The Journal of Transport and Land Use jtlu.org
Vol. 17 No. 1 [2024] pp. 557–577

Exploring factors affecting route choice of cyclists: A novel varying-contiguity spatially lagged exogenous modeling approach

Nick van Nijen
University of Twente
nick.vannijen@outlook.com

M. Baran Ulak (contributing author)
University of Twente
m.b.ulak@utwente.nl

Sander Veenstra
Witteveen+Bos
sander.veenstra@witteveenbos.com

Karst Geurs
University of Twente
k.t.geurs@utwente.nl

Abstract: Cycling is one of the main transport modes and cycling infrastructure is strongly embedded in transport infrastructure in the Netherlands. Nonetheless, the bicycle network still undergoes frequent improvements and expansions. One of the critical elements in deciding on improvements and expansions is to understand the route choice of cyclists, which helps identify bottlenecks in bicycle flows and substantiate the need for new bicycle infrastructure. Yet, the factors affecting the route choice of cyclists are still not fully understood. To address this, we develop a varying-contiguity spatially lagged exogenous (VCSLX) model and analyze the probability of a cyclist choosing a certain segment based not only on the characteristics of that segment but also considering the characteristics of its neighbors along a route. Characteristics that are included in this study are the presence of bicycle infrastructure, traffic control installations and artificial lighting, as well as pavement type, bicycle and motorized-vehicle volumes and different land-use zones. The model involves the analysis of the observed routes extracted from cycling trajectories from Fietstelweek data, as well as corresponding hypothetical shortest path routes identified from the origin-destinations of the observed trips and the cycling network. The results of the study can help to understand the factors convincing cyclists to deviate from the shortest possible routes. The study contributes to the current literature by focusing on the underexplored aspect of spatial dependencies between route segments in the route choice of cyclists.

Keywords: Bicycle route choice, varying-contiguity spatially lagged exogenous model, bicycle infrastructure, land use

Article history:
Received: October 30, 2023
Received in revised form: March 21, 2024
Accepted: April 14, 2024
Available online: July 12, 2024

1 Introduction

In the Netherlands, the bicycle is one of the main transport modes frequently used for commuting, shopping, or leisure (28% of all trips in 2022 (CBS, 2022)). Therefore, cycling infrastructure is strongly embedded in the transport infrastructure of the Netherlands. Nonetheless, the bicycle network still undergoes improvements and

---

https://doi.org/10.5198/jtlu.2024.2452
ISSN: 1938-7849 |

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

expansions to strengthen the position of cycling even more (Ensing & Janssen, 2020). It is important to understand the route choice of cyclists to decide on the improvements and expansions particularly in a country already with a complex bicycle network. Preferences of cyclists are often used to identify bottlenecks in bicycle flows and substantiate the need for new bicycle infrastructure (Veenstra, 2021). Nonetheless, the factors affecting the route choice of cyclists are still not fully understood.

Trip length and trip duration are considered to be the attributes most often cited in the literature as factors influencing the route choice of cyclists, suggesting that the shortest path would be sufficient to estimate the route between the origin and destination for cyclists (Chen, 2016; Khatri et al., 2016; Prato et al., 2018; Winters et al., 2010). However, literature indicates that various environmental and infrastructural factors are influencing the route choice of cyclists such that the route choice deviates from the shortest path. For example, the cycling routes collected during a data collection period of two weeks in the study area, municipality of Enschede, showed that only 55.1% of the observed routes fully overlap with the corresponding shortest path (Fietstelweek, 2016).

The results of stated and revealed preference surveys show that the effects of environmental factors on the route choice of cyclists vary between studies and contexts. Alattar et al. (2021) argue that one of the main limitations in previous literature concerning the route choice of cyclists is disregarding the spatial dependencies among the segments along a route. Including such spatial aspect can provide more robust results. Current literature that considers spatial dependency solely aims at explaining bicycle volumes rather than the route choice of cyclists. In that context, all first-order neighboring segments are considered in a network. However, considering all first-order neighboring segments in a network would be inappropriate for route choice analysis, as cyclists do not choose their route based on individual segments. Rather, they select multiple neighboring segments in sequence to construct their route. For example, cyclists will not select one segment with excellent bicycle infrastructure if that results in including many other segments without bicycle infrastructure in their route. They rather choose a set of spatially coherent segments with good bicycle infrastructure. This illustrates that individual segments are spatially affected by neighboring segments in a route. Lv et al. (2023) verified this effect by showing that the characteristics of adjacent segments influence the bicycle traffic flow at the targeted segment. They concluded that segments with extensive bicycle facilities do not always have high bicycle volumes and attributed this difference to the effect of adjacent segments.

The objective of this study is to explore the factors affecting the route choice of cyclists. For this purpose, we developed a varying-contiguity spatially lagged exogenous (VCSLX) model to analyze the probability of choosing a certain segment based on its characteristics and the characteristics of neighboring segments. These certain segments are parts of the observed routes (based on cycling trajectories obtained from Fietstelweek data) and corresponding hypothetical shortest path routes identified from the origin-destinations of trips and the cycling network. We assume that the cyclists would prefer the shortest possible route if there were no effect of segment characteristics. Therefore, the analysis results can help in understanding the factors convincing cyclists to deviate from the shortest possible routes. The study contributes to the current literature by focusing on an underexplored aspect of the route choice of cyclists, namely the spatial dependencies between route segments. Incorporating this component ensures that the probability of choosing a segment depends not only on its own characteristics but also on those of neighboring segments within a route.

2 Literature review

2.1 Analysis of route choice of cyclists

There exist different approaches to analyze influential factors on the route choice of cyclists, such as the stated preference surveys (Li et al., 2012; Strauss & Miranda-Moreno, 2013; Winters et al., 2010). This approach has the advantage of controlling for the experimental environment (Prato et al., 2018). However, this approach has a major downside, as it only includes the intentions of cyclists rather than their actual choices (Bernardi et al., 2018). Data on the revealed preferences of cyclists, like the observed routes from Fietstelweek, are becoming more available with the increase in the GPS track datasets and studies using such datasets can mitigate the disadvantages of stated preference studies.

Recent research has highlighted the importance of addressing spatial dependencies in route choice modelling (Alattar et al., 2021; De Jong et al., 2023). Alattar et al. (2021) note that current methods often overlook spatial aspects, despite spatial models providing more insights and better estimation for events with spatial autocorrelation, such as cycling (Lee & Sener, 2021). Alattar et al. (2021) considered two types of models that can deal with spatial dependency, namely the spatial lag model and the spatial error model. The first accounts for the spatial dependence of the dependent variable by neighboring areas, while the latter does not consider the dependent and independent variables, but rather includes an error term to estimate the spatial dependency. Both approaches are widely used in many applications (e.g., estimating bike sharing demand), describing a relation between bicycling frequency and land use characteristics and demand modelling for traffic volume prediction (Faghih-Imani & Eluru, 2016; Zhao et al., 2020). Moreover, De Jong et al. (2023) clustered segments which share the same street names in an error term to account for spatial dependency and used this clustering in both a cycling demand model and a cycling route choice model. They included examples of revealed versus shortest routes and conclude that the selection or avoidance of a route depends on some specific segment(s) and contribute this to spatial dependency.

Spatial lag and spatial error model incorporate the impact of the dependent variable or the error term respectively of neighboring units on the dependent variable. As a result, both models are strong prediction models (Alattar et al., 2021). However, their explanatory power is less compared to the spatial lagged exogenous (SLX) model. The SLX model measures the impact of the explanatory variables in neighboring units on the dependent variable (Elhorst & Vega, 2017). Therefore, the model is conditional upon the infrastructure, traffic and land use allocation factors of neighboring segments. Thus, this approach allows for the consideration of the spatial dependence of different infrastructure, traffic and land use allocation factors separately allowing for a relatively straightforward model estimation and interpretation (Goetzke, 2008).

In general, one contiguity matrix is created to symbolize the spatial arrangement of neighboring units (Goetzke, 2008). Nevertheless, this would not be appropriate in this study, since a certain segment can be part of multiple routes and its contiguity varies depending on the route it belongs to. Therefore, that segment is only affected by the characteristics of other segments in that particular route and not in other routes. Therefore, a contiguity matrix was created for every route individually. This resulted in a new type of SLX model, namely the varying-contiguity spatially lagged exogenous (VCSLX) model. A detailed representation of the model is provided in the methodology.

2.2 Environmental factor influencing the route choice of cyclists

This section focusses on the explanatory variables that explained bicycle route choice in previous studies. A brief overview of selected papers on the influence of infrastructure, traffic and land use allocation factors on the route choice of cyclists published in the 21st century is presented in Table 1. It addresses the factors included and the direction of the included factors.

Table 1. Summary of the infrastructural, traffic and land use allocation factors, and their effect according to literature

Factor/Source	(Broach et al., 2012)	(Campos-Sánchez et al., 2019)	(Chen, 2016)	(De Jong et al., 2023)	(Khatri et al., 2016)	(Koch & Dugundji, 2021)	(Li et al., 2012)	(Łukawska et al., 2023)	(Meister et al., 2023)	(Prato et al., 2018)	(Saelens et al., 2003)	(Utley et al., 2020)	(Winters et al., 2010)	(Winters et al., 2011)	(Zhao et al., 2020)	(Zimmermann et al., 2017)
Traffic control installations	±				+	−			−	−						
Cycle lane			+	+		+		+	+	+						+
Separate bike path	+	+	+	+	+	+		+	+	+	+					
Slope	−		−	−			−	−	−	−			−			−
Turns	−				−					−			−			−
Artificial lighting												+		+		
Pavement (asphalt)								+		+				+		
Pavement (paving stones)								−								
Motor vehicle intensities	−						−		−	−			−			−
Bicycle intensities							−									
Residential land use zone						−	−				·			−	+	
Commercial land use zone						+	−				·			−	+	
Greenery land use zone		+		−		+		+		+					+	
Industrial land use zone										−				·		
Land use mix											+		+		+	
Degree of urbanisation			+	−				−		+	−			−		

+ : positive effect, − : negative effect, · : not significant

2.2.1 Infrastructure-related factors

Using revealed preferences, many studies found that high-quality bicycle facilities improve the attractiveness of a certain route significantly, as they improve bicycle flow and safety. Especially when the bicycle paths are separated from motorized traffic (Broach et al., 2012; Campos-Sánchez et al., 2019; Chen, 2016; De Jong et al., 2023; Koch & Dugundji, 2021; Łukawska et al., 2023; Meister et al., 2023; Prato et al., 2018). However, also suggestive cycle lanes have a positive effect (De Jong et al., 2023; Koch & Dugundji, 2021; Meister et al., 2023; Prato et al., 2018; Zimmermann et al., 2017). Several studies generated a set of alternative routes through revealed route data and evaluated these routes using various types of logit models. For example, Broach et al. (2012) showed that cyclists are already willing to take a significant detour if the slope of a certain route is 2% or more.

Additionally, current literature is not conclusive on the effect of traffic control installations. On one hand Prato et al. (2018) and Koch and Dugundji (2021) concluded that cyclists are encouraged to take a detour to avoid traffic control installations based on their studies in Denmark and the Netherlands respectively. They attributed this to the disadvantage of stopping for cyclists. On the other hand, Khatri et al. (2016) and Broach et al. (2012) found that traffic control installations are deemed valuable in a lefthand turn with high traffic volumes, but these studies are focusing on two American cities which are designed to be more car centric.

Prato et al. (2018) also found the negative effects of unpaved infrastructure, motorized vehicle intensities and turns on the attractivity of routes. In general cyclists prefer a simple route including a low frequency of turns (Zimmermann et al., 2017). In this, left turns are especially penalized heavier than right turns as left turns are more associated with higher delays at both signalized and unsignalized intersections and they add a safety risk (Khatri et al., 2016; Prato et al., 2018).

Only stated preference studies consider artificial lighting as an influencing factor in route choice of cyclists and find that respondents are hesitant to cycling on routes which are not well lit after dark (Uttley et al., 2020; Winters et al., 2011). However, effect of artificial lighting on cyclists' route preferences is little studied.

2.2.2 Traffic-related factors

The effects of bicycle intensities on the route choice of cyclists are also little studied. Stated preference studies indicate that high bicycle volumes negatively affect the comfort of cyclists, as it is harder to overtake and keep the same speed (Li et al., 2012). Also, a negative effect of high bicycle volumes on the perceived safety of cyclists was identified in Uijtdewilligen et al. (2024). However, data limitations result in little understanding of this perception in revealed preference studies. For motorized vehicle volumes there exists evidence from both stated and revealed preference studies that high volumes are unwanted by cyclists (Broach et al., 2012; Li et al., 2012; Meister et al., 2023; Prato et al., 2018).

2.2.3 Land use-related factors

Literature shows that land use can influence the attractiveness of a bicycle routes. For most of the land use classes, no decisive conclusion is obtained by studies. For example, on the one hand revealed preference studies conclude that commercial areas are preferred by cyclists (Koch & Dugundji, 2021; Zhao et al., 2020). On the other hand, stated preference studies conclude the opposite (Li et al., 2012; Winters et al., 2011). Possibly commercial areas are not attractive to cycle through, but are the only viable option for cyclists, for example in city centers. Also, on the effect of greenery are studies not conclusive. Prato et al. (2018) concluded that forests and parks have a positive effect on the willingness to detour for cyclists. Nevertheless, Campos-Sánchez et al. (2019) found that green areas alone do not influence cyclists, but; rather, the proximity to separated cycle paths is necessary to be more attractive. When estimating relative route choice between observed routes and shortest paths in Oslo, Norway, De Jong et al. (2023) even found that greenery had a negative influence and contribute this to the unease in wayfinding through green spaces.

3 Methodology

The proposed methodology involves the estimation of the probability of choosing a certain segment based on its characteristics and the characteristics of neighboring segments in a route through a comparison of observed routes from Fietstelweek data with hypothetical shortest routes. The approach is summarized in Figure 1. For all origin-destination pairs in the observed routes from Fietstelweek, a corresponding shortest path is generated using Dijkstra's shortest path algorithm. For each of the routes, a separate contiguity matrix is created. This contiguity matrix involves the contiguity of all segments in a route.

The contiguity matrices are combined with characteristics such as traffic volumes and land use to produce the contiguity characteristics of all segments. These served as input for the varying-contiguity spatially lagged exogenous (VCSLX) model. The study focuses on the factors playing a role in convincing the cyclist to deviate from the shortest routes; thus, segments of the shortest routes that overlap with the observed routes are excluded from the analysis. The remainder of this section describes this VCSLX model in more detail and introduces the study area and corresponding data sources.

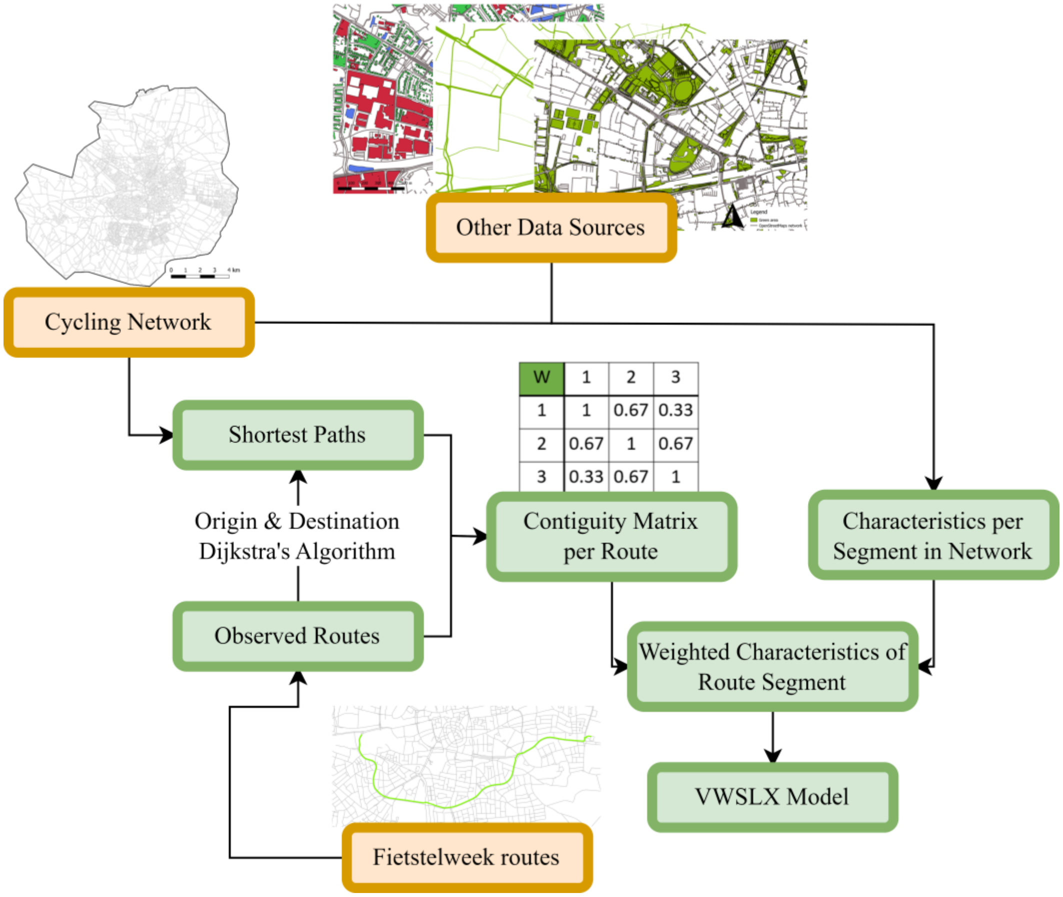


Figure 1. The approach of the study

3.1 Varying-contiguity spatially lagged exogenous model

The dependent variable in this study has a binary nature, as a segment belongs to either the observed route or the shortest path. Therefore, all observed segments are denoted as 1 and segments in the shortest routes are denoted as 0. In case a segment is in both observed and the shortest routes, it is added to the dataset twice, once with a 1 and once with a 0. The shortest route segments overlapping with segments from the observed route were used in developing the contiguity matrix but they are excluded in the model estimation. Binary dependent variables are often dealt with using logistic regression (Alattar et al., 2021). Therefore, the standard SLX model in a binary environment is estimated by the following equation:

$$Y_i = \text{logit}(P_i) = \alpha + \boldsymbol{X}\beta + \boldsymbol{WX}\gamma \quad (1)$$

Where α is the constant, β is the set of coefficients of the characteristics at a certain segment, γ is the set of coefficients of the spatial lag of the characteristics of neighboring segments, X is the characteristics of the segments and W is the contiguity matrix.

As mentioned, usually one contiguity matrix is created to represent the spatial dependencies. However, a segment is only affected by the characteristics of other segments within the same route, not in other routes. As a result, a contiguity matrix was created for every route individually, hence the varying-contiguity spatially lagged exogenous (VCSLX) model. The nature of the contiguity matrix has a substantial effect on the results of the regression analysis (Goetzke, 2008). Since it is expected that the characteristics of segments further away from a certain segment affect the attractivity less, compared to closer segments, a distance decay function based on the inverse distance is used. Although segments further away from a particular segment have less effect, still all the segments in a route are considered.

In the contiguity matrix, diagonal values are typically zero; however, it is possible to combine the effect of an independent variable with its spatial lag by adding an identity matrix to the contiguity matrix. This approach allows for a more robust estimation of variable effects and is conducted by the following equation:

$$Y_i = \text{logit}(P_i) = \alpha + (\boldsymbol{W_j} + \boldsymbol{I})\boldsymbol{X}\beta \tag{2}$$

This equation is similar to Equation 1; however, β coefficient vector represents both the characteristics of a certain segment and the characteristics of neighboring segments. I is the identity matrix with the same size as W_j, where W_j is the contiguity matrix of route j. A hypothetical contiguity matrix is schematically visualized in Figure 2 with the "greenery cover" of three segments in a route. The greenery cover of segments 1, 2 and 3 is 25%, 0% and 75% respectively. Looking from the perspective of Segment 1, Segment 3 is twice as far away as Segment 2, assuming that the lengths of the segments are the same. Segment 1 (highlighted in red) would have a spatially weighted greenery cover of 0.50 (1*0.25+0*0.67+0.33*0.75=0.50) if the data from the weighted matrix and the greenery cover were combined.

W	1	2	3
1	1	0.67	0.33
2	0.50	1	0.50
3	0.33	0.67	1

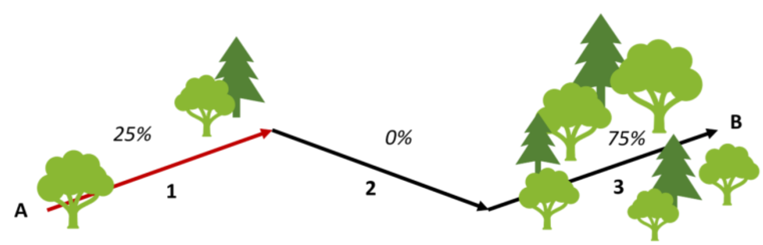
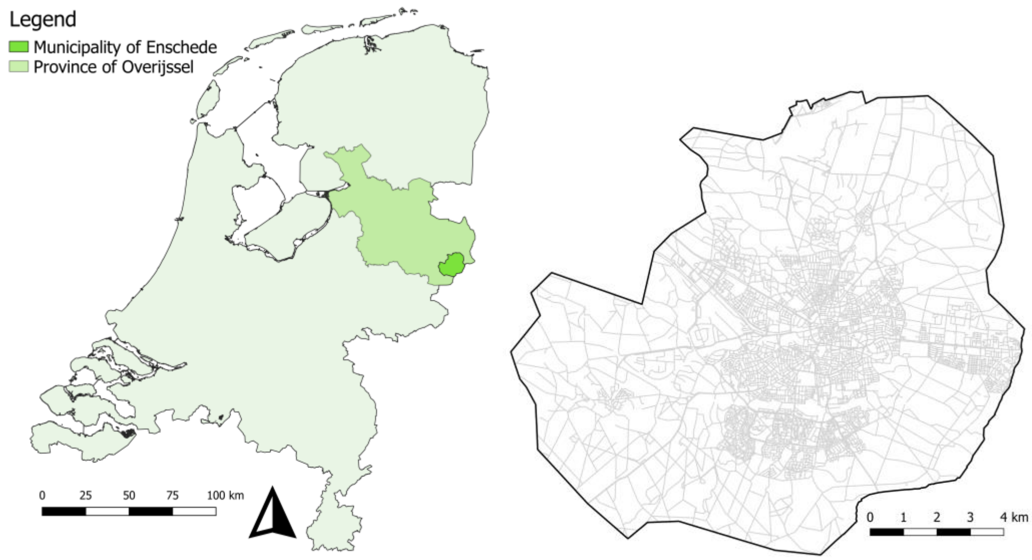


Figure 2. Schematical visualization of the weighted characteristics A) contiguity matrix, B) greenery cover per segment in a route [%] (red row shows weights for Segment 1)

3.2 Study area

The study area is the municipality of Enschede located in the province of Overijssel in the East of the Netherlands (Figure 3). The municipality of Enschede is a medium-sized city with a few neighboring villages with a total of 161,235 inhabitants in 2023 (CBS, 2023a). Enschede is strategically chosen as the Fietstelweek data initiative originates from Enschede (Fietstelweek, 2016). This results in a large availability of data for this particular municipality. In total, 8,606 observed routes are present in the Fietstelweek data for Enschede. These observed routes have at least either the origin or the destination in the municipality of Enschede. This means that cyclists can travel from other municipalities in the Twente region close to the municipality of Enschede such as Hengelo, Haaksbergen, Losser and Oldenzaal. However, only the part of routes within the municipality borders of Enschede is included in this study, because some of the data is less trustworthy or not available for this study outside the municipal boundaries. The municipality of Enschede has 1,159 km of infrastructure available for cyclists of which 23% consists of separated bicycle paths and 5% consists of cycle lanes.



A. **B.**

Figure 3. A) Municipality of Enschede in dark green in the province of Overijssel in lighter green, B) the cycling network of the municipality of Enschede

3.3 Data sources

Based on the conclusions drawn from the literature review, data was collected on the infrastructural and land use allocation factors influencing the route choice of cyclists. All the data sources are summarized in Table 2.

Table 2. Overview of the used data sources

Data	Variable	Description	Source
Route	Observed routes	-	(Fietstelweek, 2016)
Infrastructure	Traffic control installation	Binary variable whether or not present	(Fietsersbond, 2021)
	Cycle lane	Binary variable whether or not present	(Fietsersbond, 2021)
	Separate cycle path	Binary variable whether or not present	(Fietsersbond, 2021)
	Turns	Binary variable whether or not a turn angle of 60 degrees	(Fietsersbond, 2021)
	Artificial lighting	Binary variable whether or not present	(Fietsersbond, 2021)
	Pavement (asphalt)	Binary variable whether or not present	(Fietsersbond, 2021)
	Pavement (paving stones)	Binary variable whether or not present	(Fietsersbond, 2021)
Traffic	Bicycle intensities	Categorical variable: low (less than 250 bi/day), medium (250-500 bi/day) and high volumes (500+ bi/day)	(Veenstra, 2021)
	Motorized vehicle intensities	Categorical variable: low (less than 1,000 veh/day), medium (1,000-3,000 veh/day) and high volumes (3,000+ veh/day)	-
Land use	Residential land use	Ratio with buffer area; continuous variable between 0 and 1	(PBL, 2018)
	Commercial land use	Ratio with buffer area; continuous variable between 0 and 1	(PBL, 2018)
	Greenery land use	Ratio with buffer area; continuous variable between 0 and 1	(PDOK, 2021)
	Industrial land use	Ratio with buffer area; continuous variable between 0 and 1	(PBL, 2018)
	Land-use mix	Ratio with buffer area; continuous variable between 0 and 1	(PBL, 2018)
	Degree of urbanization	Categorical variable between 1 (densely urbanized (2,500+ addresses/km^2)) to 5 (not urbanized (less than 500 addresses/km^2)	(CBS, 2023b)

3.3.1 Cycling routes

The Fietstelweek dataset provides observed routes of cyclists. The most recent data is from 2016 in which bicycle users were asked to map their cycle behavior via a smartphone app using GPS during the week of the 19th of September (Koch & Dugundji, 2021). Nonetheless, this data source has its limitations. To anonymize the data, from all trips, a distance between zero and 400 meters was dropped at the beginning and end to mask the true origin and destination (Koch & Dugundji, 2021). Based on the origin and destination of the observed route, the shortest path was generated using Dijkstra's shortest path algorithm using the travel time of the segments as weights. For this average travelling speeds by Fietstelweek were used, while including average waiting times at signalized intersections as estimated by Velthuijsen (2020) on signalized intersections in Enschede.

3.3.2 Network and traffic

The Fietsersbond, a Dutch cyclist union, provided their network (Fietsersbond, 2021). The characteristics of the Dutch cycling network are updated annually by volunteers. The network is highly valued in Bernardi et al. (2018), due to the richness of its characteristics. The dataset for the municipality of Enschede includes 22,214 directional, cyclable links, which provide insights into most of the infrastructural characteristics considered in this study, such as the presence of traffic control installation, separate cycle path, cycle lane and artificial lighting and the type of surface material.

The bicycle intensities are provided via the FietsMonitor of Witteveen+Bos (Veenstra, 2021). Witteveen+Bos estimated a four-stage model which visually represents bicycle flows. This helps in identifying bottlenecks in current infrastructure and analyzing the effect of proposed measures. Next, motorized vehicle intensities are estimated using a Mobi Surround tool provided by the University of Twente. This tool is a four-stage model made as a QGIS plugin that is specifically designed for estimating the motorized vehicle volumes of the municipality of Enschede.

3.3.3 Land use

Basisregistratie Grootschalige Topografie data (PDOK, 2021) is a data source providing detailed information on all kinds of physical objects in the built environment with a resolution of 20 centimeters. For this study, this source is utilized to identify locations with various types of greenery, ranging from public parks, to pastures, sport fields and trees functioning as street furniture. Besides that, Wijken en Buurten data (CBS, 2023b) is a data source providing data on the degree of urbanization for all neighborhoods, districts, and municipalities in the Netherlands. Furthermore, RUDIFUN1 data (PBL, 2018) is a data source of Planbureau voor de leefomgeving, a Dutch governmental institution specialized in policy analysis on environment, nature and space. This data source provides information on the land use in the built environment on a building level.

3.3.4 Data processing

The Fietsersbond network already included most of the infrastructural characteristics that were included in this study and all of them were converted to a binary nature, indicating whether or not a certain characteristic is present or not. As suggested by Broach and Dill (2016) and Zimmermann et al. (2017) intensities were made categorical. Zimmermann et al. (2017) used between 10,000 and 20,000 vehicles/day as medium vehicle intensities and over 20,000 vehicles/day as heavy motorized vehicle intensities. Nonetheless, their analysis was executed in the Phoenix metropolitan area, one of the most car-centric areas in the United States with around 30 times as many inhabitants compared to Enschede. Therefore, based on the motorized vehicle intensities of Enschede, between 1,000 and 3,000 vehicles/day is chosen for medium vehicle intensities and over 3,000 vehicles/day for heavy vehicle intensities. The same categorical approach was done for bicycle intensities with 250-500 bicycles/day being the medium range and over 500 bicycles/day representing heavy intensities. The number of turns in each route is estimated using the Fietsersbond network and the observed routes or shortest paths. Turns are considered as subsequential links in an observed route or shortest path with more than a 60-degree angle between them.

To derive the type of building environment around the segments, a crow-fly buffer of 250 meters is employed. This buffer size is chosen as it has been used in previous studies and as a middle-size buffer, it is useful in both dense urban areas as well as more rural areas (Winters et al., 2010, 2011). Moreover, the buffer shape was chosen to be round, as recommended by Winters et al. (2010) and Van Nijen (2022). Then, the portion of a certain land use class within the whole buffer area is considered in the model. The land use mix is then estimated using the Shannon index (Strauss & Miranda-Moreno, 2013; Winters et al., 2010; Zhao et al., 2020). The same buffer area is used to estimate the degree of urbanization, as the average of the neighborhoods within the buffer area is considered to be the degree of urbanization. Note that a higher class number corresponds with a lower degree of urbanization.

4 Results

4.1 Varying-contiguity spatially lagged exogenous model

A total of 104,841 observed route segments and 55,358 shortest path segments are considered in the analysis. By considering these segments, the coefficients of the VCSLX model are estimated and described in Table 3. Additionally, the data was also used to fit a logistic regression model, to verify the effect of the spatial lag and identify whether spatial lags improve the model fit.

Table 3. Statistical summary of the VCSLX model and logistic regression model

	Factor	β	Std. Err.	p	β	Std. Err.	p
			VCSLX model			*Logistic regression model*	
	Constant	0.316	0.061	~0	0.064	0.047	0.169
Infrastructural	Traffic control installation	1.240	0.028	~0	0.772	0.028	~0
	Cycle lane	0.209	0.015	~0	0.308	0.020	~0
	Separate cycle path	0.076	0.012	~0	0.134	0.017	~0
	Turns	-0.171	0.013	~0	-0.102	0.014	~0
	Artificial lighting	-0.095	0.018	~0	-0.146	0.024	~0
	Pavement (asphalt)	0.092	0.019	~0	0.068	0.025	~0
	Pavement (paving stones)	-0.127	0.024	~0	-0.153	0.031	~0
Traffic	Heavy car intensities	0.116	0.015	~0	0.116	0.019	~0
	Medium car intensities	0.087	0.020	~0	0.084	0.023	~0
	Heavy bicycle intensities	-0.364	0.010	~0	-0.421	0.014	~0
	Medium bicycle intensities	-0.117	0.010	~0	-0.166	0.017	~0
Land use	Residential land use zone	-0.230	0.050	~0	-0.533	0.083	~0
	Commercial land use zone	0.599	0.060	~0	0.833	0.098	~0
	Greenery land use zone	-0.501	0.034	~0	-0.792	0.054	~0
	Industrial land use zone	-0.386	0.062	~0	-0.317	0.093	0.007
	Land use mix	-0.276	0.029	~0	-0.349	0.047	~0
	Degree of urbanisation	0.001	0.005	0.764	0.004	0.008	0.596
	MSE	0.214			0.219		
	AIC	179,702			182,824		

Table 3 shows that both models perform similarly. Nonetheless, the Mean Squared Error (MSE) and Akaike information criterion (AIC) are both lower for the VCSLX model favoring it over the model without spatial lags. To indicate the relative contribution of the factors (Table 4), standardized regression coefficients are used to eliminate the effects of different scales or measurement units. Standardized coefficients can give insights into the relative importance of the influential factors (Siegel & Wagner, 2022). Table 4 shows that the most substantial factors influencing the likelihood of choosing a segment are the presence of a traffic control installation, heavy bicycle intensities, green land use zone and land use mix.

Table 4. Standard regression coefficients

	Factor	Standardised β
	Constant	0
Infrastructural	Traffic control installation	0.593
	Cycle lane	0.215
	Separate cycle path	0.121
	Turns	-0.172
	Artificial lighting	-0.129
	Paved infrastructure (asphalt)	0.128
	Paved infrastructure (paving stones)	-0.124
Traffic	Heavy car intensities	0.125
	Medium car intensities	0.060
	Heavy bicycle intensities	-0.554
	Medium bicycle intensities	-0.112
Land use	Residential land use zone	-0.095
	Commercial land use zone	0.185
	Green land use zone	-0.403
	Industrial land use zone	-0.118
	Land use mix	-0.304
	Degree of Urbanisation	0.008

4.2 Likelihood of choosing a segment

The VCSLX model estimates the probability of a cyclist choosing a segment based on its characteristics and the characteristics of its neighbors in a route. The model can also be used to identify segments that are more likely to be chosen in the network. Such visualization can be a powerful tool for policymakers as it introduces deficiencies in the bicycle network. Especially, it can also provide information on why a certain segment is relatively less preferred when the map is combined with the underlying datasets, such as land use or traffic volumes.

This map with the likelihood of choosing a segment is created for the municipality of Enschede and a cut-out of the inner city and presented in Figure 4 and Figure 5, respectively. As the likelihood of choosing a segment is based on other segments in a route, only segments in the network that are also part of a route are considered. For segments, which occur in more than one route, the average likelihood is considered. Although only 55% of the network is covered by segments in either the observed routes, the shortest path, or both, the spatial distribution of included segments is sufficient to create this map, as it excludes only local roads.

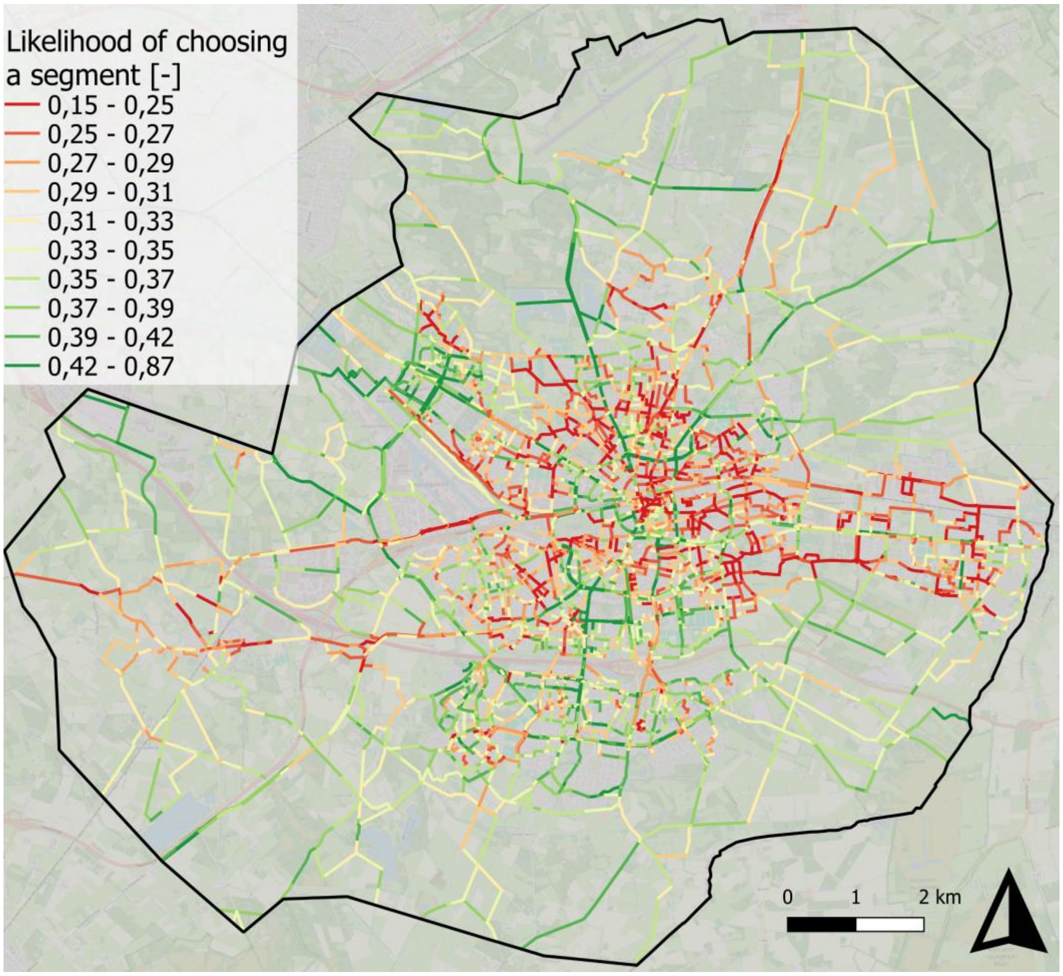


Figure 4. The likelihood of choosing a segment, classification by equal counts of the municipality of Enschede

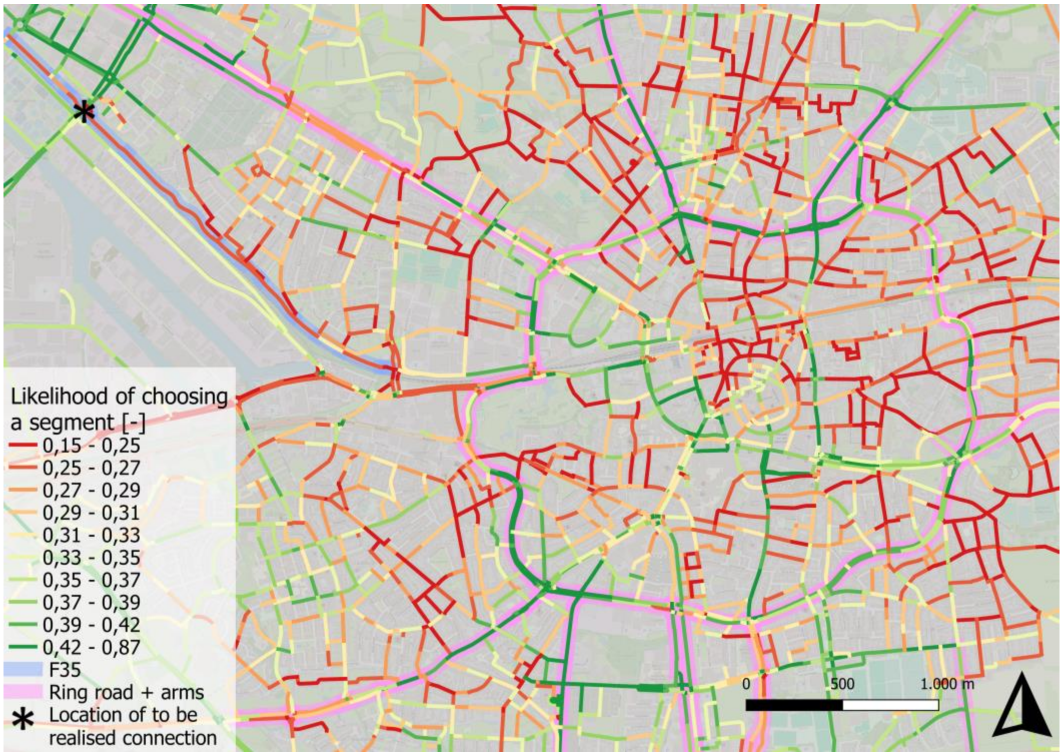


Figure 5. The likelihood of choosing a segment, classification by equal counts of the inner city of Enschede

5 Discussion

The VCSLX model revealed the effects of several traffic, infrastructure, and land use variables on the route choice of cyclists. The most substantial factors on the likelihood of choosing a segment are the presence of traffic control installation, heavy bicycle intensities, green land use zone and land use mix. These results are also illustrated in Figure 4 and Figure 5.

A notable observation from Figure 5 is the preference for the Enschede's ring road and its arms (highlighted in pink in Figure 5), although it is not known for its high-quality infrastructure and comfortable cycling. The ring road is associated with high motorized vehicle intensities, less bicycle intensities, less greenery, relatively many stops due to a high density of traffic control installations, but also few turns. Considering the latter, the ring road is often used as an identification mark when navigating through the city, and cyclists prefer simple routes without turns (Zimmermann et al., 2017). De Jong et al. (2023) indeed state that green spaces are associated with difficulties in wayfinding for cyclists. Thus, the cyclists of the Fietstelweek favor simple routes over routes with more greenery and less motorized vehicle intensities. Another interesting aspect of Figure 5 is the fact that the bicycle highway F35 (highlighted in blue in Figure 5) is less preferred, despite its high asphalt quality and greenery cover, contrary to the parallel route along the Hengelosestraat. Following the VCSLX model, this would be expected, but the result is counterintuitive. What could play a role in this is that when the Fietstelweek data was generated (2016) a major bridge connection (pointed out with an asterisk in Figure 5) connecting Hengelo with Enschede was not yet finished.

While the majority of the results of the VCSLX model are intuitive and in accordance with the literature (Table 5), some of the findings are contrary to expectations. For example, artificial lighting is found to negatively affect cyclist route choice. This finding seems counterintuitive, yet artificial lighting only influences the route choice of cyclists after nightfall and in this study, no time component is considered. Therefore, artificial lighting might be more accurately explained by the approaches adopted in previous studies. For instance, Uttley et al. (2020) looked at the reduction in bicycle flows in darkness and how artificial lighting can curb this reduction. Their conclusions show that the route choice of cyclists is only indirectly affected. Moreover, Winters et al. (2011) executed a stated preference study considering among other things artificial lighting. However, this approach only indicates the intentions of the cyclists rather than their actual choices.

The presence of land use mix, similar to artificial lighting, negatively affects the route choice of cyclists. Land use mix is used in a direct way in this study. Approaches adopted by studies considering land use mix identify a difference in cycling frequency (Saelens et al., 2003; Zhao et al., 2020) and compared this frequency with the frequency of using private motorized vehicles (Winters et al., 2010). Both approaches only indirectly include the route choice of cyclists and might explain why areas with a high land use mix are less preferred. Another explanation could be that areas with high land use mixture are associated with relatively high density of conflicts (Asadi et al., 2022). Also, three-quarters of the included types of land use in the land use mix have a negative influence, so it is reasonable to expect that a combination of land uses also has a negative influence.

Table 5. Literature consistency of the VCSLX model

Factor	Expected sign	Found sign	Consistency
Traffic control installation	+/-	+	✓
Cycle lane	+	+	✓
Separate cycle path	+	+	✓
Turns	-	-	✓
Artificial lighting	+	-	✗
Paved infrastructure (asphalt)	+	+	✓
Paved infrastructure (paving stones)	-	-	✓
Heavy car intensities	-	+	✗
Medium car intensities	-	+	✗
Heavy bicycle intensities	-	-	✓
Medium bicycle intensities	-	-	✓
Residential land use zone	+/±/-	-	✓
Commercial land use zone	+/±/-	+	✓
Greenery land use zone	+/±/-	-	✓
Industrial land use zone	±/-	-	✓
Land use mix	+	-	✗
Degree of urbanisation	+/-	±	✗

5.1 Limitations and future work

The observed routes of Fietstelweek provided great insight into the actual routes cyclists use; however, the data is relatively old (2016) compared to most other sources (2022/2023) used in this study. This means that renovations, extensions, or reductions in the current bicycle network included in more recent data sets were not present when the observed routes were generated. That is, a substantial portion of the bicycle infrastructure is renovated or newly created based on the vision of the municipality of Enschede. Therefore, the changes in the infrastructure may affect the outcome of the analysis.

The data of the Fietstelweek used in this study is also not fully complete, since the timing of a trip was not included. However, the timing of a trip is estimated to substantially influence the route choice. As mentioned, this could affect the estimated coefficient for artificial lighting, as it is expected that artificial lighting only affects the route choice after nightfall. Moreover, motorized vehicle intensities, as well as bicycle intensities, are highly affected by the timing of the day, as peak hours result in higher intensities. Nonetheless, this aspect is not included in this study and as it varies largely over the day, it is recommended to include it in a future study.

It is worth noting that the participants in the Fietstelweek were mostly experienced cyclists, resulting in a relatively large share of longer trips. According to the Fietstelweek data of 2016, the average distance per trip was 4.44 kilometers. However, the average distance per trip by bicycle according to (CBS, 2017) was 3.56 kilometers in 2016. Another downside is that no demographic data on, for example, gender and age, is included in the dataset and therefore also not included in this study, even though these aspects influence cycling behavior and route choice for cyclists (Broach et al., 2012; Broach & Dill, 2016; Segadilha & Sanches, 2014; Winters et al., 2010).

There is also a difference between the network of the observed routes (open street map) and the network of Fietsersbond used for shortest paths. When converting the observed routes to the Fietsersbond network, it became clear that not all observed routes were generated matching the network perfectly, since a substantial number was connected to perpendicular roads, often only open for other modes of transport, or other insufficiencies. This is most likely caused by the fact that GPS data is not precise enough for the level of detail of urban infrastructure networks (Khatri et al., 2016; Zimmermann et al., 2017). As a result, only 61% of the observed routes of Fietstelweek could be used in the analysis. Also, Prato et al. (2018) could use only 68% of their GPS-generated trips as a result of map-matching problems.

Future research might consider timing of travelling and demographic data on cyclists to create a more specific understanding of cyclists' the route choice. Also, the VCSLX model shows promising results, but this study was only executed in one study area, while the estimates could be fully different in another study area. Executing research with the VCSLX model in other study areas can increase the robustness of the method.

6 Conclusion

The route choice of cyclists has been often analyzed in the literature without considering the spatial correlations of segment features (i.e., traffic, infrastructure, etc.) within a route, although these features affect the preference of cyclists towards a certain segment. To overcome this limitation, we developed and employed a varying-contiguity spatially lagged exogenous model. This model allowed us to analyze the probability of choosing a segment based not only on its characteristics, but also the characteristics of neighboring segments within a route.

Literature shows that there are fifteen major factors affecting the route choice of cyclists. Fourteen of these factors were analyzed using the VCSLX model (only slope is excluded due to study area characteristics). Analysis results demonstrated that the VCSLX model outperforms the logistic regression model, thereby confirming the importance of accounting for spatial correlations. The findings indicate that cyclists prefer direct routes (e.g., Enschede Ring Road) despite adverse features such as high motorized vehicle intensities, low greenery, and a high density of traffic control installations. This is because direct routes can make navigation through the city relatively easy due to a low frequency of turns. Furthermore, we found that residential areas are less preferable verifying the findings of previous studies. However, a counterintuitive finding is that the F35 bicycle highway is not deemed favorable, despite its high-quality asphalt. This may be because the F35 was still partly under construction during the period when the Fietstelweek observed routes were generated.

The developed model was also used to estimate the likelihood of a cyclist choosing a segment. Results show that the presence of bicycle facilities like cycle lanes and separate cycle paths positively affects segment preference. In contrast, the likelihood of choosing a segment is negatively influenced by the frequency of turns, heavy bicycle intensities, green and residential land use as well as industrial land use and land use mix. These findings can guide policymakers when deciding on improvements in the cycling network and in designing new bicycle infrastructure.

Author contribution

Nick van Nijen: Conceptualization, methodology, validation, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, visualization. M. Baran Ulak: Conceptualization, methodology, resources, writing—review and editing, supervision, project administration. Sander Veenstra: conceptualization, methodology, resources. Karst Geurs: Writing—review and editing.

References

Alattar, M. A., Cottrill, C., & Beecroft, M. (2021). Modelling cyclists' route choice using Strava and OSMnx: A case study of the City of Glasgow. *Transportation Research Interdisciplinary Perspectives*, *9*, 100301. https://doi.org/10.1016/J.TRIP.2021.100301

Asadi, M., Ulak, M. B., Geurs, K. T., Weijermars, W., & Schepers, P. (2022). A comprehensive analysis of the relationships between the built environment and traffic safety in the Dutch urban areas. *Accident Analysis & Prevention*, *172*, 106683. https://doi.org/10.1016/J.AAP.2022.106683

Bernardi, S., Geurs, K., & Puello, L. L. P. (2018). Modelling route choice of Dutch cyclists using smartphone data. *Journal of Transport and Land Use*, *11*(1), 883–900. https://doi.org/10.5198/JTLU.2018.1143

Broach, J., & Dill, J. (2016). Using predicted bicyclist and pedestrian route choice to enhance mode choice models. *Transportation Research Record*, *2564*, 52–59. https://doi.org/10.3141/2564-06

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, *46*(10), 1730–1740. https://doi.org/10.1016/J.TRA.2012.07.005

Campos-Sánchez, F. S., Valenzuela-Montes, L. M., & Abarca-Álvarez, F. J. (2019). Evidence of green areas, cycle infrastructure and attractive destinations working together in development on urban cycling. *Sustainability*, *11*(17), 4730. https://doi.org/10.3390/SU11174730

CBS. (2017). *Onderzoek verplaatsingen in Nederland (OViN) 2016 plausibiliteitsrapportage*. Retrieved from https://www.cbs.nl/-/media/_pdf/2018/28/2017ep32-plausibiliteitsrapportage-ovin-2016.pdf

CBS. (2022). *Hoeveel fietsen inwoners van Nederland?* The Hague, Netherlands: Centraal Bureau Voor de Statistiek. https://www.cbs.nl/nl-nl/visualisaties/verkeer-en-vervoer/personen/fietsen

CBS. (2023a). *Inwoners per gemeente*. The Hague, Netherlands: Centraal Bureau Voor de Statistiek. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/regionaal/inwoners

CBS. (2023b, December 15). *Kerncijfers wijken en buurten - Dataset of 2022*. The Hague, Netherlands: Centraal Bureau Voor de Statistiek. https://www.cbs.nl/nl-nl/maatwerk/2023/50/kerncijfers-wijken-en-buurten-2022

Chen, P. (2016). Built environment effects on bicyclists' route preferences: A GPS data analysis. In *Bicycling and the built environment: Route choice and road safety* (pp. 19–54). Seattle, WA: University of Washington.

De Jong, T., Böcker, L., & Weber, C. (2023). Road infrastructures, spatial surroundings, and the demand and route choices for cycling: Evidence from a GPS-based mode detection study from Oslo, Norway. *Environment and Planning B: Urban Analytics and City Science*, *50*(8), 2133–2150. https://doi.org/10.1177/23998083221141431/FORMAT/EPUB

Elhorst, J. P., & Vega, S. H. (2017). The SLX model: Extensions and the sensitivity of spatial spillovers to W. *Papeles de Economía Española*, *152*, 34–50.

Ensing, N., & Janssen, K. (2020). *Actieplan Fiets 2020-2022*. Retrieved from https://gemeentebestuur.haarlem.nl/bestuurlijke-stukken/20200219065-2-Bijlage-1-Actieplan-Fiets-2020-2022-1.pdf

Faghih-Imani, A., & Eluru, N. (2016). Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. *Journal of Transport Geography*, *54*, 218–227. https://doi.org/10.1016/j.jtrangeo.2016.06.008

Fietsersbond. (2021). *Fietsersbond – Dataset of 2021*. Retrieved from https://www.fietsersbond.nl/

Fietstelweek. (2016). *Nationale fietstelweek 2015, 2016 en 2017 – Dataset of 2016*. Retrieved from http://opendata.cyclingintelligence.eu/

Goetzke, F. (2008). Network effects in public transit use: Evidence from a spatially autoregressive mode choice model for New York. *Urban Studies*, *45*(2), 407–417. https://doi.org/10.1177/0042098007085970

Khatri, R., Cherry, C. R., Nambisan, S. S., & Han, L. D. (2016). Modeling route choice of utilitarian bikeshare users with GPS data. *Transportation Research Record, 2587*(1), 141–149. https://doi.org/10.3141/2587-17

Koch, T., & Dugundji, E. R. (2021). Taste variation in environmental features of bicycle routes. *Proceedings of the 14th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, *IWCTS 2021*, article 2, (pp. 1–10). https://doi.org/10.1145/3486629.3490697

Lee, K., & Sener, I. N. (2021). Strava Metro data for bicycle monitoring: A literature review. *Transport Reviews*, *41*(1), 27–47. https://doi.org/10.1080/01441647.2020.1798558

Li, Z., Wang, W., & Ragland, D. R. (2012). Physical environments influencing bicyclists' perception of comfort on separated and on-street bicycle facilities. *Transportation Research Part D*, *17*, 256–261. https://doi.org/10.1016/j.trd.2011.12.001

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. *Transportation Research Part A: Policy and Practice*, *176*, 103834. https://doi.org/10.1016/J.TRA.2023.103834

Lv, H., Li, H., Sze, N. N., Zhang, Z., Ren, G., & Zhang, Y. (2023). Analysis of factors influencing cycling count: A micro-level study using road segment count data in London. *International Journal of Sustainable Transportation*, *17*(10), 1113–1128. https://doi.org/10.1080/15568318.2022.2152404

Meister, A., Felder, M., Schmid, B., & Axhausen, K. W. (2023). Route choice modeling for cyclists on urban networks. *Transportation Research Part A: Policy and Practice*, *173*, 103723. https://doi.org/10.1016/J.TRA.2023.103723

PBL. (2018). PBL Dataportaal, RUDIFUN1, Dataset of 2018. Retrieved from https://dataportaal.pbl.nl/RUDIFUN1

PDOK. (2021). Basisregistratie Grootschalige Topografie (BGT), Dataset of 2021. Retrieved from https://www.pdok.nl/introductie/-/article/basisregistratie-grootschalige-topografie-bgt

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. *International Journal of Sustainable Transportation*, *12*(10), 770–781. https://doi.org/10.1080/15568318.2018.1437236

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, *25*(2), 80–91. https://doi.org/10.1207/S15324796ABM2502_03

Segadilha, A. B. P., & Sanches, S. da P. (2014). Identification of factors that influence cyclistś route choice. *Procedia Social and Behavioral Sciences*, *160*, 372–380. https://doi.org/10.1016/J.SBSPRO.2014.12.149

Siegel, A. F., & Wagner, M. R. (2022). Multiple regression: Predicting one variable from several others. In *Practical business statistics* (pp. 371–431). Cambridge, MA: Academic Press. https://doi.org/10.1016/B978-0-12-820025-4.00012-9

Strauss, J., & Miranda-Moreno, L. F. (2013). Spatial modeling of bicycle activity at signalized intersections. *Journal of Transport and Land Use*, *6*(2), 47–58. https://doi.org/10.5198/JTLU.V6I2.296

Uijtdewilligen, T., Baran Ulak, M., Wijlhuizen, G. J., & Geurs, K. T. (2024). Effects of crowding on route preferences and perceived safety of urban cyclists in the Netherlands. *Transport Research Part A: Policy and Practice, 183,* 104030.

Uttley, J., Fotios, S., & Lovelace, R. (2020). Road lighting density and brightness linked with increased cycling rates after-dark. *PLoS ONE*, *15*(5). e0233105. https://doi.org/10.1371/JOURNAL.PONE.0233105

Van Nijen, N. (2022). *The influence of infrastructure and land use allocation on the route choice of cyclists*. Enschede, Netherlands: University of Twente. https://essay.utwente.nl/91571/1/Nijen_Nick_van.pdf

Veenstra, S. (2021). FietsMonitor/Witteveen+Bos, dataset of 2021. Retrieved from https://digitalsolutions.witteveenbos.com/ruimtelijke-ontwikkeling-wonen-en-mobiliteit/mobiliteit/fietsmonitor

Velthuijsen, T. (2020). *Calculating cycling delay at signalized intersections using smartphone data*. Enschede, Netherlands: University of Twente. https://essay.utwente.nl/85605/1/Velthuijsen-Tim.pdf

Winters, M., Brauer, M., Setton, E. M., & Teschke, K. (2010). Built environment influences on healthy transportation choices: Bicycling versus driving. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, *87*(6), 969–993. https://doi.org/10.1007/S11524-010-9509-6

Winters, M., Davidson, G., Kao, D., & Teschke, K. (2011). Motivators and deterrents of bicycling: Comparing influences on decisions to ride. *Transportation*, *38*(1), 153–168. https://doi.org/10.1007/S11116-010-9284-Y

Zhao, Y., Lin, Q., Ke, S., & Yu, Y. (2020). Impact of land use on bicycle usage: A big data-based spatial approach to inform transport planning. *Journal of Transport and Land Use*, *13*(1), 299–316. https://doi.org/10.5198/JTLU.2020.1499

Zimmermann, M., Mai, T., & Frejinger, E. (2017). Bike route choice modeling using GPS data without choice sets of paths. *Transportation Research Part C: Emerging Technologies*, *75*, 183–196. https://doi.org/10.1016/J.TRC.2016.12.009