routes specifically designed for bicycle travel

(Loidl and Hochmair, 2018; Schirck-Matthews et al., 2023). In our hybrid method, up to three routes from the Google API are included in the final choice set, alongside those generated using the BFS-LE approach. Overall for each trip, we generate up to 13 alternative routes. The origin-destination pairs are derived from the collected GPS data. See Fig. 3 for an illustration of several sampled observed and generated alternative routes.

4.3. Choice set evaluation

Two of the measures used to assess the performance of a choice set generation procedure are the coverage and consistency indices (Prato and Bekhor, 2007). Coverage is defined as the percentage of observations for which an algorithm generates a route that satisfies a particular threshold for the overlap measure and can be formulated as follows.

$$\max_{r} \sum_{t \in T} I(O_{t,r} \geq \delta) \tag{2}$$

where $I(.)$ denotes the coverage function which equals 1 if its argument is true and 0 otherwise. δ denotes the overlap threshold which is commonly used at 80 % (Prato and Bekhor, 2007; Halldórsdóttir et al., 2014; Ghanayim and Bekhor, 2018). T denotes all trips.

$O_{t,r}$ denotes the overlap measure, as determined by the choice set generation algorithm r, between the generated choice set and the observed trip route t.

$$O_{t,r} = \frac{L_{t,r}}{L_t} \tag{3}$$

where L_t denotes the length of the observed route for trip t and $L_{t,r}$ denotes the overlapping length between the generated route and the observed route for trip t by algorithm r. Ideally, a choice set generation method would perfectly reproduce the observed behavior by replicating each route link by link, resulting in 100 % coverage for a 100 % overlap threshold. However, actual choice set generation methods only partially replicate the observed behavior and generate varying numbers of routes. The consistency index measures the behavioral consistency of these methods by accounting for the total overlap across all observations. The consistency index of algorithm r can be formulated as follows.

$$CI_r = \frac{\sum_{t \in T} O_{t,r}^{max}}{|T| O^{max}} \tag{4}$$

where $O_{t,r}^{\max}$ denotes the maximum overlap measure from the generated route by algorithm r for the observed choice set of trip t. This refers to a single route among multiple generated routes that has the maximum overlap with the observed route on trip t. $O^{\max}$ denotes 100 % overlap over all the observations for the ideal algorithm. T denotes all trips.

To ensure that all trips considered are utilitarian bicycle trips (riding a bicycle where the origin or destination serves a utility for the person making the trip such as commute, shopping, picking up children, etc.) and that the choice set contains realistic route alternatives, three criteria were used to filter trips for estimating the route choice model. To ensure that all trips considered are utilitarian bicycle trips and that the choice set contains realistic route alternatives, the previously introduced criteria were applied to filter trips for estimating the route choice model. An extensive sensitivity analysis was conducted on these criteria to strike a balance between retaining a larger number of trips and ensuring that the choice model outcomes remained plausible. For instance, the distance parameter in a route choice model is expected to be positive, otherwise this would imply that cyclists prefer longer routes purely for the sake of distance-a behavior inconsistent with utilitarian trips. First, the trip length L_n had to be less than 30 km. It is uncommon for utilitarian bike trips to exceed 30 km (Lißner and Huber, 2021), so this threshold ensured that we captured most or all utilitarian bike trips while filtering out recreational trips, which are often longer. The second criterion was that the detour factor DF_t had to be less than 3. The detour factor, defined as the actual trip length (network distance between the origin and destination covered by the cyclist) divided by the beeline distance (Euclidean distance between the origin and destination), is critical for identifying leisure trips (Lißner and Huber, 2021) and is formulated as follows.

$$DF_t = \frac{L_t}{L_t^{euc}} \tag{5}$$

where L_t^{euc} denotes the Euclidean distance between the origin and destination of the trip t.

This is because utilitarian bike trips typically do not involve unreasonable detours, which are more common in recreational bike trips. The final criterion involves dividing the observed distance by the average distance of the choice set. This criterion is formulated as follows.

$$DD_t = \frac{L_t}{\sum_a^A L_{t,a} / |A_t|} \tag{6}$$

where $|A_t|$ denotes the number of all generated routes or a choice set size for the trip t.

The distance difference threshold was set to less than 1.3. This ensures that the generated choice set routes are realistic alternatives to the observed route. These thresholds were empirically derived after thoroughly investigating the GPS data. With the three criteria in place-a trip length threshold of 30 km, a detour factor threshold of 3, and an average distance difference threshold of 1.3-our final dataset comprises 12,224 trips. The dataset achieves a coverage measure of 30 % at the 80 % overlap threshold and a consistency index of 0.52, as shown in Table 3. The generated choice set performed adequately, including up to 13 alternatives per trip. While this

Table 3
Comparison of coverage measures and consistency indices to evaluate the quality of choice set generation methods proposed in past studies and this study. The table highlights the ability of each method to capture observed choices at varying overlap thresholds (100 %, 90 %, 80 %, and 70 %) and assesses the consistency of the generated sets with observed data.

Methods	Percentage coverage at different overlap thresholds				Consistency
	100 %	90 %	80 %	70 %	Index
BFS-LE in this study before filtering with 17,495 OD pairs	4.39	11.25	25.89	48.96	0.47
BFS-LE in this study before filtering with 12,224 OD pairs	6.10	12.74	30.37	57.97	0.53
Link elimination on the road network from Prato and Bekhor (2007)	58.47	58.47	69.92	81.78	0.872
An algorithm on the road network from Broach et al. (2010)	22.5	29.4	42.3	54.6	
BFS-LE on the road network with three cost attributes from Halldórsdóttir et al. (2014)	67.9	74.8	80.1	84.8	0.895
An algorithm on the bicycle network from Ghanayim and Bekhor (2018)	73.5	77.5	85.0	89.5	
BFS-LE on bicycle network from Ton et al. (2018)	0.9	1.9	3.3	6.2	0.270
An algorithm on the bicycle network from Łukawska et al. (2023)	21	38	56	70	
BFS-LE on bicycle network from Meister et al. (2023)	9	-	-	44	0.63

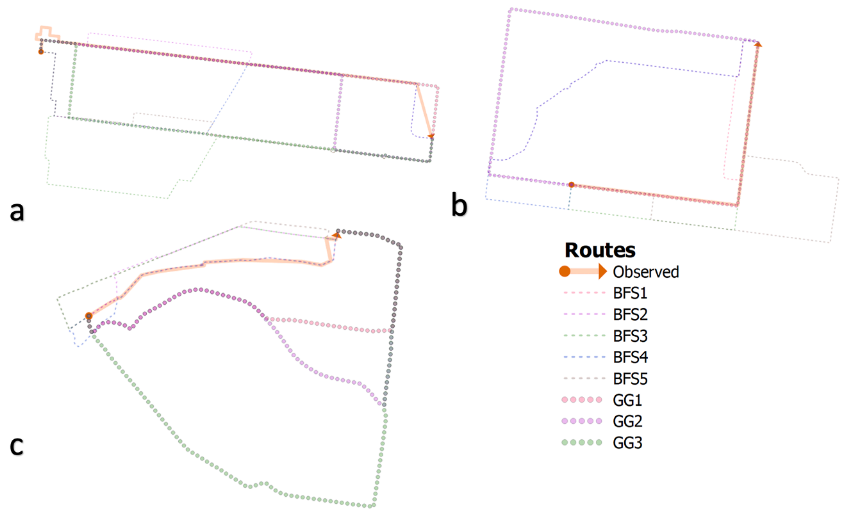

Fig. 3. Examples of observed and generated alternative cycling routes. An orange solid line with an arrow represents the observed trajectory. Dashed lines with various colors represent five alternative routes generated from BFS-LE algorithm ('BFS'). Dotted lines with various colors represent three alternative routes from Google Direction API ('GG'). (a) Sample trip 1 (b) Sample trip 2 (c) Sample trip 3.

is smaller than in some other studies, such as Łukawska et al. (2023), which generated up to 31 routes per trip, the difference primarily reflects variations in modeling choices, particularly in trip filtering criteria and the number of alternatives deemed meaningful for inclusion. For instance, Łukawska et al. (2023) applied stricter filters by excluding trips shorter than 500 m and using a detour factor threshold of $\frac{\pi}{2}$. A larger choice set is not inherently better; rather, it depends on the context, data characteristics, and the intended trade-off between computational efficiency and behavioural realism. Similarly, Halldórsdóttir et al. (2014) used a larger choice set with up to 20 routes per trip.

4.4. Descriptive analysis

The descriptive statistics of the observed chosen routes are presented in Table 4. These statistics are compared between the observed routes and the generated alternatives in the choice set. A description of all attributes is provided in Table 2. The proportions of different types of bicycle infrastructure are calculated as the percentage of the route's total distance that occurs on a specific type of infrastructure. The length of each Level of Traffic Stress (LTS) category is determined based on the distance traveled within that particular LTS. Fig. 4 illustrates that the observed routes and choice sets have similar distributions across most route attributes.

Further analysis indicates that men's average cycling distance is longer, at 6.10 km, compared to women's average of 4.8 km. In a previous study, the average distance for chosen e-bike routes was higher, with a mean of 3.5 km compared to 2.7 km for traditional bikes (Meister et al., 2023). However, in this study, we found no significant difference between the average distances for e-bike and traditional bike routes, with both having a mean of 5.6 km. Cyclists classified as "Interested but concerned" tend to take slightly shorter trips, with a mean distance of 5.5 km, while the "Enthused and confident" or "Strong and fearless" groups average 6 km.

Meister et al. (2023) found that the average slope of the routes taken by e-bikes was steeper, at 0.05 %, compared to 0.03 % for traditional bikes. In contrast, our study did not find a significant difference in the mean slope for e-bikes (0.031 %) compared to

Table 4
Descriptive statistics of attributes across 12,224 cycling trips. The mean, median, and maximum values are reported for the observed routes and the alternatives.

Total number of trips				12,224		
Total number of cyclists				646		

Attribute	Observed			Alternatives in the choice set		
	Mean	Median	Max	Mean	Median	Max
Distance (km)	5.58	3.36	29.87	5.23	3.21	61.08
Prop. of Arterial roads- Painted bike lane (%)	4.74	0.73	79.32	6.92	2.17	78.70
Prop. of Arterial roads- Mixed traffic (%)	11.81	7.07	100	12.55	7.93	90.40
Prop. of Local roads- Mixed traffic or sharrow (%)	42.24	37.54	100	42.28	37.78	100
Prop. of Collector roads - Mixed traffic (%)	9.52	5.17	100	7.92	4.46	92.15
Prop. of Protected bike lane (%)	0.88	0	64.61	0.83	0	59.83
Prop. of Off-road bike path (%)	20.23	11.85	100	16.17	10.04	100
Distance on LTS1 (km)	1.77	0.49	23.75	1.25	0.43	30.95
Distance on LTS2 (km)	2.06	1.43	16.86	2.12	1.49	16.80
Distance on LTS3 (km)	0.85	0.47	9.80	0.91	0.49	14.81
Distance on LTS4 (km)	0.91	0.39	26.49	0.95	0.40	31.87
Max Slope (%)	0.03	0.02	0.93	0.06	0.04	6.44
POI (#)	39.29	15	778	52.31	17	747
Turns (#)	19.09	13	153	22.00	15	232
Path Size	0.37	0.35	0.99	0.35	0.32	0.99

traditional bikes (0.032 %). This difference may partly reflect the underlying topography, as Zurich-where the Meister et al. (2023) study was conducted-is notably hillier than Melbourne, with more pronounced elevation changes. Additionally, we observed that the "Interested but concerned" group tends to avoid routes with higher LTS levels, with mean distances on LTS 3 at 0.83 km and LTS 4 at 0.85 km. Meanwhile, the "Strong and fearless" group had mean distances of 1.19 km on LTS 3 and 1.67 km on LTS 4.

4.5. Random utility route choice model formulation

The second step in our route choice analysis involves estimating utility function coefficients using the Path Size Logit (PSL) and Mixed Logit models, both of which incorporate a path size attribute. The PSL model is an extension of the standard multinomial logit model, designed to address the potential violation of the independence of irrelevant alternatives (IIA) assumption when routes in the choice set overlap. In route choice contexts, overlapping alternatives, where multiple routes share common segments, can lead to correlation among utilities and thus invalidate the IIA property. The path size variable accounts for this by down-weighting routes that heavily overlap with others, reducing their influence in the model. While some overlap between alternatives is unavoidable and even realistic in urban cycling networks, the inclusion of the path size attribute ensures that such overlap is explicitly accounted for in the estimation. The utility of each route is then modelled as a function of its attributes, including distance, bike infrastructure, traffic density, maximum slope, points of interest, and number of turns.

The Mixed Logit is known by many names, e.g., Hybrid Logit, or Kernel Logit, or Random Parameter Logit, and consists of a probit-like component that captures the interdependencies among the alternatives and an identically distributed Gumbel error component (Ben-Akiva and Bolduc, 1996; Bekhor et al., 2002; Hensher and Greene, 2003; Sener et al., 2009). The Mixed Logit with a path size attribute is also an extension of the multinomial logit model that accounts for unobserved heterogeneity in preferences across individuals and corrects for overlapping routes by incorporating the path size attribute (Broach et al., 2012; Chen et al., 2018; Meister et al., 2023). Unlike the PSL model, which assumes fixed preferences, the Mixed Logit allows for random variation in the coefficients of the utility function, enabling a more flexible representation of individual route choice behavior. This model considers the same set of route attributes while introducing random parameters to capture variations in sensitivity to these attributes. By including the path size attribute, the Mixed Logit model addresses potential biases due to route overlap, ensuring a more robust estimation of route choice preferences.

The lack of theoretical guidance for the C-Logit (CL) model (Cascetta et al., 1996) led Ben-Akiva and Bierlaire (1999) to the proposal of the PSL model. The PSL approach is conceptually similar to the CL model, where a correction for overlapping routes is achieved by adding an attribute to the deterministic part of the utility (Bovy et al., 2008; Duncan et al., 2020). In the route choice context, we assume that an overlapping route may not be perceived as a distinct alternative, as a route may contain links shared by several routes. Consequently, the size of a route with one or more shared links may be less than one. Accordingly, we formulate the utility function for the PSL model as follows.

$$U_{i,n,t} = \beta_D X^{D}_{i,n,t} + \sum_{k\in K_{inf}} (\beta_{inf,k} X^{inf,k}_{i,n,t}) + \sum_{s\in S_{LTS}} (\beta_{LTS,s} X^{LTS,s}_{i,n,t}) + \beta_{slope} X^{slope}_{i,n,t} + \beta_{POI} X^{POI}_{i,n,t} + \beta_{turn} X^{turn}_{i,n,t} + \beta_{PS} ln(PS_{i,n,t}) + \epsilon_{i,n,t} \tag{7}$$

where i denotes the route, t denotes the trip, and n the cyclist. β_D denotes the route distance coefficient, and $X^{D}_{i,n,t}$ represents the route distance of alternative i for trip t of cyclist n. $\beta_{inf,k}$ denotes the coefficient for infrastructure class k, and $X^{inf,k}_{i,n,t}$ represents the

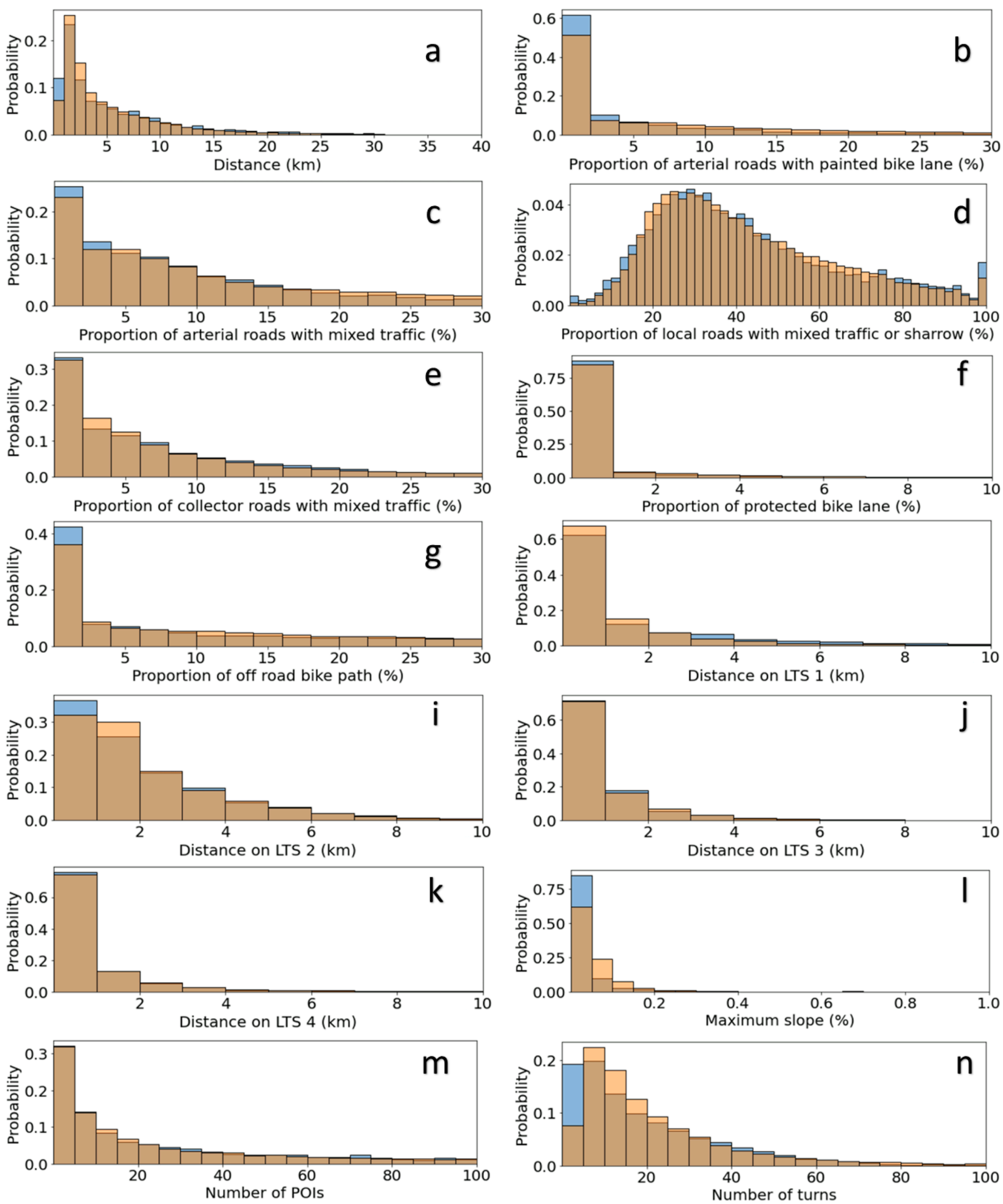

Fig. 4. Distribution of choice set route attributes, including the observed route: (a) Distance, (b) Proportion of arterial roads with painted bike lanes (%), (c) Proportion of arterial roads with mixed traffic (%), (d) Proportion of local roads with mixed traffic or sharrows (%), (e) Proportion of collector roads with mixed traffic (%), (f) Proportion of protected bike lanes (%), (g) Proportion of off-road bike paths (%), (h) Distance on LTS 1 (km), (i) Distance on LTS 2 (km), (j) Distance on LTS 3 (km), (k) Distance on LTS 4 (km), (l) Maximum slope (%), (m) Number of POIs, and (n) Number of turns. black bar charts represent the observed route, while orange bar charts represent the generated choice set. Where the observed value falls within a bin already occupied by alternatives, the orange and black bars overlap, resulting in a brown appearance..

proportion of route i for trip t of cyclist n on infrastructure class k in percentage. K_{inf} consists of six infrastructure classes: arterial roads with painted bike lanes, arterial roads with mixed traffic, local roads with mixed traffic or sharrows, collector roads with mixed traffic, protected bike lanes, and off-road bike paths. $\beta_{LTS,s}$ denotes the coefficient for LTS level s, and $X^{LTS,s}_{i,n,t}$ represents the proportion of route i for trip t of cyclist n on LTS level s in unit of distance. S_{LTS} consists of four levels described in Section 3. See Appendix B for a detailed discussion of how using proportion-based versus distance-based parameters influences the model estimates and their interpretation. β_{slope} denotes the coefficient for maximum slope, and $X^{slope}_{i,n,t}$ represents the maximum slope of route i for trip t of cyclist n. β_{POI} denotes the coefficient for points of interest (POIs), and $X^{POI}_{i,n,t}$ represents the number of POIs along route i for trip t of cyclist n. β_{turn} denotes the coefficient for turns, and $X^{turn}_{i,n,t}$ represents the number of turns along route i for trip t of cyclist n. $\epsilon_{i,n,t}$ denotes an error term with extreme-value-distributed variables for route i in the choice set $C_{n,t}$ associated with trip t of cyclist n. β_{PS} denotes the path size scaling coefficient, and the overlapping path size can be defined as follows.

$$PS_{i,n,t} = \sum_{a\in A} \frac{l_a}{L_i} \frac{1}{\sum_{j\in C_{n,t}} \delta_{aj}} \tag{8}$$

where $PS_{i,n,t}$ denotes the Path Size attribute of route i of cyclist n making trip t, l_a denotes the length of link a, A denotes the set of all links in the network, and L_i is the length of route i, μ denotes a scale factor. $C_{n,t}$ denotes the choice set associated with trip t of cyclist n. δ_{aj} denotes the link-path incidence dummy which equals 1 if link a is on a route/path j otherwise equals to 0.

The utility function is specified slightly differently for the Mixed Logit model, as follows.

$$\begin{aligned} U_{i,n,t} = {} & -\lambda_D X^D_{i,n,t} + \sum_{k\in K_{inf}} (\lambda_{inf,k} X^{inf,k}_{i,n,t}) + \sum_{s\in S_{LTS}} (\lambda_{LTS,s} X^{LTS,s}_{i,n,t}) \\ & - \lambda_{slope} X^{slope}_{i,n,t} - \lambda_{POI} X^{POI}_{i,n,t} - \lambda_{turn} X^{turn}_{i,n,t} - \beta_{PS} \ln(PS_{i,n,t}) + \epsilon_{i,n,t} \end{aligned} \tag{9}$$

$$\lambda_p = \exp(\beta_p + \beta^{std}_p \eta_p) \tag{10}$$

where λp denotes a log-normally distributed scale coefficient for parameter p to account for heterogeneity, while ηp denotes a normally distributed random component. β_p represents the coefficient for parameter p, and β^{std}_p denotes the scale coefficient for the random component. In the Mixed Logit utility function, the signs of the λ values need to be pre-specified. This is achieved by assessing the signs of the coefficients estimated from the PSL model. Additionally, note that the λ values for the infrastructure class and LTS are presented in a summarized form in the formula for brevity. Each λ, depending on the infrastructure class and LTS level, can have either a positive or negative sign that must be determined a prior. Note that the error components in the mixed PSL model are specified at the trip level to account for unobserved heterogeneity arising from trip-specific factors. We then estimate the probability of route i being selected from a choice set as follows.

$$P(i \mid C_{n,t}) = \frac{\exp(U_{i,n,t})}{\sum_{j\in C_{n,t}} \exp(U_{j,n,t})} \tag{11}$$

For the Mixed PSL model, the mean (β_p) and standard deviation (β^{std}_p) of the normal distribution for parameter p must be transformed to a lognormally distributed variable $X = \exp(Y)$. If $Y \sim N(\mu, \sigma)$, the expectation and standard deviation of the lognormal distribution can be expressed as follows (Limpert et al., 2001; Casella and Berger, 2002).

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \tag{12}$$

$$\sigma^2(X) = \exp(\sigma^2 - 1)\exp(2\mu + \sigma^2) \tag{13}$$

where μ denotes the mean of the normally distributed parameter, and σ denotes its standard deviation, as defined earlier in Eq. 10. To clarify the distinction in the formulation between the Logit and Mixed Logit models, Eq. 11 uses the utility function from Eq. 7 for the Logit model, while it uses the utility function from Eq. 9 for the Mixed Logit model.

5. Model estimation results

We present the results of the estimated general PSL and Mixed PSL models based on 12,224 trips made by 646 cyclists in Table C1. Negative coefficients represent disutility, meaning cyclists are less likely to choose routes with such attributes, while positive coefficients indicate higher utility and preference. To make the findings more interpretable, we translate coefficients into distance equivalents, reflecting how specific route attributes affect cyclists' perceived route length, as shown in Table 6.

The PSL model highlights the strong influence of distance on route choice, as cyclists generally prefer shorter routes. However, the presence of cycling infrastructure can compensate for longer distances. For example, each 1 % increase in the proportion of off-road bike paths reduces the perceived distance by approximately 80 m, reflecting cyclists' strong preference for dedicated cycling infrastructure. Similarly, a 1 % increase in protected bike lanes corresponds to a reduction of 61 m in perceived distance. In contrast, arterial roads with painted bike lanes add disutility, with each 1 % increase equivalent to increasing the route length by about 18 m, reinforcing the finding that painted lanes on high-traffic roads do not provide adequate comfort or safety for cyclists. These results align with previous studies, such as Beck et al. (2023) and Meister et al. (2023), which similarly highlight the preference for separated infrastructure over painted lanes or shared roads. Route complexity, measured by the number of turns, significantly reduces utility. Each additional turn adds an equivalent of 121 m to the perceived route distance, indicating that cyclists prefer more direct routes with fewer interruptions. This finding is consistent with Broach et al. (2012), where additional turns were also shown to increase perceived effort. Gradient also plays a major role, as steep slopes are a strong deterrent. The maximum slope coefficient indicates that cyclists perceive uphill segments as significantly increasing route disutility, emphasizing the importance of flat or gently sloping routes. Routes with low traffic stress are strongly preferred. Lower stress environments, represented by Level of Traffic Stress (LTS) 1 and 2, enhance route attractiveness, while high-stress routes (LTS 3 and 4) reduce utility. A kilometer on an LTS 1 route, for instance, is perceived as much shorter than one on a higher-stress route, reinforcing the value of calm and cyclist-friendly environments for encouraging cycling.

The results align closely with prior studies on cyclist behavior. For example, Broach et al. (2012) found that bike paths reduce perceived distance by approximately 16 %, while Meister et al. (2023) reported that separated paths equate to a 36 % reduction in perceived distance. Our findings were consistent with these prior studies, showing that cyclists perceive off-road bike paths and protected lanes as major contributors to reduced effort, while turns and steep gradients remain substantial deterrents. Similarly, Fitch and Handy (2020) estimated that increasing the proportion of bicycle infrastructure along a route from 0 % to 30 % is equivalent to reducing the trip distance by 20 % for bicycle lanes and 22 % for off-street bike paths.

The estimated coefficient for the path size term is negative, implying a positive contribution to utility and a preference for overlapping routes. While this may seem counterintuitive, similar findings have been reported in the literature. Frejinger and Bierlaire (2007) suggest that overlapping routes may offer behavioral benefits such as flexibility and the possibility of switching routes. A similar interpretation was proposed by Hoogendoorn-Lanser et al. (2005) in the context of multimodal route choice. The result may also reflect the composition of the choice set, where many observed routes share segments with other alternatives.

A commonly used measure for evaluating the performance of discrete choice models is first-preference recovery (FPR) (Hauser, 1978; Bass et al., 2011; Łukawska et al., 2023). FPR determines the percentage of instances where the chosen route is also the route with the highest modeled utility. For perfect overlap, the chosen route and the route with the highest probability should be identical, resulting in a 100 % overlap. FPR is defined as the ratio of the probability of the chosen route to the probability of the route with the highest utility. The FPR is formulated as follows.

$$FPR = \frac{\sum_{C_{n,t} \in T} \frac{\sum_{i \in C_{n,t}} P(i|C_{n,t})\delta_i}{\max_{i \in C_{n,t}} P(i|C_{n,t})}}{|T|} \tag{14}$$

where δ_i equals 1 for the chosen route and 0 otherwise. $P(i|C_{n,t})$ denotes the probability of route i in the choice set $C_{n,t}$ for cyclist n and trip t. T denotes all trips. The FPR value for 1223 trips from the test set is 75.24 %, indicating that the estimated PSL model performs well. For comparison, previous studies reported FPR values of 15 % (Hood et al., 2011), 21.1 % (Ton et al., 2018), 88 % (Sobhani et al., 2019), and 44.24 % (Łukawska et al., 2023).

Next, we discuss the Mixed PSL model estimation results, where log-normal distributions are assumed for the scale parameters, incorporating random components to account for cyclists' taste heterogeneity-that is, variation in how different individuals value route attributes. Unlike standard logit models that estimate a single fixed coefficient for each attribute, the Mixed PSL model estimates both the mean and standard deviation of each coefficient, capturing unobserved heterogeneity in preferences across the population. A significant standard deviation indicates that cyclists vary meaningfully in their sensitivity to that attribute.

The inclusion of these random components enhances model performance, as shown by improvements in the final log-likelihood and Akaike Information Criterion (AIC). While the PSL model estimates presented in Table C1 offer insights into whether cyclists perceive specific attributes positively or negatively on average, the coefficients estimated in the Mixed PSL model cannot be directly compared or interpreted in the same way. To make meaningful comparisons, the Mixed PSL model coefficients are transformed to the same scale as the distance equivalent. This transformation uses the mean and standard deviation of their respective distributions, as shown later in Table 6. By aligning coefficients to the distance scale, the model allows for the estimation of the distribution of the distance parameter, enabling an analysis of preference heterogeneity around the mean distance attribute. To further explore systematic heterogeneity, differences in preferences that can be linked to observable characteristics such as gender, age, or cycling confidence, we also estimate separate PSL models for each subgroup in Section 6. This two-tiered approach provides deeper insight into both random and structured variations in cyclists' preferences.

Note that in the mixed PSL model, the distance term is incorporated through a fixed negative transformation, as shown in Eq. 9, ensuring that its effect on utility is always negative by design. This means the sign of the estimated coefficient does not reflect the direction of the utility impact in the same way as in the standard PSL models. While Table 5. presents coefficients from all models side by side for consistency, care should be taken when interpreting the sign of distance-related variables in the mixed PSL model. See Appendix C for a direct comparison of the General PSL and Mixed PSL model coefficients, using Eq. 12.

Overall, the Mixed Logit model estimates provide empirical evidence of taste heterogeneity among cyclists. The significance of several standard deviation coefficients indicates that preferences for specific attributes, such as infrastructure classes and slope, vary widely across individual cyclists. While the coefficients themselves cannot be directly interpreted, their significance demonstrates the existence of individual-level differences in how cyclists perceive and prioritize these attributes. For instance, the variation in preferences for bicycle infrastructure highlights that some cyclists strongly prefer or avoid certain classes of infrastructure, while others may have more moderate or divergent preferences. Similarly, the heterogeneity in sensitivity to slope suggests that cyclists exhibit different levels of tolerance to steepness. The model results do not indicate significant taste heterogeneity for distance and LTS, as the standard deviation coefficients for these attributes were not statistically significant. This suggests that cyclists generally exhibit consistent preferences for distance and LTS, with limited variation across individuals. While these attributes remain critical factors in route choice, the lack of significant heterogeneity implies that their influence is relatively uniform within the sampled population.

Table 5
Cycling route choice model estimation results (N = 11,001): General PSL and Mixed PSL models.

Attribute	Multinomial Logit		Path Size Logit (PSL)		Mixed PSL	
	β	t-value	β	t-value	β	t-value
Distance (km)	−0.16	−20.70**	−1.67	−4.53**	0.15	−2.50*
Distance stdv (km)					−0.12	−1.95*
Prop. of Arterial roads– Painted bikelane			−2.77	−8.54**	1.54	14.75**
Prop. of Arterial roads– Mixed traffic			2.07	8.73**	−6.38	−9.25**
Prop. of Local roads– Mixed traffic or sharrow			3.83	22.50**	1.21	17.90**
Prop. of Collector roads – Mixed traffic			6.03	26.90**	1.35	17.93**
Prop. of Protected bikelane			2.35	3.60**	0.90	5.98**
Prop. of Off-road bike path			5.04	22.19**	1.18	12.91**
Infrastructure stdv					1.46	18.39**
Distance on LTS1 (km)			1.83	4.94**	0.42	141.96**
Distance on LTS2 (km)			1.94	5.32**	0.39	161.18**
Distance on LTS3 (km)			1.51	4.12**	0.01	94.31**
Distance on LTS4 (km)			8e-4	2.08*	−16.05	−26.31**
LTS stdv (km)					0.01	0.23
Max Slope (%)			−1.66	−12.92**	1.32	151.35**
Max Slope stdv (%)					2.21	30.92**
POI (1000s)			−6.83	−15.22**	1.83	13.37**
POI stdv (1000s)					1.07	9.19**
Turns (100s)			−7.76	−34.54**	2.33	53.33**
Turns stdv (100s)					−0.80	−14.29**
Path Size			−1.65	−15.19**	3.89	20.91**
Rho-squared-bar	0.007		0.23		0.31	
Final log-likelihood	−25,429		−19,658		−17,698	
Akaike Information Criterion	50,859		39,345		35,437	

Note: ** indicates significant values at the 1 % level (** $p < 0.01$). * indicates significant values at the 10 % level (* $p < 0.1$).

However, the observed taste heterogeneity for other attributes highlights the importance of considering individual differences when analyzing route choice behavior, as further explored through exogenous segmentation models described in the next section.

6. Exogenous segmentation models

In this section, we explore heterogeneity in cyclist preferences by estimating exogenous models across subgroups defined by gender, age group, Geller typology, and bicycle type (traditional vs. e-bike). Key factors such as maximum slope, number of turns, and types of cycling infrastructure often have varying impacts across these groups, while Level of Traffic Stress (LTS) shows relatively marginal effects, consistent with the results from the general Mixed PSL model.

To enable direct comparison of model coefficients across groups, we present distance equivalence values in Table 6 (and illustrated later in Fig. 6). These values represent the amount of route length (in meters) required for a given attribute to have the same utility impact as one meter of travel distance in the general model. Positive values indicate stronger preference for the attribute within the subgroup, while negative values suggest a tendency to avoid it.

The values and standard deviations reported in the Mixed PSL model demonstrate substantial heterogeneity in cyclists' sensitivity to route attributes. For instance, the coefficient for off-road bike paths is -80 (significant at the 1 % level), indicating a strong overall preference, while the associated standard deviation of +65 suggests wide variability across user groups. Table 6 expresses these effects in terms of equivalent distance: for female cyclists, a 1 % increase in the proportion of off-road paths is valued similarly to a 17-meter reduction in route length. For male cyclists, the equivalent is even greater (31 m) indicating stronger preference for off-road infrastructure.

6.1. Gender

We first examine the effect of gender by segmenting the sampled population into men (67 %) and women (33 %) subgroups, as illustrated in Fig. 5(a). The estimated PSL models for both groups are summarized in Table 7, with the equivalent distances provided in Table 6 and illustrated in Fig. 6(a).

The analysis highlights several gender differences in route preferences among cyclists. Women cyclists exhibit greater avoidance of arterial roads with painted bike lanes (+18 m) compared to men cyclists (+10 m), suggesting differing levels of comfort with this type of infrastructure. Local roads in mixed traffic or sharrows are preferred by both genders, with women showing a weaker preference (-15 m) compared to men (-25 m). Similarly, both women (-21 m) and men (-41 m) show a preference for collector roads in mixed traffic, with men demonstrating a stronger preference.

Protected bike lanes are preferred by both genders, with men exhibiting a stronger preference (-20 m) compared to women (-3 m). Off-road bike paths are also preferred by both groups, though men's preference is greater (-31 m) than women's (-17 m). Both genders

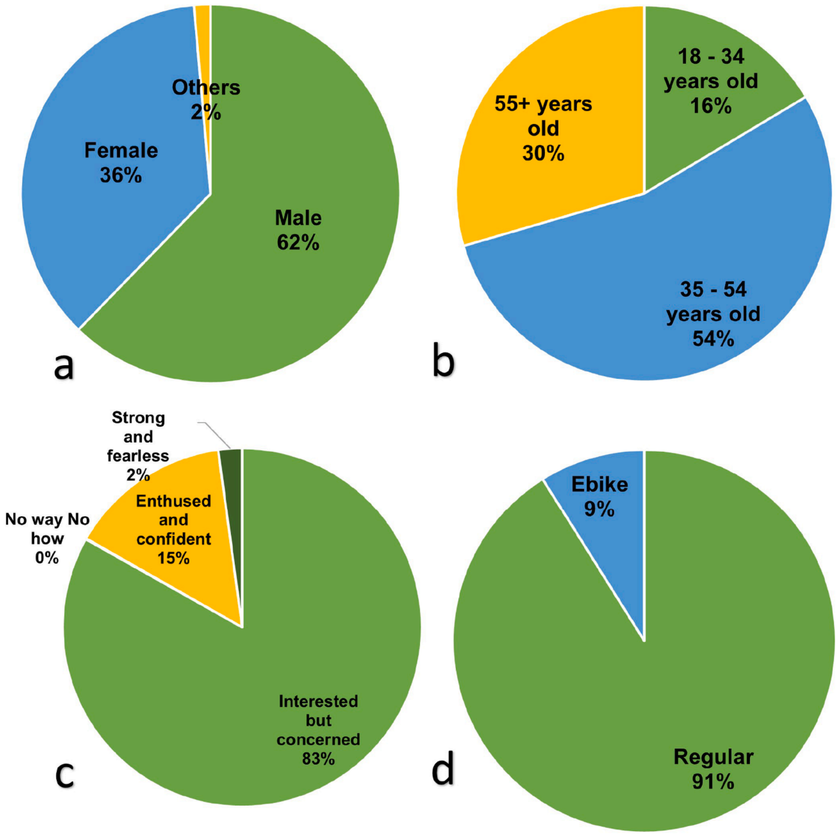

Fig. 5. Demographic segmentation across 646 participating cyclists based on (a) Gender (b) Age group (c) Geller typology (d) Bicycle type.

show a consistent preference for low-stress environments (LTS-1 and LTS-2) and avoidance of high-stress environments (LTS-4), with women demonstrating slightly stronger avoidance of LTS-4 routes compared to men.

In terms of route characteristics, men exhibit higher sensitivity to routes with a greater number of turns (+56 m) compared to women (+23 m). Women also show greater avoidance of steep slopes (+12.2 m) compared to men (+6.8 m). Both genders exhibit similar avoidance of routes with a higher density of POIs, with women reporting +3.1 m and men +3.3 m.

6.2. Age groups

We next explore the effect of age by segmenting the sampled population into three age groups: 18–34 years old (19 %), 35–54 years old (52 %), and 55+ years old (29 %) as shown in Fig. 5(b). The estimated PSL models for the three population segments are shown in Table 8, and the equivalent to distance measurements are shown in Table 6 and illustrated in Fig. 6(b).

The analysis reveals distinct route preferences across age groups, with significant variations in how cyclists respond to distance, slope, and infrastructure characteristics. For the youngest group (18-34 years), the distance attribute was found to be positive but not statistically significant. In contrast, both the middle-aged (35-54 years) and older (55+ years) groups show significant sensitivity to distance, indicating a stronger preference for shorter, more direct routes.

Sensitivity to slopes shows a counter-intuitive pattern. The middle-aged group exhibits greater avoidance of slopes (+9.2 m) compared to the older group (+4.4 m). This suggests differences in how these groups navigate sloped routes, with middle-aged cyclists showing a stronger aversion to steep inclines.

Avoidance of arterial roads with painted bike lanes decreases with age. Younger cyclists (18-34 years) show the highest avoidance (+17 m), followed by the middle-aged group (+15 m) and the older group (+5 m). For protected bike lanes, middle-aged cyclists show a clear preference (–14 m), while older cyclists exhibit a non-significant positive response (+9 m).

Off-road bike paths are strongly preferred by both the middle-aged (–24 m) and older (-20 m) groups, with the middle-aged group showing a slightly stronger preference. For routes with a higher number of POIs, middle-aged cyclists (+3.3 m) exhibit higher sensitivity compared to the older group (+2.7 m). Both groups show an aversion to routes with frequent turns, with older cyclists showing greater sensitivity (+36 m) compared to the middle-aged group (+29 m).

Table 6
Comparison of the equivalent distance metrics [m] estimated from the PSL route choice models across different sub-groups by gender, age group, Geller typology, and bicycle type, against the equivalent distance metrics estimated from the general Mixed PSL model.

Variable (unit)	General	Gender		Age groups		Geller Typology		Bicycle Type	
	(Mixed PSL)	Male	Female	35–54	55 +	Enthused. OR Strong.	Inter. concerned	Traditional	E-bike
Distance mean (m)	+ 1*	+ 1**	+ 1**	+ 1**	+ 1**	+ 1**	+ 1**	+ 1**	+ 1**
Distance stdv (m)	+ 0.12*								
Prop. of Arterial roads–Painted bikelane	+ 115	+ 10**	+ 18**	+ 15**	+ 5**	+ 4*	+ 20**	+ 17**	+ 5**
Prop. of Arterial roads–Mixed traffic	0**	−20**	−4**	−7**	−12**	−2	−16**	−13**	−2**
Prop. of Local roads–Mixed traffic or sharrow	−83**	−25**	−15**	−18**	−18**	−8**	−28**	−25**	−4**
Prop. of Collector roads – Mixed traffic	−96**	−41**	−21**	−27**	−23**	−18**	−41**	−39**	−7**
Prop. of Protected bikelane	−61**	−20**	−3*	−14**	+ 9	+ 18*	−21**	−14**	0
Prop. of Off-road bike path	−80**	−31**	−17**	−24**	−20**	−5**	−38**	−33**	−4**
Infrastructure stdv	+ 65								
Distance on LTS 1 (m)	−1.3**	−1.1**	−1.1**	−1.0**	−1.1**	−1.1**	−1.1**	−1.1**	−1.1**
Distance on LTS 2 (m)	−1.3**	−1.2**	−1.1**	−1.1**	−1.2**	−1.1**	−1.2**	−1.2**	−1.0**
Distance on LTS 3 (m)	−0.9**	−0.9**	−1.0**	−1.0**	−0.9**	−0.9**	−0.9**	−0.9**	−1.0**
Distance on LTS 4 (m)	0**	0*	0	0**	0*	0**	0	0*	0**
Distance on LTS stdv	0								
Max Slope mean (1000 s of %)	+ 37**	+ 6.8**	+ 12.2**	+ 9.2**	+ 4.4**	+ 7.4**	+ 10.6**	+ 10.3**	+ 2.8**
Max Slope std (1000 s of %)	+ 420**								
POI mean (#)	+ 9.4**	+ 3.3**	+ 3.1**	+ 3.3**	+ 2.7**	+ 1.2**	+ 5.1**	+ 4.9**	−0.1**
POI stdv (#)	+ 13.7**								
Turns mean (#)	+ 121**	+ 56**	+ 23**	+ 29**	+ 36**	+ 21**	+ 53**	+ 52**	+ 8**
Turns stdv (#)	+ 114**								

Note: ** indicates significant values at the 1 % level (** $p < 0.01$). * indicates significant values at the 10 % level (* $p < 0.1$).

6.3. Geller typology

We further explore cyclists' behavior inspired by Geller's typology by dividing the sampled population into different groups, as shown in Fig. 5(c). Due to sample size limitations, we were only able to classify cyclists into two groups: "Enthused and Confident" or "Strong and Fearless", and "Interested but Concerned". The estimated PSL models for both groups are presented in Table 9, and the equivalent distance metrics are shown in Table 6 and illustrated in Fig. 6(c).

The analysis reveals clear differences in route preferences between the Interested but Concerned group and the Enthused and Confident or Strong and Fearless group. The Interested but Concerned cyclists show a strong preference for off-road bike paths (-38 m) and protected bike lanes (-21 m), indicating high value placed on physically separated infrastructure. In contrast, the Enthused and Confident or Strong and Fearless cyclists exhibit a weaker preference for off-road bike paths (-5 m) and even avoidance of protected bike lanes (+ 18 m). This reversal in sign underscores a fundamental difference in how infrastructure is perceived among different cyclists. While more cautious riders seek the comfort and safety of segregation, more confident cyclists may view protected lanes as limiting their speed or route flexibility.

Sensitivity to slopes differs between the groups. The Interested but Concerned group shows higher sensitivity to slopes (+ 10.6 m) compared to the Enthused and Confident or Strong and Fearless group (+ 7.4 m). Similarly, the Interested but Concerned group demonstrates greater sensitivity to turns (+ 53 m) than the Enthused and Confident or Strong and Fearless group (+ 21 m).

Routes with higher numbers of POIs also reveal differing preferences. The Enthused and Confident or Strong and Fearless cyclists exhibit stronger avoidance of POI-dense areas (+ 5.1 m) compared to the Interested but Concerned group (+ 1.2 m).

Overall, the Interested but Concerned group exhibits stronger preferences for segregated infrastructure and lower complexity in routes, while the Enthused and Confident or Strong and Fearless group shows more tolerance for challenging conditions and less reliance on dedicated cycling facilities.

6.4. Electric bicycle

In this section, we further explore the behaviour of cyclists by segmenting the sampled population into traditional bike riders (90 %) and e-bike riders (10 %) as shown in Fig. 5(d). The estimated PSL models for both groups are summarized in Table 10 and the equivalent to distance measurements are shown in Table 6 and illustrated in Fig. 6(d).

The analysis reveals clear differences in sensitivity to route characteristics between traditional bike riders and e-bike riders. Traditional bike riders exhibit significantly higher sensitivity to steep slopes (+ 10.3 m) compared to e-bike riders (+ 2.8 m). Similarly,

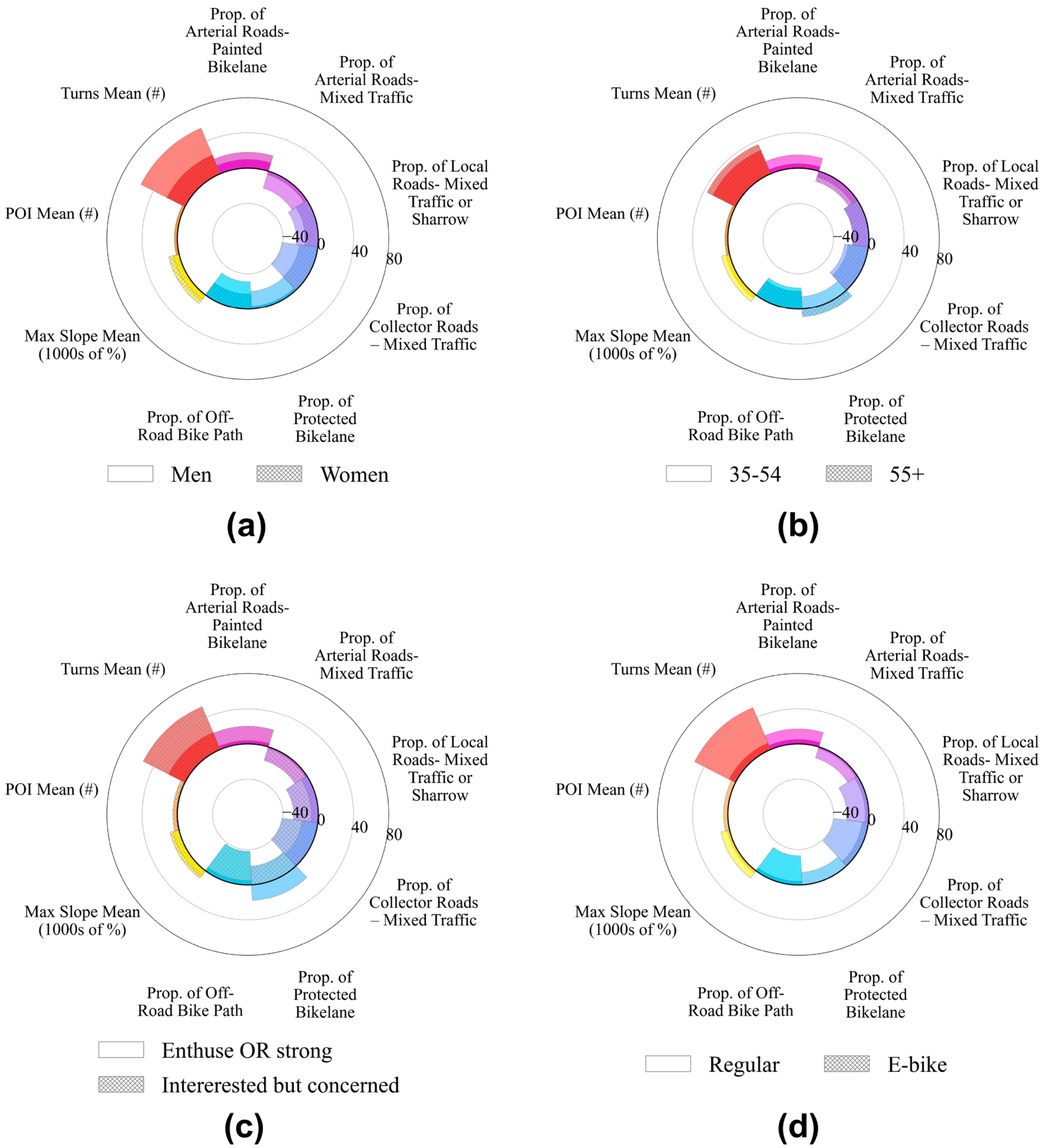

Fig. 6. Visual comparison of the equivalent distance metrics [m] based on (a) Gender, (b) Age group, (c) Geller typology, and (d) Bicycle type. Negative values indicate a reduction in perceived route distance, while positive values indicate an increase in perceived route distance.

traditional bike riders are more affected by frequent turns (+52 m) than e-bike riders (+8 m). Routes with a higher density of POIs also show greater avoidance among traditional bike riders (+4.9 m), whereas e-bike riders exhibit minimal sensitivity to these areas (−0.1 m).

Table 7
Cycling route choice PSL model estimation results: Exogenous segmentation by gender.

Attribute	Male		Female	
	β	t-value	β	t-value
Distance (km)	−1.53	−3.48**	−2.75	−2.93**
Prop. of Arterial roads- Painted bikelane	−1.60	−3.86**	−4.97	−9.38**
Prop. of Arterial roads- Mixed traffic	3.00	10.48**	1.01	2.20*
Prop. of Local roads- Mixed traffic or sharrow	3.77	17.24**	4.24	15.31**
Prop. of Collector roads - Mixed traffic	6.29	22.53**	5.80	14.92**
Prop. of Protected bikelane	3.03	3.25**	0.92	0.92
Prop. of Off-road bike path	4.78	17.00**	4.76	11.09**
Distance on LTS 1 (km)	1.72	3.89**	3.10	3.25**
Distance on LTS 2 (km)	1.78	4.07**	3.08	3.28**
Distance on LTS 3 (km)	1.37	3.12**	2.63	2.81**
Distance on LTS 4 (km)	0.0	1.49	0.0	1.71
Max Slope (%)	−1.04	−8.76**	−3.36	−10.93**
POI (#)	−5.04	−9.92**	−8.50	−10.08**
Turns (#)	−8.62	−29.65**	−6.25	−16.78**
Path Size	−1.56	−11.91**	−2.21	−10.74**
Rho-squared-bar	0.217		0.293	
Final log-likelihood	−12,400		−6,613	
Akaike Information Criterion	24,829		13,257	
Number of trips	6,861		3,980	

Note: ** indicates significant values at the 1 % level (** $p < 0.01$). * indicates significant values at the 10 % level (* $p < 0.1$).

Table 8
Cycling route choice PSL model estimation results: Exogenous segmentation by age.

Attribute	18–34 years old		35–54 years old		55 + years old	
	β	t-value	β	t-value	β	t-value
Distance (km)	0.33	0.48	−2.20	−4.83**	−2.85	−2.93**
Prop. of Arterial roads- Painted bikelane	−3.58	−4.27**	−3.21	−7.73**	−1.34	−1.96**
Prop. of Arterial roads- Mixed traffic	1.91	3.78**	1.50	4.39**	3.42	7.48**
Prop. of Local roads- Mixed traffic or sharrow	1.30	3.46**	3.92	16.79**	5.11	14.67**
Prop. of Collector roads - Mixed traffic	5.79	11.33**	5.92	19.21**	6.59	14.97**
Prop. of Protected bikelane	2.99	2.61**	3.07	3.66**	−2.66	−1.92*
Prop. of Off-road bike path	2.98	5.54**	5.24	16.51**	5.81	13.86**
Distance on LTS 1 (km)	−0.19	−0.29	2.28	4.95**	3.20	3.27**
Distance on LTS 2 (km)	−0.29	−0.44	2.43	5.35**	3.40	3.48**
Distance on LTS 3 (km)	−0.57	−0.83	2.11	4.64**	2.62	2.71**
Distance on LTS 4 (km)	0.0	−2.18*	0.0	2.93**	0.0	2.02*
Max Slope (%)	−1.54	−4.63**	−2.03	−9.27**	−1.25	−8.58**
POI (1000 s of #)	−5.99	−6.77**	−7.16	−12.29**	−7.70	−6.03**
Turns (100 s of #)	−8.27	−14.43**	−6.48	−22.49**	−10.12	−21.20**
Path Size	−1.95	−7.30**	−1.78	−11.88**	−1.56	−7.95**
Rho-squared-bar	0.217		0.221		0.283	
Final log-likelihood	−3,325		−10,748		−5,392	
Akaike Information Criterion	6,680		21,526		10,814	
Number of trips	1,823		5,921		3,257	

Note: ** indicates significant values at the 1 % level (** $p < 0.01$). * indicates significant values at the 10 % level (* $p < 0.1$).

In terms of infrastructure preferences, distinct differences emerge between e-bike and traditional bike riders, reflecting their varying dependencies on route characteristics and infrastructure quality. Traditional bike riders display a stronger preference for off-road bike paths (−33 m) compared to e-bike riders (−4 m), highlighting the greater reliance of traditional bike riders on dedicated cycling infrastructure for comfort and safety.

Traditional bike riders also exhibit a clear preference for protected bike lanes (−14 m), whereas e-bike riders demonstrate no statistically significant preference for this type of infrastructure.

Both groups strongly avoid arterial roads with painted bike lanes, though traditional bike riders show a higher sensitivity (+ 17 m) compared to e-bike riders (+ 5 m). This shared aversion highlights the inadequacy of painted bike lanes on high-traffic roads in providing sufficient safety and comfort for cyclists.

For local roads with mixed traffic or sharrows, traditional bike riders perceive a stronger benefit (−25 m) compared to e-bike riders (−4 m). Similarly, collector roads with mixed traffic are associated with a greater perceived reduction in distance for traditional bike

Table 9
Cycling route choice PSL model estimation results: Exogenous segmentation by Geller typology.

Attribute	Enthused and confident OR strong and fearless		Interested but concerned	
	β	t-value	β	t-value
Distance (km)	−3.22	−3.33**	−1.48	−3.85**
Prop. of Arterial roads- Painted bikelane	−1.41	−1.92*	−2.94	−8.12**
Prop. of Arterial roads- Mixed traffic	0.55	1.06	2.43	9.02**
Prop. of Local roads- Mixed traffic or sharrow	2.68	6.72**	4.10	21.60**
Prop. of Collector roads - Mixed traffic	5.90	11.27**	6.10	24.54**
Prop. of Protected bikelane	−5.93	−2.12*	3.12	4.61**
Prop. of Off-road bike path	1.67	3.02**	5.65	22.35**
Distance on LTS 1 (km)	3.55	3.63**	1.61	4.16**
Distance on LTS 2 (km)	3.48	3.63**	1.75	4.58**
Distance on LTS 3 (km)	2.87	3.00**	1.35	3.53**
Distance on LTS 4 (km)	0.0	2.77**	0.0	1.27
Max Slope (%)	−2.38	−5.46**	−1.57	−11.87**
POI (1000 s of #)	−3.95	−4.08**	−7.53	−14.62**
Turns (100 s of #)	−6.83	−12.03*	−7.85	−32.02**
Path Size	−2.07	−7.45**	−1.57	−13.21**
Rho-squared-bar	0.200		0.242	
Final log-likelihood	−3,372		−16,160	
Akaike Information Criterion	6,775		32,349	
Number of trips	1,840		9,154	

Note: ** indicates significant values at the 1 % level (** p < 0.01). * indicates significant values at the 10 % level (* p < 0.1).

Table 10
Cycling route choice PSL model estimation results: Exogenous segmentation by bicycle type.

Attribute	Traditional bike		Electric bike	
	β	t-value	β	t-value
Distance (km)	−1.50	−4.03**	−10.86	−4.43**
Prop. of Arterial roads- Painted bikelane	−2.62	−7.74**	−5.51	−4.34**
Prop. of Arterial roads- Mixed traffic	2.02	8.09**	2.00	2.55**
Prop. of Local roads- Mixed traffic or sharrow	3.70	20.52**	4.82	9.20**
Prop. of Collector roads - Mixed traffic	5.88	24.88**	7.15	9.83**
Prop. of Protected bikelane	2.17	3.10**	−0.25	−0.13
Prop. of Off-road bike path	4.94	20.99**	4.65	4.82**
Distance on LTS 1 (km)	1.65	4.40**	11.55	4.55**
Distance on LTS 2 (km)	1.78	4.80**	11.25	4.58**
Distance on LTS 3 (km)	1.35	3.64**	10.60	4.36**
Distance on LTS 4 (km)	0.0	1.62	0.01	4.16**
Max Slope (%)	−1.54	−11.95**	−3.03	−5.80**
POI (1000 s of #)	−7.36	−15.50**	1.06	0.72
Turns (100 s of #)	−7.73	−33.02**	−8.55	−10.56**
Path Size	−1.70	−15.06**	−1.86	−4.37**
Rho-squared-bar	0.225		0.318	
Final log-likelihood	−18,042		−1,543	
Akaike Information Criterion	36,114		3,117	
Number of trips	10,023		978	

Note: ** indicates significant values at the 1 % level (** p < 0.01). * indicates significant values at the 10 % level (* p < 0.1).

riders (−39 m) compared to e-bike riders (−7 m). These results indicate that while both groups value lower-stress mixed-traffic roads, traditional bike riders derive substantially greater utility from these environments.

7. Discussion

This study offers critical empirical insights into the factors influencing cyclists' route choices, revealing both expected patterns and new dynamics across gender, age, Geller typology, and bike types. The study provides evidence for the existence of significant taste heterogeneity in cycling route choice preferences. These findings have important implications for sustainable urban planning, particularly in designing infrastructure that caters to the diverse needs of cyclists and encourages greater uptake. A key finding is that perceptions of safety, as reflected in infrastructure choices, play a significant role in shaping cyclists' behavior.

The study reveals that arterial roads with painted bike lanes are more strongly avoided by cyclists than arterial roads with mixed traffic, a finding that may seem counterintuitive but likely reflects concerns over safety consistent with a few past studies in the literature (Beck et al., 2016, 2019). Painted lanes, while technically designated as cycling infrastructure, do not offer physical separation and often place riders close to fast-moving vehicles. As a result, cyclists may avoid these segments more than arterial roads with no marked infrastructure, which may be more prevalent or offer greater continuity. The preference for local roads with mixed traffic or sharrows further underscores the value cyclists place on lower-traffic environments over infrastructure that offers little or no perceived protection.

An interesting observation emerging from the analysis is that infrastructure-related attributes exhibit substantial heterogeneity across user groups, as seen in the high standard deviations in the Mixed PSL model. In contrast, LTS attributes show comparatively little variation in preference. This suggests that cyclists respond more strongly to specific physical design elements (e.g., protected vs. painted lanes) than to generalized stress levels, emphasizing the importance of visible, high-quality infrastructure in shaping behavior.

7.1. Gender-based route preferences

Building on the results presented in Section 6.1, we interpret gendered route preferences through the lens of safety perception, physical exertion, and infrastructure availability. Women's stronger avoidance of painted bike lanes, for instance, aligns with previous studies suggesting heightened sensitivity to traffic proximity and concerns about inadequate protection (Aldred et al., 2017; Pearson et al., 2023).

While both genders show a preference for protected bike lanes and off-road paths, men exhibit a stronger preference for these facilities. At first glance, this may seem counter-intuitive, but it could reflect differences in route planning strategies. For instance, men may be more willing to detour from the shortest path to access higher-quality infrastructure, or may have greater flexibility in destination or time constraints. On the other hand, women's weaker preferences could be shaped by the availability and connectivity of these facilities. If off-road paths or protected bike lanes do not connect to commonly accessed destinations, their practical value, and thus their influence on route choice, may be limited.

Gender differences also emerge in sensitivity to slope and route complexity. Women demonstrate stronger avoidance of steep slopes, likely reflecting a greater emphasis on comfort and accessibility. Interestingly, men are more sensitive to the number of turns, suggesting a preference for more direct and efficient routes, while women may prioritize other aspects such as perceived safety, even if it means navigating a less direct path (Aldred et al., 2017).

Although both groups tend to avoid routes with high POI densities, the small difference in sensitivity may reflect slightly different route priorities, efficiency for men and safety or predictability for women (Pearson et al., 2023). These differences suggest that gender plays a meaningful role in shaping how cyclists weigh trade-offs between comfort, safety, directness, and infrastructure availability.

7.2. Age-based route preferences

Age-related differences in route preferences, as presented in Section 6.2, suggest the influence of physical capacity, infrastructure familiarity, and safety perceptions. While younger cyclists may prioritize flexibility and exploration, older cyclists tend to prefer routes with fewer turns and lower gradients, reflecting concerns over exertion and stability. Results suggest that infrastructure design that minimizes physical demands is particularly important for aging populations.

Younger cyclists show limited sensitivity to distance and slope, suggesting that physical demands and route directness may be less critical for this group. This could reflect a greater physical capacity or a more recreational or exploratory cycling pattern. In contrast, middle-aged and older cyclists demonstrate stronger preferences for shorter, flatter, and more direct routes, highlighting the importance of minimizing physical strain in supporting cycling uptake among these groups. Interestingly, middle-aged cyclists exhibit stronger slope aversion than older cyclists, despite having greater physical capacity. This may reflect different motivations-middle-aged individuals may cycle for mixed utilitarian purposes (e.g., mixing commuting and recreation), placing higher value on energy efficiency and convenience, while older cyclists may prioritize other factors such as safety or familiarity, even if routes are not physically optimal.

Differences in infrastructure preferences also emerge. Middle-aged cyclists show a clear preference for protected bike lanes, while older cyclists do not respond significantly to this feature. This could indicate that older riders tend to avoid high-traffic routes altogether rather than rely on infrastructure for protection. Both groups show strong preferences for off-road bike paths, reinforcing the importance of safe, low-traffic environments for older populations.

Reactions to POI-dense areas also vary slightly. Middle-aged cyclists tend to avoid these routes more, possibly to maintain uninterrupted movement or reduce interaction with pedestrians. Older cyclists appear more tolerant of such environments, which may offer familiarity, access to amenities, or perceived safety due to slower surrounding traffic.

Finally, both middle-aged and older cyclists avoid routes with frequent turns, though older cyclists show greater sensitivity. This likely reflects physical and cognitive effort required to navigate complex routes, emphasizing the value of direct, low-turn designs in infrastructure aimed at aging populations.

7.3. Route preferences based on cyclist comfort and confidence

The Geller typology results, as presented in Section 6.3, reinforce the role of cyclist confidence in shaping infrastructure preferences. “Interested but Concerned” riders tend to avoid complex or unprotected environments and seek clearly delineated infrastructure. In contrast, “Enthused and Confident” or “Strong and Fearless” cyclists show a greater tolerance for risk, often prioritizing speed and directness over physical separation.

These behavioural differences reflect broader trends noted in earlier studies (Pearson et al., 2022; Dill and McNeil, 2013; Hosford et al., 2020), which show that Interested but Concerned cyclists are more cautious and require infrastructure that minimizes perceived risk. Their preference for off-road paths and protected bike lanes, as well as their higher sensitivity to slopes and route complexity, suggests a need for environments that feel safe and predictable.

In contrast, confident cyclists appear less reliant on segregated infrastructure and more accepting of challenging conditions. Their avoidance of protected bike lanes and POI-dense areas may reflect a desire to maintain higher speeds and avoid disruptions. These findings support the idea that some forms of infrastructure, while intended to improve safety, may be perceived as constraining by more experienced riders. This contrast highlights the importance of inclusive infrastructure planning. Designing networks that provide both safe, low-stress routes and flexible, direct connections is key to accommodating the needs of a diverse cycling population.

7.4. Route preferences for traditional bike versus e-bikes

Comparing traditional and e-bike users highlights how motor assistance influences infrastructure reliance and physical sensitivity. As shown in Section 6.4, traditional cyclists are more affected by features such as steep slopes, frequent turns, and high POI density, reflecting the greater physical demands of non-motorized cycling. Without motor support, interruptions and inclines present greater effort and can act as deterrents.

E-bike riders, on the other hand, navigate a broader range of environments with greater ease, likely due to smoother acceleration, reduced fatigue, and the ability to maintain consistent speed in more complex conditions. This helps explain their lower sensitivity to route complexity and reduced reliance on segregated infrastructure. These distinctions have implications for infrastructure design. While both groups benefit from safe, connected cycling networks, traditional bike riders rely more critically on physically separated or low-stress environments for comfort and usability.

Some of these differences should be interpreted with caution. The relatively small sample size of e-bike users may limit the generalizability of these findings. Furthermore, if e-bike riders disproportionately travel in areas with limited or disconnected infrastructure, their observed behavior may reflect contextual constraints rather than inherent preferences. Future work should investigate the role of spatial infrastructure distribution in shaping route choice and validate these findings across more diverse populations.

8. Conclusion

This study provides a comprehensive analysis of cyclist route choice behavior, utilizing large-scale RP data from Melbourne to explore taste heterogeneity through PSL and mixed PSL models and segmentation analyses across various demographic and typological groups. By integrating detailed route characteristics and cyclist attributes, the study sheds light on how infrastructure and route features influence decisions, offering critical insights into the diverse preferences of urban cyclists.

The results show a strong and significant presence of taste heterogeneity, and the segmentation analysis reveals some of these differences. A major finding is the consistent preference for segregated infrastructure, such as protected bike lanes and off-road paths, across all cyclist groups. While the level of reliance on such infrastructure varies, these facilities consistently emerge as critical for promoting safety, comfort, and accessibility. Women, older cyclists, and the “Interested but Concerned” group exhibit particularly strong preferences for segregated infrastructure, emphasizing its role in encouraging cycling participation among risk-averse groups. In contrast, confident cyclists and e-bike riders display greater adaptability, prioritizing route efficiency over infrastructure quality.

The implications for policy and planning underscore the need for inclusive infrastructure design that accommodates diverse user groups. Expanding the network of protected bike lanes and off-road paths can address the safety concerns of risk-averse cyclists, while also supporting confident cyclists through direct routes.

Despite the consistent findings, this study has some limitations. The small sample size of e-bike riders and the under-representation of certain groups constrain the generalizability of some results. Additionally, the availability of infrastructure in the RP data can limit the accuracy of the route choice models in capturing some of the true preferences. For example, suppose certain groups, such as female riders, predominantly travel in areas with limited separated bike infrastructure. In that case, their preferences may be underestimated due to the lack of exposure to those options in the RP data. Future research should address these gaps by expanding the data, incorporating dynamic factors like traffic interactions, and integrating contextual data such as trip purposes or local infrastructure quality. Conducting a combined revealed preference and stated preference analysis could offer valuable insights into the preferences of groups not represented in RP data, such as people not currently riding, and help evaluate potential responses to infrastructure variations. Longitudinal studies examining how cyclists’ preferences evolve with infrastructure changes would provide deeper insights into the effectiveness of investments and behavioral shifts over time. This approach could help refine infrastructure planning and support the role of cycling as a sustainable and equitable mode of transportation in urban networks.

Our findings also draw attention to the influence of model specification choices. Whether route attributes are expressed in absolute terms (e.g., distance) or relative measures (e.g., proportions) can alter the strength and even the direction of parameter estimates. See Appendix B, Tables B.1 and B.2. This sensitivity points to complex interdependencies between route features, suggesting that preferences cannot always be cleanly separated across alternative formulations. Exploring these dynamics in greater depth represents an important direction for future research.

CRediT authorship contribution statement

Tanapon Lilasathapornkit: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation; **Debjit Bhowmick:** Writing – review & editing, Methodology, Formal analysis, Data curation; **Ben Beck:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation; **Hao Wu:** Writing – review & editing, Software, Methodology; **Christopher Pettit:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition; **Kerry Nice:** Writing – review & editing, Investigation; **Sachith Seneviratne:** Writing – review & editing, Investigation; **Mohit Gupta:** Writing – review & editing; **Hai L. Vu:** Writing – review & editing, Investigation; **Trisalyn Nelson:** Writing – review & editing, Investigation, Formal analysis; **Meead Saberi:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Data availability

The authors do not have permission to share data.

Acknowledgement

This research was mainly funded by the Australian Government through the Australian Research Council (DP210102089). Meead Saberi and Tanapon Lilasathapornkit were also supported by the Australian Research Council (DP220102382). Ben Beck was was supported by an Australian Research Council Future Fellowship (FT210100183).

Appendix A. Sensitivity analysis of distance difference threshold

This section presents the results of a sensitivity analysis conducted to support the empirical derivation of the criteria used to retain utilitarian trips in the model. The analysis aimed to identify distance difference (DD_n) thresholds that preserve a large portion of the dataset while ensuring the model's behavior remains consistent with theoretical expectations, specifically, that the estimated distance coefficient (β_D) remains negative, reflecting a preference for shorter routes.

Table A.1 shows model estimation results across different values of DD_n. When DD_n exceeds 1.2, β_D becomes positive, indicating an implausible preference for longer routes. Based on this, a threshold below 1.2 was selected to ensure realistic model outcomes and retain the behavioural validity of the utility function.

Table A.1
Logit model outputs for different distance difference thresholds applied during trip filtering.

Variable	$DD_n = 0.7$		$DD_n = 1.0$		$DD_n = 1.2$		$DD_n = 1.4$	
	β	p-value	β	p-value	β	p-value	β	p-value
Distance (km)	−7.48	0	−0.053	0	0.045	0	0.150	0
Infrastructure[a]	0.129	0.11	0.188	0	0.112	0	0.090	0
Speed limit[b] (km/h)	6E−4	0.97	−0.032	0	−0.061	0	−0.066	0
AADT	1.9E-4	0.003	−5E-5	6E-5	−1.4E-4	0	−1.5E-4	0
POI (#/m)	−19.8	0.07	0.814	0.68	0.12	0.94	−1.344	0.35
Final Log Likelihood	−331		−10,612		−18,771		−20,881	
Akaike Information Criterion	672		21,236		37,553		41,773	
Number of trips	450		6,511		11,446		12,899	

[a] Infrastructure was classified into nine ordinal categories based on facility quality, with higher values representing better infrastructure (e.g., protected bike lanes, off-road paths), and lower values representing minimal or no cycling facilities (e.g., mixed traffic).
[b] Average speed limit for motor vehicles along the route.
Note: Assumes $L_n = 30$ km and $DF_t = 3$.

Appendix B. Comparison between proportion and distance-based variables

This section presents a comparison between models using proportion-based and distance-based route variables. While proportion-based specifications allow relative comparisons along the route (e.g., percentage of a specific facility), they result in utility coefficients with mixed units (e.g., 1 % increase = X meters). In contrast, distance-based specifications yield coefficients with consistent units and potentially more interpretable marginal effects.

Table B.1
Logit model estimates comparing proportion-based and distance-based representations of LTS.

Variable	LTS Dist. only		Dist. + LTS Dist.		Dist. + LTS Prop.	
	β	p-value	β	p-value	β	p-value
Distance (km)			−2.33	0	−0.4	0
Distance of LTS1 (km)	−0.26	0	2.08	3.1E-15		
Distance of LTS2 (km)	−0.56	0	1.77	1.1E-11		
Distance of LTS3 (km)	−0.79	0	1.53	4.5E-09		
Distance of LTS4 (km)	−1.07	0	1.26	0		
Prop. of LTS1					3.55	0.22
Prop. of LTS2					0.69	0.81
Prop. of LTS3					−0.06	0.98
Prop. of LTS4					−2.24	0.44
Final Log Likelihood	−26,015		−25,973		−25,469	
Akaike Information Criterion	52,038		51,956		52,948	
Number of trips	12,224		12,224		12,224	

Table B.2
Logit model estimates comparing proportion-based and distance-based representations of cycling infrastructure.

Variable	Dist. Infra. only		Dist. + Dist. Infra		Dist. + Prop. Infra	
	β	p-value	β	p-value	β	p-value
Distance (km)			−1.09	0.09	−0.3	0
Distance of Arterial roads - Painted bikelane (km)	−1.3	0	−0.21	0.75		
Distance of Arterial roads - Mixed traffic (km)	−1.12	0	−0.04	0.95		
Distance of Other (km)	−1	0	0.09	0.89		
Distance of Collector roads - Painted bikelane (km)	−0.91	0	0.18	0.78		
Distance of Local road - Painted bikelane (km)	−0.98	0	0.11	0.86		
Distance of Local roads - Mixed traffic or sharrow (km)	−0.11	0	0.98	0.12		
Distance of Collector roads - Mixed traffic (km)	0.31	0	0.78	0.22		
Distance of Protected bikelane (km)	0.27	0	1.36	0.03		
Distance of Off-road bike path (km)	−0.13	0	0.96	0.13		
Prop. of Arterial roads - Painted bikelane					−4.37	9.4E-06
Prop. of Arterial roads - Mixed traffic					−2.77	0.01
Prop. of Other					−2.29	0.05
Prop. of Collector roads - Painted bikelane					−1.7	0.12
Prop. of Local road - Painted bikelane					−0.72	0.52
Prop. of Local roads - Mixed traffic or sharrow					1.88	0.09
Prop. of Collector roads - Mixed traffic					2.41	0.03
Prop. of Protected bikelane					3.76	2.1E-03
Prop. of Off-road bike path					3.8	5.4E-04
Final Log Likelihood	−24,791		−24,791		−25,459	
Akaike Information Criterion	49,601		49,602		50,938	
Number of trips	12,224		12,224		12,224	

Table B.1 compares model results for different representations of Level of Traffic Stress (LTS) variables. The model using proportions of LTS types yielded statistically insignificant coefficients, whereas the model with LTS distances performed better. Table B.2 compares model results for bicycle infrastructure types. In this case, the model using distance variables produced mostly insignificant coefficients, while the model with proportion variables performed better in terms of statistical significance and interpretability.

Appendix C. Direct comparison of PSL versus Mixed PSL

Table C1 below shows a direct comparison of parameters from PSL and mixed PSL using Eq. 12.

Table C1
Cycling route choice model estimation results (N = 11,001): Direct comparison between the General PSL and Mixed PSL models.

Attribute	Path Size Logit		Mixed Logit			
	β	t-value	β	t-value	$exp(\beta_p + 0.5(\beta_p^{std})^2)$	sign
Distance (km)	−1.67	−4.53**	0.15	−2.509*	−1.17	negative
Distance st.dev (1000s)			−0.12	−1.95*		
Prop. of Arterial roads- Painted bikelane	−2.77	−8.54**	1.54	14.75**	−13.54	negative
Prop. of Arterial roads- Mixed traffic	2.07	8.73**	−6.38	−9.25**	0.00	positive
Prop. of Local roads- Mixed traffic or sharrow	3.83	22.50**	1.21	17.90**	9.74	positive
Prop. of Collector roads - Mixed traffic	6.03	26.90**	1.35	17.93**	11.20	positive
Prop. of Protected bikelane	2.35	3.60**	0.90	5.98**	7.14	positive
Prop. of Off-road bike path	5.04	22.19**	1.18	12.91**	9.45	positive
Infrastructure st.dev			1.46	18.39**		
Distance on LTS1 (km)	1.83	4.94**	0.42	141.96**	1.52	positive
Distance on LTS2 (km)	1.94	5.32**	0.39	161.18**	1.48	positive
Distance on LTS3 (km)	1.51	4.12**	0.01	94.31**	1.01	positive
Distance on LTS4 (km)	8e-4	2.08*	−16.05	−26.31**	0.00	positive
LTS st.dev (1000s)			0.01	0.23		
Max Slope (%)	−1.66	−12.92**	1.32	151.35**	−43.04	negative
Max Slope st.dev			2.21	30.92**		
POI (1000s)	−6.83	−15.22**	1.83	13.37**	−11.05	negative
POI st.dev			1.07	9.19**		
Turns (100s)	−7.76	−34.54**	2.33	53.33**	−14.15	negative
Turns st.dev (10s)			−0.80	−14.29**		
Path Size	−1.65	−15.19**	3.89	20.91**	−3.89	negative
Rho-squared-bar	0.23		0.31			
Final log-likelihood	−19,658		−17,698			
Akaike Information Criterion	39,345		35,437			

Note: ** indicates significant values at the 1 % level (** $p < 0.01$). * indicates significant values at the 10 % level (* $p < 0.1$).

References

Akar, G., Clifton, K.J., 2009. Influence of individual perceptions and bicycle infrastructure on decision to bike. Transp. Res. Rec. 2140 (1), 165–172.
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. Transp. Rev. 37 (1), 29–55.
Arellana, J., Saltarín, M., Larrañaga, A.M., González, V.I., Henao, C.A., 2020. Developing an urban bikeability index for different types of cyclists as a tool to prioritise bicycle infrastructure investments. Transp. Res. Part A 139, 310–334.
Bass, P., Donoso, P., Munizaga, M., 2011. A model to assess public transport demand stability. Transp. Res. Part A 45 (8), 755–764.
Beck, B., Chong, D., Olivier, J., Perkins, M., Tsay, A., Rushford, A., Li, L., Cameron, P., Fry, R., Johnson, M., 2019. How much space do drivers provide when passing cyclists? understanding the impact of motor vehicle and infrastructure characteristics on passing distance. Accident Anal. Prevent. 128, 253–260.
Beck, B., Stevenson, M., Newstead, S., Cameron, P., Judson, R., Edwards, E.R., Bucknill, A., Johnson, M., Gabbe, B., 2016. Bicycling crash characteristics: an in-depth crash investigation study. Accident Anal. Prevent. 96, 219–227.
Beck, B., Winters, M., Nelson, T., Pettit, C., Leao, S.Z., Saberi, M., Thompson, J., Seneviratne, S., Nice, K., Stevenson, M., 2023. Developing urban biking typologies: quantifying the complex interactions of bicycle ridership, bicycle network and built environment characteristics. Environ. Plann. B 50 (1), 7–23.
Bekhor, S., Ben-Akiva, M.E., Ramming, M.S., 2002. Adaptation of logit kernel to route choice situation. Transp. Res. Rec. 1805 (1), 78–85.
Ben-Akiva, M., Bergman, M.J., Daly, A.J., Ramaswamy, R., 1984. Modelling inter urban route choice behaviour. In: Papers presented during the Ninth International Symposium on Transportation and Traffic Theory held in Delft the Netherlands, 11–13 July 1984., pp. 299–330.
Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. In: Handbook of Transportation Science. Springer, pp. 5–33.
Ben-Akiva, M., Bolduc, D., 1996. Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure. Working Paper. Massachusetts Institute of Technology.
Bernardi, S., La Paix Puello, L., Geurs, K., 2018. Modelling route choice of Dutch cyclists using smartphone data. J. Transp. Land Use 11 (1), 883–900.
Bhowmick, D., Dai, D., Saberi, M., Nelson, T., Stevenson, M., Seneviratne, S., Nice, K., Pettit, C., Vu, H.L., Beck, B., 2025. Collecting population-representative bike-riding GPS data to understand bike-riding activity and patterns using smartphones and bluetooth beacons. Travel Behav. Soc. 38, 100919.
Bhowmick, D., Saberi, M., Stevenson, M., Thompson, J., Winters, M., Nelson, T., Leao, S.Z., Seneviratne, S., Pettit, C., Vu, H.L., Nice, K., Beck, B., 2022. A systematic scoping review of methods for estimating link-level bicycling volumes. Transp. Rev., 1–30. https://doi.org/10.1080/01441647.2022.2147240
Boeing, G., 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Comput. Environ. Urban Syst. 65, 126–139.
Bovy, P. H.L., 2009. On modelling route choice sets in transportation networks: a synthesis. Transp. Rev. 29 (1), 43–68.
Bovy, P. H.L., Bekhor, S., Prato, C.G., 2008. The factor of revisited path size: alternative derivation. Transp. Res. Rec. 2076 (1), 132–140.
Bovy, P. H.L., Fiorenzo-Catalano, S., 2007. Stochastic route choice set generation: behavioral and probabilistic foundations. Transportmetrica 3 (3), 173–189.
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 46 (10), 1730–1740.
Broach, J., Gliebe, J., Dill, J., 2010. Calibrated labeling method for generating bicyclist route choice sets incorporating unbiased attribute variation. Transp. Res. Rec. 2197 (1), 89–97.
Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified logit route choice model overcoming path overlapping problems. specification and some calibration results for interurban networks. In: Transportation and Traffic Theory. Proceedings of The 13th International Symposium On Transportation And Traffic Theory, Lyon, France, 24–26 July 1996, pp. 697–711.
Casella, G., Berger, R.L., 2002. Statistical Inference. Thomson Learning, Pacific Grove, CA. 2nd ed.
Chen, P., Shen, Q., Childress, S., 2018. A GPS data-based analysis of built environment influences on bicyclist route preferences. Int. J. Sustain. Transp. 12 (3), 218–231.
Cho, S.-H., Shin, D., 2022. Estimation of route choice behaviors of bike-sharing users as first-and last-mile trips for introduction of mobility-as-a-service (maas). KSCE J. Civ. Eng. 26 (7), 3102–3113.

Cubells, J., Miralles-Guasch, C., Marquet, O., 2023a. E-scooter and bike-share route choice and detours: modelling the influence of built environment and sociodemographic factors. J. Transp. Geogr. 111, 103664.
Cubells, J., Miralles-Guasch, C., Marquet, O., 2023b. Gendered travel behaviour in micromobility? travel speed and route choice through the lens of intersecting identities. J. Transp. Geogr. 106, 103502. https://doi.org/10.1016/j.jtrangeo.2022.103502
Dane, G., Feng, T., Luub, F., Arentze, T., 2020. Route choice decisions of E-bike users: analysis of GPS tracking data in the Netherlands. In: Geospatial Technologies for Local and Regional Development: Proceedings of the 22nd AGILE Conference on Geographic Information Science 22. Springer, pp. 109–124.
Del Rosario, L., Wu, H., Lee, J.B., Roberts, L., Arnold, T., Mathur, S., Pettit, C., 2024. Assessing the monetary value of active transport and e-micromobility: a systematic review. Transp. Res. Interdiscip. Perspect. 27, 101243.
Desjardins, E., Higgins, C.D., Scott, D.M., Apatu, E., Páez, A., 2022. Correlates of bicycling trip flows in Hamilton, Ontario: fastest, quietest, or balanced routes? Transportation 49 (3), 867–895.
Dill, J., Gliebe, J., 2008. Understanding and Measuring Bicycling Behavior: A Focus on Travel Time and Route Choice. Technical Report. Oregon Transportation Research and Education Consortium (OTREC).
Dill, J., McNeil, N., 2013. Four types of cyclists? examination of typology for better understanding of bicycling behavior and potential. Transp. Res. Rec. 2387 (1), 129–138.
Duncan, L.C., Watling, D.P., Connors, R.D., Rasmussen, T.K., Nielsen, O.A., 2020. Path size logit route choice models: issues with current models, a new internally consistent approach, and parameter estimation on a large-scale network with GPS data. Transp. Res. Part B 135, 1–40.
Fitch, D.T., Handy, S.L., 2020. Road environments and bicyclist route choice: the cases of davis and san francisco, CA. J. Transp. Geogr. 85, 102705.
Fosgerau, M., Łukawska, M., Paulsen, M., Rasmussen, T.K., 2023. Bikeability and the induced demand for cycling. Proc. Natl. Acad. Sci. 120 (16), e2220515120.
Foti, F., Waddell, P., Luxen, D., 2012. A generalized computational framework for accessibility: from the pedestrian to the metropolitan scale. In: Proceedings of the 4th TRB Conference on Innovations in Travel Modeling. Transportation Research Board.
Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. Transp. Res. Part B 41 (3), 363–378.
Garrard, J., Rissel, C., Bauman, A., 2012. Health benefits of cycling. City Cycl. 31, 31–56.
Ghanayim, M., Bekhor, S., 2018. Modelling bicycle route choice using data from a GPS-assisted household survey. Eur. J. Transp. Infrastruct. Res. 18 (2).
Götschi, T., Garrard, J., Giles-Corti, B., 2016. Cycling as a part of daily life: a review of health perspectives. Transp. Rev. 36 (1), 45–71.
Halldórsdóttir, K., Rieser-Schussler, N., Axhausen, K.W., Nielsen, O.A., Prato, C.G., 2014. Efficiency of choice set generation methods for bicycle routes. Eur. J. Transp. Infrastruct. Res. 14 (4), 332–348.
Hauser, J.R., 1978. Testing the accuracy, usefulness, and significance of probabilistic choice models: an information-theoretic approach. Oper. Res. 26 (3), 406–421.
Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. Transportation 30, 133–176.
Hess, S., Quddus, M., Rieser-Schüssler, N., Daly, A., 2015. Developing advanced route choice models for heavy goods vehicles using GPS data. Transp. Res. Part E 77, 29–44.
Hood, J., Sall, E., Charlton, B., 2011. A GPS-based bicycle route choice model for San Francisco, California. Transp. Lett. 3 (1), 63–75.
Hoogendoorn-Lanser, S., van Nes, R., Bovy, P., 2005. Path size modeling in multimodal route choice analysis. Transp. Res. Rec. 1921 (1), 27–34. https://doi.org/10.1177/0361198105192100104
Hosford, K., Laberee, K., Fuller, D., Kestens, Y., Winters, M., 2020. Are they really interested but concerned? a mixed methods exploration of the geller bicyclist typology. Transp. Res. part F 75, 26–36.
Koch, T., Knapen, L., Dugundji, E., 2019. Path complexity for observed and predicted bicyclist routes. Procedia Comput. Sci. 151, 393–400.
Larsen, J., El-Geneidy, A., 2011. A travel behavior analysis of urban cycling facilities in montréal, canada. Transp. Res. Part D 16 (2), 172–177.
Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability-normal or log-normal: that is the question. Bioscience 51 (5), 341–352.
Lin, J.-J., Wei, Y.-H., 2018. Assessing area-wide bikeability: a grey analytic network process. Transp. Res. Part A 113, 381–396.
Lißner, S., Huber, S., 2021. Facing the needs for clean bicycle data–a bicycle-specific approach of GPS data processing. Eur. Rransp. Res. Rev. 13, 1–14.
Loidl, M., Hochmair, H.H., 2018. Do online bicycle routing portals adequately address prevalent safety concerns? Safety 4 (1), 9.
Lu, W., Scott, D.M., Dalumpines, R., 2018. Understanding bike share cyclist route choice using GPS data: comparing dominant routes and shortest paths. J. Transp. Geogr. 71, 172–181.
Łukawska, M., 2024. Quantitative modelling of cyclists' route choice behaviour on utilitarian trips based on GPS data: associated factors and behavioural implications. Transp. Rev., 44(5), 1–32.
Łukawska, M., Paulsen, M., Rasmussen, T.K., Jensen, A.F., Nielsen, O.A., 2023. A joint bicycle route choice model for various cycling frequencies and trip distances based on a large crowdsourced GPS dataset. Transp. Res. Part A 176, 103834.
Meister, A., Felder, M., Schmid, B., Axhausen, K.W., 2023. Route choice modeling for cyclists on urban networks. Transp. Res. Part A 173, 103723.
Meister, A., Liang, Z., Felder, M., Axhausen, K.W., 2024. Comparative study of route choice models for cyclists. J. Cycl. Micromob. Res. 2, 100018. https://doi.org/10.1016/j.jcmr.2024.100018
Menghini, G., Carrasco, N., Schüssler, N., Axhausen, K.W., 2010. Route choice of cyclists in Zurich. Transp. Res. Part A 44 (9), 754–765.
Misra, A., Watkins, K., 2018. Modeling cyclist route choice using revealed preference data: an age and gender perspective. Transp. Res. Rec. 2672 (3), 145–154.
Monsere, C., Dill, J., McNeil, N., Clifton, K.J., Foster, N., Goddard, T., Berkow, M., Gilpin, J., Voros, K., van Hengel, D., et al., 2014. Lessons from the Green Lanes: Evaluating Protected Bike Lanes in the US. Technical Report. Portland State University.
Newson, P., Krumm, J., 2009. Hidden Markov Map Matching through Noise and Sparseness. Association for Computing Machinery, New York, NY, USA, p. 336–343. https://doi.org/10.1145/1653771.1653818
Park, Y., Akar, G., 2019. Why do bicyclists take detours? a multilevel regression model using smartphone GPS data. J. Transp. Geogr. 74, 191–200.
Pearson, L., Dipnall, J., Gabbe, B., Braaf, S., White, S., Backhouse, M., Beck, B., 2022. The potential for bike riding across entire cities: quantifying spatial variation in interest in bike riding. J. Transp. Health 24, 101290.
Pearson, L., Reeder, S., Gabbe, B., Beck, B., 2023. What a girl wants: a mixed-methods study of gender differences in the barriers to and enablers of riding a bike in Australia. Transp. Res. Part F 94, 453–465. https://doi.org/10.1016/j.trf.2023.03.010
Pettit, C., Lee, B., Wu, H., Roberts, L., 2024. Improving the Bikeability of Our Cities. Technical Report. City Futures Research Centre, School of Built Environment, University of New South Wales.
Prato, C.G., Bekhor, S., 2007. Modeling route choice behavior: how relevant is the composition of choice set? Transp. Res. Rec. 2003 (1), 64–73.
Prato, C.G., Halldórsdóttir, K., Nielsen, O.A., 2018. Evaluation of land-use and transport network effects on cyclists' route choices in the copenhagen region in value-of-distance space. Int. J. Sustain. Transp. 12 (10), 770–781.
Pucher, J., Buehler, R., 2017. Cycling towards a more sustainable transport future. Transp. Rev. 37 (6), 689–694.
Rieser-Schüssler, N., Balmer, M., Axhausen, K.W., 2013. Route choice sets for very high-resolution data. Transportmetrica A 9 (9), 825–845.
Ryu, S., Chen, A., Su, J., Choi, K., 2021. A multi-class, multi-criteria bicycle traffic assignment model. Int. J. Sustai. Transp. 15 (7), 524–540.
Ryu, S., Su, J., Chen, A., Choi, K., 2019. Estimating bicycle demand of a small community. KSCE J. Civ. Eng. 23 (6), 2690–2701. https://doi.org/10.1007/s12205-019-0415-5
Schirck-Matthews, A., Hochmair, H.H., Strelnikova, D., Juhász, L., 2023. Bicycle trips in endomondo, google maps, and mapquest: a comparison between South Florida and North Holland. Transp. Lett. 15 (4), 308–320.
Sener, I.N., Eluru, N., Bhat, C.R., 2009. An analysis of bicycle route choice preferences in texas, US. Transportation 36, 511–539.
Sevtsuk, A., Basu, R., Li, X., Kalvo, R., 2021. A big data approach to understanding pedestrian route choice preferences: evidence from San Francisco. Travel Behav. Soc. 25, 41–51.

Sobhani, A., Aliabadi, H.A., Farooq, B., 2019. Metropolis-Hasting based expanded path size logit model for cyclists' route choice using GPS data. Int. J. Transp. Sci. Technol. 8 (2), 161–175.

Tahlyan, D., Pinjari, A.R., 2020. Performance evaluation of choice set generation algorithms for analyzing truck route choice: insights from spatial aggregation for the breadth first search link elimination (BFS-LE) algorithm. Transportmetrica A 16 (3), 1030–1061.

Ton, D., Cats, O., Duives, D., Hoogendoorn, S., 2017. How do people cycle in Amsterdam, Netherlands?: estimating cyclists' route choice determinants with gps data from an urban area. Transp. Res. Rec. 2662 (1), 75–82.

Ton, D., Duives, D., Cats, O., Hoogendoorn, S., 2018. Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. Travel Behav. Soc. 13, 105–117.

Vicmap Spatial Services Branch, 2022. Vicmap Elevation LiDAR DEMs Collection: Product Description. Department of Transport and Planning. PO Box 527, Melbourne VIC 3001 Australia.

Wallentin, G., Loidl, M., 2015. Agent-based bicycle traffic model for Salzburg city. GI_Forum - J. Geogr. Inf. Science 3, 558–566. https://doi.org/10.1553/giscience2015s558

Wang, H., Moylan, E., Levinson, D., 2023. Route choice set generation on high-resolution networks. Transp. Res. Rec., 2678(5), 03611981231188775.

Water, V. D. o. E.L., 2021. Planning Vicmap elevation. https://www.land.vic.gov.au/maps-and-spatial/spatial-data/vicmap-catalogue/vicmap-elevation.

Wibowo, B.S., Aditya, R.B., Harianto, T.F., 2021. Harnessing open data and technology for the study of accessibility: the case of Indonesia's capital site candidate. Spatium, 46, 46–53.

Zhuwaki, N.T., Coetzee, J., 2021. Network accessibility study to evaluate the extent of public transport coverage in the Harare metropolitan area. In: Proceedings of the Southern African Transport Conference 2021.

Zimmermann, M., Mai, T., Frejinger, E., 2017. Bike route choice modeling using GPS data without choice sets of paths. Transp. Res. Part C 75, 183–196.