Khavarian, Khashayar; Vosough, Shaghayegh; Roncoli, Claudio

Bike users' route choice behaviour: Expectations from electric bikes versus reality in Greater Helsinki

Published in:
Journal of Cycling and Micromobility Research

DOI:
10.1016/j.jcmr.2024.100045

Published: 01/12/2024

Document Version
Publisher's PDF, also known as Version of record

Journal of Cycling and Micromobility Research 2 (2024) 100045

Contents lists available at ScienceDirect

Journal of Cycling and Micromobility Research

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

Bike users' route choice behaviour: Expectations from electric bikes versus reality in Greater Helsinki

Khashayar Khavarian [a], Shaghayegh Vosough [b,*], Claudio Roncoli [b]

[a] Department of Civil Engineering, Sharif University of Technology, Iran
[b] Department of Built Environment, Aalto University, Finland

ARTICLE INFO

Keywords:
Cycling
E-bike
Route choice
Discrete choice modelling
Classification approach
Expectation–reality gap

ABSTRACT

Electric pedal-assist bikes (e-bikes) are an emerging technology that aims to enhance cycling by incorporating battery-powered motors activated while pedalling. To promote cycling effectively, it is crucial to understand the factors that influence cyclists' route choice behaviour. This study investigates individual route choice behaviour among cyclists, taking into account their bike type (i.e., e-bikes and regular bikes). Data collected through a stated preference (SP) survey in Finland is analysed using discrete choice models to compare the differences between e-bike and regular bike users' route choice behaviour. The study also compares the outputs of multinomial and mixed Logit models for both e-bike and regular bike users to address the impact of error correlation in SP data. Furthermore, by employing a classification approach, the study examines the differences between the expected and actual behavioural changes upon using e-bikes, referred to as the expectation–reality gap, in terms of route choice behaviour. Our research findings highlight certain factors that consistently promote cycling among both regular bike and e-bike users, specifically, low interaction with traffic, fewer intersections, and the presence of separated bike facilities. Also, our findings imply that the SP survey is well-designed to capture the preferences of the individuals. Hence, the observations are not severely correlated, i.e., errors can be assumed to be independently and identically distributed. Furthermore, we show that regular bike and e-bike users with similar characteristics do not share similar beliefs regarding the effects of e-bikes on their cycling habits.

1. Introduction

1.1. Motivation

Promoting active modes of travel provides many advantages, from decreasing air pollution to declining obesity cases and related diseases (Sałabun et al., 2019; Anderson et al., 2022). According to the World Health Organization (WHO) guideline (World Health Organization, 2020), individuals between the ages of 18 and 64 should engage in 150-300 min of moderate-intensity aerobic physical activity per week. Encouraging the adoption of active transportation modes can assist the general population in fulfilling the WHO's health recommendations. One popular active mode is cycling.

Many people hesitate to choose cycling as their commuting mode due to the perceived physical effort involved. That is why researchers propose workplace showers as a solution, which has been proven to boost the number of cycling commuters to work (Buehler, 2012). Anyhow, with advancements in technology, there are solutions available that are expected to decrease the physical effort of cycling as well as save time by increasing the speed. Pedal-assisted electric bikes (in short *e-bikes*) can help cyclists while pedalling, especially on routes with hills (Arning et al., 2023), compared to regular bikes (in short *r-bikes*). Gojanovic et al. (2011) shows that e-bikes, despite making cycling trips easier, still have positive effects on individuals' health and well-being. Yet, some researchers imply that easing up cycling pedalling effort is not enough to promote an active mode to the rest of the currently passive transport users unless planned properly (Kroesen, 2017; Haustein and Nielsen, 2016).

Although bicycle usage is promoted in many European countries, governments are still actively striving to further increase the cycling share; yet, different patterns regarding cycling have been observed. For instance, France, Italy, and Germany have witnessed a more than 10% increase in cycling demand, in 2020, compared to 2019, while Finland and Ireland lost more than 10% of their weekday cyclists in the same period (Counter, Eco, 2021). The average travel distance in Helsinki is more than 2.2 km using only city bikes which is a public shared bicycle system (Khachatryan, 2021), while the average length of cycling in European countries is 3 km and a large share of it is devoted to work

* Corresponding author.
E-mail address: shaghayegh.vosough@aalto.fi (S. Vosough).

https://doi.org/10.1016/j.jcmr.2024.100045
Received 10 October 2023; Received in revised form 22 July 2024; Accepted 23 August 2024
Available online 30 August 2024

trips.[1] Understanding the factors that underlie the variations in cycling cultures among different countries is complex (Haustein and Nielsen, 2016; Goel et al., 2022). Hence, more promoting actions are required to increase the cycling share.

One approach to promote cycling in urban areas is through effective infrastructure planning tailored to accommodate cyclists; that is, planners must understand the factors influencing cycling behaviour, particularly route choice (Broach et al., 2012; Huber et al., 2021), since transportation infrastructure characteristics have a major effect on this decision. Researchers usually employ route choice modelling for active modes (e.g., bicycles) to assess the infrastructure characteristics' impacts on the mode users (Segadilha and Sanches, 2014; Bernardi et al., 2018).

Discrete choice models are a tool widely used to model the route choice behaviour (Aloulou, 2018). Research employing these models has found that cyclists' route preferences are influenced by a multitude of factors, with variables such as distance, gradient, and road characteristics typically emerging as the most significant ones (Stinson and Bhat, 2003; Majumdar and Mitra, 2017), theorising that safety and pedalling efforts are the main concerns in reducing the cycling demand when the infrastructure, traffic laws, and affordability of bikes are in place (Hull and O'Holleran, 2014).

However, there is always a disconnection between what individuals anticipate or expect to happen in a particular situation and what actually occurs in reality. This is referred to as the *expectation–reality gap*, which is investigated in various fields (Stroh et al., 1998; Moffett et al., 2000; Prashar et al., 2022; Larsen et al., 2022). In the context of our study, the beliefs of r-bike users (in short *r-bikers*) about the changes in their cycling habits with e-bike usage may differ from the actual habits experienced by e-bike users (in short *e-bikers*). In fact, if potential e-bike users have certain expectations about the convenience and accessibility of routes, but the reality does not match up, it could act as a barrier to bike adoption, especially e-bikes. Investigating this gap, which has not been investigated before, can identify and address misconceptions, leading to a more effective promotion plan for bike usage as a viable transportation option.

1.2. Aims and contribution

The main contribution of this research is to test the hypothesis that e-bikes bring about significant changes in cyclists' behaviour, particularly their route choice decisions. To address this gap, we first, investigate factors affecting cyclists' route choice decisions that may be used in infrastructure planning, using separate discrete choice models for cyclists with r-bikes and e-bikes. Additionally, the model specification that may also interfere with factors' effects on route choice is investigated by employing two different types of models, i.e., multinomial Logit (MNL) model and Mixed Logit (MXL) model with panel effect, referred to as PMXL. The MNL assumes independently and identically distributed (IID) errors, while the PMXL accounts for the panel effect of the data by relaxing this assumption. Since each choice situation presented to respondents is considered an observation, the error terms of discrete choice models may not be IID (Axhausen et al., 2006). Thus, comparing the MNL and PMXL models allows us to better understand the influence of the model specification on the results and ensure the robustness of our main findings. Similar types of models with the panel effects have been employed by Meister et al. (2023), and in other studies that are looking for the model specification effects in their results (Brownstone et al., 2000).

Furthermore, this study contributes to a better understanding of the alignment between the r-bikers' expectations from e-bike adaptation with the experienced habit changes of e-bikers, using a classification approach. This aspect has not been thoroughly investigated in previous literature and represents a unique contribution to understanding cyclists' behaviour when using e-bikes. The classification approach categorises the r-bikers and e-bikers population based on their characteristics and their belief in their change of habits upon using e-bikes. By comparing the anticipated changes in cycling habits that r-bikers believe e-bikes would bring with the actual habits e-bikers have experienced in real life, this study provides insights for city planners and policymakers to make informed decisions and tailor the policies for promoting e-bike usage. This study focuses particularly on the Helsinki region by employing one source of stated preference data in the Greater Helsinki area.

To summarise, the contribution of this research to the literature is threefold:

1. Investigating factors affecting route choice of e-bikers, and comparing them with r-bikers to identify the main affecting factors of cycling promotion with less pedalling effort, with a focus on the Greater Helsinki area, Finland;
2. Analysing the impacts of model specification on the research findings by utilising models with and without panel effect of data;
3. Bridging the expectation–reality gap by comparing characteristics of r-bikers who expect to change their habits due to owning an e-bike with the e-bikers who actually reported changes in their habits.

The rest of the paper is organised as follows. Section 2 reviews the literature on factors affecting cyclists' route choice, and the data and models used in previous studies, followed by Section 3 introducing the data and method we employ to investigate factors affecting cyclists' route choice and analyse the differences between expected and real changes happen due to e-bike usage. Then, the outputs of the estimated models and the results' interpretations are presented in Section 4. The conclusions are drawn in Section 5.

2. Literature review

The literature on cyclists' route choice behaviour is pretty rich, and reviews of influential factors can be found in several studies, including Hull and O'Holleran (2014), Tarkkala (2022), Huber et al. (2021), Tarkkala et al. (2023).

Regarding cycling route choice studies, scientific literature offers two data collecting methods: stated preference (SP) and field experiments or revealed preference (RP). An SP experiment asks individuals to express their choice or their action in hypothetical situations, while an RP experiment records the real-world choice or action of people and estimates their selected and not selected option attributes a-posteriori. While collecting RP data has traditionally been costly (Hensher et al., 2005; de Dios Ortúzar and Willumsen, 2011), it remains expensive for certain studies, particularly as companies collecting data and providing transport services through smartphones and other devices often charge substantial fees for access to their data. However, more recent studies (Scott et al., 2021; Fosgerau et al., 2023) based on large RP data sets have overcome this burden but still face practical constraints, such as the inability to record users' characteristics alongside their trips' features unless a complementary survey is conducted to gather such information. Moreover, representativeness and privacy are other limitations of RP surveys, especially those based on smartphone applications (Nelson et al., 2021). On the other hand, SP data is cheaper to collect and allows for testing different combinations of influential factors. However, SP data usually suffers from hypothetical bias, which is the distortion due to unrealistic answers of individuals. In this study, we use SP to control the variables, allowing us to investigate the effect of cyclists', road, and traffic characteristics on route choice behaviour.

The studies that used SP data provided different hierarchies regarding the effective variables on the route choice and are not conclusive.

[1] https://road-safety.transport.ec.europa.eu/eu-road-safety-policy/priorities/safe-road-use/cyclists/walking-and-cycling-transport-modes_en

For instance, Majumdar and Mitra (2017) claimed perceived safety is the most important factor, while (Stinson and Bhat, 2003) reported travel time as the primary factor. Travel time can be considered a combination of length, slope, and the surrounding traffic. In their study, they provided values for the aforementioned variables but length.

It is shown that different infrastructure-related factors can affect the cycling demand, including high-level factors such as directness, attractiveness, safety, and comfort of the path and cycling routes (Hull and O'Holleran, 2014). These high-level factors are not directly measurable; thus, measurable dimensions such as the length of the route, steepness, separated cycling path, or cycling through vehicle traffic are employed to evaluate cycling infrastructure quantitatively (Schoner and Levinson, 2014).

Menghini et al. (2010) using a large sample of GPS observations, which is an RP-based study and overcomes the limitations of SP surveying, claimed that the most important factor affecting the route choice of cyclists is reported to be the route's length. The overall findings of RP-based literature presented in Huber et al. (2021) show that distance, slope, number of intersections, and motorised traffic reduce the probability of choosing a specific route while special cycling infrastructure and safety increase the probability. Some of these findings have been investigated in many other studies such as Broach et al. (2012), Hood et al. (2013), Casello and Usyukov (2014), and Ton et al. (2017).

Scott et al. (2021) based on 12-month trajectories of 750 shared bikes showed that, if no other factor is considered in the route choice, then the long distance would reduce the selection probability. On the other hand, they showed that people would detour from the shortest path and select a longer one if there were other interesting factors involved such as designated cycling infrastructure. They also showed selecting major roads with or without special bike lanes is more probable than minor roads with bike lanes. This again proves that, besides safety, some other factors such as steep slopes and turns cause this choice of route. However, these RP-based studies overlook the effect of cyclists' characteristics on their route choices.

We now review two streams of literature on e-bikes: route choice and mode choice. Practically, cyclists' mode choice depends, to some extent, on their route choice. For instance, many people may cycle to their destination if they find a safe cycling route (Hwang and Guhathakurta, 2023); hence, we cover the findings of both streams in this section.

Studies dedicated to the route choice of e-bikers or the change in the attitude of cyclists using e-bikes are limited, but provide interesting insights. A scoping study analysing research conducted on e-bikes up to 2019 showed that there are only 76 studies related to the overall travel behaviour change of cyclists after using e-bikes, motives the e-biking and general attitude towards e-bike and r-bike usage (Bourne et al., 2020). Another scoping study through literature from 1946 to 2021 resulted in 107 studies related to different features of e-bike implications. It is reported that age, gender, income, weight, and lack of infrastructure are barriers to e-bike promotion (Jenkins et al., 2022). Chavis and Martinez (2021) found that e-bikes increase the length that cyclists ride, while they also reported shorter travel times for e-bikers than for r-bikers, which means a significant increase in speed is observed. Moreover, with the increase in e-bike numbers, major roads were more frequently selected by cyclists than minor roads.

The mode choice and its related behaviour for e-bikers are studied more extensively than their route choices. It is found that the changes in e-bikers travel behaviour are influenced by the primary mode of transport before e-bike adaptation (Castro et al., 2019). The mode e-bikes are replacing is studied for various reasons. For instance, from an environmental point of view, Cherry et al. (2009) stated that the e-bike effect on emission depends on the technology of the power plant of its region and the modes it is replacing. Nevertheless, the research about the share that e-bikes are absorbing from other modes of transport shows contradictory results (Haustein and Nielsen, 2016) or has plausible shortcomings (Kroesen, 2017). For instance, Kroesen (2017) concluded that e-bikes are the main substitute for regular bikes, although direct questions from the e-bike owners suggest that they are substituting their cars and transit trips with their e-bikes. Andersson et al. (2021) concluded that e-bikes would replace 21% of car trips yet their sample is only restricted to regular personal vehicle users of a company who has willingly decided to participate in the research about e-bikes and no general conclusion must be made. Haustein and Nielsen (2016) suggested that e-bikes replacing regular bikes (in short r-bikes) is dominant in countries like Denmark and the Netherlands, i.e., countries with rich cycling cultures, while it is not common, e.g., in Sweden. These studies imply that although e-bikes would ease up cycling pedalling effort, this is not enough to promote an active mode to the rest of the currently passive transport users unless planned properly.

Rérat (2021) surveyed more than 2000 e-bikers and almost 11000 r-bikers in Switzerland, revealing an increased usage of e-bikes by females (50% of e-bikers vs. 40% of r-bikers), as well as an increase in average age in the e-bikers. Castro et al. (2019) also found that e-bikers are on average older than r-bikers, yet observed no significant difference due to sex or education level. Regarding the season of cycling, it was observed that the e-bikers were almost abandoning their bikes in winter, switching to public transport or other motorised modes, probably due to the fact that e-bikes are used more frequently for longer trips than r-bikes (Rérat, 2021).

We have selectively reviewed general route choice studies of cyclists since our study aims to show the difference in route choice decisions between e-bikers and r-bikers, and studies comparing route choice of these two groups are scarce. For instance, Meister et al. (2023) performed a study on the effect that an e-bike has on the route choice in comparison to the route choice decision of r-bikers. They used GPS data and generated alternative choices for each observation based on the real network of Zurich. To address the scale effect, they estimated their models with variables normalised by the distance between the origin and destination, employing a path-sized logit model. They implemented a MXL model to show the effect of e-bikes on route choice, yet some of their findings were counter-intuitive. Dane et al. (2020) developed mixed logit models for r-bikers and e-bikers to show the difference in factors affecting their route choices. However, their estimated model uses length as the only alternative specific variable. Hence, their study results in a positive coefficient for distance which means that longer trips for both e-bikers and r-bikers are probable. They used the interaction of different social-demographic variables with length to account for differences among people or taste heterogeneity.

Concluding this literature review, little attention has been paid to the e-bike's impacts on the route choice of cyclists. Although general studies are available, they have not assessed the change due to electrification and did not compare their results with route choice models for r-bikes. In summary, research investigating the effects of e-bike usage has not fully identified the key factors driving the promotion of both r-bikes and e-bikes. We claim that properly managing effective variables of cyclists' route choice promotes this transport mode because the decision to cycle is not only influenced by the mode characteristics, but it is mostly a joint decision influenced by both available modes and route characteristics (Broach and Dill, 2016). As an instance, Zhu et al. (2020) demonstrated the need for a better understanding of the factors affecting r-bikers' and e-bikers' choices by providing a complementary equilibrium model between the mode and route choice of e-bikers. Hence, this research addresses the variables influencing cycling route choice for both e-bikers and r-bikers which, to the best of the authors' knowledge, has not been extensively explored before.

3. Methodology

3.1. Data collection

To analyse the route choice of cyclists concerning the type of their bikes, this study uses SP data collected in Finland. The study area

Table 1
Examined variables and their levels in survey.

Row	Factor	Levels
1	Bike Facility	Mixed Traffic (no path), Painted Lane, Adjacent Path, separated path
2	Road type	Local, main, arterial
3	Vehicle traffic	Light, moderate, substantial, heavy
4	Presence of Controlled intersection	No signals, few signals, many signals
5	Route gradients	No hill, moderate hills, steep hills,

is the Greater Helsinki region, which comprises 15 municipalities, including the Finnish capital, with around 1.55 million population in 2023 (HSL, 2024). Around 4.7 million daily trips across the region were reported on a normal weekday in 2018 (Brandt et al., 2019), and the share of sustainable modes, including walking, cycling, and public transportation, was found to be 60%. The share of cycling as a primary mode was found to be 9% among daily trips which was around 420,000 trips per day by bikes. Due to the COVID-19 pandemic, travel habits underwent significant shifts in the region, similar to other cities in the world (de Palma et al., 2022). According to the 2021 data, the share of sustainable modes decreased to 52%; however, by 2023, this share rebounded to 62% (HSL, 2024), with cycling contributing to 9% of total trips (Eriksson, 2024).

The data is gathered using a survey assessing the following general factors: the presence or type of a bike facility, the road type, the vehicle traffic, the presence of controlled intersections along the route, the route gradients, and its length. Each factor is discretised to different levels which are summarised in Table 1. Length variable is also categorised starting from 3 km to 10 km with half a kilometre steps. The possible number of choice options with depicted discretisation without the inclusion of the length variable is 432 ($4^2 \times 3^3$). However, not all these options are implemented in the surveying procedure: fractional factorial designs generated by SPSS software resulted in the 32 options, split into four blocks with each block having eight unlabelled choice questions between two alternatives for each individual to respond.

Computer-assisted web interview was selected as the method to collect SP data from active cyclists over 15 years old. The survey was designed with Webropol software (Webropol, 2021), in three languages: Finnish, English, and Swedish. The survey was offered online for one month during September 2021, and 1029 respondents filled out the questionnaire. Fig. 1 depicts one of the hypothetical choice situations used in the survey. More details about the survey design and the data can be found in Tarkkala (2022).

The characteristics of the sample population including trip purposes, age groups, their experience in riding a bike, and the time of year they cycle, are depicted in charts of Fig. 2. The e-bikers' share of the respondents is 9.6% which is similar to the market share, i.e., 9% (Kuva, 2020). Moreover, the share of female respondents from the filled questionnaire is 49.3% which is a fair share regarding the target population composition.

3.2. Discrete choice modelling

The choice situation among routes can be framed as a discrete choice decision based on maximum utility or minimum disutility/cost (Ben-Akiva and Lerman, 1985). This modelling approach estimates the probability of each option being selected by the decision unit which can be an individual or a group of people. In this research, two different types of discrete choice models are implemented: MNL and PMXL, which enable the investigation of the model specification impacts by comparing the models' results.

MNL estimates the probability of choosing each route based on a linear combination of factors forming a utility value. However, since it is impossible to capture completely the utility value, a utility function composed of two parts is employed: the deterministic part, V_{in}, and the error part, ϵ_{in}, where i and n refer to alternative and individual, respectively. The MNL formulation is as follows

$$U_{in} = V_{in} + \epsilon_{in} = \beta_i + \beta_{i1} X_{in1} + \beta_{i2} X_{in2} + \cdots + \epsilon_{in}, \tag{1}$$

where β_{ik} is the coefficient related to kth variable representing individual or alternative characteristics, X_{ink}. Assuming ϵ_{in} follows the Gumbel distribution, which is a special case of the Generalised Extreme Value (GEV) distribution, the probability of each alternative selection is derived via

$$P_{in} = \frac{\exp(V_{in})}{\sum_j \exp(V_{jn})}. \tag{2}$$

These formulas are the direct results of assuming the random/error part, ϵ_{in}, being independently and identically distributed (IID) extreme value (McFAdden, 1974).

The MNL cannot account for correlation among error terms, a common characteristic of SP gathered data. Hence, to assess the correlation's effect on final outputs, an MXL model capable of accounting for the panel effect, i.e., correlation among error terms, is employed. The panel effect arises in data where multiple observations are collected from the same individual over time or across different choice situations. As each respondent in our study provided choices across eight situations, a panel effect may exist. Thus, we employ the MXL model with panel effects, denoted as PMXL, which is an MXL model where the IID assumption is relaxed by introducing one or more random parameters (McFadden and Train, 2000); this eventually changes the error term to $\Sigma_k \delta_{ik} X_{ink} + \epsilon_{in}$, as follows

$$U_{in} = V_{in} + \epsilon_{in} = \beta_i + \beta_{i1} X_{in1} + \delta_{i1} X_{in1} + \beta_{i2} X_{in2} + \delta_{i2} X_{in2} + \cdots + \epsilon_{in} \tag{3}$$

Different assumptions regarding the distribution of δ_{ik} are possible, with the most common being a normal distribution (Train, 2009). This change in error term distribution is addressed through repeated simulations, and the expected probability of each alternative selection when there is only one random parameter, e.g., δ_{i1} assumed is derived by approximately estimating the following equation (Train, 2009)

$$P_{in} = \int_{-\infty}^{\infty} \frac{\exp(V_{in})}{\sum_j \exp(V_{jn})} f(\delta_{i1}) d\delta_{i1}. \tag{4}$$

Extending equation (4) to accommodate more random parameters is a straightforward process. The above integral represents an average taken over all possible values of random parameters, i.e., δ_{ik}. If these random parameters are assumed to follow a joint distribution, their joint distribution density function would replace $f(\delta_{ik})$ which is the case for PMXL. On the other hand, if they are considered to be independent, the product of their individual density functions would be used.

The maximum likelihood method is used for model estimation of both the MNL and PMXL, treating models for r-bikers and e-bikers separately. The MNL of r-bike and e-bike users are compared to identify the prominent factors affecting the route choice of individuals, while cycling effort is reduced due to the electrification of bikes. A similar comparison is made between PMXL models for r-bikers' and e-bikers' route choices. The results obtained through MNL and PMXL are then compared to show the significance of the impact of error correlation among observations. This comparison evaluates the effect of error correlation on the effective route choice factors and recognises whether significant correlations among observations from the SP survey are present or not.

3.3. Ranking the importance of variables

We examine and rank the importance of variables for accurately distinguishing among different types of r-bikers who anticipate habitual changes in their behaviour upon replacing their regular bikes with e-bikes, as well as the types of e-bikers who actually experience these changes. To accomplish this, we use random forest (RF) (Breiman,

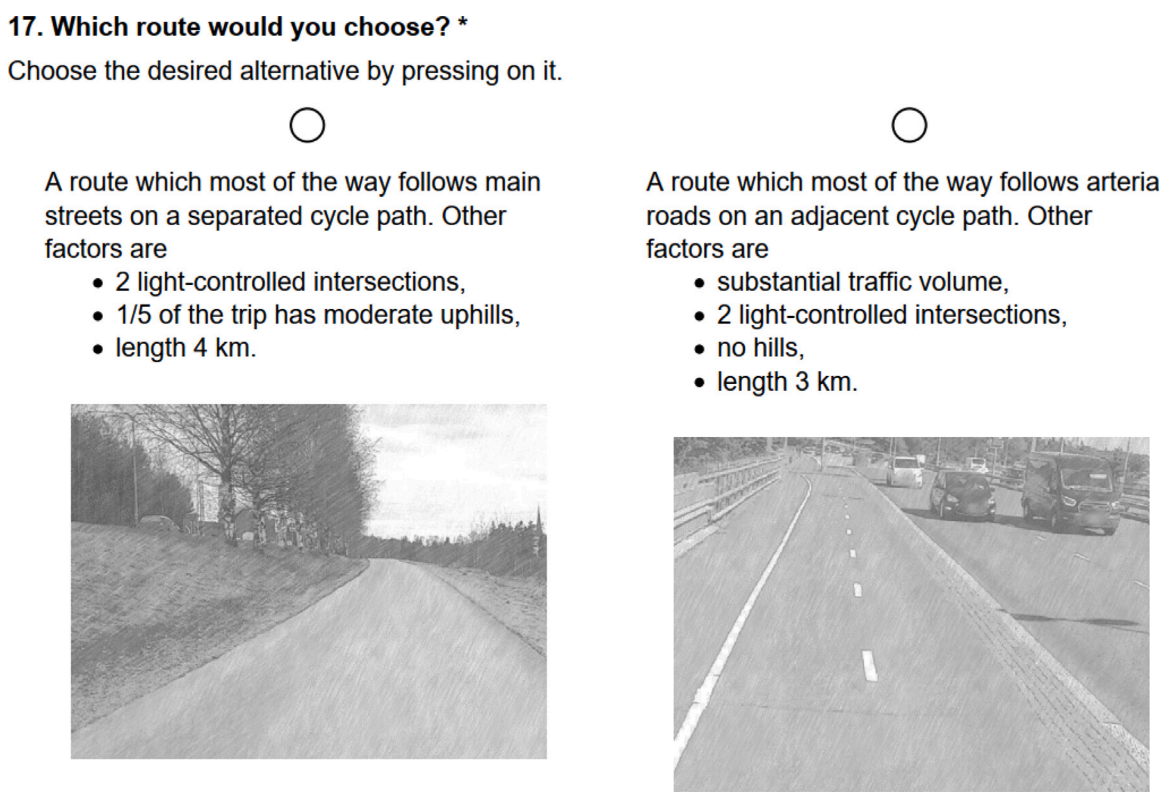

Fig. 1. An example of a choice situation used in the survey.

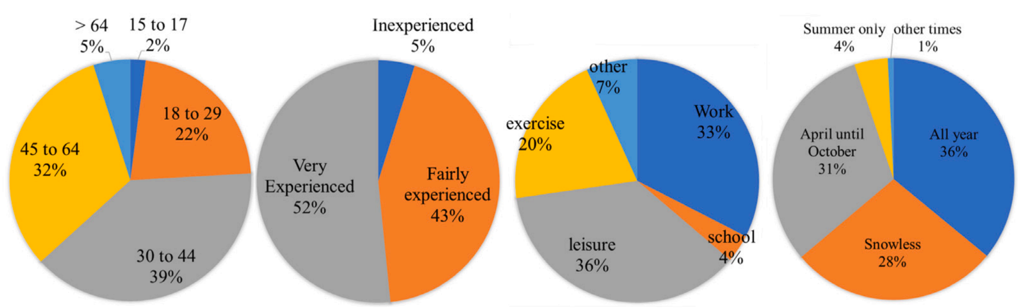

Fig. 2. Understudy population characteristics; (a) Age, (b) Cycling experience, (c) Trip purpose, and (d) Time of the year cycling.

2001), which is a classification approach widely used for variable importance ranking. RF is a collection of decision trees, where the aggregated vote of the trees forms the basis for decision-making. The RF provides significant variables along with their importance ranking. We employ RF to identify the features that demonstrate expected and experienced habit changes while using e-bikes.

In the RF procedure, to create each tree, a random subset of the data is utilised as the training data. Then, a random subset of features is selected to determine the branching of the tree at each node. This process continues by splitting new nodes based on a randomly selected feature until all the categories are defined within a tree.

The development of an RF model relies on two parameters: the number of trees, T, and the number of features, F, selected for branching. A schematic example of an RF is shown in Fig. 3.

Once all the trees are built, the importance of each variable in the classification can be measured by some indices such as: (i) Mean Decrease in Accuracy (MDA) and (ii) Mean Decrease in Gini Coefficient (MDG). The MDA index calculates the average decrease in accuracy across all trees, while the mean decrease in the Gini Coefficient measures the homogeneity contribution of each variable to nodes of trees. The choice of the index may vary depending on the specific implementation or the problem at hand. These measures rank the variables' importance values to exhibit more influential variables in the decision-making process or prediction ability of the model. Since the focus of this research is on identifying influential variables that contribute to the accurate classification of different types of bikers based on their anticipated or experienced behaviour changes upon using e-bikes, MDA serves as a suitable metric, as it is a commonly used index when predictive accuracy is the main concern (Song et al., 2021; Wang and Kim, 2019; Liaw et al., 2002). For a more in-depth understanding of the mathematical aspects of RF and the variable importance measures, readers are referred to Biau and Scornet (2016).

4. Results

Two sets of models (MNL and PMXL) are estimated using Stata 17 (StataCorp, 2021) for r-bike (with 930 observations), e-bike (with 99

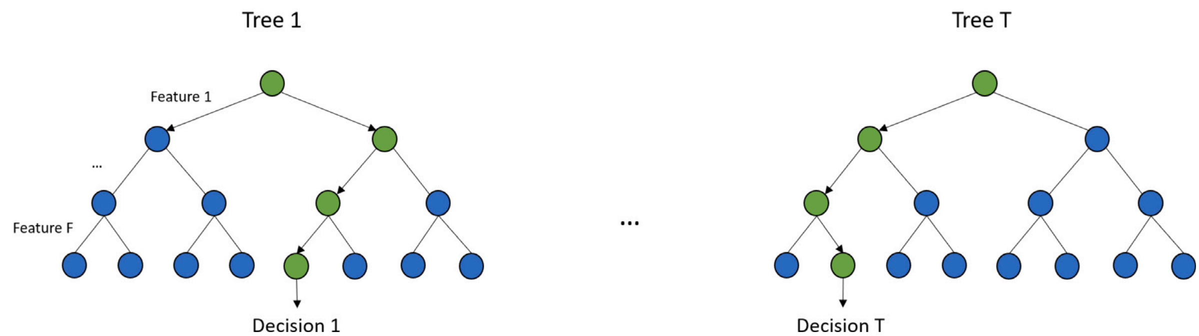

Fig. 3. Example of classification by random forest.

observations) route choices, and all the gathered data pooled together. The model coefficients are presented in Table 2 and their marginal effects are presented in Table 3. In PMXLs, the random parameter is the coefficient of the route's length variable, which is treated as a continuous variable within our estimation. This choice is justifiable since everyone has different preferences and constraints, including varying levels of fitness. The fact that they have different levels of tolerance for effort can influence how much importance they place on other variables during decision making. For example, the trade-off between the route's lengths and other characteristics like traffic level, greenery, and perceived safety can differ among individuals. Thus, some people might prefer longer routes because of the scenery (Koch and Dugundji, 2021). Since we assume that all individuals' coefficients of length are negative, we are considering a log-normal distribution, which requires using the length values with a negative sign.

Then, Logit models of r-bikers and e-bikers are compared to identify the prominent factors affecting the route choice of individuals, while the pedalling effort of cycling is removed due to the electrification of bikes. Using pooled models would verify the significance of changes observed between models of these segments. Another comparison is made between MNLs and PMXLs for r-bikes and e-bikes to evaluate the effect of error correlation on the effective route choice factors.

In the final step, the expected and the actual route choice changes reported by r-bikers and e-bikers, respectively, are presented and the outputs of the classification approaches are discussed to determine the factors associated with the expectation–reality gap.

4.1. R-bikers vs. E-bikers

Both sets of models provide similar results to previous studies regarding the route choice behaviour of r-bikers, consisting of the negative influence of length and steepness on the probability of choosing a route.[2] Moreover, less interaction with traffic through low adjacent traffic and the provision of completely separated bike facilities are the main factors that remained effective in r-bikers' and e-bikers' route choices. Some new insights are also observed for r-bikers: female cyclists avoid vehicular traffic and prefer controlled intersections in their routes more than men. The coefficient's sign of the variable representing the number of controlled devices in a route shows that r-bikers and e-bikers are generally reluctant to take routes with many controlled intersections. On the other hand, the coefficient of the dummy variable representing the presence of controlled intersection is positive for pooled models and r-bikers model, indicating that r-bikers prefer routes with controlled intersections; however, this preference diminishes with repeated exposure to controlled intersections.

[2] It is worth mentioning that in PMXL for r-bikers, the coefficient of a negative length value is equal to $exp(-0.188+0.690\epsilon)$, which is a positive value for any ϵ. We have used negative length values to ensure the negative effect of length on the route choice. Hence, the margin effect would be positive for a negative length of the route. A similar structure exists for other PMXLs.

R-biker models show bikers' preference towards main streets and arterial over small streets. Although more information is required to elucidate this observation, it can be potentially associated with the cycling speed, duration, or length variables. For instance, fast cycling along main streets is possible while cycling speeds on routes made up of small streets may be lower. In previous studies, the willingness of r-bikers to cycle on main streets has also been connected to other variables such as better lighting or less probability of coercion and robbery (Majumdar and Mitra, 2017). Although the variable representing the interaction of respondent being female and route consisting of arterial streets is preserved in the r-bikers model, its coefficient significance is not strong enough to show a difference among male and female r-bikers.

On the other hand, female e-bikers prefer small streets over major ones. The factors that may create safety perception for female e-bikers in small streets include the lower speed of vehicles, minor streets being less crowded than majors, and/or the greater number of controlled intersections. Thus, female e-bikers would choose routes with small streets more probably if they are more concerned about the speed of other vehicles. This difference is present among female and male e-bikers since e-bikes provide ease and a sense of confidence for cyclists which results in male e-bikers preferring to ride faster.

Female and older people's confidence would also increase and change their route choice decision if riding an e-bike. For instance, we observed that r-bikers older than 65 years find the traffic more disturbing than other r-bikers while e-bikers older than 65 years do not get bothered by heavy traffic. A similar attitude towards traffic situations is observed in female cyclists. Comparing models based on segmented data and pooled shows that pooled data is not capable of drawing these conclusions. Some other differences e-bikes make in cyclist route choice behaviour include the following:

- The male cyclists' preference towards main streets mitigates due to e-bikes while the corresponding coefficient for females stays the same (negative) as r-bikers.
- Although the length of the trip and hills are significant factors for both r-bikers and e-bikers, yet, as expected, the impacts of these variables are much milder for e-bikers.
- E-bikers prefer bike lanes, which is not the case for r-bikers.
- Female r-bikers have significant preferences for traffic avoidance, compared to men, while not all e-bikers like heavy and mixed traffic similarly.

Comparing pooled models with segmented ones verifies that the effect of hills becomes milder due to electric assist for bikes, as evidenced by the positive sign of its coefficient, yet it cannot show the significance of a change in route length.

Another finding is based on the results of the random parameter of the PMXLs. The large value of the standard deviation coefficient of the length variable indicates that heterogeneity in perception is significantly present in both r-bikers' and e-bikers' route choices. However, the confidence interval for e-bikers ([−2.31, 0.43]) shows more

Table 2
MNLs and PMXLs for E-bikers' and R-bikers' route choice.

Row	Variables	MNL			PMXL		
		R-Bike	E-bike	Pooled	R-Bike	E-bike	Pooled
1	Route consists of main streets	0.302*** (5.75)		0.273*** (5.52)	0.361*** (6.06)		0.329*** (5.83)
2	Route consists of arterial streets	0.461*** (6.51)		0.436*** (6.31)	0.461*** (7.77)		0.447*** (7.65)
3	Route is mixed with vehicular traffic	−0.798*** (−13.60)	−1.059*** (−7.42)	−0.797*** (−13.98)	−0.966*** (−14.09)	−1.174*** (−4.57)	−0.961*** (−14.48)
4	Route is on a bike lane					0.527* (1.92)	
5	Traffic of streets near the bike facility is at moderate level	−0.606*** (−10.01)	−0.406** (−2.18)	−0.591*** (−10.23)	−0.655*** (−10.18)	−0.592*** (−2.64)	−0.650*** (−10.53)
6	Traffic of streets near the bike facility is at heavy level	−1.579*** (−19.11)	−1.333*** (−7.59)	−1.585*** (−20.16)	−1.786*** (−18.92)	−1.901*** (−7.36)	−1.786*** (−19.81)
7	Traffic of streets near the bike facility is at substantial level	−1.242*** (−18.25)	−0.964*** (−5.47)	−1.223*** (−19.10)	−1.297*** (−17.05)	−1.045*** (−4.87)	−1.273*** (−17.81)
8	Number of controlled intersections in the route	−0.467*** (−13.01)	−0.349*** (−5.08)	−0.462*** (−13.71)	−0.538*** (−13.35)	−0.461*** (−5.62)	−0.537*** (−14.13)
9	Dummy variable representing presence of control devices at intersections of the route	0.675*** (8.26)		0.625*** (8.06)	0.749*** (8.18)		0.704*** (8.09)
10	Dummy variable representing presence of hills in the route	−1.091*** (−22.80)	−0.673*** (−5.35)	−1.081*** (−22.88)	−1.112*** (−20.59)	−0.731*** (−4.79)	−1.112*** (−20.69)
11	Interaction of respondent being female and route consisting of arterial Streets	−0.145 (−1.56)	−0.497*** (−2.42)	−0.145 (−1.64)		−0.443* (−1.93)	
12	Interaction of respondent being female and route's vehicular traffic being at heavy level	−0.381*** (−3.65)		−0.337*** (−3.31)	−0.516*** (−4.34)		−0.516*** (−4.51)
13	Interaction of respondent being female and number of controlled intersections in the route	0.116*** (3.35)	0.212** (2.06)	0.124*** (3.81)	0.130*** (3.49)	0.280** (2.38)	0.143*** (4.02)
14	Interaction of respondent being female and cycling mixed with vehicles	−0.486*** (−5.71)		−0.466*** (−5.67)	−0.569*** (−5.74)		−0.574*** (−6.00)
15	Interaction of respondent age being more than 65 and the route consisting of bike lane		0.942** (2.37)				
16	Interaction of respondent age being more than 65 and the vehicular traffic being heavy	−0.551** (−2.11)			−0.777** (−2.21)		−0.496* (−1.76)
17	Negative Value of Length of the route	−0.857*** (−32.73)	−0.749*** (−10.10)	−0.901*** (−27.06)	−0.1888*** (−4.20)	−0.352*** (−1.90)	−0.211*** (−4.80)
18	Interaction of respondent being female and length of the route			0.07* (1.70)			
20	Interaction of respondent having e-bike and route consisting of arterial streets			−0.319** (−2.30)			−0.396*** (−2.67)
21	Interaction of respondent having e-bike and route consisting of bike lane			0.380*** (2.68)			0.547*** (3.47)
22	Interaction of respondent having e-bike and dummy representing presence of hills in the route			0.377*** (2.94)			0.378*** (2.63)
23	Interaction of respondent having e-bike and route's length			0.109 (1.57)			
24	Standard deviation for the length of the route (log-normal distribution)				0.690***	0.998***	0.724***
25	Constant	0.079** (2.16)	0.179* (1.75)	0.097*** (2.80)	−0.098** (−2.83)	−0.251** (−2.29)	−0.112 (−3.02)
26	ρ^2	0.312	0.221	0.306	0.317	0.271	0.320
27	AIC	7112	875	7955	7068	820	7768
28	BIC	7226	921	8103	7186	871	7901

* Significance level: 90%
** Significance level: 95%
*** Significance level: 99%

dispersion than the r-bikers ([−2.17,0.17]), implying that the e-bike increases the variation of people's opinions toward the length of cycling. Comparing the model fitness levels based on Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) indicates that the PMXL model is better representative of the choice situation.

Ranking variables based on their marginal effect on the probability of the choice, Table 3, reveals that traffic levels and their interaction with other variables emerge as the most significant factors influencing route choice. Following closely are hills, route length, and the presence of cycling mixed with or separated from vehicular traffic, all exerting notable influence. A comparison between r-bike and e-bike models demonstrates that, while the effects of variables may vary, the ranking remains consistent. Additionally, note that ranking variables by the magnitude of coefficients yields a similar hierarchy of influential factors.

4.2. PMXL vs. MNL

In general, the outputs of the two types of models are quite aligned, especially for models based on segmented data. The PMXLs results verify the findings from MNLs, yet some small differences are present between MNL and PMXL which are:

- Lane bikes are favourable for e-bikers but not r-bikers based on PMXL which are not significant using MNL.
- Based on the MNL model, people older than 65 with e-bikes would prefer bike lanes more than other e-bikers, while this is not confirmed by the PMXL model.

It should be noted that no significant difference is found between the coefficients' signs of the two types of models. These findings demonstrate that the SP survey specifically designed for this research can

Table 3
Marginal effects of variables on route choice probability.

Row	Variables	MNL			PMXL		
		R-Bike	E-bike	Pooled	R-Bike	E-bike	Pooled
1	Route consists of main streets	0.075***		0.068***	0.052***		0.047***
2	Route consists of arterial streets	0.115***		0.109***	0.064***		0.059***
3	Route is mixed with vehicular traffic	−0.199***	−0.265***	−0.199***	−0.114***	−0.167***	−0.126***
4	Route is on a bike lane					0.074*	
5	Traffic of streets near the bike facility is at moderate level	−0.151***	−0.102**	−0.147***	−0.095***	−0.083***	−0.097***
6	Traffic of streets near the bike facility is at heavy level	−0.395***	−0.333***	−0.396***	−0.249***	−0.267***	−0.296***
7	Traffic of streets near the bike facility is at substantial level	−0.310***	−0.241***	−0.306***	−0.183***	−0.147***	−0.296***
8	Number of controlled intersections in the route	−0.116***	−0.087***	−0.116***	−0.075***	−0.065***	−0.086***
9	Dummy variable representing presence of control devices at intersections of the route	0.168***		0.156***	0.106***		0.092***
10	Dummy variable representing presence of hills in the route	−0.273***	−0.168***	−0.270***	−0.154***	−0.103***	−0.195***
11	Interaction of respondent being female and route consisting of arterial Streets	−0.036	−0.124**	−0.036		−0.071**	
12	Interaction of respondent being female and route's vehicular traffic being at heavy level	−0.095***		−0.084***	−0.071***		−0.065***
13	Interaction of respondent being female and number of controlled intersections in the route	0.029***	0.053**	0.031***	0.018***	0.034**	0.020***
14	Interaction of respondent being female and and cycling mixed with vehicles	−0.122***		−0.116***	−0.071***		−0.065***
15	Interaction of respondent age being more than 65 and the route consisting of bike lane		0.235**				
16	Interaction of respondent age being more than 65 and the vehicular traffic being heavy	−0.137**			−0.105**		−0.082**
17	Negative Value of Length of the route	−0.215***	−0.187***	−0.225***	0.124***	0.115***	0.122**
18	Interaction of respondent being female and length of the route			0.017* (1.70)			
20	Interaction of respondent having e-bike and route consisting of arterial streets			−0.079**			−0.051***
21	Interaction of respondent having e-bike and route consisting of bike lane			0.094***			0.603***
22	Interaction of respondent having e-bike and dummy representing presence of hills in the route			0.094***			0.063**
23	Interaction of respondent having e-bike and route's length			0.027*			

* Significance level: 90%
** Significance level: 95%
*** Significance level: 99%

capture the preferences of the individuals so that the errors in the responses are not strongly correlated. Still, this outcome is specific to the parameters tested in our study and can be influenced by many factors including the choices and variables classification; hence, it should not be interpreted as evidence for disproving all studies on the correlation of errors among observations using SP surveys.

4.3. Expectation vs. Reality

The questionnaire used in this research had a multiple-choice question asking about the habits of cyclists with and without e-bikes. We asked e-bikers if any of the following sentences describe their cycling habits.

- I cycle longer distances than before,
- I do not avoid hills as much as before,
- Trips tend to be faster than before, and
- I cycle more on asphalt-free roads than before.

Results show that 75% of e-bikers among our respondents claimed that they cycle longer than when they had r-bikes and 62% claimed that they do not avoid hills as much as before. Astonishingly, only 67% of e-bikers among our respondents reported faster cycling. Only 8% of e-bikers claimed they cycle more on asphalt-free roads than before, indicating that e-bikes are not widely preferred for off-road or unpaved routes.

Similar questions are asked from the current r-bikers to see how they expect to change after buying an e-bike. The 71% of current r-bikers who are considering buying an e-bike believe that they would cycle longer, 58% think they would not avoid hills like they are currently doing, while only 41% believe that trips would be faster. And 10% of r-bikers who consider buying an e-bike expect to cycle more on asphalt-free roads with e-bikes.

Here, to understand the chance of r-bikers' expectation occurrence, we classify the population based on their characteristics and their habit change. This classification aims to identify whether individuals with similar characteristics share similar beliefs regarding the effects of e-bikes on their cycling habits.

Two RF classifications are created, one based on e-bikers' responses and the other based on r-bikers' responses, to compare the r-bikers' expectation of e-bike effects on their route choice and cycling habits with the reality of changes reported by current e-bikers. As the RF outputs are not very sensitive to the number of trees in the forest and number of variables for branching (Liaw et al., 2002), we follow the recommendations that suggest the number of trees ranging from 64 to 128 is suitable for most cases (Oshiro et al., 2012) and the square root of the total number of features available in data is a good approximation for branching (Liaw et al., 2002). Considering our problem, which involves 22 available features for classification, it is advisable to select 5 variables for branching at each tree. The social demographic data gathered through the survey consists of the respondents' gender, age, level of cycling experience, time of year for cycling, the purpose of cycling, and cycling frequency. The changes in cycling habits we consider are cycling longer and not avoiding hills, i.e., taking uphills with e-bikes. The free R package (Liaw et al., 2002) is employed for implementing the RF method.

4.3.1. Longer cycling distance

The first change of habit that is investigated is to cycle longer distances with e-bikes. The importance values of the first five features measured based on MDA are presented in Table 4. The MDA index shows that r-bikers who cycle with the purpose of exercising believe that they would cycle longer distances with e-bikes. This belief appears to be rational since e-bikes provide assistance and make cycling less strenuous, which results in longer cycling distances. The main feature in the classification of the e-bikers who cycle longer distances is also found to be exercise purposes. This suggests that exercise-oriented e-bikers are more likely to cycle longer distances, in reality, aligning with the expectations of r-bikers. In our study, 35 e-bikers reported exercise as their trip purpose while 439 r-bikers reported exercise as their trip purpose.

Additionally, the gender of r-bikers is identified as the next most important factor. Female r-bikers expect that with e-bikes, they would cycle longer distances. However, it is noteworthy that being female is not among the top factors influencing the classification of e-bikers who actually cycle longer distances. This indicates that while female r-bikers have higher expectations of increased cycling distances with e-bikes, female e-bikers may not experience this change to the same extent in reality.

The subsequent most important factors are related to the purpose of the trips with the bike, specifically, commuting to work, leisure trips, and commuting for study. The MDA analysis suggests that r-bikers who travel with these specific purposes expect to cycle longer distances when using e-bikes. Regarding leisure trips, the results in Table 4 indicate that e-bikers who cycle for leisure purposes do, in fact, take longer routes in reality. This aligns with the rational expectation that e-bike users might cycle more for leisure since the assistance from the e-bike makes it easier and less physically demanding to pedal. On the other hand, e-bikers who commute to study do not experience taking longer trips with e-bikes. This is likely because school or university locations are fixed, and there is little flexibility in choosing longer routes for study-related trips. Furthermore, e-bikers who commute to work report that they believe they would be cycling longer distances, as commuting for work is the fourth most important variable in e-bikers' classification. This could indicate either adapting their route choice to take longer routes or choosing to cycle for trips where they normally would not cycle. However, drawing conclusions on these aspects requires further investigations.

The RF based on assisted cyclists shows that e-bikers aged 45 to 64 years have experienced taking longer routes with e-bikes while no age-related impact is seen in the expectations of r-bikers for taking longer routes if they use e-bikes. In our study, e-bikers aged 45 to 64 comprise 43 percent of the e-bikers, totalling 43 individuals. Based on the RF analysis for e-bikers, it is evident that individuals aged 45 to 64 who use e-bikes have experienced taking longer routes compared to their regular biking habits. These findings highlight the potential benefits of e-bike usage for older individuals, as it allows them to comfortably take longer routes, potentially increasing their mobility and encouraging more active transportation. However, when looking at

Table 4
Variables' importance estimated by RF for R-bikers/E-bikers expected/experienced cycling longer while using E-bikes.

Rank	R-bikers		E-bikers	
	Feature	MDA	Feature	MDA
1	Cycling purpose: Exercise	60.23	Cycling purpose: Exercise	31.82
2	Being Female	55.44	Cycling purpose: Leisure	26.28
3	Cycling purpose: Work	51.35	Age: 45 to 64	25.17
4	Cycling purpose: Leisure	46.79	Cycling purpose: Work	24.61
5	Cycling purpose: Study	44.34	Being Female	23.99

r-bikers, there is no age-related impact on their expectations for taking longer routes if they were to use e-bikes. This means that the age of r-bikers does not significantly influence their expectations regarding the increase in cycling distances with e-bikes.

These findings shed light on the role of cycling purposes, age, and gender in shaping expectations and actual route choices of e-bike users. Understanding these factors can aid in developing targeted strategies for promoting e-bike usage among specific user groups and encouraging longer cycling distances to enhance the appeal and effectiveness of e-bikes as a sustainable and user-friendly transportation option.

4.3.2. Avoiding hills

Based on the RF analysis, as shown in Table 5, examining r-bikers' expectations about route steepness (avoiding hills) reveals that female r-bikers strongly believe that with e-bikes, they would be less concerned about hills. This is supported by the MDA index, which indicates that being female is the most important factor in decision trees of r-bikers' expectations. Female e-bikers also experienced not avoiding hills in reality, and being female is the second most important factor in the RF analysis for e-bikers. This suggests that the perception of female cyclists aligns with their actual experiences when using e-bikes, indicating consistency between expectations and reality.

Among r-bikers, those who exercise by cycling believe that they would be less concerned about hills with e-bikes. This expectation is rational since the pedal-assist of e-bikes would make their exercise sessions easier. This aligns with the reality that e-bikers who cycle for exercise experience not avoiding hills compared to before using e-bikes as the most important factor in this classification is found to be exercising with an e-bike.

Other trip purposes such as commuting to study, leisure activities, and commuting to work also influence the expectations of r-bikers regarding avoiding hills if they use e-bikes. However, interestingly, these factors are not as important in the real world. E-bikers with the same trip purposes did not report any change in their habits regarding riding uphills compared to before using e-bikes. This suggests that while trip purposes may influence initial expectations, they might not have a significant impact on actual behaviour when using e-bikes.

On the other hand, the RF analysis reveals that e-bikers aged older than 45 years and those who cycle frequently (i.e., every day) experience not being concerned about hills when using e-bikes. This suggests that older individuals and frequent cyclists may find the assistance provided by e-bikes particularly beneficial in overcoming uphill challenges. However, older r-bikers might not be aware of this benefit since they have not yet experienced the assistance provided by e-bikes. This lack of knowledge could influence their expectations, leading to not anticipating the extent to which e-bikes can ease uphill cycling. Thus, by providing information to them, policymakers can contribute to engaging the elderly in cycling by making their cycling experiences more enjoyable and less physically demanding. As was mentioned in Section 4.3.1, 43 percent of e-bikers are aged between 45 to 64 while 11 percent (i.e., 11 individuals) are aged 65 and older.

Table 5
Variables' importance estimated by RF for R-bikers/E-bikers expected/experienced not avoiding hills while using E-bikes.

Rank	R-bikers		E-bikers	
	Feature	MDA	Feature	MDA
1	Being Female	56.45	Cycling purpose: Exercise	32.25
2	Cycling purpose: Exercise	51.18	Being Female	28.63
3	Cycling purpose: Study	47.98	Age: More than 65	27.98
4	Cycling purpose: Leisure	46.76	Age: 45 to 64	26.54
5	Cycling purpose: Work	44.53	Cycling every day	26.34

5. Conclusion and discussion

This research investigates the differences in r-bikers' and e-bikers' route choice behaviour. Two sets of models, the MNL and PMXL, are estimated based on data collected from a stated preference survey in the Greater Helsinki area, Finland. Both sets of models verify the previous findings in the literature, demonstrating that the probability of choosing a route for cyclists is negatively influenced by trip length and steepness. Notably, riding an e-bike reduces the importance of the length of the trip and steepness, and e-bikers care less about the type of facility and road type, i.e., major or minor streets, that they are cycling along. Still, providing dedicated routes with no interruptions from vehicular traffic can be introduced as the main effective factor in bike promotion. Concluding on the differences between e-bikers' and r-bikers' route choices, it is apparent that e-bikes would remove the sensitivity of old people and female cyclists to heavy vehicular traffic alongside their cycling path. Moreover, e-bikes would also make older people prefer bike lanes over other facilities. It also would remove the excess sensitivity of female cyclists toward mixed cycling with other vehicles. These findings suggest that e-bikes would make the cyclists' preference towards routes more uniform, which can result in a more predictable outcome after policy implementation.

Regarding the model specification, we observed that there is no substantial difference between MNLs and PMXLs for e-bikers, in our case. This implies that the SP survey is designed and conducted properly in which errors in the responses are not strongly correlated and can be reasonably assumed to possess the IID property. If the IID assumption holds, respondents' preferences do not cause a correlation among the error terms. Hence, in the presence of IID property, the MNLs offer comparable outputs to PMXLs without increasing the complexity of the estimation process. However, in cases where the errors feature correlation, the potential enhancement in model fit provided by PMXLs, allowing for the existence of heterogeneity among responses may be offset by the increased computational burden.

The assessment of habit changes resulting from the reduced pedalling effort with e-bike reveals that r-bikers indicate a willingness to cycle longer distances, especially when cycling for exercise or engaging in delivery duty. Moreover, female r-bikers are found to be most promising about their cycling length after transitioning to an e-bike. The analysis based on e-bikes' reported habit changes verifies these changes, implying that individuals in this group successfully achieved their intended changes through e-bikes. Conversely, r-bikers' expectations of increased tolerance towards steepness are found to be unrealistic since factors influencing this reported change among e-bikers differ significantly from those of r-bikers'.

A combination of both approaches yields valuable insights that can enlighten transportation planners and policymakers about the impacts and transformations brought about by e-bikes. For instance, the logit models reveal females' willingness to undertake longer routes compared to men. Additionally, the RF analysis provides insight into the relative importance of gender compared to other cyclist characteristics, particularly for e-bikers and r-bikers. In another instance, the logit models demonstrate the e-bikers' inclination to tackle hills compared to r-bikers, and RF analysis provides insight into which characteristics of e-bikers influence hill acceptance, which are trip purpose, age, gender, and frequency of cycling.

A major limitation of this research, which is consistent with previous literature on cycling behaviour, and is also present in the RP studies based on GPS tracking of cyclists (Dane et al., 2020; Menghini et al., 2010; Huber et al., 2021), is that all respondents are already cyclists. Hence, the results cannot be directly extrapolated to non-cyclists regarding their preferences and obstacles towards biking. However, focusing on cyclists is needed in this research due to the fact that we were looking for the differences created by e-bike implementation. Moreover, adding more randomness to the PMXL by considering other variables' possible taste heterogeneity among the population is believed to result in a more realistic model. Similar efforts have been found by Koch and Dugundji (2021) with cycling environment-related variables. Additionally, researchers ideally should combine the use of RP and SP to mitigate some of the disadvantages while benefiting from strengths (Ben-Akiva et al., 1994), but applying both methods may prove to be excessively demanding which may explain why only one of the methods is commonly applied in the majority of route choice literature. Furthermore, we compare the current e-bikers beliefs regarding their behaviour with the expectations of current r-bikers about their possible future behaviour. Since these two groups of people are not precisely similar, there can be questions regarding the findings based on the differences in population. A before and after study based on a controlled group of people can be a future extension of the current research.

CRediT authorship contribution statement

Khashayar Khavarian: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Shaghayegh Vosough:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Claudio Roncoli:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

We gratefully acknowledge the Helsinki Region Transport Authority (HSL) for their support in collecting and providing the data used in this study. We also thank Konsta Tarkkala for his help with the data. The work of Shaghayegh Vosough and Claudio Roncoli was partly funded by the Academy of Finland project AIforLEssAuto (no. 347200).

References

Aloulou, F., 2018. The application of discrete choice models in transport. In: Statistics-Growing Data Sets and Growing. p. 85.

Anderson, C.C., Clarkson, D.E., Howie, V.A., Withyman, C.J., Vandelanotte, C., 2022. Health and well-being benefits of e-bike commuting for inactive, overweight people living in regional australia. Health Promot. J. Aust.: Official J. Aust. Assoc. Health Promot. Prof. 33 (S1), 349–357.

Andersson, A.S.F., Adell, E., Hiselius, L.W., 2021. What is the substitution effect of e-bikes? a randomised controlled trial. Transp. Res. Part D: Transp. Environ. 90, 102648.

Arning, L., Silva, C., Kaths, H., 2023. Review of current practice and research on e-bikes in transport models. Transp. Res. Rec. 03611981231168848.

Axhausen, K.W., Hess, S., König, A., Abay, G., Bates, J., Bierlaire, M., 2006. State of the art estimates of the swiss value of travel time savings. Arbeitsberichte Verkehrs-Raumplanung 383.
Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H., Rao, V., 1994. Combining revealed and stated preferences data. Market. Lett. 5, 335–349.
Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application To Travel Demand, vol. 9, MIT Press.
Bernardi, S., Geurs, K., Puello, L.L.P., 2018. Modelling route choice of dutch cyclists using smartphone data. J. Transp. Land Use 11 (1), 883–900.
Biau, G., Scornet, E., 2016. A random forest guided tour. Test 25, 197–227.
Bourne, J.E., Cooper, A.R., Kelly, P., Kinnear, F.J., England, C., Leary, S., Page, A., 2020. The impact of e-cycling on travel behaviour: A scoping review. J. Transp. Health 19, 100910.
Brandt, E., Kantele, S., Räty, P., 2019. Liikkumistottumukset Helsingin seudulla 2018 (travel habits in the Helsinki region in 2018). In: HSL Helsingin seudun liikenne (HSL Helsinki Region Transport). https://www.kauniainen.fi/wp-content/uploads/2021/04/hsl_julkaisu_9_2019_netti1.pdf.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Broach, J., Dill, J., 2016. Using predicted bicyclist and pedestrian route choice to enhance mode choice models. Transp. Res. Rec. 2564 (1), 52–59.
Broach, J., Dill, J., Gilebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transp. Res. Part A: Policy Pract. 46 (10), 1730–1740.
Brownstone, D., Bunch, D.S., Train, K., 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. Transp. Res. B 34 (5), 315–338.
Buehler, R., 2012. Determinants of bicycle commuting in the washington, dc region: The role of bicycle parking, cyclist showers, and free car parking at work. Transp. Res. Part D: Transp. Environ. 17 (7), 525–531.
Casello, J.M., Usyukov, V., 2014. Modeling cyclists' route choice based on gps data. Transp. Res. Rec. 2430 (1), 155–161.
Castro, A., Gaupp-Berghausen, M., Dons, E., Standaert, A., Laeremans, M., Clark, A., Anaya-Boig, E., Cole-Hunter, T., Avila-Palencia, I., Rojas-Rueda, D., et al., 2019. Physical activity of electric bicycle users compared to conventional bicycle users and non-cyclists: Insights based on health and transport data from an online survey in seven european cities. Transp. Res. Interdiscip. Perspect. 1, 100017.
Chavis, C., Martinez, V., 2021. E-Bikes' Effect on Mode and Route Choice: a Case Study of Richmond, VA Bike Share. Technical report, Urban Mobility & Equity Center, Morgan State University.
Cherry, C.R., Weinert, J.X., Xinmiao, Y., 2009. Comparative environmental impacts of electric bikes in China. Transp. Res. Part D: Transp. Environ. 14 (5), 281–290.
Counter, Eco, 2021. Bike count dashboard: tracking the growth of cycling by country. https://www.eco-counter.com/cycling-data-tracker/.
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: Lecture Notes in Geoinformation and Cartography, pp. 109–124.
de Dios Ortúzar, J., Willumsen, L.G., 2011. Modelling Transport. John Wiley & Sons.
de Palma, A., Vosough, S., Liao, F., 2022. An overview of effects of covid-19 on mobility and lifestyle: 18 months since the outbreak. Transp. Res. Part A: Policy Pract. 159, 372–397.
Eriksson, H., 2024. Hsl:n vastuullisuusraportti 2023 (hsl's responsibility report 2023). In: HSL Helsingin seudun liikenne (HSL Helsinki Region Transport). https://hslfi.azureedge.net/globalassets/julkaisuarkisto/2024/vastuullisuusraportti_190324.pdf.
Fosgerau, M., Łukawska, M., Paulsen, M., Rasmussen, T.K., 2023. Bikeability and the induced demand for cycling. Proc. Natl. Acad. Sci. 120 (16), e2220515120.
Goel, R., Goodman, A., Aldred, R., Nakamura, R., Tatah, L., Garcia, L.M.T., Zapata-Diomedi, B., de Sa, T.H., Tiwari, G., de Nazelle, A., et al., 2022. Cycling behaviour in 17 countries across 6 continents: levels of cycling, who cycles, for what purpose, and how far? Transp. Rev. 42 (1), 58–81.
Gojanovic, B., Welker, J., Iglesias, K., Daucourt, C., Gremion, G., 2011. Electric bicycles as a new active transportation modality to promote health. Med. Sci. Sports Exerc. 43 (11), 2204–2210.
Haustein, S., Nielsen, T.A.S., 2016. European mobility cultures: A survey-based cluster analysis across 28 european countries. J. Transp. Geogr. 54, 173–180.
Hensher, D.A., Rose, J.M., Rose, J.M., Greene, W.H., 2005. Applied Choice Analysis: A Primer. Cambridge University Press.
Hood, J., Sall, E., Charlton, B., 2013. A GPS-based bicycle route choice model for San Francisco, California. Transp. Lett. 3 (1), 63–75.
HSL, 2024. Helsinki region land use, housing and transport plan. In: MAL 2023: Summary. HSL Helsinki Region Transport. https://hslfi.azureedge.net/globalassets/hsl/mal/mal-julkaisut/2023/mal_tiivistelmaraportti_210x260_eng_2023_web.pdf.
Huber, S., Lißner, S., Lindemann, P., Muthmann, K., Schnabel, A., Friedl, J., 2021. Modelling bicycle route choice in german cities using open data, mnl and the bikesim web-app. In: 7th International Conference on Models and Technologies for Intelligent Transportation Systems. MT-ITS.
Hull, A., O'Holleran, C., 2014. Bicycle infrastructure: Can good design encourage cycling? Urban, Plan. Transp. Res. 2 (1), 369–406.
Hwang, U., Guhathakurta, S., 2023. Exploring the impact of bike lanes on transportation mode choice: A simulation-based, route-level impact analysis. Sustainable Cities Soc. 89, 104318.
Jenkins, M., Lustosa, L., Chia, V., Wildish, S., Tan, M., Hoornweg, D., Lloyd, M., Dogra, S., 2022. What do we know about pedal assist e-bikes? A scoping review to inform future directions. Transp. Policy 128, 25–37.
Khachatryan, T., 2021. Helsinki city bikes: exploratory data analysis. https://towardsdatascience.com/helsinki-city-bikes-exploratory-data-analysis-e241ce5096db.
Koch, T., Dugundji, E.R., 2021. Taste variation in environmental features of bicycle routes. In: Proceedings of the 14th ACM SIGSPATIAL International Workshop on Computational Transportation Science. pp. 1–10.
Kroesen, M., 2017. To what extent do e-bikes substitute travel by other modes? Evidence from the Netherlands. Transp. Res. Part D: Transp. Environ. 53, 377–387.
Kuva, H., 2020. Smaller volumes but same e-bike trend in Finland. https://digimagazine.bike-eu.com/market_reports/finland.
Larsen, M.H., Hansson, K.E., Larsen, E.H., Fridh, M.K., Petersen, N.N., Mellblom, A.V., Ruud, E., Larsen, H.B., Lie, H.C., 2022. The gap between expectations and reality: A qualitative study of psychosocial challenges of young childhood cancer survivors from the paccs study. Eur. J. Cancer Care 31 (6), e13696.
Liaw, A., Wiener, M., et al., 2002. Classification and regression by random forest. R News 2 (3), 18–22.
Majumdar, B.B., Mitra, S., 2017. Valuing factors influencing bicycle route choice using a stated-preference survey. J. Urban Plann. Dev. 143 (3), 04017001.
McFAdden, D., 1974. Conditional logit analysis for qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. pp. 105–142.
McFadden, D., Train, K., 2000. Mixed mnl models for discrete response. J. Appl. Econometrics 15 (5), 447–470.
Meister, A., Felder, M., Schmid, B., Axhausen, K.W., 2023. Route choice modeling for cyclists on urban networks. Transp. Res. Part A: Policy Pract. 173, 103723.
Menghini, G., Carrasco, N., Schussler, N., Axhausen, K.W., 2010. Route choice of cyclists in zurich. Transp. Res. Part A: Policy Pract. 44 (9), 754–765.
Moffett, J.A.K., Newbronner, E., Waddell, G., Croucher, K., Spear, S., 2000. Public perceptions about low back pain and its management: A gap between expectations and reality? Health Expect. 3 (3), 161–168.
Nelson, T., Ferster, C., Laberee, K., Fuller, D., Winters, M., 2021. Crowdsourced data for bicycling research and practice. Transp. Rev. 41 (1), 97–114.
Oshiro, T.M., Perez, P.S., Baranauskas, J.A., 2012. How many trees in a random forest?. In: Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, Proceedings, vol. 8, Springer, pp. 154–168.
Prashar, A., Gupta, P., Jeyaraj, A., Dwivedi, Y.K., 2022. Expectation-reality gap in information technology discontinuance intention. J. Comput. Inf. Syst. 1–13.
Rérat, P., 2021. The rise of the e-bike: Towards an extension of the practice of cycling? In: Mobilities.
Sałabun, W., Palczewski, K., Wątróbski, J., 2019. Multicriteria approach to sustainable transport evaluation under incomplete knowledge: Electric bikes case study. Sustainability 11 (12), 3314.
Schoner, J.E., Levinson, D.M., 2014. The missing link: Bicycle infrastructure networks and ridership in 74 US cities. Transportation 41 (6), 1187–1204.
Scott, D., Lu, W., Brown, M., 2021. Route choice of bike share users: Leveraging GPS data to derive choice sets. J. Transp. Geogr. 90, 102903.
Segadilha, A.B.P., Sanches, S.D.P., 2014. Identification of factors that influence cyclistś route choice. Procedia - Soc. Behav. Sci. 160, 372–380.
Song, S., He, R., Shi, Z., Zhang, W., 2021. Variable importance measure system based on advanced random forest. CMES Comput. Model. Eng. Sci. 128 (1).
StataCorp, 2021. Stata statistical software: release 17.
Stinson, M.A., Bhat, C.R., 2003. Commuter bicyclist route choice: Analysis using a stated preference survey. Transp. Res. Rec. (1828), 107–115.
Stroh, L.K., Gregersen, H.B., Black, J.S., 1998. Closing the gap: Expectations versus reality among repatriates. J. World Bus. 33 (2), 111–124.
Tarkkala, K., 2022. Application of cyclists' route choice preferences in travel demand modelling.
Tarkkala, K., Vosough, S., West, J., Roncoli, C., 2023. Evaluating the influence of cyclists' route choice incorporation into travel demand modelling: A case study in Greater Helsinki. submitted for publication.
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.
Train, K.E., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press, Cambridge.
Wang, X., Kim, S.H., 2019. Prediction and factor identification for crash severity: Comparison of discrete choice and tree-based models. Transp. Res. Rec. 2673 (9), 640–653.
Webropol, 2021. Webropol powerful insight. https://webropol.fi/.
World Health Organization, 2020. WHO guidelines on physical activity and sedentary behaviour: web annex: evidence profiles. https://apps.who.int/iris/handle/10665/336657.
Zhu, S., Ma, J., Tang, T., Shi, Q., 2020. A combined modal and route choice behavioral complementarity equilibrium model with users of vehicles and electric bicycles. Int. J. Environ. Res. Public Health 17 (10), 3704.