UHASSELT
KNOWLEDGE IN ACTION

School voor Mobiliteitswetenschappen

master in de mobiliteitswetenschappen

Masterthesis

Bicycle counts and built environment characteristics

Tim Vervoort

Scriptie ingediend tot het behalen van de graad van master in de mobiliteitswetenschappen, afstudeerrichting mobiliteitsmanagement

PROMOTOR :
Prof. dr. Luc INT PANIS

COPROMOTOR :
dr. Evi DONS

2019
2020

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

School voor Mobiliteitswetenschappen

master in de mobiliteitswetenschappen

Masterthesis

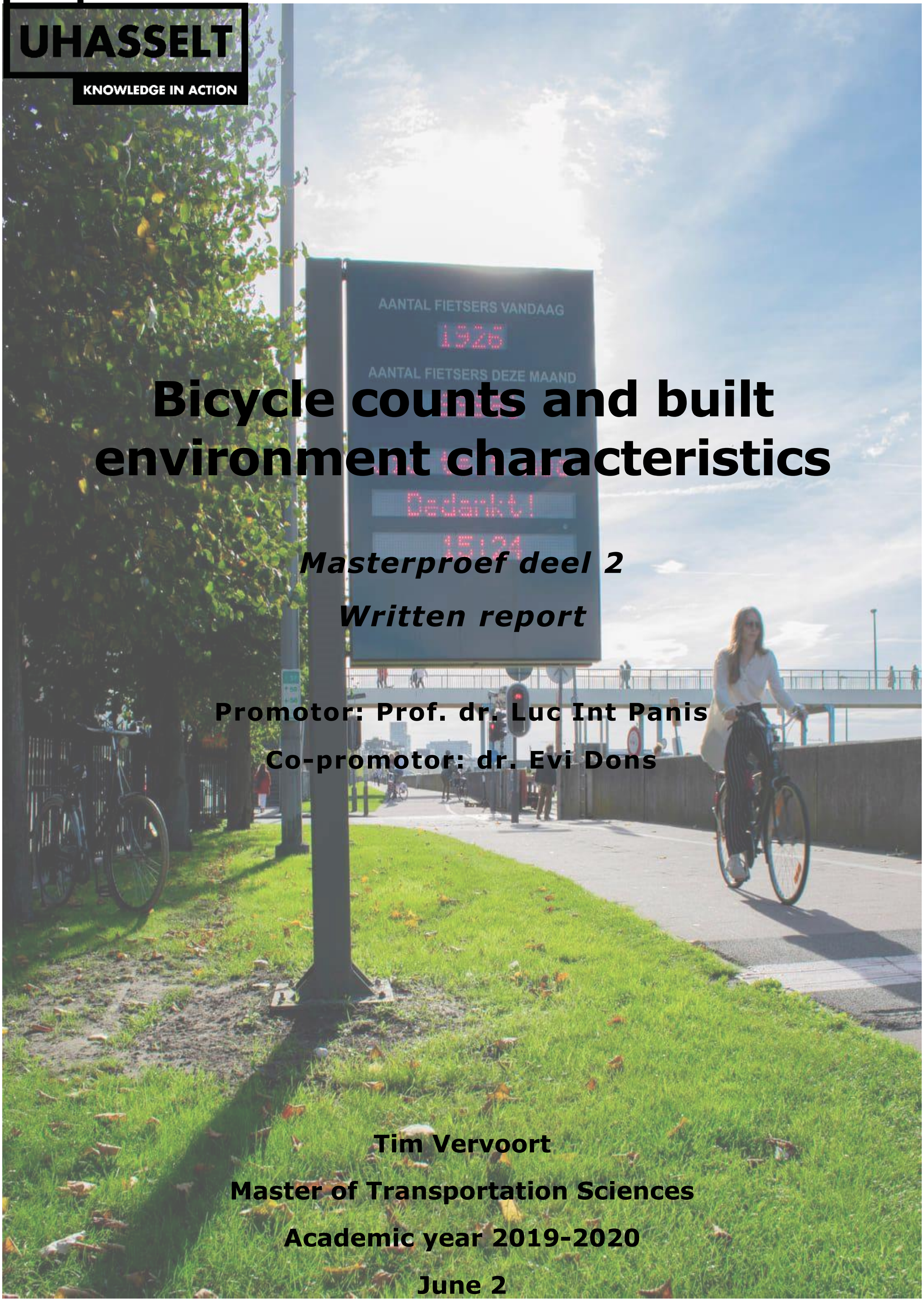
Bicycle counts and built environment characteristics

Tim Vervoort

Scriptie ingediend tot het behalen van de graad van master in de mobiliteitswetenschappen, afstudeerrichting mobiliteitsmanagement

PROMOTOR :
Prof. dr. Luc INT PANIS

COPROMOTOR :
dr. Evi DONS



Bicycle counts and built environment characteristics

Masterproef deel 2

Written report

Promotor: Prof. dr. Luc Int Panis

Co-promotor: dr. Evi Dons

Tim Vervoort

Master of Transportation Sciences

Academic year 2019-2020

June 2

Preface

Dear reader

Thank you for showing interest in this report. This Master’s Thesis is my final assignment in the Master of Transportation Sciences. And just like my Bachelor’s Thesis, it is about cyclists. As I try to do all my trips using sustainable modes, I care a lot about cycling infrastructure and, more generally, the attention cyclists get in transportation policies. What if we could make cycling policies more evidence-based? Or what if we could adjust the demand for cycling and the built environment to each other? Maybe my thesis can give a start to answer these questions. The model I developed aims to predict the number of cyclists on a location, based on the built environment characteristics of that location. As you will be able to read here, the model is quite promising!

Although I wrote this thesis on my own, it would not have been what it is now without the help of some people. I would especially like to thank my promotor, prof. dr. Luc Int Panis, and co-promotor, dr. Evi Dons. Their knowledge and their willingness to discuss my progress each week were truly valuable for this study. Their feedback was always constructive and their enthusiasm motivated me each time. Steven Soetens and Jelle Dekeyser also deserve my thanks, for providing me with their bicycle count data. In these strange times of COVID-19 lockdown and social isolation, I would also like to thank my friends and family for their support.

And thanks to the uncountable Skype-hours with my videocall-buddy, my stress level remained acceptable.

I hope you enjoy reading this report!

Tim Vervoort

This master’s thesis was written during the COVID-19 crisis in 2020. This global health crisis has had an impact on the proces of writing and analyzing, on performing the study and on the results, because:

- Follow-up with the team of promotors could only take place digitally;

- Discussing the results with stakeholders and practitioners through physical meetings was impossible. These discussions would have been useful for feedback or identification of use cases.

Summary

Introduction, background and aims

Bicycle traffic volume measurements in Flanders happen rather arbitrary and this leads to a knowledge gap regarding built environment (BE) characteristics that influence bicycle use. Studies show that there is an association, but studies focusing on Flanders have not been performed yet. Additionally, bicycle volume data that are available are mostly a posteriori collected data. Being able to predict bicycle traffic volumes would be useful because it helps supporting knowledge-based policy decisions. This master's thesis explored the possibilities and accuracy of applying a Land Use Regression (LUR) model to bicycle counts in Flanders. The objective was to develop a LUR model using existing bicycle count data and BE characteristics as predictor variables.

Methods

A literature review of 30 papers identified 107 BE characteristics that potentially influenced bicycle use. Sociodemographic variables were included in the literature review. Of these 107 characteristics, 51 were available as a GIS-layer in Flanders.

Bicycle counts were retrieved from 26 permanent and 11 temporary counting points in the province of Antwerp for the year 2019. Using buffers in GIS (radii 100, 300, 500, 1000, 2000, 4000, 6000 m) predictor variables were generated from these BE characteristics for each counting point. Initially, 294 variables were available for model development. These data were checked for correlation, with no variables being allowed to correlate more than 95% and no more than 60% if they were divided into the same category. Model development started with 112 remaining variables.

A linear regression model was developed using the *supervised forward linear regression* method in R. By leaving out one of the counting points that was close to the Dutch border and developing the model again, a sensitivity analysis was performed.

The performance of the model was checked through leave-one-out cross validation and external validation. Leave-one-out cross validation used the same data the model was developed with. The parameter coefficients were estimated again while leaving out one counting point at a time and repeating this 37 times. External validation consisted of estimating the number of cyclists at 83 new counting points in the provinces of Antwerp and Flemish Brabant, using the original model.

Results

All 112 variables were presented to the model and ten iterations led to an initial model with nine variables. After checking for significance, five variables remained. The final model consists of: the number of buildings within 1000 m of a counting point, the length of highways within 1000 m, the length of cycleways within 500 m, the number of different types of land use within 2000 m, and the weighted average percentage of registered partnerships within 500 m of a counting point. This model had an R^2 of 0.74.

The sensitivity analysis resulted in the same model parameters, with only one parameter being different from the original model. Leave-one-out cross validation resulted in an R^2 of 0.58. External validation yielded an R^2 of 0.41. Performance of the model (external validation in the province of Antwerp) resulted in an R^2 of 0.52, while its transferability (external validation in the province of Flemish Brabant) yielded an R^2 of 0.00008.

Conclusions and future research

Future research could focus on finetuning the model by using more accurate data and by dividing the bicycle counting points into different categories. Furthermore, this study is of societal relevance. A LUR model that predicts bicycle traffic volumes can be used as a tool to make policy decisions regarding bicycle infrastructure investments. The model could be used to identify areas that need attention. Due to the (very) poor transferability, the model in this study can only be used in the area it was developed in, being the province of Antwerp. Scientific relevance can be found in the fact that this study is the first one to apply the land use regression method to bicycle counts in, to the best of the author's knowledge, Europe.

Nederlandstalige samenvatting

Inleiding, achtergrond en doelen

In Vlaanderen gebeuren fietstellingen op een eerder arbitraire basis. Hierdoor is er een gebrek aan kennis betreffende bepaalde omgevingskarakteristieken die het fietsgebruik zouden kunnen beïnvloeden. Er zijn wel onderzoeken die dit verband aantonen, maar aan studies specifiek in Vlaanderen ontbreekt het nog. Daarbij komt nog dat de fietsteldata die beschikbaar zijn, meestal data zijn die achteraf verzameld werden. Het zou nochtans handig zijn om het aantal fietsers op een bepaalde plek op voorhand te kunnen voorspellen, omwille van de kansen die dit biedt om beleidsbeslissingen te ondersteunen. Deze masterthesis onderzocht de mogelijkheden en nauwkeurigheid van het gebruik van een *Land Use Regression* (LUR) model om fietsersaantallen te voorspellen in Vlaanderen. Het doel was daarbij om een LUR-model te ontwikkelen op basis van bestaande fietstellingen, met omgevingskarakteristieken als voorspellingsvariabelen.

Methodologie

Een literatuurstudie in 30 papers, die ook de socio-demografische variabelen betrok, identificeerde 107 variabelen als mogelijk van invloed op het fietsgebruik. Van deze 107 waren er in Vlaanderen 51 beschikbaar als GIS-datalaag.

De fietsteldata van 2019 van 26 permanente en 11 tijdelijke telpunten in de provincie Antwerpen werden voor het onderzoek gebruikt. De voorspellingsvariabelen werden voor elk telpunt berekend door middel van de bufferfunctie in GIS, waarbij er stralen van 100, 300, 500, 1000, 2000, 4000, en 6000 m gehanteerd werden. Dit leidde tot een aanvankelijk aantal van 294 variabelen die gebruikt konden worden voor de ontwikkeling van het model. De correlatie van deze variabelen werd gecontroleerd, waarbij variabelen niet meer dan 95% met elkaar mochten controleren, of zelfs niet meer dan 60% als ze tot dezelfde categorie behoorden. Uiteindelijk begon het ontwikkelen van het model met 112 variabelen.

Er werd een lineair regressiemodel opgesteld in R aan de hand van de *supervised forward linear regression*-methode. Door een telpunt, dat dichtbij de Nederlandse grens gelegen was, weg te laten en het model opnieuw te ontwikkelen, werd een sensitiviteitstoets uitgevoerd.

Door middel van een *leave-one-out cross* validatie en een externe validatie werd de prestatie van het model nagegaan. De *leave-one-out cross* validatie gebruikte dezelfde dataset als diegene die gebruikt werd voor de ontwikkeling van het model. De coëfficiënten van de verschillende parameters werden opnieuw geschat, waarbij er telkens een telpunt werd weggelaten. Dit werd 37 keer herhaald. De externe validatie had als doel het aantal fietsers op 83 nieuwe telpunten in de provincies Antwerpen en Vlaams Brabant te schatten aan de hand van het originele model.

Resultaten

Elk van de 112 variabelen werd aan het model aangeboden. Tien iteraties leidden tot een initieel model dat bestond uit negen variabelen, waarvan er na het controleren op significantie nog vijf overbleven. Het uiteindelijke model bestaat uit: het aantal gebouwen in een straal van 1000 m van het telpunt, de lengte aan snelwegen binnen 1000 m, de lengte aan fietspaden binnen 500 m, het aantal verschillende soorten landgebruik binnen 2000 m en het gewogen gemiddelde percentage aan wettelijk samenwonenden binnen 500 m van een telpunt. De R^2 van dit model bedraagt 0.74.

De sensitiviteitstoets leverde een model op met dezelfde parameters, waarvan er slechts één verschilt van het originele model. De leave-one-out cross validatie zorgde voor een R^2 van 0.58, terwijl de externe validatie een R^2 opleverde van 0.41. Het opsplitsen van de externe validatie in enerzijds de prestatie van het model (externe validatie in de provincie Antwerpen) en anderzijds de overdraagbaarheid van het model (externe validatie in Vlaams-Brabant), leverde een R^2 op van 0.52 en 0.00008 respectievelijk.

Conclusies en toekomstig onderzoek

Toekomstig onderzoek kan zich richten op het verder scherpstellen van het model aan de hand van meer accurate data enerzijds en het onderverdelen van de verschillende soorten telpunten anderzijds. Verder heeft dit onderzoek een maatschappelijke relevantie omdat het model gebruikt kan worden als tool om beleidsbeslissingen rond investeringen met betrekking tot fietsinfrastructuur te ondersteunen. Het model kan gebruikt worden om locaties die bijkomende aandacht nodig hebben te identificeren. Als gevolg van de slechte overdraagbaarheid komt uit dit onderzoek een model voort dat enkel gebruikt kan worden binnen het gebied waar het ontwikkeld is, namelijk de provincie Antwerpen. De wetenschappelijke relevantie van dit onderzoek kan gevonden worden in het feit dat het het eerste is dat de LUR-methode toepast op fietstellingen in, voor zover de auteur weet, Europa.

Content

List of figures and tables

1 Introduction

Cycling and overall attention for bicycles in Flanders has significantly increased in the past few years. In comparison to 2016, the number of cyclists in 2018 has clearly increased, not only during rush hours or weekends, but at any time of the day and week. Specifically in the Province of Antwerp, there were 14% more cyclists in 2018 compared to 2017. An increase of 21% has been measured compared to the beginning of the data collection in 2015. Cycling clearly has become more popular, both for functional and leisure trips (Fietsberaad Vlaanderen, 2019; Provincie Antwerpen, 2019b; VAB, 2019).

Together with the increasing volume of bicycle traffic, the means and attention for bicycle infrastructure increase as well. The Flemish Government's Coalition Agreement prioritizes investments in bicycle infrastructure, investing €300 million in the next five years. The aim is to overcome any possible resistance that could hold cyclists from using their bicycle (Vlaamse Regering, 2019).

It is not easy to find exact figures on bicycle traffic in the literature, and it is even harder to find figures in order to make a valid comparison with previous years. An important cause for this problem is that bicycle counts in Flanders happen inconsistently. The Province of Antwerp has 18 permanent counting points, divided over 70 municipalities. Separate projects exist as well, such as the *Telraam* project which uses cameras to count bicycle traffic volumes among other traffic modes (Provincie Antwerpen, 2019a; Telraam, 2019).

The above results in two issues. First, the rather arbitrary measurement of bicycle traffic volumes leads to a knowledge gap on Built Environment (BE) characteristics that influence the willingness to use the bicycle. Several studies all over the world have shown that there is an association between BE characteristics and bicycle use. Studies specifically focusing on Flanders have not been performed yet. As a consequence, there is no knowledge as to which characteristics influence bicycle use the most. This knowledge is useful as it can help motivate policy decisions and can be used to help allocate means and resources for bicycle investments. (Gao et al., 2018; Le et al., 2018; Mertens et al., 2017; Yang et al., 2019).

The second issue is that, when bicycle volume data are available, these are mostly data that were collected a posteriori. Being able to predict bicycle traffic volumes would be useful to support knowledge-based policy decisions. The need for a prediction tool exists, and a technique that is promising is the Land Use Regression (LUR) modeling method. LUR modeling is a commonly used technique to predict air pollution concentrations and their health effects. It is often used in epidemiological studies and is based on a number of predictor variables that are usually traffic-, population- and land use-related. A Geographic Information System (GIS) is used to obtain these variables. A LUR comprises of a number of components (Boniardi et al., 2019; Dons et al., 2013a; Hoek et al., 2008; Int Panis, 2018):

1. **Measuring/monitoring dependent variables**. Dependent variables are usually measured or monitored at 20 to 100 locations spread over the research area. In this study, the dependent variables were the bicycle counts.
2. **Collecting independent variables**. These variables have to be available for use in GIS. In this study, the independent variables were the BE characteristics.
3. **Estimating and validating the model**. Using linear regression, the model parameters are estimated. To test its accuracy, the model will then be validated.
4. **Applying the model**.

LUR models predict air pollution concentrations such as NO_2, NO_x, $PM_{2.5}$ and VOCs surprisingly good in a variety of situations (Hoek et al., 2008). It is expected that LUR performance in a bicycle traffic-related context will be good as well.

2 Problem statement

In Flanders, the number of permanent bicycle counting points is limited and unknown, as both the Flemish Region, the provinces and the individual municipalities can install such points. As a consequence, bicycle traffic data are, if available in the first place, hard to compare. It is not possible to find out the number of cyclists on randomly selected locations, since this is only possible for locations where bicycle counts took place.

BE characteristics influence bicycle use. It is not clear however, which factors do so in Flanders, which have the most significant correlation, and most importantly, which of these factors are available. Literature in Flanders that handles the application of the LUR method to predict bicycle volumes is non-existent.

The above forms the main problem this study aimed to resolve: it is not clear how many cyclists are to be expected at any given location in Flanders. This makes it difficult to make efficient policy decisions, to take precise measures, or to identify locations with a lot of cycling potential.

3 Objectives and research questions

3.1 Objectives

The main objective of this research was to identify which BE characteristics influence bicycle use, and if these characteristics can be used in a LUR model in Flanders to predict bicycle traffic volumes at any given location.

The following sub-objectives were defined as well:

- Identify characteristics with the greatest influence.
- Verify how a LUR model differs from a classic four step traffic model.
- Verify whether a LUR model can be used for other than air quality-related purposes.

3.2 Research questions

The main research question for this study was: Can a land use regression model be used to predict bicycle traffic volumes and potential in Flanders, using built environment characteristics as predictor variables?

Additional sub-questions were defined.

1. Which built environment characteristics influence bicycle use?

2. Which of the characteristics defined in sub-question 1 can be translated into data that are available in Flanders?

3. What are the similarities and differences between a classic four step traffic model and a land use regression model?

4. How is a land use regression model that aims to predict bicycle traffic volumes developed?

5. Which of the characteristics identified in sub-question 1 are most important regarding bicycle use?

6. To what extent is the model able to predict bicycle traffic volumes?

4 Methodology

4.1 Research design

Definitions and assumptions

Bicycles were defined as any vehicle that requires the user to pedal. This includes electric bicycles with electric support up to 25 km/h, and speed pedelecs, with support up to 45 km/h.

Bicycle sharing systems were not considered in the literature review. In the Province of Antwerp, only the city of Antwerp has a bicycle sharing scheme, together with a few municipalities that offer Mobit or Blue-bike (Blue-bike, 2019; Mobit, 2019). Since bicycle counts do not make a distinction between privately-owned bicycles and other types of bicycles, sharing stations were presented to the LUR model.

Focus area

This study focused on Flanders without the Brussels Capital Region. The LUR model was developed using Province of Antwerp's bicycle count data (Provincie Antwerpen, 2019a). After finishing the model, its performance was compared with bicycle count data from the provinces of Antwerp and Flemish Brabant, using data retrieved from *Dataplatform Fiets* (Dataplatform fiets, 2019).

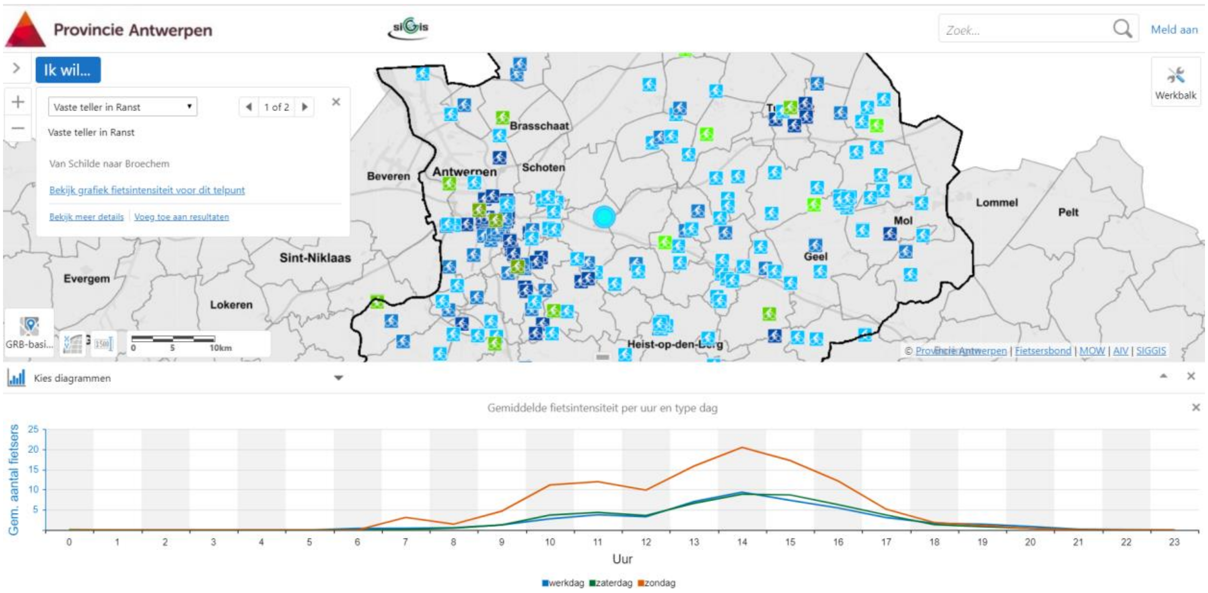


Figure 1. Province of Antwerp's Fietsbarometer open data platform (Provincie Antwerpen, 2019a).

4.2 Procedure

Literature review

Research questions 1 to 4 fully or partially relied on acquiring information from the literature. These questions were answered through a literature review, in which the exploratory literature review from the plan of approach was expanded. Literature was retrieved using mainly Hasselt University library's online platform.

The following search strings were used to find BE characteristics that influence bicycle use: *bicycle determinants, bicycle use determinants, bicycle built environment, land use bicycle, environmental characteristics bicycle use*, and *built environment characteristics bicycle use*. Results concerning bicycle sharing systems were filtered out. The interpretation of BE characteristics for this study was rather broad. Features such as educational status or age of the population were considered BE characteristics as well. For each sub variable, a hypothesis was mentioned whether the influence on bicycle use will be positive or negative. It was based on the papers the variable was mentioned in.

To find literature on LUR modeling and four step models, the following search strings were used: *land use regression model, air quality modeling*, and *four-step model*. Additionally, Google Scholar was used with the search string *land use regression.* These papers were supplemented with papers from the reference lists of the found papers, papers that were already known, and papers provided by the promotor and co-promotor of the master's thesis.

Data collection

Both dependent (bicycle count data) and independent (BE characteristics – land use variables) GIS-compatible data had to be collected. These data sets were searched for on websites of federal (Belgian), regional (Flemish) or other governmental institutions, as well as through search strings in Google.

Model development

Model development followed the approach outlined by Beelen et al. For the ESCAPE-project, Beelen et al. developed LUR models to explain differences in air pollution levels. Beelen et al. was cited 396 times and was therefore considered a well-suited approach for this study (Beelen et al., 2013).

To develop the model, the bicycle count data had to be standardized first. Land use variables had to be calculated, using buffers and intersect functions in GIS. To prevent overfitting, the land use data had to be cleaned and checked for correlations.

Next, a linear regression model was developed. Starting with zero variables and adding one variable with each iteration, a variable was added to the model if it yielded the highest increase in adjusted R^2. If, after *n* iterations, the model's adjusted R^2 did not increase with more than 1%, no more variables were added and the estimates (the coefficients and intercept) were calculated. The estimates had to be statistically significant at 95%. If not, these variables were removed from the model and the estimates were calculated again, until each variable in the model was significant at 95% (Beelen et al., 2013).

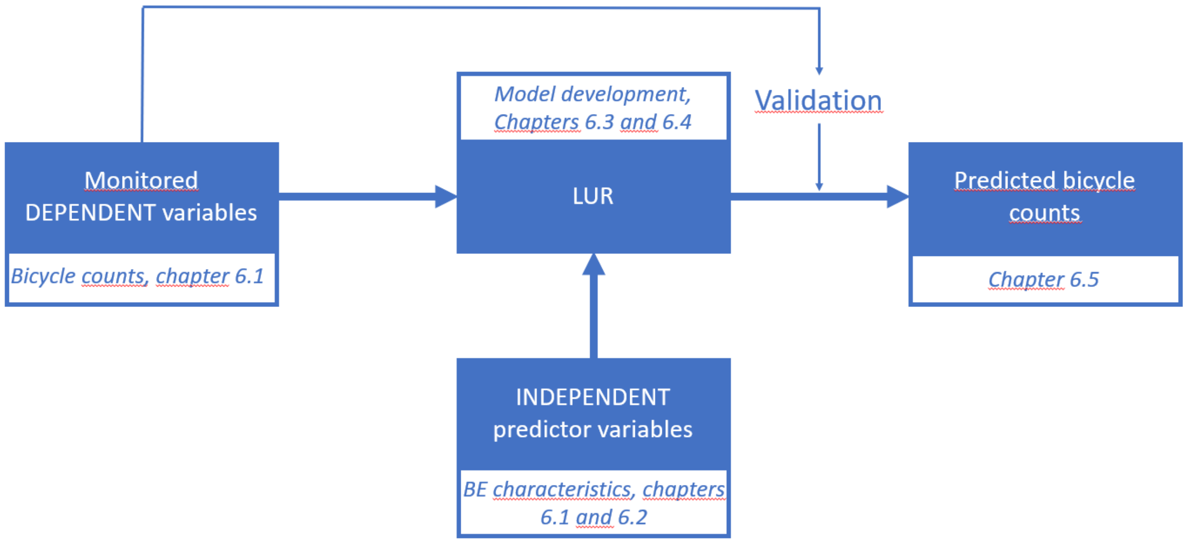


Figure 2. Schematic overview of the study.

4.3 Data

Dependent variables: bicycle count data

Because of time restrictions, the study relied on existing data. No data were measured specifically for the purpose of this study. Two main sets of data were collected: the dependent data, which are the bicycle counts, and the independent data, which are the BE characteristics or land use variables.

Two sets of bicycle counts were retrieved; one for developing the model and one for the external validation of the model. The development of the model relied on data from the Province of Antwerp. Since 2015, in collaboration with its cities and municipalities, the province collects bicycle count data through a number of permanent counters. For 2019, data from 20 permanent counters were available, as shown in Figure 3.

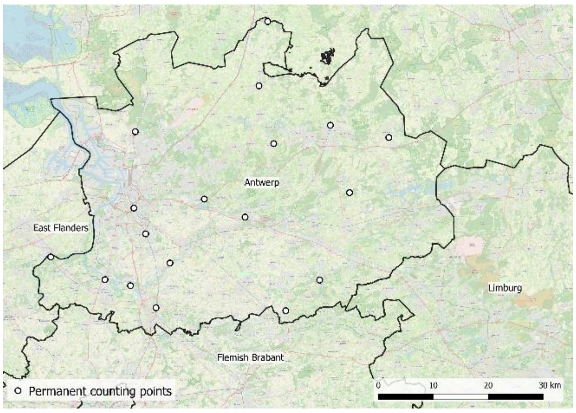


Figure 3. Permanent counting points available for this study.

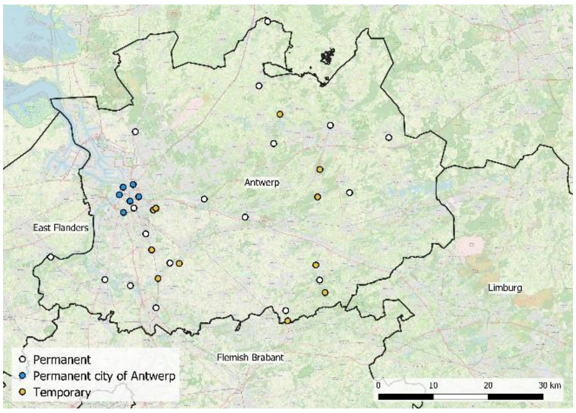


Figure 4. All counting points available for this study.

For each counting point, the average number of cyclists in both directions for each day of the week was available for each month of the year. As stated in the introduction on page 14, a minimum of 20 locations is required. Moreover, Figure 3 shows that, although they are equally spread throughout the province, the density of locations is rather low. Therefore, additional data were collected, from 11 temporary counting points. Also, bicycle counts from 6 permanent bicycle counters in the city of Antwerp were retrieved.

The bicycle count data collection scheme performed by the Province of Antwerp was a benchmark in Europe when it started in 2014. They use both temporal and permanent counting methods and they collect both functional and recreational bicycle data. The project is spread throughout the whole province (Provincie Antwerpen, 2014).

The permanent counters count bicycle traffic 24 hours per day and seven days per week. Optical fiber technology is used. The temporal counters use counting hoses, and in each municipality it is aimed to perform counts on three locations. These locations are strategically chosen and are situated on different routes of the Regional Functional Bike network, as well as on different types of infrastructure (Provincie Antwerpen, 2014).

For the external validation, another set of bicycle counts had to be obtained. *Dataplatform fiets* collects bicycle count data all over Flanders, and its latest data set dates from 2018. This data set contains 197 counting points, spread over Flanders (Dataplatform fiets, 2019).

Independent variables: built environment characteristics

Based on the literature review, data were collected regarding built environment in Flanders. Seven sources provided these data. Most of them were publicly available on websites of governmental institutions. Data that could not be found on governmental websites, were provided by *Open Street Map* (OSM). OSM is an open source platform, containing information that is fully provided by volunteers (OpenStreetMap, 2019).

4.4 Data analysis

Bicycle count data standardization

The basic dependent variable for the development of the model would be the number of cyclists in both directions, on an average day of the year, regardless of week or weekend days. Thus, for each day in 2019, the number of cyclists per counting point had to be calculated. A distinction was made between the permanent counting points and the temporary counting points. Permanent counting points had counted cyclists each day of the year, whereas temporary counting points only counted during a specific period.

The average number of cyclists for each day of the week, for each month was available for 26 permanent counting points in 2019. Therefore, there were bicycle counts for each day of 12 average weeks. Each day in every month of 2019 was assigned the corresponding number of cyclists. For example, January 2019 had four Mondays, so every Monday in January had the same number of cyclists. Then, the yearly average number of cyclists for each counting point was calculated. Also, the median for each day of the year was calculated over all counting points.

		Counting points				Median for each day (permanent counting pts)
		AREGV07	BERGV333	WILGV222	ANT02	PERM_MEDIAN
Number of cyclists per day	dinsdag 1 januari 2019	67	2.777	284	1.143	319
	woensdag 2 januari 2019	271	2.788	230	1.141	317
	donderdag 3 januari 2019	278	3.122	259	1.212	297
	zondag 29 december 2019	99	867	478	1.397	270
	maandag 30 december 2019	221	2.364	1.025	2.396	380
	dinsdag 31 december 2019	272	2.535	1.108	2.518	480
	YEARLY AVERAGE	374	2.805	1.042	1.739	677

Figure 5. Standardization for permanent counting points.

In addition to the counting points in the paragraph above, there were also 11 temporary counting points. The data provided were the number of cyclists for the specific days counting took place. For each counting point, the average daily number of cyclists for the counting period was calculated. Also, the average for the short temporary period of counting was calculated from the daily medians of the permanent counting points. This value was then divided by the yearly average median. This yielded a factor that indicated by how much the temporary counting period over- or underestimated the yearly number of cyclists. Finally, the average daily number of cyclists for each temporary counting point was divided by this factor. Now, each temporary counting point had a standardized yearly average daily number of cyclists.

		Median for each day (permanent counting pts)	Temporary counting points						
		PERM_MEDIAN	BRB02	RIJ02	KAS01	KAS02	HRS01	HRS02	HRS03
Number of cyclists per day	dinsdag 1 januari 2019	319							
	vrijdag 24 mei 2019	760			486		1184		
	zaterdag 25 mei 2019	561			229		798		
	zondag 26 mei 2019	772			236		889		
	maandag 27 mei 2019	786			357		1073		
	dinsdag 28 mei 2019	846			288		714		
	woensdag 29 mei 2019	1.047			434				
	donderdag 30 mei 2019	869			239				1213
	vrijdag 31 mei 2019	760			338				586
	zaterdag 1 juni 2019	743			467				1240
	zondag 2 juni 2019	1.156			549				1502
	maandag 3 juni 2019	918			321				427
	dinsdag 4 juni 2019	1.075							611
	dinsdag 31 december 2019	480							
	YEARLY AVERAGE	677	778	295	359	155	1.061	996	545
	FACTOR: average median period/yearly average median		0,789222	1,371678	1,238554	1,373842	1,315498	1,217221	1,26425
	STANDARDIZED YEARLY AVERAGE: Yearly average * factor		985,5696	215,108	289,4871	113,0303	806,8203	818,289	431,303

Figure 6. Standardization for temporary counting points.

Calculating land use variables

The land use variables were calculated using QGIS. Around each counting point, 7 buffers were created with sizes (radii) of 100, 300, 500, 1000, 2000, 4000, and 6000 meters. Through the intersect function, the value of each variable within a certain buffer distance of the counting point was determined.

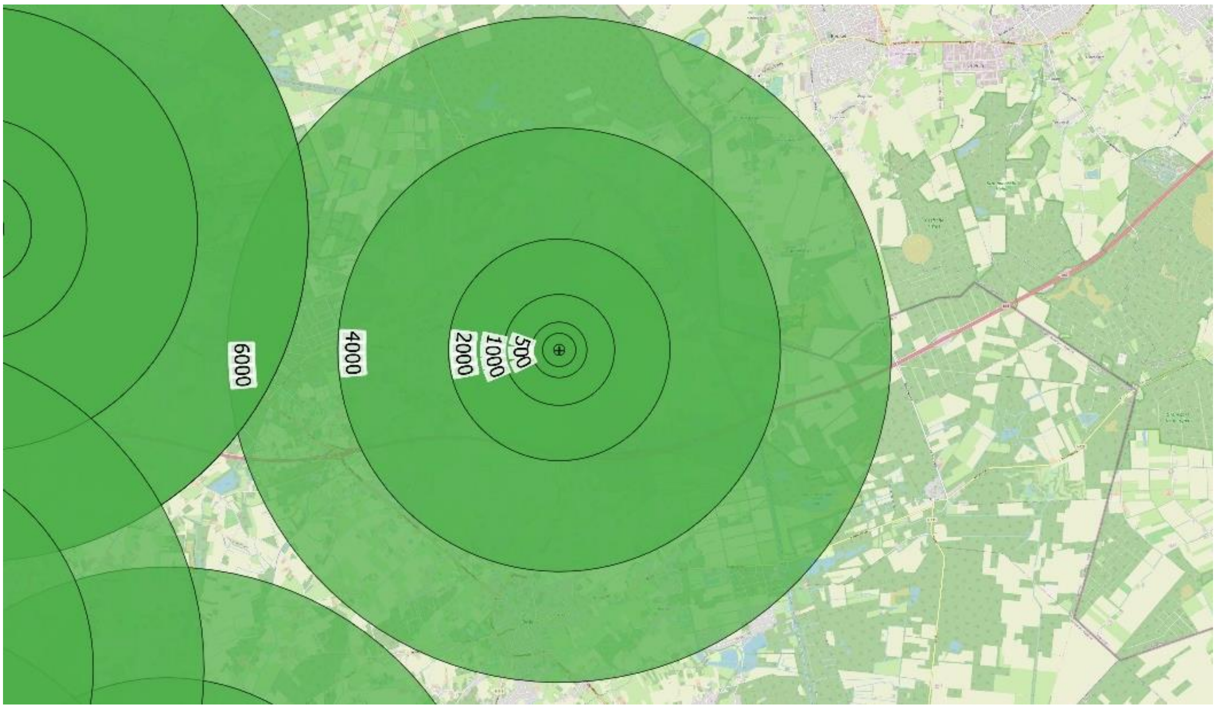


Figure 7. Seven buffers were created around each counting point.

Data cleaning

To prevent overfitting, some variables were omitted. If a variable had '0' as value for more than 32 of the 37 counting points, the value was omitted. Also, incomplete data sets were removed from the variables list.

Although first included in the variables list, variables regarding ethnicity were removed. The literature identified ethnicity as possibly of influence on bicycle rates. However, this literature is mainly based on large American cities, where ethnic distinctions between neighborhoods are larger than in Flanders. This study did not have enough counting points to clarify this distinction. Moreover, the percentage of inhabitants could be a proxy for urbanization.

Correlation between independent variables

A correlation matrix of all variables, including the number of cyclists, was generated in R. Variables that correlated more than 95% with each other, were considered as measuring the same thing. If two variables correlated more than 95% with each other, they were checked for their correlation with the number of cyclists. The one that correlated least with the number of cyclists, was removed.

After this first check for correlation, within-category correlation was checked for. The remaining variables were divided into a number of categories. These categories were determined subjectively, and were rather broad. For each category, the variable that correlated most with the number of cyclists was selected. This variable was the significant variable for its category. Then, each variable in each category that correlated for more than 60% with the significant variable, was removed. Henderson et al. (2007) used a similar technique (Henderson et al., 2007).

The remaining variables were all presented to the model.

Model development

The model was developed in R, following the *supervised linear forward regression method* as applied by Beelen et al. This method develops the model bottom-up, starting without any variables and adding a new variable with each iteration. After each iteration, a variable was added to the model if 1) its slope corresponded to the slope mentioned in the hypothesis; and 2) it had the highest adjusted R^2. This was repeated as long as the increase in adjusted R^2 was higher than 1%. Variables that entered the model, were still presented to the model in different buffer sizes in the following iterations. At the point where no additional variables were added, the significance of each variable was checked. Variables that were not statistically significant at 95% were omitted. The coefficients of each variable were then calculated again. This was repeated until only the significant variables remained (Beelen et al., 2013).

Two regression analysis checks were performed. Cook's D indicates the counting points that strongly influence the model. Counting points should not have a Cook's

D higher than 1. The Variance Inflation Factors (VIF) were calculated as well, and these should not exceed 3 (Beelen et al., 2013).

One counting point was situated close to the Belgian-Dutch border. Since the variables that were used to develop the model were based on Belgian or Flemish data sets, no Dutch data were included. As only a small part of each of the seven buffers was situated on the Belgian side of the border, the model was developed again without this counting point.

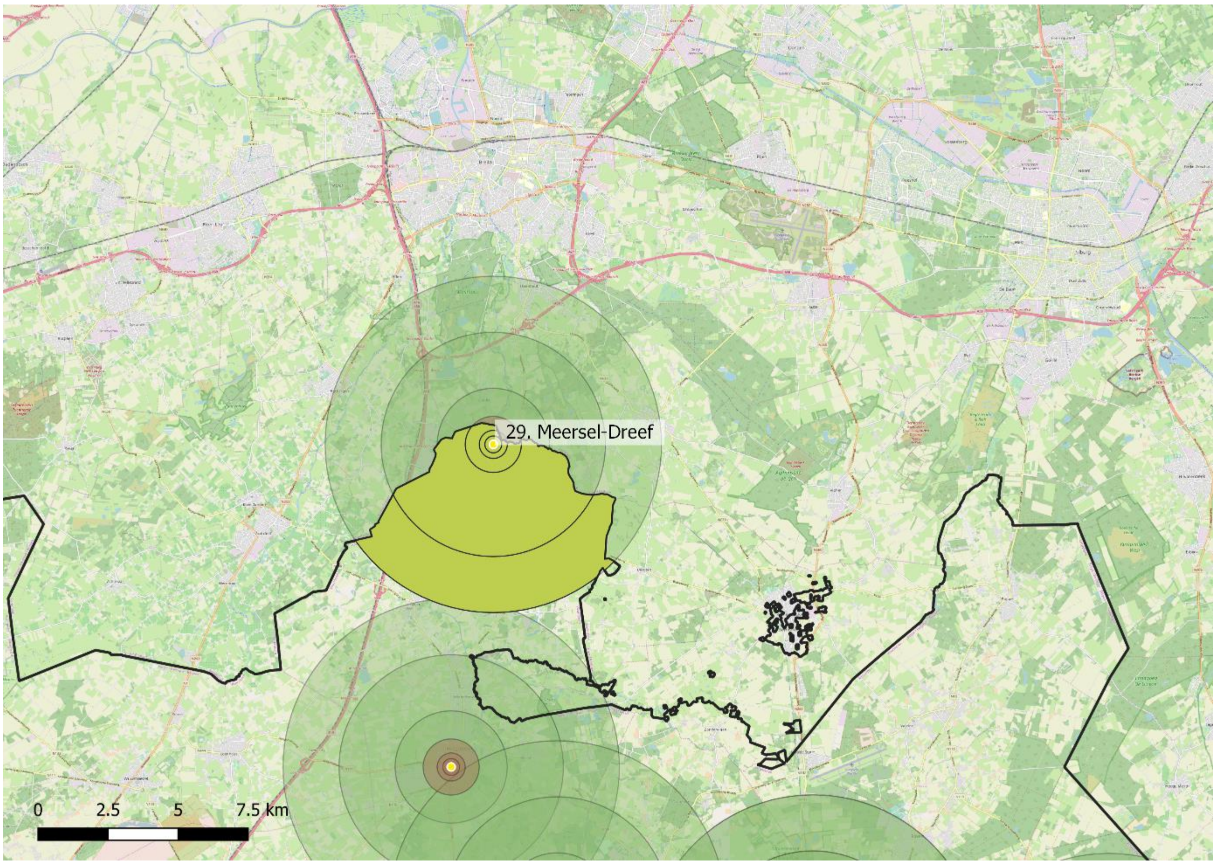


Figure 8. Approximately 2/3 of the buffers are situated on the Dutch side of the border.

This redevelopment of the model can be considered a sensitivity test, which would check whether a model developed without this counting point would differ from the initial model. There were no new checks for correlation, the 36 remaining counting points and their variables were immediately offered to the model. The same procedure of model development was followed.

Leave-one-out cross validation

Evaluation of the model happened through leave-one-out cross validation and external validation. Leave-one-out cross validation happens with the data set the model has been developed with. Leaving out one counting point at a time, the parameter estimates (coefficients and intercept) were estimated again. With these new estimates, the number of cyclists was calculated for the point that was left out. This was repeated 37 times, leaving another counting point out each time.

External validation

External validation is the method where the model's performance is tested on an external data set, in this case the 197 counting points from *Dataplatfom Fiets*. To avoid coincidental over- or undercounts, due to meteorological circumstances for example, only those points that had counted for 14 days or more were used for the external validation. This resulted in 83 counting points, all located in the provinces of Antwerp and Flemish Brabant, as shown in Figure 9.

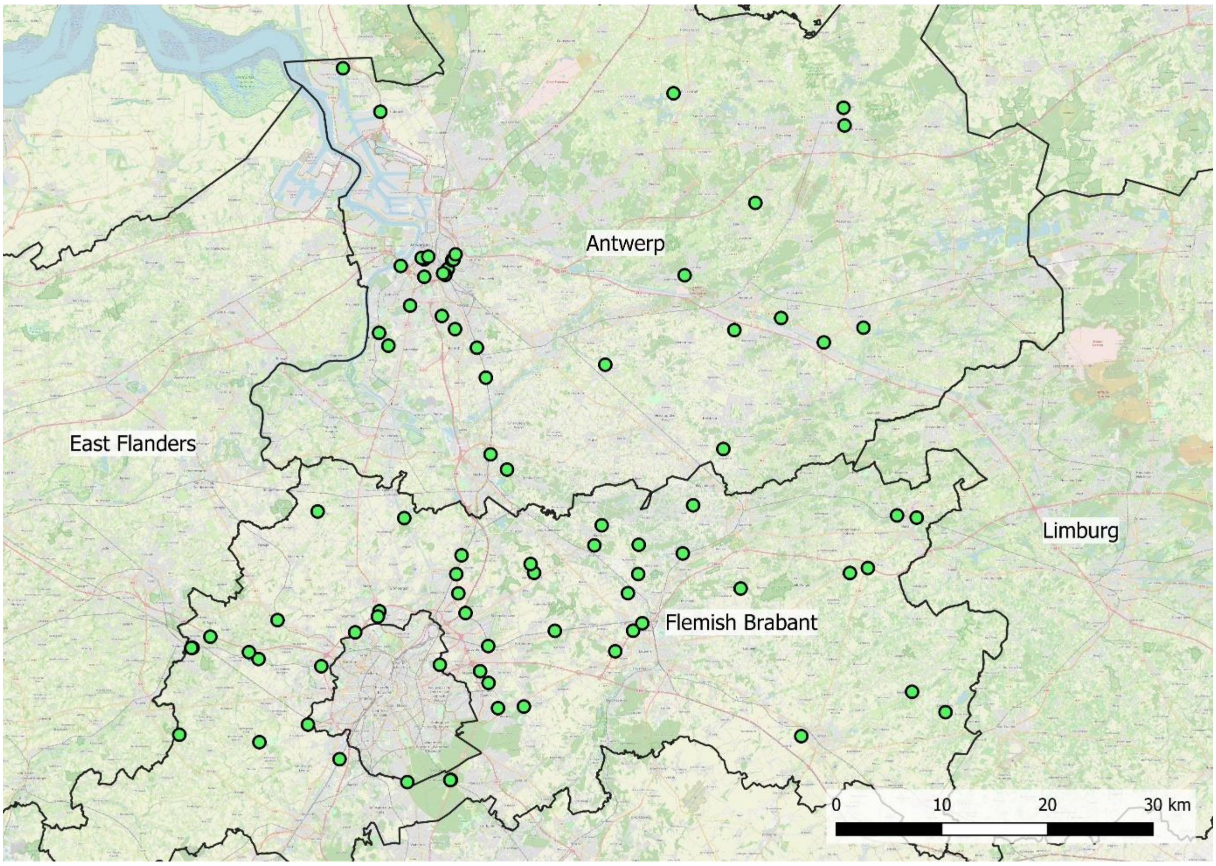


Figure 9. 83 Counting points were used for external validation.

The bicycle count data were standardized following the same method described on pages 24 and 25, and with the same permanent counting points as described there. Because these counting points are only situated in the province of Antwerp and date from 2019, external validation happened with both raw and standardized data. By doing so, R^2 from both data sets could be checked. Since external validation consists of an estimation of bicycle numbers, it could impossibly result in negative values. Therefore, the minimum number of cyclists was fixed on the lowest measure minus 50%. The values for each counting point were estimated by filling in the model's equation with the values of the counting point's variables.

External validation tests both the performance of the model and its transferability. To make a distinction between these two, the counting points from the province of Antwerp and those from the province of Flemish Brabant were considered two distinctive data sets after the general external validation. The estimated values remained equal to those in the general external validation. R^2 was calculated for both data sets, with the one from the province of Antwerp describing the performance of the model and the one from the province of Flemish Brabant describing the model's transferability. This was done with the standardized data set.

Software

Analyses of the data, which is mainly data cleaning and preparation and the development of the model, took place mainly using three programs.

To generate the values for each BE characteristic variable, QGIS was used. QGIS is geographic information system software and allows to perform spatial analyses of data.

The development of the model was done through R. R is programming software, and a script was used to develop the model using the supervised forward regression method.

All results and data sets were saved and cleaned in Excel spreadsheets. Simple calculations were done in Excel as well.

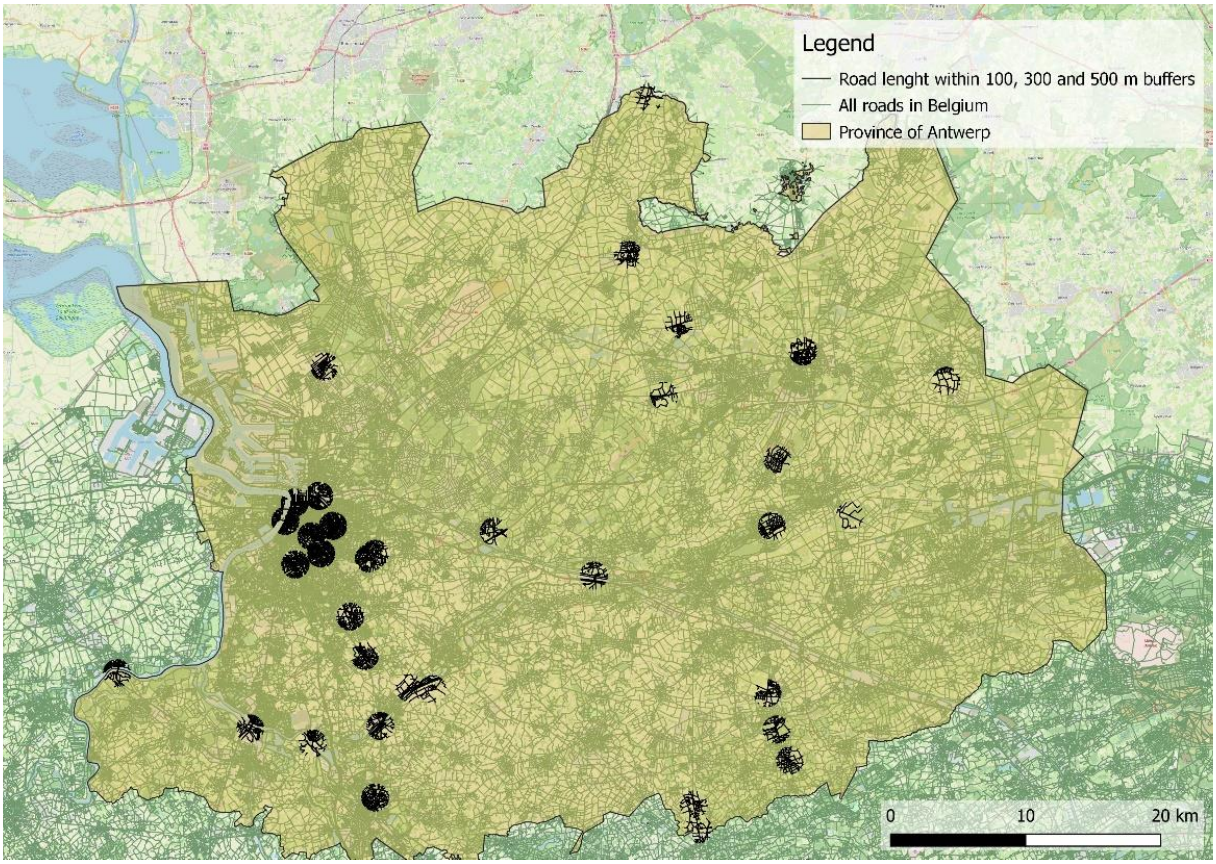


Figure 10. Example of analysis in GIS.

5 Literature review

5.1 Context

A great share of the literature on BE characteristics influencing bicycle traffic was published in the US. Together with Australia, the cycling culture in this country strongly differs from the one in (Western) Europe. In Europe, cycling as a means of commuting is very common, whereas in the US and Australia cycling is more popular as a means of recreation. This means that some findings from the literature have to be interpreted with caution (Heinen et al., 2010; Moudon et al., 2005; Rietveld & Daniel, 2004).

5.2 Built environment characteristics

In total, 30 papers were reviewed concerning BE characteristics. Most of them identify factors that significantly influence bicycle use through statistical analyses or models. Three review papers were consulted as well. North America was the study area in the majority of the consulted works, followed by Europe. Three papers had a study area in Australia, one in South America and one in China. The study area is important because, as stated above, cycling cultures strongly differ between countries and continents.

Table 1 gives an overview of the results from the literature. Table 2 summarizes all significant variables mentioned in Table 1. They are categorized in variables and sub variables. The significant variables are related to *land use and urban spatial structure, bicycle network, relative position with respect to cars, personal safety, traffic safety, motorized vehicle's network, parking, socio-demographic characteristics, the presence of public transport, public space, cycling culture*, and *cycling policies.*

Most variables are mentioned by more than one paper. *Destination accessibility and trip length* is the sub variable that was mentioned by the highest number of authors. It has a significant influence on bicycle use in 11 papers. This sub variable is categorized under land use and urban spatial structure, since destination accessibility is a direct derivative of land use planning.

For some sub variables, the hypothesis is inconclusive (+/-), meaning that the variable had different effects in different papers. For example, Rietveld and Daniel (2004) state that *City size* can have both positive and negative effects on bicycle use in The Netherlands. Vandenbulcke et al. (2011) however, finds that the number of cyclists is highest in the largest towns in Belgium.

Table 1. Overview of the consulted literature in chronological order.

Paper	Type	Context/Setting	Data sample	Significant factors	Notes
(Rietveld & Daniel, 2004)	Paper based on bivariate and multivariate analyses	103 Dutch municipalities	Bicycle use	Number of stops cyclists have to make, road use hindrances, relative position with respect to cars, travel time (spatial structure, adequacy cycling infrastructure, detours, waiting time at crossings), physical needs and comfort (quality of infrastructure, weather, flatness of surface), traffic safety, risk of theft or vandalism, parking cost, personal security.	Most of the inner-municipality variation in bicycle use is related to physical aspects such as altitude differences and city size. Cycling is mainly considered as a recreational tool in the USA. Cycling is determined by culture – immigrants do not use the bicycle as often as native people from the Netherlands.
(Moudon et al., 2005)	Paper based on a model	Urbanized Kings County, Washington	Cycling behavior data	Presence of agglomerations of offices, clinics/hospitals, and fast food restaurants.	Cycling happens more often **for recreational purposes** than for transportation. Decision to cycle rests largely on personal and not on environmental factors. Bike lanes, traffic conditions and street connectivity remain insignificant.
(Hunt & Abraham, 2007)	Paper based on a model	Edmonton, Canada	Stated preference questionnaire with current cyclists	Trip length, cycling facilities (mixed traffic, lanes, paths), secure parking.	
(Parkin et al., 2008)	Paper based on a logistic regression model	8.800 Welsh electoral wards	Whole population	Physical condition of the highway, rainfall, temperature, hilliness, proportion of off-road bicycle routes, car ownership.	Cultural norms with respect to cycling may be different across ethnic groups within society.

Table 1 (continued).

(Heinen et al., 2010)	Review			Distance, function mixture, storage facilities, block size and density, presence of bicycle infrastructure and its continuity, traffic lights and stop signs, land use, parking facilities, showers at work, weather, car ownership, bicycle ownership, adjacent car parking, hilliness.	Cycling for commuting is likely to be influenced by other determinants than cycling for recreational purposes
(Owen et al., 2010)	Paper based on logistic regression models	Adelaide, Australia and Ghent, Belgium	Objectively measured data from Adelaide, perceived data from Ghent	Gender, neighborhood walkability, employment.	Living in a high-walkable neighborhood was associated with significantly higher odds of bicycle use for transport in both cities
(Pucher et al., 2010)	Review			Overall bikeability measures, on-road bicycle lanes, two-way travel on one-way streets, shared bus/bike lanes, off-street paths, signed bicycle routes, bicycle boulevards, cycle tracks, colored lanes, shared lane markings, bike boxes, bicycle phases – traffic signals, maintenance of facilities, wayfinding signage, techniques to shorten cyclists' routes, traffic calming, home zones, car-free zones, complete streets, bike parking, showers at workplaces, bicycle stations.	
(Winters et al., 2010)	Paper based on a model	Metro Vancouver, Canada	Bike and car trips by adults	Topography, intersection density, highway, arterial road, road markings or signage for cyclists, traffic calming, cyclist-activated lights, density, land use mix.	

Table 1 (continued).

(Kirner Providelo & da Penha Sanches, 2011)	Paper based on focus group	Rio Claro, Sao Paulo, Brazil	Focus group	Lane width, motor vehicle speed, visibility at intersections, presence of intersections, street trees.	
(Vandenbulcke et al., 2011)	Paper based on a model	The 589 municipalities in Belgium	Bicycle use	Relief, traffic volumes, bike accidents, town size, distance travelled, demographic and socio-economic determinants (age, income, gender, education, family commitments), cultural, societal and environmental determinants (weather, urban spatial structure, infrastructure, hilliness, population and job densities, connectivity), policy-related determinants (land use and transport planning, pro-cycling policies).	More cycling in one municipality can stimulate cycling in neighboring municipalities. Increased safety is often valued higher by cyclists than other factors. Heavy traffic **does not** have an influence on cycling in Flanders, because many cyclists create high visibility and feeling of security
(Buehler & Pucher, 2012)	Paper based on statistical analyses	90 of the 100 largest US cities	Dataset on length of bike lanes and paths	Bike lane supply, bike path supply, cycling safety, college students, car access, sprawl index, gasoline price.	Percentage of college students in the city is significant predictor of bike commuting. There is no statistically significant difference between paths and lanes
(Sallis et al., 2013)	Paper based on multivariable models	Seattle, Washington and Baltimore, Maryland, USA	Survey conducted on 1780 adults (aged 20-65)	Region, age, vehicle ownership, BMI, ethnicity, education, marital status, land use mix, pedestrian/traffic safety.	

Table 1 (continued).

(Zhao, 2013)	Paper based on statistical analyses	Beijing, China	Household interview survey	Destination accessibility, urban design at the community level, diversity of land use, number of exclusive bicycle lanes, number of street crossings, number of mainroad or expressway crossings.	Number of street crossings has more significant effects on cycling than street density. Residential density, one of the most important factors affecting cycling in Europe and North America, has no significant effects in Beijing.
(Cui et al., 2014)	Paper based on models	Maryland, USA	Data from national, state and local planning organizations	Population-, household-, workers-,employment density, school enrollment, number of retail locations, number of recreational locations, congestion and free flow speed, Amtrak presence.	When land use diversity increases, especially with transit stations, grocery stores, and retail stores, people tend to rely on non-automobile modes more frequently.
(Broberg & Sarjala, 2015)	Paper based on a model	Helsinki region	School journeys	Density of major roads, connectivity, intersection density, proportion of land covered by single-family housing, population and housing densities, proximity, traffic safety, car ownership.	
(Heesch et al., 2015)	Paper based on statistical analyses	Brisbane, Australia	Study of physical activity, sedentary behavior, and health in adults aged 40-65	Disadvantagedness level of the neighborhood, network distance, bike path length, tree coverage, street lights, network distance to coast, connectivity.	

Table 1 (continued).

(Piatkowski & Marshall, 2015)	Paper based on a model	Denver, Colorado, USA	Survey data	Gender, age, race, household size, education level, household income level, car availability, trip distance, link-to-node ratio, intersection density, safety and infrastructure, security and comfort, relative convenience.	Barriers to cycling are different for people who already cycle compared to those who do not cycle yet, but are interested. Infrastructure improvements and presence of bicycle facilities is frequently associated with increased bicycle commuting.
(Schoner et al., 2015)	Paper based on a model	City center and suburban neighborhoods in the Minneapolis area, USA	Survey data	Job accessibility, commute distance, employer's parking policy, bike routes beyond the neighborhood, bike lanes.	Establishing causality between travel behavior and the built environment is challenging. Bicycle infrastructure might work as a magnet rather than a catalyst: people who are motivated to cycle, tend to move to places with bicycle infrastructure.
(Buehler & Dill, 2016)	Review			Bikeway networks, type of cycling infrastructure (separated tracks or bike lanes), traffic volume, car traffic speeds, car parking, intersections, bicycle-specific traffic control devices (bike boxes, bike traffic signals, bicycle signal activation.	

Table 1 (continued).

(Mertens et al., 2016)	Paper based on conjoint analysis		1950 middle-aged adults' answers to a web-based questionnaire	Separated cycle paths, speed limit, traffic density, evenness of surface, vegetation, upkeep, walkability, connectivity, residential density, land use mix diversity, access to shops/services/work.	Consistent relationships between macro scale factors (walkability, access to shops/services/work, degree of urbanization) and transport-related cycling in adults. Providing streets with separated cycle path increases appeal for cycling for transport. When a separated path is already present, comfort and aesthetic measures become more important
(Zahabi et al., 2016)	Paper based on a model	Montreal, Canada	Automobile and bicycle trip information from OD-surveys	Bicycle infrastructure accessibility, age.	Residential self-selection bias: people who like to cycle are more likely to choose to live in locations that are most amenable to cycling (i.e. with high population density), and the effect of built environment variables may also be capturing this self-selection
(Chen et al., 2017)	Paper based on a model	Seattle, Washington, USA	Bicycle counts at 50 locations	Land use, landscape (presence of water, hilliness), workplaces, bicycle infrastructure, neighborhood racial and age compositions, trip distance (city size, density, road connectivity, block size, destination accessibility, compact urban environment…), car related costs, amenities at destination, weather, season, peak travel hour, weekends.	Built environment factors can be categorized as either functional, safety, or aesthetic-related. Majority of trips had **recreational** purpose.

Table 1 (continued).

(Mertens et al., 2017)	Cross-sectional study	Ghent (BE), Randstad (NL), Paris (FR), Budapest (H), London (UK)	Data from online survey	Traffic calming features, number of bicycle lanes, speed limit, trees, litter, age, gender, educational level.	Two distinct dimensions of the physical environment: objective and perceived attributes (perceived attributes may be biased). Living in a neighborhood with more traffic calming features or fewer bicycle lanes was associated with being less likely to engage in cycling for transport.
(Osama et al., 2017)	Paper based on a model	134 Traffic Analysis Zones in Vancouver, Canada	Land use and road facility data	Length of bike network, bike network connectivity, bike network coverage, continuity, recreational bike networks, slope, bike network linearity, residential zoned areas, recreational zoned areas, arterial roads, collector roads.	Cyclists are highly sensitive to distance.
(Sun et al., 2017)	Paper based on a model	Glasgow	Cycling trips through Strava Metro app.	Road length, road connectivity, residential land use, volume of motor vehicles.	This paper focuses on **recreational cycling**.
(Aziz et al., 2018)	Paper based on a model	New York City, USA	Regional Household Travel Survey data	Gender, ethnicity, sidewalk width, total bike lane length, traffic safety, land use related to vacant lots; open space; parking facilities or industry.	Three major elements in the built environment are transportation infrastructure, land use pattern, and urban design. Higher fractions of open and vacant land use decreases likelihood of using bike mode. Perception of traffic safety plays a significant role in mode choice decision.

Table 1 (continued).

(Le et al., 2018)	Paper based on a model	20 US metropolitan areas	Bicycle and pedestrian counts from National Bicycle and Pedestrian Documentation Project	Easy access to water, high job density, high rates of active commuting, bike lanes, shared lane markings, off-street trails.	
(Cervero et al., 2019)	Paper based on a model	36 cities and towns in Britain	Cycle trips, retrieved from UK Census	Circuity, on-road stress, land use mix, landscape, cycling culture, weather and topography.	There is no single factor, even in cities with remarkably high commuter cycling.
(Sarjala, 2019)	Paper based on a model	2 neighborhoods in Tampere, Finland	73 commute routes	Institutional land use, hilliness, intersection density, age of buildings, forests, height of buildings.	Most significant associations are found mainly with the smallest buffer (15m). Intersections and long, steep slopes are to be avoided. Institutional land uses, slight hilliness, dwellings, forests , high buildings and variation in land use are preferred features along commute routes.
(Weliwitiya et al., 2019)	Paper based on 8 generalized linear models	Melbourne, Australia	Bicycle access counts at 207 metropolitan rail stations	Train frequency, availability of secure bicycle parking, elevation, patronage, percentage of local roads, land use mix, bicycle crash count density.	

Table 2. (Sub) variables found to be significant in the literature.

Variable	Sub variables	Effect hypothesis	Papers
Land use and urban spatial structure	Land use mix	+	(Cervero et al., 2019; Heinen et al., 2010; Mertens et al., 2016; Sallis et al., 2013; Vandenbulcke et al., 2011; Weliwitiya et al., 2019; Winters et al., 2010; Zhao, 2013)
	Job density	+	(Chen et al., 2017; Cui et al., 2014; Vandenbulcke et al., 2011)
	Population density	+	(Broberg & Sarjala, 2015; Cui et al., 2014; Mertens et al., 2016; Osama et al., 2017; Vandenbulcke et al., 2011)
	Residential zoned areas	+	(Broberg & Sarjala, 2015; Osama et al., 2017; Pucher et al., 2010; Sun et al., 2017)
	Institutional land use	+	(Sarjala, 2019)
	Destination accessibility & trip length	+	(Broberg & Sarjala, 2015; Chen et al., 2017; Cui et al., 2014; Hunt & Abraham, 2007; Mertens et al., 2016; Moudon et al., 2005; Piatkowski & Marshall, 2015; Rietveld & Daniel, 2004; Schoner et al., 2015; Vandenbulcke et al., 2011; Zhao, 2013)
	Land use related to vacant lots	-	(Aziz et al., 2018)
	Land use related to industry	-	(Aziz et al., 2018)
	Land use related to open space	-	(Aziz et al., 2018)
	Land use related to parking facilities	-	(Aziz et al., 2018)
	City size	+/-	(Chen et al., 2017; Rietveld & Daniel, 2004; Vandenbulcke et al., 2011)
	Density	+	(Buehler & Pucher, 2012; Chen et al., 2017; Heinen et al., 2010; Winters et al., 2010)
	Block size & intersection density	+/-	(Broberg & Sarjala, 2015; Chen et al., 2017; Heinen et al., 2010; Piatkowski & Marshall, 2015; Sarjala, 2019; Winters et al., 2010)
	Building age	-	(Sarjala, 2019)
	Building height	+/-	(Sarjala, 2019)

Table 2 (continued).

Bicycle network	Presence of bicycle infrastructure	+	(Buehler & Pucher, 2012; Chen et al., 2017; Heinen et al., 2010; Piatkowski & Marshall, 2015; Vandenbulcke et al., 2011; Winters et al., 2010; Zahabi et al., 2016)
	Physical condition and maintenance of the infrastructure	+	(Mertens et al., 2016; Parkin et al., 2008; Pucher et al., 2010; Rietveld & Daniel, 2004)
	Hindrances in road use	-	(Rietveld & Daniel, 2004)
	Length of bike network	+	(Aziz et al., 2018; Heesch et al., 2015; Osama et al., 2017)
	Circuity	-	(Cervero et al., 2019; Pucher et al., 2010; Rietveld & Daniel, 2004)
	Continuity	+	(Buehler & Dill, 2016; Heinen et al., 2010; Osama et al., 2017)
	Connectivity	+	(Broberg & Sarjala, 2015; Heesch et al., 2015; Mertens et al., 2016; Osama et al., 2017; Piatkowski & Marshall, 2015; Vandenbulcke et al., 2011)
	Coverage	+	(Osama et al., 2017)
	Linearity	+	(Osama et al., 2017)
	Signed bicycle routes	+	(Pucher et al., 2010)
	Bike routes beyond the neighborhood	+	(Schoner et al., 2015)
	Distance to the coast	-	(Heesch et al., 2015)
	Presence of a recreational bike network	+	(Osama et al., 2017)
	Waiting time at crossings	-	(Rietveld & Daniel, 2004)
	Number of stops cyclists have to make	-	(Heinen et al., 2010; Rietveld & Daniel, 2004)
	Bike boxes	+	(Buehler & Dill, 2016; Pucher et al., 2010)
	Bike traffic signals	-	(Buehler & Dill, 2016; Pucher et al., 2010)
	Bike signal activation	+	(Buehler & Dill, 2016; Winters et al., 2010)
	Bike parking facilities	+	(Heinen et al., 2010; Pucher et al., 2010; Weliwitiya et al., 2019)
	Storage facilities	+	(Heinen et al., 2010)
	Amenities at destination	+	(Chen et al., 2017)

Table 2 (continued).

	Showers at destination	+	(Heinen et al., 2010; Pucher et al., 2010)
Relative position with respect to cars	Relative position with respect to cars	+/-	(Rietveld & Daniel, 2004)
	Proportion of off-road bicycle routes	+	(Parkin et al., 2008)
	Separated tracks & paths	+	(Buehler & Dill, 2016; Buehler & Pucher, 2012; Heesch et al., 2015; Hunt & Abraham, 2007; Mertens et al., 2016; Pucher et al., 2010)
	Bike lanes	+	(Buehler & Dill, 2016; Buehler & Pucher, 2012; Hunt & Abraham, 2007; Mertens et al., 2016; Pucher et al., 2010; Schoner et al., 2015; Zhao, 2013)
	Bike boulevards	+	(Pucher et al., 2010)
	Shared bus/bike lanes	+	(Pucher et al., 2010)
	Mixed traffic	-	(Hunt & Abraham, 2007)
	Two-way travel on one-way streets	+	(Pucher et al., 2010)
	Shared lane markings	+	(Pucher et al., 2010)
	Colored lanes	+	(Pucher et al., 2010)
Personal safety	Personal safety	+	(Piatkowski & Marshall, 2015; Rietveld & Daniel, 2004)
	Risk of theft	-	(Hunt & Abraham, 2007; Rietveld & Daniel, 2004)
	Risk of vandalism	-	(Rietveld & Daniel, 2004)
Traffic safety	Traffic safety	+	(Aziz et al., 2018; Broberg & Sarjala, 2015; Buehler & Pucher, 2012; Piatkowski & Marshall, 2015; Rietveld & Daniel, 2004; Sallis et al., 2013)
	On-road stress	-	(Cervero et al., 2019)
	Bicycle accidents	-	(Vandenbulcke et al., 2011; Weliwitiya et al., 2019)
	Visibility at intersections	+	(Kirner Providelo & da Penha Sanches, 2011)
Motorized vehicles' network	Road connectivity	+/-	(Chen et al., 2017; Sun et al., 2017)
	Road length	+/-	(Sun et al., 2017)
	Highways	-	(Broberg & Sarjala, 2015; Winters et al., 2010)

Table 2 (continued).

	Arterial roads	-	(Broberg & Sarjala, 2015; Osama et al., 2017; Winters et al., 2010)
	Collector roads	-	(Broberg & Sarjala, 2015; Osama et al., 2017)
	Percentage of local roads	+/-	(Weliwitiya et al., 2019)
	Intersections	+	(Buehler & Dill, 2016; Kirner Providelo & da Penha Sanches, 2011; Zhao, 2013)
	Mainroad or expressway crossings	-	(Zhao, 2013)
	Lane width	-	(Kirner Providelo & da Penha Sanches, 2011)
	Car traffic speeds	+/-	(Buehler & Dill, 2016; Kirner Providelo & da Penha Sanches, 2011; Mertens et al., 2016, 2017)
	Congestion speeds	+/-	(Cui et al., 2014)
	Free flow speeds	+/-	(Cui et al., 2014)
	Traffic calming features	+	(Mertens et al., 2017; Pucher et al., 2010)
	Car related costs	+	(Chen et al., 2017)
	Gasoline price	+	(Buehler & Pucher, 2012)
	Traffic volumes	-	(Buehler & Dill, 2016; Sun et al., 2017; Vandenbulcke et al., 2011)
	Traffic density	-	(Mertens et al., 2016)
Parking	Car Parking	-	(Buehler & Dill, 2016)
	Adjacent car parking	-	(Heinen et al., 2010)
	Parking cost	+	(Heinen et al., 2010)
	Employer's parking policy	+/-	(Schoner et al., 2015)
Socio-demographic	Ethnicity	+/-	(Aziz et al., 2018; Chen et al., 2017; Piatkowski & Marshall, 2015; Sallis et al., 2013)
	Age	+/-	(Chen et al., 2017; Mertens et al., 2017; Piatkowski & Marshall, 2015; Sallis et al., 2013; Vandenbulcke et al., 2011; Zahabi et al., 2016)
	Income	+/-	(Piatkowski & Marshall, 2015; Vandenbulcke et al., 2011)

Table 2 (continued).

	Gender	+/-	(Aziz et al., 2018; Mertens et al., 2017; Owen et al., 2010; Piatkowski & Marshall, 2015; Vandenbulcke et al., 2011)
	Body Mass Index	-	(Sallis et al., 2013)
	Education	+/-	(Cui et al., 2014; Mertens et al., 2017; Piatkowski & Marshall, 2015; Sallis et al., 2013; Vandenbulcke et al., 2011)
	Marital status	+/-	(Sallis et al., 2013)
	Family commitments	-	(Vandenbulcke et al., 2011)
	Household size	+	(Piatkowski & Marshall, 2015)
	Employment	+/-	(Owen et al., 2010)
	Number of college students	+	(Buehler & Pucher, 2012)
	Car ownership	-	(Broberg & Sarjala, 2015; Heinen et al., 2010; Parkin et al., 2008; Sallis et al., 2013)
	Car access	-	(Buehler & Pucher, 2012; Piatkowski & Marshall, 2015)
	Bicycle ownership	+	(Heinen et al., 2010)
	Neighborhood disadvantagedness	+	(Heesch et al., 2015)
Presence of PT	Amtrak presence (long-distance trains)	-	(Cui et al., 2014)
	Train frequency	+	(Weliwitiya et al., 2019)
Public space	Weather and season	+/-	(Cervero et al., 2019; Chen et al., 2017; Heinen et al., 2010; Parkin et al., 2008; Rietveld & Daniel, 2004; Vandenbulcke et al., 2011)
	Topography and hilliness	-	(Cervero et al., 2019; Heinen et al., 2010; Osama et al., 2017; Parkin et al., 2008; Sarjala, 2019; Vandenbulcke et al., 2011; Weliwitiya et al., 2019; Winters et al., 2010)
	Presence of water	+	(Chen et al., 2017)
	Urban design	+	(Zhao, 2013)
	Vegetation	+	(Heesch et al., 2015; Kirner Providelo & da Penha Sanches, 2011; Mertens et al., 2016, 2017)

Table 2 (continued).

	Street lights	+	(Heesch et al., 2015)
	Upkeep	+	(Mertens et al., 2016, 2017)
	Complete streets	+	(Pucher et al., 2010)
	Car free zones	+	(Pucher et al., 2010)
	Bikeability	+	(Pucher et al., 2010)
	Walkability	+	(Mertens et al., 2016; Owen et al., 2010)
	Sidewalk width	+	(Aziz et al., 2018)
	Forests	+	(Sarjala, 2019)
Cycling culture	Cycling culture	+	(Cervero et al., 2019)
Policies	Pro-cycling policies	+	(Vandenbulcke et al., 2011)

107 BE characteristics had a possible significant influence on bicycle use, according to the literature. They are listed in Table 2. 51 of these 107 so-called predictors were available in Flanders and the province of Antwerp. They are listed in Table 3, which is quite similar to Table 2. Table 3 mentions the data source, the predictor and its hypothesized effect, together with the corresponding variable that was available for developing the model.

Table 3. Predictor variables from Table 2 and corresponding variables available for Flanders and the Province of Antwerp.

Source	Predictor	Effect	Available variables	Notes
OSM	Land use mix	+	Different types of land use	
OSM	Job density	+	Points of interest; industrial land use	
Geo.be	Population density	+	Inhabitants per statistical sector	
OSM	Residential zoned areas	+	Residential land use	
OSM	Destination accessibility & trip length	+	Points of interest	
OSM	Land use related to industry	-	Industrial land use	
OSM	Land use related to parking	-	Parking facilities	
OSM	Density	+	Buildings	
OSM	Block size and intersection density	+/-	Intersections	
Statbel	Building age	-	Percentage of buildings built after 2001, after 1991, before 1991	Data on municipal level. 2011 data.
OSM, Geopunt	Length of bike network	+	Cycle paths, lanes, tracks	
OSM	Connectivity	+	Intersections	
Geopunt	Separated tracks & paths	+	Separated tracks	Only for AWV bicycle infrastructure
Geopunt	Bike lanes	+	Tracks that are not separated	Only for AWV bicycle infrastructure
AWV	Traffic safety	-	Locations with most severe accidents	
OSM	Road connectivity	+/-	Intersections	
OSM	Road length	+/-	Roads	
OSM	Highways	-	Highways	
OSM	Intersections	+	Intersections	
Geopunt	Car traffic speeds	+/-	Speed limit for each road	
Geopunt	Traffic volumes and density	-	Road categories	Only for AWV roads

Table 3 (continued).

OSM	Car parking	-	Parking facilities	
Statbel	Ethnicity	+/-	People immigrated after 1980; foreigners; people born abroad	Data on municipal level. 2011 data.
Statbel	Age	+/-	Average age of population	Data on municipal level. 2011 data.
Statbel	Income	+/-	Average income	Data on municipal level.
Statbel	Gender	+/-	Ratio female/male	Data on municipal level. 2011 data.
Statbel	Education	+/-	People older than 20 with high degree	Data on municipal level. 2011 data.
Statbel	Marital status	+/-	Married people; People with registered partnerships	Data on municipal level. 2011 data.
Statbel	Household size	+	Average household size	Data on municipal level. 2011 data.
Statbel	Employment	+/-	Percentage of employees	Data on municipal level. 2011 data.
OSM	Number of college students	+	Schools, universities, colleges	
OSM	Amtrak presence (Long distance trains)	+	Train stations	
OSM	Topography and hilliness	-	Bridges; tunnels	
Corine	Presence of water	+	Water bodies	
Corine	Vegetation	+	Vegetation types	
OSM	Street lights	+	Street lamps	
OSM	Car free zones	+	Pedestrian-only streets	
Corine	Forests	+	Forests	
Stad Antwerpen	Bike sharing	+	Bike sharing stations	

5.3 LUR modeling

Land Use Regression (LUR) modeling is a commonly used technique to predict air pollution concentrations. It is often used in epidemiological studies and is based on a number of predictor variables, which are usually traffic-, population- and land use-related. To obtain these variables, a GIS is used. A LUR was first used in the SAVIAH study by Briggs et al. (1997), who called the technique regression mapping. The objective was to predict respiratory diseases in children, based on air pollution concentrations. Regression mapping seems to be a more suitable name for the technique, as the predictor variables that are used are not always related to land use (Boniardi et al., 2019; Briggs et al., 1997; Dons et al., 2013b; Hoek et al., 2008; Int Panis, 2018).

LUR models have become very popular in the past years, thanks to the developments in Geographic Information Systems (GIS). The SAVIAH study was the first one to compile air pollution-related variables, such as traffic measures, population density, land use and altitude variables, in a GIS. Using various buffers and linear regression, the variables were calculated and the model was developed (Briggs et al., 1997; Hoek et al., 2008).

As stated before, a LUR consists of a number of steps, each of them is explained below.

The first part of composing a LUR is **data monitoring**. Existing data have to be collected as input for the model. They are considered the dependent variables in the model. Either routine monitoring or purpose-designed monitoring can be used. Routine monitoring networks are existing networks that have been installed for another purpose than the study itself. In the case of bicycle counts, routine monitoring could exist of fixed counting devices that count bicycles continuously. Usually, routine networks lack density and cannot enable modelling of variability on small scales. Therefore, most studies monitor data for the specific purpose of the model. For bicycle counts, these are the typical installations with rubber hoses. An important advantage of purpose-designed monitoring is the fact that investigators can control the locations where they want to monitor the data. Disadvantages are the cost and limited temporal coverage of the data monitoring. In contrast to routine monitoring, purpose-designed monitoring typically happens during one to four periods of generally seven to 14 days (Hoek et al., 2008; Ryan & LeMasters, 2006).

Most studies were based on between 20 and 100 monitoring sites. As the size of the city and population should be taken into account, 40 to 80 monitoring sites seems a reasonable number. However, the locations of the monitoring sites may be more important than the number of sites used, as the latter seems not to be correlated with the model R^2. Thus, data with very low or high numbers of cyclists should not be ignored, as they define the outcomes of the model (Hoek et al., 2008; Int Panis, 2018; Ryan & LeMasters, 2006).

The second step of composing a LUR model is the collection of **predictor variables**. They are considered the independent variables. For this report, this has happened in chapter 6.2. Initially, most studies start with a large set of potential predictor variables, even up to 140 predictors. The final models however only contain a small number of these predictors (Hoek et al., 2008; Ryan & LeMasters, 2006).

How these predictor variables are defined depends mainly on the availability of the data and the unique features of the study area. Potential problems with the collection of these data are the accessibility, completeness and precision of the predictor variables. Especially traffic intensity data often tend to be problematic to acquire. Often, they are not accessible, or if they are, only for major roads. This implies that counts have to be extrapolated along and between road links or that traffic models can be used to assign traffic counts to roads of which data are not available. In cases where traffic counts were completely unavailable, the length of specific road types has proved to be a suited alternative (Hoek et al., 2008; Ryan & LeMasters, 2006).

When the predictor variables are obtained, they are usually computed using buffer functions in GIS. Doing so, the selection of the buffer size is critical, as it determines the amount of surrounding traffic, road length and land cover which may explain the variability of the dependent variable. It thus defines the performance of the model. Usually, the buffer radii are based on the decay of the modeled pollutant. (Hoek et al., 2008; Ryan & LeMasters, 2006).

The third step is the actual **development and validation of the model**. To develop the model, most studies use linear regression techniques by constructing a regression equation based on the predictor variables. Some studies predefine the sign of the regression slopes for specific variables. This increases the applicability of the model beyond the investigation site. Some studies also use different predictors for different spatial scales (Briggs et al., 1997; Hoek et al., 2008).

The model validation itself is a crucial part of applying a LUR. It indicates how good the model predicts the values of the dependent variable. Most models consider annual average concentrations Two methods are commonly applied (Dons et al., 2013b; Hoek et al., 2008; Meng Wang et al., 2012):

- <u>Leave-one-out cross-validation</u>. With this type of validation, the model is developed for all sites except for one. The predicted values are then compared with the actual measured ones at the left-out site. After repeating this for all sites, the overall level of fit between the predicted and observed values is computed. The result can then be considered as a measure of model performance.
- <u>Hold-out validation:</u> the group of monitoring sites is subdivided into a training data set, that will be used for the model development, and a smaller group of sites, that will be used for validation of the model. This enables less intensive computer processing, yet subdividing the sites a priori may be a disadvantage.
- A third validation method can be to compare the LUR estimates with the data from fixed monitors (Int Panis, 2018).

The last step is the **application of the model**.

Whether the models are transferable between different areas, will most likely depend on the availability of equivalent variables in these areas. Models may not be transferable between areas that have a different structure in terms of land use. In the SAVIAH study, a separate model was used for each city, based on the same features: traffic volume, land cover and topography predictors, and similar buffers (Briggs et al., 1997; Hoek et al., 2008).

For air pollution (NO_2) LUR models, the R^2 value, which explains the variation of the prediction model, is typically between 50 and 80%. A robust model is characterized by similar R^2 values for both the model and the cross-validation. In studies that compare LUR to other modelling methods, LUR typically has much higher R^2 than these other methods (Hoek et al., 2008; Ryan & LeMasters, 2006).

Land Use Regression modeling is an empirical method, which means it can be readily adapted to local circumstances and availability. Studies that focus on a small spatial scale with little monitored data will benefit from applying LUR. Thus, they allow for the accounting for small scale variability in the dependent variable (Briggs et al., 1997; Ryan & LeMasters, 2006).

LUR models have some limitations. Specifically for air pollutants, the ability to separate the different impacts of different pollutants is limited. Furthermore, LUR models may not be able to predict extremely local variations, but this is mainly related to the precision of the input data. Finally, the short temporal coverage of purpose-designed monitoring campaigns does not allow the precise calculation of absolute concentrations (Hoek et al., 2008).

Comparison with a traditional four-step model

In this study, bicycle counts will be forecasted using a LUR model. However, other methods exist to predict travel demand. The most commonly used is the traditional four-step travel demand model.

A traditional four-step model, or gravity model, predicts traffic volumes. It is often used to support infrastructure decisions and consists of four unique steps. These steps are the trip generation, trip distribution, modal split and traffic assignment. Each step applies to a traffic analysis zone (TAZ). Trip generation determines the number of trips every TAZ produces or attracts, regardless of the travel mode. Trip distribution matches trip productions to trip attractions while considering both the time or cost of connections and the spatial distribution of these productions and attractions. The third step determines the travel mode that is used for each trip. Traffic assignment assigns trips to the networks of each mode. This step determines the routes that are taken (Anderson et al. , 2006; Clifton et al., 2016; Park et al., 2019).

Existing travel patterns, obtained from surveys, deliver input for estimating the model. Calibration and validation happens by comparing the predicted trips to the actual travel patterns from the survey (Park et al., 2019).

An important limitation of the four-step model are its TAZs. The model performs rather inaccurate on predicting trips within the same zone, as all trips in a zone depart from the TAZ’s centroid. This results in an aggregation bias where the data set for each TAZ is aggregated and cannot be influenced by intrazonal land use or other characteristics (Clifton et al., 2016; Park et al., 2019).

Similar to the LUR model that will be developed for this study, a four-step model is used to predict traffic volumes. Both methods consist of four steps of which some are rather similar. For both models, the first step, either data monitoring or trip generation, relies on existing data. Both models have to be validated using existing data.

An important difference between four-step and LUR models is the spatial scale. Whereas four-step models apply to TAZs and are consequently spatially rather aggregated, LUR models can be applied to a very small scale. This makes the LUR method more suited for this study, as the aim is to predict bicycle counts on specific locations.

6 Results

6.1 Data cleaning and preparation

Bicycle count data were available for 26 permanent counting points and 11 temporary counting points. Descriptive statistics of these measurements are shown in Table 4. The table also shows descriptive statistics for the different categories of counting points, both permanent and temporary. A distinction between bike highways, urban counting points and rural counting points was made as well. 32% of the counting points was located on a bicycle highway, while 46% of the counting points was located in a rural area.

Table 4. Descriptive statistics of bicycle counts by category.

Type	Variable	Number (%)	Min	Max	Average	Median
Overall	N	37 (100%)	113.03	2804.83	827.47	697.31
Permanent	N	26 (70%)	183.19	2804.83	858.41	676.73
Temporary	N	11 (30%)	113.03	1626.66	754.33	806.82
Bike Hwy	N	12 (32%)	473	1827.21	929.41	814.77
Urban	N	8 (22%)	463.19	2804.83	1349.05	1307.62
Rural	N	17 (46%)	113.03	1413.92	510.05	394.71

Through buffering and the intersect function in QGIS, these 37 counting points received a value for each of the 294 land use variables. How to interpret these variables can be found in Table 18 on page 93.

The variables were then cleaned to prevent overfitting of the model. 21 variables had '0' as a value for more than 32 of the 37 counting points. These were mainly variables with small buffer sizes (100-500 m). Additionally, each of the 25 AWV bicycle infrastructure variables was omitted, as AWV only manages a small part of bicycle infrastructure in the province of Antwerp. Instead, cyclepath data from OSM were used. Finally, a road length variable was added with buffer sizes of 100, 300, 500, 1000, 2000, 4000 and 6000 m. After this cleaning process, 224 variables remained.

6.2 Correlation between independent variables

After checking for variables that correlated at least 95% with each other, 135 variables remained. All variables were then divided into five categories specifically chosen for this data set: Land use variables, Sociodemographic variables, Infrastructure variables, Collective mode variables, and Traffic variables.

After omitting variables that correlated for 60% or more with the significant variable of a category, a total of 112 variables remained. Descriptive statistics of these variables are shown in Table 5.

Table 5. Descriptive statistics of independent variables (see Table 18 in appendix for the meaning of these variables).

Variable	Min	Max	Average	Median
l_cycleway100	0	644	312.24	290.00
l_cycleway300	0	3618	1227.32	1256.00
l_cycleway500	0	7392	2334.86	2205.00
brdg_100	0	1	0.30	0
brdg_300	0	1	0.54	1
brdg_500	0	1	0.59	1
brdg_1000	0	1	0.81	1
tun1000	0	1	0.59	1
zp500	0	3	0.27	0
park300	0	36039	2368.05	0.00
park500	0	207991	13831.70	0.00
park1000	0	1035105	80363.49	38242.00
park2000	0	1925456	276684.30	97381.00
ind300	0	126733	10215.00	0.00
ind500	0	210546	28886.00	0.00
ind1000	0	873678	102488.00	10217.00
ind2000	0	2785733	434939.00	169622.00
ind4000	10781	7546295	2087532.00	1436021.00
ind6000	392185	19017926	5399359.00	4222663.00
resi100	0	31398	8999.60	0.00
station1000	0	1	0.30	0.00
station4000	0	7	2.35	2.00
avg_popdens4000	191	2749	952.65	642.20
agri100	0	31399	8977.40	0.00
agri500	0	632146	207954.00	133564.00
for500	0	651353	81271.00	0.00
for1000	0	2310473	324622.00	0.00
lowveg1000	0	2785209	358308.60	0.00
water300	0	132940	17361.14	0.00
water500	0	285614	52392.54	0.00
water1000	0	978633	16453.16	0.00
parking100	0	8511	422.70	0.00
parking300	0	14208	2141.35	1096.00
parking_b1000	0	2	0.16	0.00
sch500	0	2	0.19	0.00
sch1000	0	6	0.84	0.00
sch2000	0	16	3.32	2.00
sch6000	0	46	15.49	9.00
h_edu2000	0	2	0.35	0.00
h_edu4000	0	4	0.78	0.00
h_edu6000	0	4	1.03	0.00
jobs100	0	12	0.76	0.00
poi100	0	12	1.24	0.00
poi1000	1	1337	118.87	48.00
int500	4	1328	392.00	276.00
l_hwy1000	0	12684	1058.38	0.00
prim300	0	1838	204.81	0.00
prim1000	0	8429	1108.27	0.00
prim4000	0	49950	16689.19	15062.00
sec300	0	1212	209.00	0.00
sec1000	0	6026	1292.65	0.00

Table 5 (continued).

Variable	Min	Max	Average	Median
sec2000	0.00	15242.00	4547.43	3776.00
sec4000	0.00	35750.00	17834.05	17129.00
sec6000	0.00	71406.00	37304.95	38670.00
velo300	0	4	0.32	0
velo1000	0	32	3.76	0
avg_inc500	46217.00	74460.45	53811.76	53046.00
reg_partner100	0.02	0.04	0.03	0.03
reg_partner500	0.02	0.04	0.03	0.03
married1000	0.48	0.58	0.54	0.56
r_femalemale500	94.89	107.19	100.96	101.65
r_femalemale1000	94.89	105.65	100.70	101.65
avg_age100	39.76	43.30	41.26	41.49
avg_age500	39.76	43.40	41.35	41.51
u15_100	0.14	0.18	0.16	0.16
u15_1000	0.14	0.18	0.16	0.16
65plus_300	0.15	0.21	0.18	0.17
before1991_500	0.02	0.22	0.10	0.08
after1991_1000	0.09	0.37	0.22	0.24
after2001_300	0.03	0.18	0.10	0.11
avg_inc4000	41103.62	72883.78	53924.28	54136.65
avg_inc6000	47475.80	72657.25	55203.70	54909.60
empl_r2000	0.78	0.88	0.85	0.86
empl_r4000	0.78	0.93	0.85	0.86
empl_r6000	0.78	0.87	0.85	0.85
avg_hh4000	1.95	2.72	2.37	2.40
reg_partner2000	0.03	0.04	0.03	0.03
reg_partner4000	0.02	0.30	0.04	0.03
reg_partner6000	0.02	0.06	0.03	0.03
married4000	0.48	0.65	0.54	0.56
married6000	0.48	0.78	0.55	0.56
20hi_edu2000	0.21	0.41	0.27	0.26
20hi_edu4000	0.22	0.49	0.28	0.26
20hi_edu6000	0.22	0.37	0.28	0.27
r_femalemale4000	76.38	114.54	100.58	101.67
r_femalemale6000	94.89	132.77	102.09	102.49
avg_age2000	39.78	43.45	41.25	41.39
avg_age4000	30.99	46.32	41.10	41.35
avg_age6000	40.14	54.60	41.75	41.40
u15_2000	0.14	0.18	0.16	0.16
u15_6000	0.14	0.24	0.16	0.16
15-65_2000	0.64	0.68	0.66	0.66
15-65_4000	0.64	0.75	0.67	0.66
15-65_6000	0.65	0.93	0.67	0.65
65plus_2000	0.15	0.21	0.18	0.17
65plus_6000	0.15	0.20	0.18	0.18
before1991_2000	0.02	0.21	0.11	0.09
before1991_4000	0.02	0.43	0.11	0.10
before1991_6000	0.02	0.19	0.10	0.10
after1991_4000	0.09	0.43	0.23	0.25
after1991_6000	0.10	0.42	0.23	0.25
after2001_4000	0.03	0.34	0.11	0.10

Table 5 (continued).

Variable	Min	Max	Average	Median
nbuildings1000	43	9386	2594.97	1569
distnear_st	236.08	18708.98	3772.01	2395.31
lu100	0	5	2.27	2
lu300	1	7	4.19	4
lu500	2	11	5.76	5
lu1000	5	12	7.86	8
lu2000	7	15	11.24	11
lu4000	9	17	14.51	15
lu6000	11	19	16.24	16

6.3 Model development

Ten iterations led to a linear regression model with nine variables, as described in Table 6.

Table 6. Initial model. The significance symbols translate as follows: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Variable	Estimate	Std. error	t-value	pr(>|t|)	significance
Intercept	-3066	1578	-1.942	0.062596	.
nbuildings1000	0.06659	0.03134	2.125	0.042883	*
l_hwy1000	-0.1090	0.02186	-4.986	0.0000316	***
avg_popdens4000	0.2515	0.1353	1.860	0.073879	.
l_cycleway500	0.1543	0.4032	3.828	0.000696	***
lu2000	105.0	31.06	3.382	0.002211	**
before1991_4000	-2144	798.7	-2.684	0.012268	*
poi1000	0.5441	0.2640	2.061	0.049074	*
reg_partner500	-29600	13410	-2.207	0.036011	*
empl_r4000	3724	1956	1.904	0.067557	.

This model had a multiple R^2 of 0.83 and an adjusted R^2 of 0.78. However, two variables (avg_popdens4000 and empl_r4000) are not statistically significant at 95%. Thus, the parameters were estimated again without these variables and because of this re-estimating, another variable (before1991_4000) was found to be not significant at 95%. This is shown in Table 7.

Table 7. Initial model without insignificant variables (1). The significance symbols translate as follows: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Variable	Estimate	Std. error	t-value	pr(>|t|)	significance
Intercept	50.40	552.1	0.091	0.927883	
nbuildings1000	0.09834	0.02901	3.390	0.002031	**
l_hwy1000	-0.01097	0.02374	-4.622	0.0000725	***
l_cycleway500	0.01768	0.04368	4.047	0.000352	***
lu2000	121.5	33.76	3.600	0.001171	**
before1991_4000	-1264	813.9	-1.552	0.131390	
poi1000	0.6712	0.2854	2.351	0.025711	*
reg_partner500	-33810	13400	-2.523	0.017383	*

Because of this insignificance, the model parameters were estimated a third time, resulting in the values listed in Table 8. One variable was not significant at 95% again, this time the variable poi1000.

Table 8. Initial model without insignificant variables (2). The significance symbols translate as follows: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Variable	Estimate	Std. error	t-value	pr(>|t|)	Significance
Intercept	216.9	554.1	0.391	0.698240	
nbuildings1000	0.1022	0.02957	3.455	0.001665	**
l_hwy1000	-0.1096	0.02429	-4.511	0.0000924	***
l_cycleway500	0.01513	0.04141	3.653	0.000981	***
lu2000	103.8	32.5	3.193	0.003299	**
poi1000	0.05075	0.2714	1.870	0.071293	.
reg_partner500	-35080	13690	-2.564	0.015612	*

The parameters were estimated for the fourth time, resulting in the final model with five variables. This model, shown in Table 9, has no insignificant (at 95%) variables. This model had an R^2 of 0.74 and an adjusted R^2 of 0.69. No variables in the model have a VIF higher than 3, as can be seen in Table 10. Cook's D (Table 11) indicated that four counting points (6, 7, 9, and 18) might have a strong influence on the model. Of these counting points, none has a Cook's D higher than 1. The highest value for Cook's D is 0.59 for point 9, which is situated near a train station in the city center of Antwerp. The plots in Figures 11 and 12 (Residuals vs. fitted values and a Q-Q plot) of the model were generated as output in R.

Thus, the final model is one with five variables: the number of buildings within 1000 m of a counting point, the length of highways within 1000 m, the length of cycleways within 500 m, the number of different types of land use within 2000 m, and the weighted average percentage of registered partnerships within 500 m of a counting point.

Table 9. Final model, without insignificant variables. The significance symbols translate as follows: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Variable	Estimate	Std. error	t-value	pr(>|t|)	Significance
Intercept	492.1	56.38	0.761	0.45231	
nbuildings1000	0.1341	0.02510	5.343	0.00000804	***
l_hwy1000	-0.1160	0.02500	-4.639	0.0000603	***
l_cycleway500	0.1507	0.04305	3.501	0.00143	**
lu2000	87.71	32.58	2.692	0.01135	*
reg_partner500	-36530	14200	-2.572	0.01512	*

Table 10. Each of the model's variables and their VIF.

Variable	nbuildings1000	l_hwy1000	l_cycleway500	lu2000	reg_partner500
VIF	1.451315	1.765258	1.618721	1.240188	1.091672

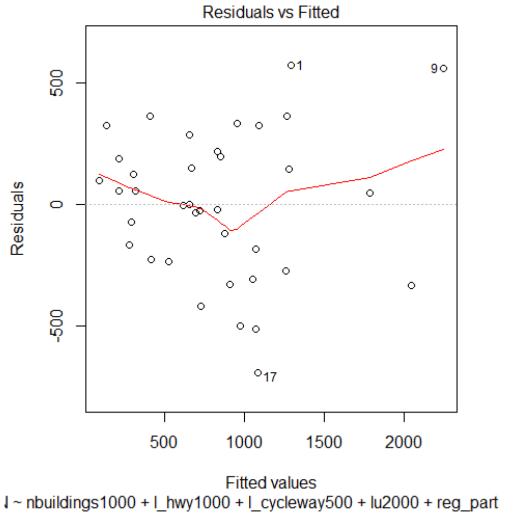


Figure 11. Residuals and fitted values.

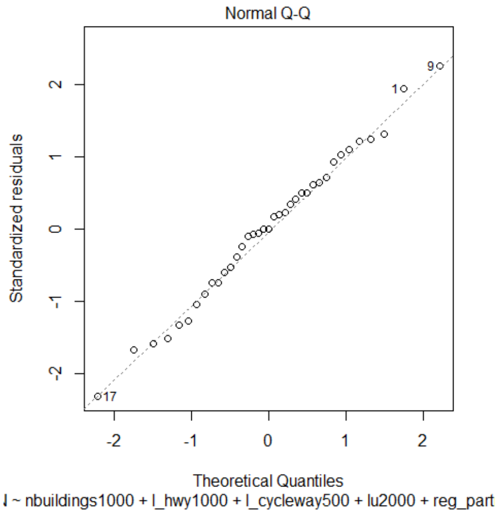


Figure 12. Q-Q plot.

Table 11. Counting points' Cook's D. Influential counting points are marked in bold.

Counting point	Cook's D	Counting point	Cook's D	Counting point	Cook's D
1	0.123	14	0.0239	26	0.0461
2	0.147	15	0.0230	27	0.00777
3	0.201	16	0.00000113	28	0.00695
4	0.00273	17	0.160	29	0.0112
5	0.0313	**18**	**0.00166**	30	0.00305
6	**0.506**	19	0.0000289	31	0.000167
7	**0.00567**	20	0.00588	32	0.00171
8	0.00155	21	0.00313	33	0.00836
9	**0.590**	22	0.000163	34	0.0398
10	0.0000000253	23	0.00598	35	0.0653
11	0.0116	24	0.00594	36	0.00834
12	0.0480	25	0.00629	37	0.0106
13	0.0252				

6.4 Sensitivity analysis

In the sensitivity analysis, the model was developed again without point 29 (Meersel-Dreef), which was located close to the Dutch border. Again, ten iterations led to a linear regression model with nine variables. These variables are listed in Table 12 and are the same variables as those in Table 6, page 56.

Table 12. Sensitivity analysis parameter estimates (1). The significance symbols translate as follows: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Variable	Estimate	Std. error	t-value	pr(>|t|)	significance
Intercept	-3663	1851	-1.979	0.058456	.
nbuildings1000	0.06524	0.03176	2.054	0.050132	.
l_hwy1000	-0.1078	0.02219	-4.859	0.0000488	***
avg_popdens4000	0.2613	0.1376	1.899	0.068778	.
l_cycleway500	0.1556	0.4082	3.812	0.000762	***
lu2000	107.8	31.70	3.400	0.002185	**
before1991_4000	-2267	830.3	-2.730	0.011212	*
poi1000	0.5505	0.2672	2.061	0.049484	*
reg_partner500	-28510	13670	-2.086	0.046976	*
empl_r4000	4345	2204	1.971	0.059403	.

Multiple R^2 for this model was 0.83 with an adjusted R^2 of 0.77. Three variables (nbuildings1000, avgpopdens4000, and empl_r4000) were not statistically significant at 95%. The parameters were estimated again, leaving out these three variables. The results are shown in Table 13.

Table 13. Sensitivity analysis parameter estimates (2). The significance symbols translate as follows: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Variable	Estimate	Std. error	t-value	pr(>|t|)	Significance
Intercept	-336.6	657.2	-0.512	0.312375	
l_hwy1000	-0.08931	0.02701	-3.306	0.002527	**
l_cycleway500	0.01981	0.05106	3.879	0.000556	***
lu2000	174.6	35.36	4.939	0.0000301	***
before1991_4000	-1463	957.5	-1.528	0.137236	
poi1000	1.208	0.2781	4.43	0.000157	***
reg_partner500	-35770	16010	-2.234	0.033330	*

Yet again, one of the variables in this model (before1991_4000) turned out not to be statistically significant at 95%. The model was developed once again, without this variable, since it was not significantly different from zero. This resulted in a model with five variables, as shown in Table 14.

Table 14. Sensitivity analysis parameter estimates (3). The significance symbols translate as follows: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Variable	Estimate	Std. error	t-value	pr(>|t|)	Significance
Intercept	-142.5	659.0	-0.216	0.830306	
l_hwy1000	-0.08828	0.02760	-3.199	0.003249	**
l_cycleway500	0.1689	0.04841	3.489	0.001520	**
lu2000	155.7	33.85	4.600	0.0000719	***
poi1000	1.040	0.2611	3.982	0.000401	***
reg_partner500	-37600	16310	-2.305	0.028275	*

All variables are statistically significant at 95%. Compared to the model with 37 counting points in 6.3, the model has an equal number of variables. The only difference is that here, poi1000 remains as one of the parameters. In the original model on the contrary, nbuildings1000 remains.

6.5 Validation

Leave-one-out cross validation

Leave-one-out cross validation happened with the same data as the ones that were used to develop the model, yielded an R^2 of 0.58. Table 15 shows each counting point, its measured number of cyclists and its estimated number of cyclists. The leave-one-out cross validation resulted in a negative estimated value for point 3. Figure 13 shows the scatter plot for this table.

Table 15. Results of leave-one-out cross validation.

Point	Measured	Estimated	Point	Measured	Estimated
1	1868.00	1179.17	**20**	818.29	644.86
2	1710.02	2223.16	**21**	431.30	290.47
3	463.19	56.52	**22**	697.31	723.75
4	752.36	885.53	**23**	289.49	540.24
5	473.00	1007.28	**24**	113.03	297.35
6	578.80	1330.52	**25**	400.44	190.76
7	375.04	280.35	**26**	742.65	1125.40
8	264.88	195.74	**27**	1413.92	1073.23
9	2804.83	1853.04	**28**	1424.50	1252.91
10	656.11	655.80	**29**	183.19	437.82
11	943.52	632.37	**30**	192.25	74.67
12	1629.66	1194.19	**31**	656.15	693.10
13	985.57	1306.81	**32**	215.11	300.70
14	1286.58	911.45	**33**	1045.19	825.25
15	307.88	758.94	**34**	768.99	345.98
16	609.91	611.92	**35**	558.63	1140.34
17	388.98	1204.80	**36**	886.89	1094.53
18	1827.21	1763.90	**37**	1046.59	800.43
19	806.82	826.19			

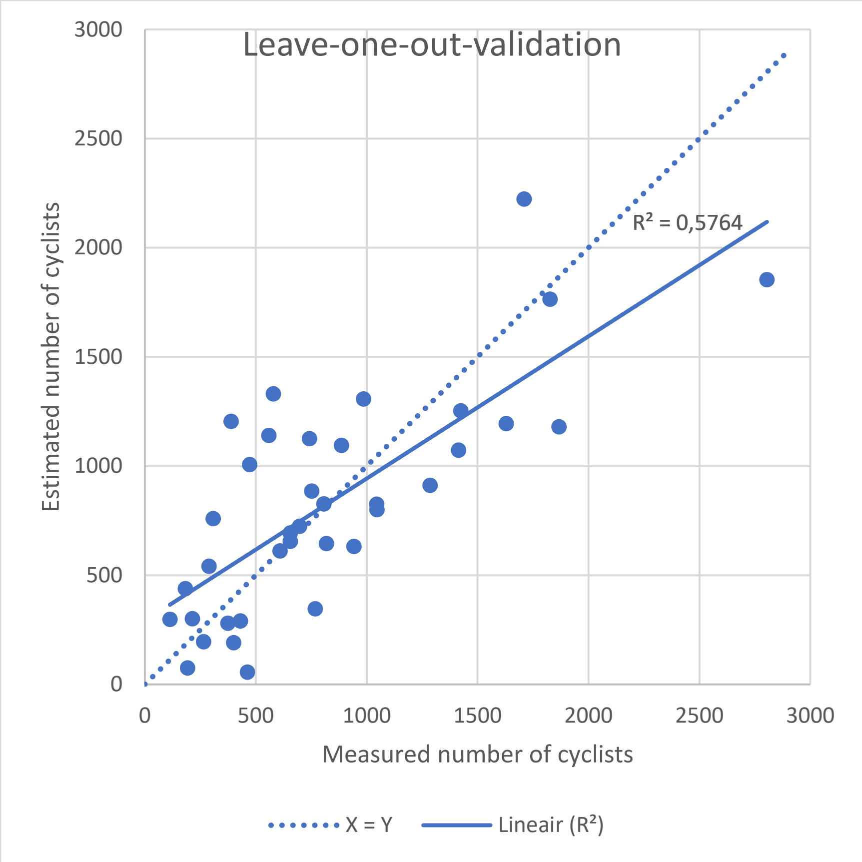


Figure 13. Scatter plot for leave-one-out validation.

External validation

The 83 counting points that were used for external validation, had a median of 35 counting days and an average of 31. Validation of the raw data (non-standardized) yielded an R^2 of 0.40. The scatter plot is shown in Figure 14. External validation with the same data, but standardized according to 2019 permanent counting points in the province of Antwerp, yielded an R^2 of 0.41. This is slightly higher than in the case of the raw data. The scatter plot of the standardized data is shown in Figure 15. The external validation data set was split into bicycle counts in Antwerp (performance), and Flemish Brabant (transferability). Figure 16 describes the model's performance, thus estimates of the number of cyclists on the counting points in the province of Antwerp. This yields an R^2 of 0.52. The transferability is shown in Figure 17. The model has a remarkably low R^2 (0.00008) here.

Table 16. Descriptive statistics of the independent variables used for external validation. The first row shows the independent variable's values from model development.

Variable	Min	Max	Average	Median
N (model)	113.03	2804.83	827.47	697.31
N (raw)	11.23	4101.75	757.26	482.57
N (standardized)	10.77	3617.78	646.84	422.88

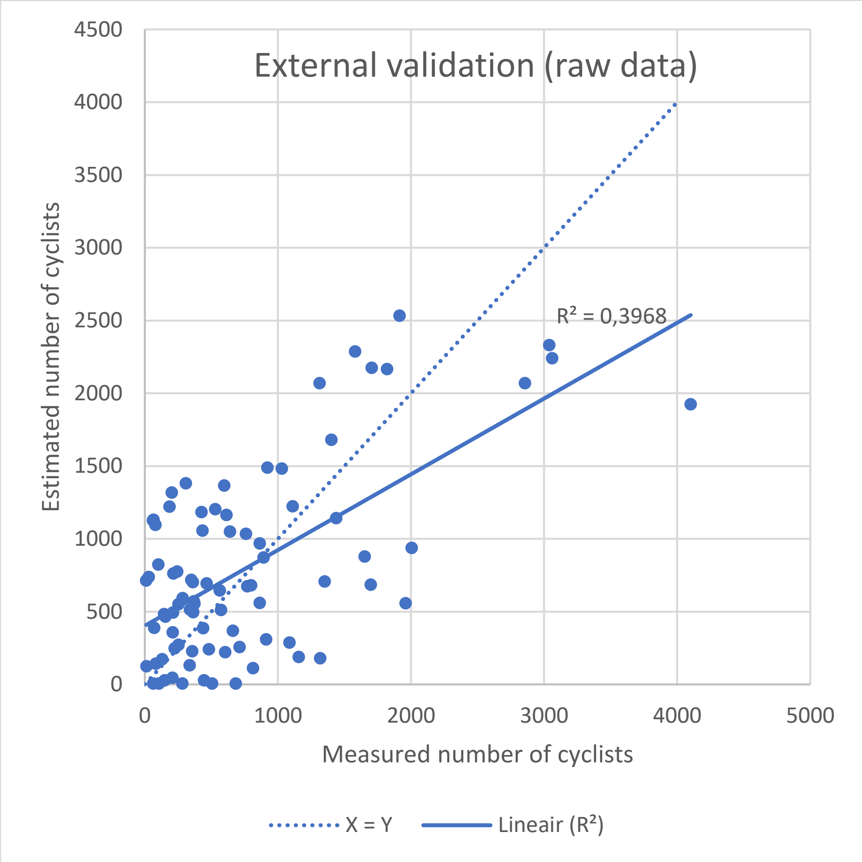


Figure 14. Scatter plot of external validation with RAW data.

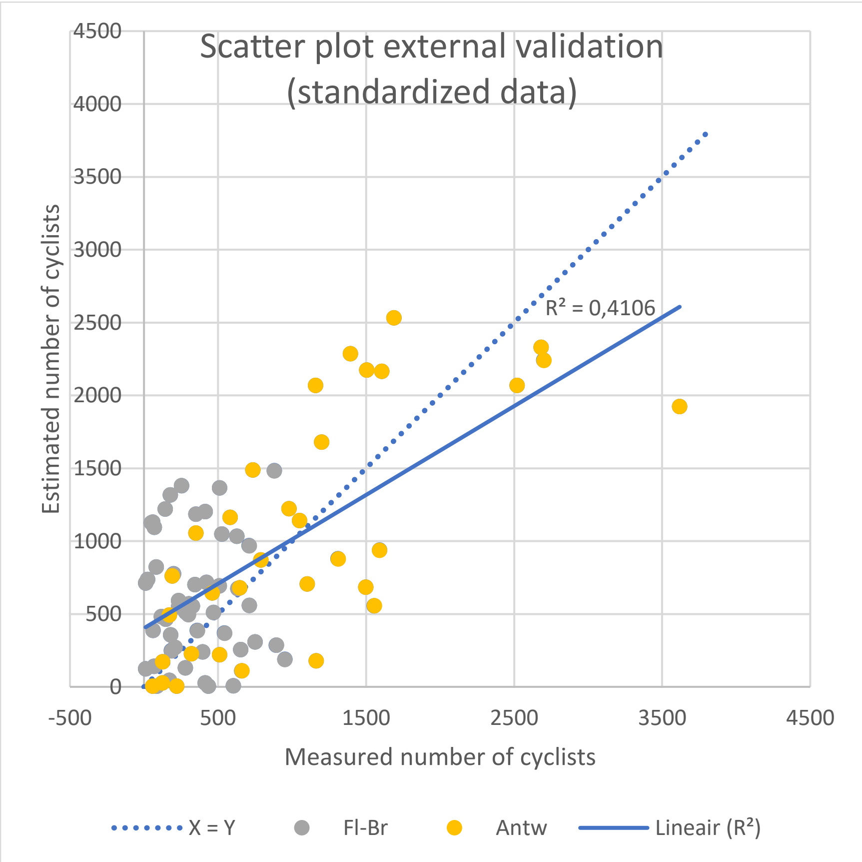


Figure 15. Scatter plot of external validation with STANDARDIZED data.

Table 17. Results of external validation. Points marked with a '*' are located in Flemish Brabant and were used to test transferability.

Point		Estimated		Point		Estimated	
	Measured	Raw	Standard		Measured	Raw	Standard
1	564.55	643.86	643.86	**43***	101.84	823.10	823.10
2*	14.56	123.67	123.62	**44***	11.23	713.20	713.20
3	1111.95	1223.11	1223.11	**45***	155.59	463.76	463.76
4	1914.50	2532.15	2532.13	**46**	1653.05	878.16	878.16
5	1318.75	179.20	179.20	**47***	771.78	673.73	673.73
6	1582.10	2286.91	2286.91	**48***	29.81	736.36	736.36
7	1822.35	2165.22	2165.22	**49***	1157.03	187.39	187.39
8	1706.75	2174.11	2174.11	**50***	362.37	701.78	701.78
9	605.11	219.69	219.69	**51***	466.81	693.48	693.48
10	2856.75	2068.65	2068.65	**52**	131.80	171.09	171.09
11	3062.40	2240.48	2240.48	**53***	341.49	512.98	512.98
12	1438.83	1141.39	1141.39	**54***	662.73	368.16	368.16
13	3040.90	2330.21	2330.21	**55**	2005.47	936.86	936.86
14	4101.75	1923.43	1923.43	**56***	426.70	1183.79	1183.79
15	1315.20	2067.70	2067.70	**57***	761.73	1033.80	1033.79
16*	309.11	1379.45	1379.45	**58**	1353.08	706.34	706.34
17*	71.38	387.53	387.53	**59**	1698.75	683.89	683.89
18*	80.62	1094.39	1094.40	**60**	63.71	5.61	5.39
19	216.31	761.76	761.76	**61***	912.32	306.97	306.97
20	613.56	1163.06	1163.06	**62***	446.92	26.93	26.94
21*	1030.23	1481.88	1481.88	**63***	685.05	5.61	5.39
22*	597.58	1364.76	1364.76	**64***	482.57	239.84	239.84
23	1961.53	556.07	556.07	**65***	368.49	570.12	570.12
24	921.43	1487.44	1487.44	**66***	864.24	968.33	968.33
25	1403.11	1679.23	1679.23	**67**	283.71	5.61	5.39
26	435.21	1056.02	1056.02	**68***	255.03	269.84	269.84
27	800.43	678.56	678.56	**69***	438.32	385.72	385.72
28*	285.81	591.85	591.85	**70***	349.61	716.97	716.97
29*	146.40	482.38	482.38	**71***	211.26	356.77	356.77
30	893.95	870.61	870.61	**72***	62.46	1124.96	1124.96
31*	253.00	549.24	549.24	**73***	106.97	5.61	5.39
32*	573.98	510.98	510.98	**74***	864.19	558.59	558.59
33	357.43	226.99	226.99	**75***	531.52	1203.40	1203.40
34*	1088.16	285.17	285.17	**76***	188.00	1220.13	1220.13
35*	712.10	254.59	254.59	**77***	243.81	773.99	773.99
36*	507.58	5.61	5.39	**78***	67.46	1130.65	1130.65
37*	84.51	140.56	140.56	**79***	204.38	1316.28	1316.28
38*	338.86	130.83	130.83	**80***	365.49	495.54	495.54
39*	208.24	43.90	43.90	**81***	640.49	1047.83	1047.83
40	814.85	109.44	109.44	**82***	226.24	247.32	247.33
41	213.36	493.02	493.02	**83***	373.67	553.32	553.32
42	150.29	25.74	25.74				

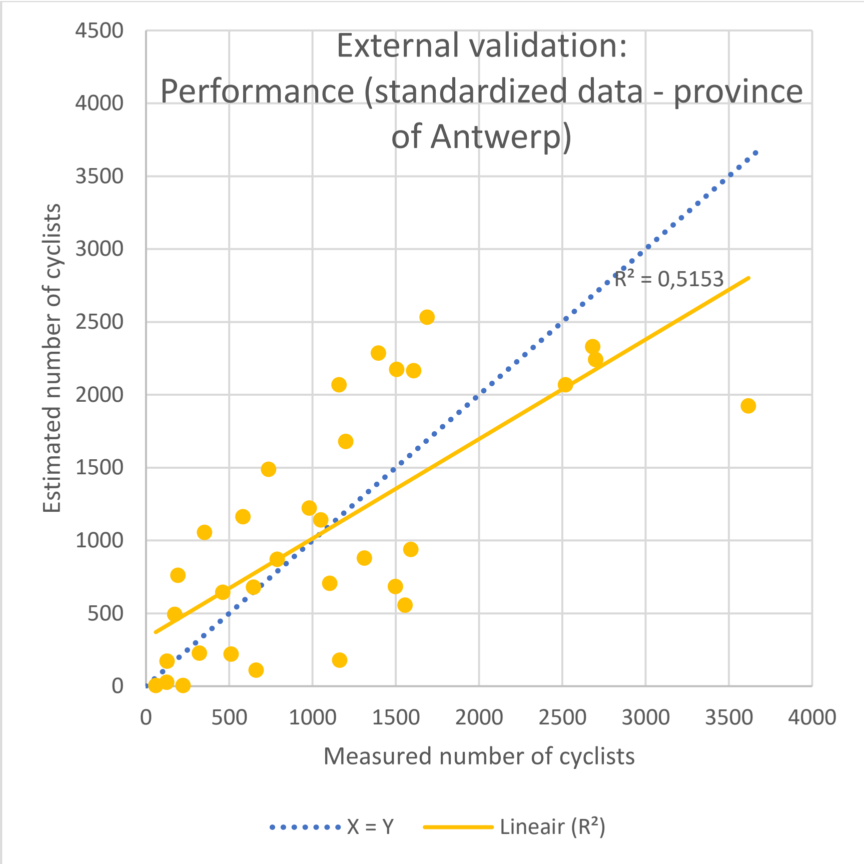


Figure 16. Scatter plot of model performance.

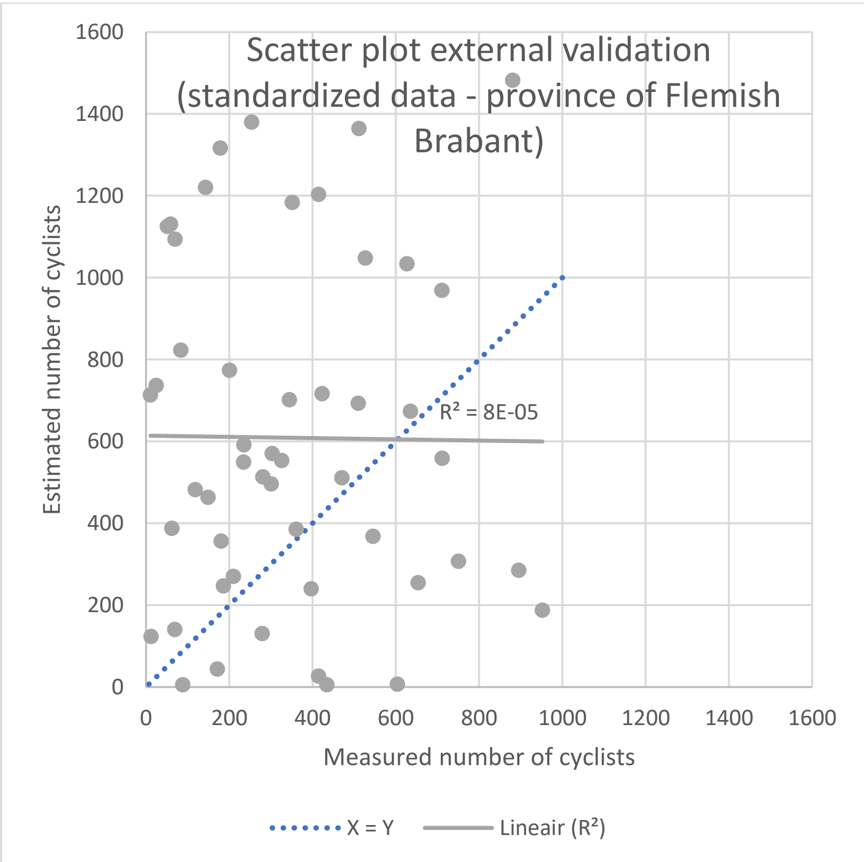


Figure 17. Scatter plot of model transferability.

7 Discussion

This study applied the land use regression modeling method to bicycle volume counts. Permanent and temporary bicycle counting point data from the province and city of Antwerp were used as dependent variables, while land use data from seven sources were used as independent variables. Supervised linear forward regression was used to develop a linear regression model that predicts the number of bicycles. A model with five variables was obtained:

$$N = 492.1 + 0.1341(nbuildings1000) - 0.1160(l\_hwy1000) + 0.1507(l\_cycleway500) + 87.71(lu2000) - 36530\,(reg\_partner500)$$

The number of cyclists at any given point is formed by an intercept, the number of buildings within 1000 m of a counting point, the length of highways within 1000 m, the length of cycleways within 500 m, the number of different types of land use within 2000 m, and the weighted average percentage of registered partnerships within 500 m of a counting point.

This model had an R^2 of 0.74. Leave-one-out cross validation resulted in an R^2 of 0.58, and external validation yielded an R^2 of 0.41. External validation was split into performance (using only the counting points in the province of Antwerp), which resulted in an R^2 of 0.52, and transferability (counting points in the province of Flemish Brabant), with an R^2 of 0.00008. Although the model performs good, it lacks transferability.

7.1 Methodology

Two major sets of data were used in this study. The first set, the dependent variables, consisted of the bicycle counts, collected through permanent and temporary counting points. The used counting methods were optical fiber technology and counting hoses, respectively. These methods are reliable, yet counting hoses have an important limitation, being that they have to be installed at an off-street path to prevent motorized traffic from being counted as well. In this study, the geographical spread of counting points over the province of Antwerp could have been better. There is a gap in the central north, south and the south east areas of the province. That is why X and Y coordinates were removed from the variables list, they were considered proxy variables for urbanization (El Esawey et al., 2015; Hyde-Wright et al., 2014).

The second data set consisted of the independent land use variables. 20 of these, 94 if buffer distances are considered separately, were retrieved from OSM. Since OSM data are entirely provided by volunteers and thus not verified by any governmental instance, this raised some accuracy questions. Several studies have found however that OSM data sets are usually complete and accurate. They can be used for research purposes. Only in Karlsruhe in 2019, OSM data appeared not to be suited to model work-related trips in a travel demand model. With this remark in mind, close attention was paid during the data analyses. By doing so, it was found that OSM was not useable to indicate all traffic lights in the study area.

(Basiri et al., 2016; Briem et al., 2019; Kloog et al., 2018; OpenStreetMap, 2019; M. Wang et al., 2013).

Another set of independent (land use related) variables were sociodemographic data. They were collected through Statbel, which is the Belgian statistical office. These are 2011 census data, which means they were nearly ten years old at the time of this study. The assumption was made that these data were still representative for the study area, in relative terms if not in absolute terms. Some slight changes in sociodemographic characteristics can be expected, but this would not significantly alter the results (Statbel, 2020a).

Buffer sizes were chosen subjectively. To make sure all relevant effects were included in the model, multiple buffer sizes were used. The maximum buffer size was set to 6000 m, as variables beyond this distance were considered not to be relevant for bicycle volumes. Other studies used similar buffer sizes, but some used even smaller ones. Air pollution studies that use the LUR method sometimes use buffers with a radius of 50 or 40 m. In this study, the smallest buffer had a radius of 100 m (Gilbert et al., 2012; Hochadel et al., 2006; Ross et al., 2006).

The independent variables were cleaned according to two correlation requirements. Individual variables should at least differ 5% in what they measure. In a category, they should differ at least 40%. These requirements were determined to make sure all variables were minimally distinct from one another.

The initial goal for external validation was to apply the model to counting points spread all over Flanders. This would have been possible, since as Figure 18 shows, the data set contained counting points in each province. Yet to prevent coincidental variability in the counts, only those points were chosen with counts for at least 14 days. As a result, only counting points in the provinces of Antwerp and Flemish Brabant remained for external validation.

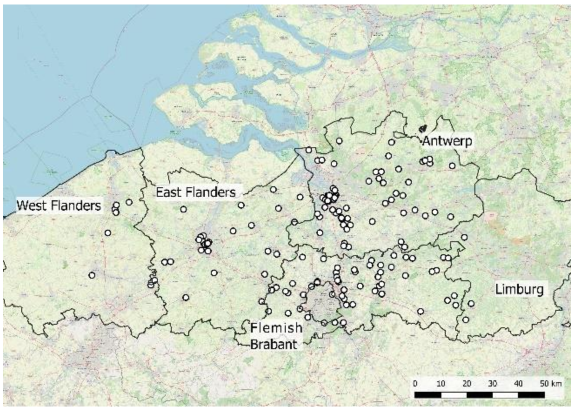


Figure 18. All counting points from Dataplatform Fiets

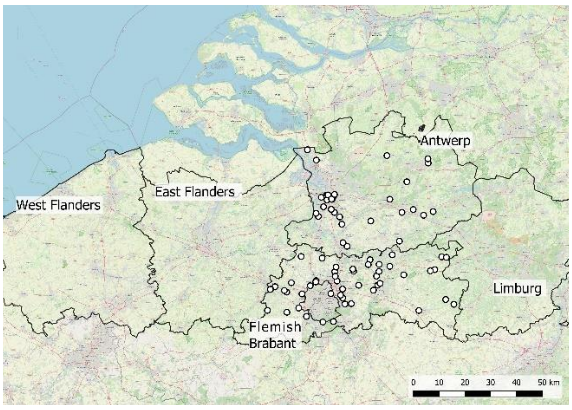


Figure 19. Counting points that counted 14 days or more.

7.2 Results

The model consists of five variables. Apart from the weighted averaged percentage of registered partnerships, all variables are distinct land use variables. These registered partnerships are a somewhat unexpected variable. Registered partnerships represent the number of people engaged in an act of legally cohabiting (Statbel, 2020b). In this model, it is possible that they are a proxy variable for another BE characteristic, however there are no strong correlations with other variables in the data set. The number of buildings on the other hand correlates strongly (around 80%) with the average population density, for example.

After the sensitivity analysis, the decision was made to keep the model the same, thus including point 29, that was close to the Dutch border. Two arguments support this decision. First, the number of available counting points was rather low, with only 37 counting points while the literature suggests a range of 20 to 100 points. Leaving out a point was not desirable. Second, the sensitivity analysis shows that the model did not drastically change. Again, ten iterations lead to the same nine variables. After checking for significance, four of the five remaining variables are the same as in the original model. The variables that were different in both models, poi1000 as opposed to nbuildings1000, have a correlation of 53.83% with each other. Yet, there are also arguments against this decision. Right now, the model is developed using a point for which 2/3 of each variable is unknown. If data layers for the Netherlands would have been available, the model's parameters could have been different.

The original model has an R^2 of 0.74, which means it explains 74% of the variation in the bicycle count data it was developed with. Compared to other model's R^2s, this is a rather high value. The Santa Monica bicycle model is a direct demand regression model based on observed counts. It was developed for the City of Santa Monica, near Los Angeles, California. This model estimates the evening peak bicycle and pedestrian volumes at an intersection and the bicycle model consists of an intercept and four variables, being the employment density, land use mix, proximity to bike routes and if an intersection is four-way. These variables are very similar to the ones in the model in Table 9. The R^2 of the Santa Monica model is 0.401. Another example is the Seamless bicycle model, which has an R^2 of 0.439. Its study area was San Diego, California. This study aimed to develop a model that predicts bicycle volumes at intersections during the morning peak period (7 to 9 A.M.). Similarly to the model in this study and the Santa Monica bicycle model, it was based on manual counts taken at 80 intersections. This model only has two variables, being the length of bicycle path and the employment density, both within a quarter of a mile of the intersection. Finally, Hankey et al reported an R^2 of 0.52. The direct demand model was developed using data from 101 locations in Blacksburg, Virginia. Again, both pedestrian and bicycle volumes were predicted. The bicycle model consists of five variables, being centrality, major roads, population density, on-street facilities and household incomes, which was negatively associated with bicycle volumes (Hankey et al., 2017; Transportation Research Board of the National Academies, 2014).

Both leave-one-out cross validation and an external validation were performed. The leave-one-out cross validation, with an R^2 of 0.58, did not result in any unexpected or strange values. There was one negative value, because of the high amount of highway meters. This warrants for the use of a cut off value to prevent estimates of bicycle counts getting negative. For example, Henderson et al. (2007) propose to truncate values to 120% of the maximum registered value. The same method could be applied at the lower end of the distribution (Henderson et al., 2007).

External validation was performed using different data than the ones the model was developed with. The external validation yielded an R^2 of 0.41. This is a drop of almost 30% compared to the original model, and more than 15% compared to the leave-one-out validation. External validation with raw, non-standardized data increased the R^2 from 0.40 to 0.41 and proved to be useful. Moreover, standardizing the bicycle counts was useful to include the evolution in bicycle rates from 2018 to 2019. Some outlying points probably contribute to this low R^2. An example is point 14 (ANT30) which counted 3617.78 cyclists, yet the model only estimated 1923.43, most probably because of the high amount of highway meters within 1000 m of this point. As can be seen in the scatter plot in Figures 14 and 15, this point pulls the R^2 away from the ideal 1-1 line. Another contribution to this R^2 is the point cloud that emerged at the bottom of the scatter plot in Figures 14 and 15. This cloud indicates points were the model either over- or underestimated the number of cyclists, and is one of the reasons of the model's bad transferability.

A first explanation for the external validation's R^2 could be found in the independent variables. It is possible that not all data were equally accurate. This could have caused the model to select some variables rather than other ones, resulting in other parameters. As stated in 7.1, socio-demographic data originate from 2011, while the bicycle counts are from 2019. This might cause some disturbance on the validity of the data. Another explanation could be the counting points themselves. Perhaps the counting points consist of two or more sub categories. 24 of the 83 counting points (28.92%) that were used in the external validation are situated on a bicycle highway. For model development, 12 of the 37 counting points (32.43%) were situated on a bicycle highway. Another sub category can be seen in the distinction between urban and rural counting points. The low R^2 of external validation could also be explained through the fact that some variables that were found in the literature were not available for the development of the model.

External validation in the case of this study was a combination of testing the model's performance, i.e. estimating bicycle counts on new locations in the province of Antwerp, and transferability, i.e. estimating bicycle counts in the province of Flemish Brabant. The performance test of the model was good, with an R^2 of 0.52. This confirms that the model performs very good in the province of Antwerp. The transferability of the model on the other hand, is very bad, with an R^2 of 0.00008. Some explanations can be found in the data set. The bicycle counts in Flemish Brabant are low compared to the ones in the province of Antwerp (see Table 17). This makes it harder to generate accurate estimates and as a result, a point cloud occurs, as can be seen in Figure 17. It is possible that, if some counting points with higher numbers of cyclists would have been included in the data set, it would yield a better R^2. Moreover, the ratio between highway length and cycleway length differs from the one in the province of Antwerp. In Antwerp, there is on average only 600 m of highway for each kilometer of cycleway near a counting point, while in Flemish Brabant this is 940 m.

The model developed in this paper can not be transferred to another study area than the province of Antwerp. The Transportation Research Board of the National Academies states that if a model has to be used in another study area, the parameters should be re-estimated. This was not done in this study, since the parameters from Table 9 were used to estimate the number of cyclists for the 83 counting points in the provinces of Antwerp and Flemish Brabant (Transportation Research Board of the National Academies, 2014).

7.3 Comparison with other bicycle count prediction methods

A report by the Transportation Research Board of the National Academies (2014) lists a number of methods to predict or estimate bicycle and pedestrian traffic. Two groups of methods exist. There are comprehensive four-step trip-based travel forecasting models. These models are choice based, which means they predict the choice cyclists will make. They are described in chapter 5.3. Depending on how large the TAZs are, these models will perform poorly regarding bicycle trips, as cyclists have a smaller range than car users. A large share of bicycle trips will probably happen within the same TAZ, leaving the accuracy low (Transportation Research Board of the National Academies, 2014).

The other group of models are the facility demand type models. They are count based and rely on existent data. Three types of models exist. Route choice models give quantitative information on how a cyclist weighs different characteristics of the environment against each other. The average cyclist would be willing to cycle one mile more on bike lanes if this avoids half a mile of mixed traffic, for example. Similarly to the above category of models, these models do not predict the number of bicycles as such, but the route these cyclists will take. Network simulation models simulate movement potentials between spaces and correlate them with counts. It is very difficult to evaluate the performance and validity of this type of models. Finally, the direct demand models use regression equations to model the number of cyclists. Typical performance values are already discussed in chapter 7.2, and show that the model developed in this chapter has a similar R^2 for external validation. The Transportation Research Board of the National Academies suggests that a smaller research area results in a better goodness of fit (Transportation Research Board of the National Academies, 2014).

An important note that has to be made is the fact that the model that was developed is a linear model. This means that the model would estimate the same number of additional cyclists, regardless of the current number of cyclists. A location with zero cyclists today would be given the same potential as that same location if it had already an average number of cyclists of 3000. The intercept intervenes a bit on this effect, but it should nevertheless be taken into account.

8 Implications

This study aimed to develop a model that is able to predict the number of cyclists at any given location in Flanders. To the best of the author's knowledge, it is the first time that this kind of model has been developed to predict bicycle volumes in Europe, and thus this study is of scientific relevance.

The model predicts the number of cyclists based on five distinct variables. This means that, if the number of buildings within 1000 m, the length of highway within 1000 m, the length of cycleways within 500 m, the number of different types of land use within 2000 m and the percentage of registered partnerships within 500 m of the counting point are known for any given point, the model is able to explain 74% of the variability in the number of cyclists. This rate could probably go up after finetuning of the model, e.g. dividing the counting points into different categories, as described in the discussion.

Due to the poor transferability of the model, it can only be used in the province of Antwerp, where it has a very good performance R^2.

A number of use cases is presented here to illustrate some of the contexts in which the model could be of support to policy makers in the province of Antwerp.

Use case 1. The model predicts that location X has 500 cyclists more than what is actually counted. This overestimation highlights the strong cycling potential of location X, yet certain reasons (e.g. bad infrastructure) cause cyclists to avoid this location. The model's prediction could then be an argument to invest in better cycling infrastructure on this location.

Use case 2. Policy makers want to compare two locations (A and B). Location A has a high number of bicycle accidents, while location B appears to be a lot safer, with noticeably fewer bicycle accidents. To perform a valid comparison, the number of cyclists at each location has to be included in the analysis since more cyclists means a higher chance for accidents. There are however no bicycle count data available for locations A and B. In this case, the model could estimate the number of cyclists on both locations, which could then be used to calculate the accident risk for both A and B.

Use case 3. A local government wants to stimulate bicycle use in a certain area of the city. Two scenarios have been composed, the first one allows new types of land use in the environment, while the second one consists of extensive investments in cycleways. The model could be used to predict the number of cyclists on several locations in the area in both scenarios. It helps to understand the results of certain policy measures.

9 Limitations and future research

Like the discussion suggests, this study has some limitations that should be taken into account and that future studies could overcome. First of all, the relatively low number of bicycle count sites could be considered a limitation of this study. Future studies that have enough means – both financial and in terms of time – should collect bicycle and land use data themselves, at a sufficient amount of locations. This has several advantages. It allows to select the number of locations according to the literature, with around 40 counting points at minimum, but up to 100 locations if the area is more diversified (Hankey et al., 2017; Le et al., 2018; Transportation Research Board of the National Academies, 2014). Moreover, it would enable to develop a more accurate model. Right now, the possibility exists for example that mopeds who were on the bicycle infrastructure were included in the bicycle count data as well. This would also allow to count for a long enough period and thus overcome the problem of only being able to use a limited number of counting points, as was the case with the external validation.

It is possible that the use of existing data also influenced the transferability of the model, resulting in a model that can only be used in the province of Antwerp. Future research could try to identify the reasons why the transferability of this model is so poor.

Another option is to develop a model using the *Dataplatform Fiets* data and to perform an external validation on the data provided by the province of Antwerp. If existing data have to be used for either model development or validation, other bicycle count initiatives could provide input data as well, such as *Straatvinken* or *Telraam* (Straatvinken, 2020; Telraam, 2019).

The same limitation can be seen with the independent variables. Not all data are 100% accurate and reliable. To make future research easier, governmental data layers that are easily accessible (open data) would be extremely valuable for this type of studies. CORINE land cover data is a good example of GIS-data that are reliable and easily accessible.

The model that was developed in this study includes the weighted average percentage of registered partnerships. This variable, although a result from the literature, might be a parameter for another BE characteristic. In the variables list, there may be additional variables that are proxies for other BE characteristics. Future research could take a closer look at these variables and make sure each variable is unique.

In the analysis, his study did not make a distinction between different types of counting points. Yet, this distinction was subtly present. A distinction can be made between bicycle highways, rural locations and urban locations. It would be interesting to have a look at these types of counting points separately and develop models for each kind. This would probably improve the performance of the model and increase the R^2. Other model variants can be made as well. It would be interesting to separate weekend- and weekdays from each other, as weekdays know a lot of commuter traffic while during weekends, people cycle more often as a way of recreation. This distinction could lead to other variables being significant, making the model more accurate. Similarly, a model for peak hours and off-peak hours could be developed.

Another remark made in the discussion were some extreme points, such as point 14. They have an influence on how R^2 is formed and therefore an additional sensitivity analysis without the extreme points can give more insight in how the model reacts.

10 Conclusions

A land use regression model can be used to predict bicycle traffic volumes in Flanders, based on built environment (BE) characteristics as predictor variables and if the model is applied in the same study area as it was developed in.

A literature review in 30 papers showed that 107 BE characteristics can have a possible influence on bicycle use. Built environment was interpreted broadly, so that it also included sociodemographic variables. These variables can be found in Tables 1 and 2 in chapter 5.2. Of these variables, 51 were available for this study in the province of Antwerp and Flanders. They are listed in Table 3 in chapter 5.2. The direction of effect of each of the variables was taken from the literature as well, although sometimes the literature was inconclusive about the effects.

There are some differences between a land use regression model and a traditional four-step travel demand model. The latter predicts traffic volumes based on route choice of the users of the network. This type of model consists of four unique steps and uses existing travel patterns as input for model estimation. A four-step model works with Traffic Analysis Zones (TAZs), which causes some spatial aggregation. This is an important difference between this type of model and a land use regression model. Four-step models rely on TAZs and are therefore spatially rather aggregated, whereas land use regression models usually have a much better spatial resolution.

The land use regression model in this study was developed through different steps. First, data cleaning and preparation for both dependent (bicycle counts) and independent (land use variables) data was performed. Bicycle count data for each of the 37 counting points were standardized to obtain daily values, averaged over a whole year. Land use variables were calculated for each counting point and thoroughly checked for correlation. This resulted in 112 initial land use variables. Both sets of data were then put together to develop the model using supervised forward linear regression, which resulted in a linear regression equation with five variables. The number of bicycles (N) can be predicted by using the following equation:

$$N = 492.1 + 0.1341(nbuildings1000) - 0.1160(l\_hwy1000) + 0.1507(l\_cycleway500) + 87.71(lu2000) - 36530\,(reg\_partner500)$$

These variables represent the number of buildings within 1000 m of a counting point, the length of highways within 1000 m, the length of cycleways within 500 m, the number of different types of land use within 2000 m, and the weighted average percentage of registered partnerships within 500 m of a counting point, respectively. The R^2 of this model is 0.74. It always predicts the daily number of cyclists in both directions, averaged over a whole year.

The model’s performance was checked through a sensitivity analysis, a leave-one-out cross validation and an external validation. The sensitivity analysis showed that when the model was redeveloped with 36 instead of 37 counting points (leaving out one counting point close to the Dutch border), this did not drastically change the model. Leave-one-out validation of this model happened with the same data it was developed with. Leaving out one point at a time, estimating the coefficients again and then calculating the number of cyclists for the left-out point, yielded an R^2 of 0.58. For external validation, 83 counting points from an external data set were used. The model was applied to predict the number of cyclists for each one of these counting points, which resulted in an R^2 of 0.41. This R^2 value is similar to or slightly lower than R^2 of similar studies.

When external validation is split into performance and transferability, the model performs very good in the area it was developed in (province of Antwerp), with an R^2 of 0.52. Transferability on the other hand is remarkably poor, with an R^2 of only 0.00008. As a result, the model can only be used in the province of Antwerp.

This study is of societal and scientific relevance. The societal relevance lies in the fact that policy decisions regarding bicycle infrastructure investments can be supported by applying this model. The output of the model can also be used to identify areas that need attention, for example areas where the actual number of cyclists is lower than what would be predicted by the model, as this could indicate a bicycle potential that is not fully used. The model could also help identify land use interventions that would increase the number of cyclists.

Scientific relevance can be found in the fact that this study is the first one to apply the land use regression method to bicycle counts in, to the best of the authors knowledge, Europe. This study could be a starting point for future studies in Flanders and Europe that could finetune the model and explore the possibilities of using land use regression for bicycle or other traffic related purposes.

Acknowledgements

Special thanks to Steven Soetens from the Province of Antwerp and Jelle De Keyser from the city of Antwerp for providing the bicycle count data. Thanks to their quick responses and clear explanation of the data set, collecting the data did not cause any delays. Also special thanks once again to dr. Evi Dons, co-promotor of this master's thesis, for opening up her office one week long and explaining the basics of both R and QGIS. Without this valuable knowledge, this master's thesis would have been much harder to perform.

References

Anderson, M. D., Sharfi, K., & Gholston, S. E. (2006). *Direct Demand Forecasting Model for Small Urban Communities Using Multiple Linear Regression*.

Aziz, H. M. A., Nagle, N. N., Morton, A. M., Hilliard, M. R., White, D. A., & Stewart, R. N. (2018). Exploring the impact of walk–bike infrastructure, safety perception, and built-environment on active transportation mode choice: A random parameter model using New York City commuter data. *Transportation*, *45*(5), 1207–1229. https://doi.org/10.1007/s11116-017-9760-8

Basiri, A., Jackson, M., Amirian, P., Pourabdollah, A., Sester, M., Winstanley, A., Moore, T., & Zhang, L. (2016). Quality assessment of OpenStreetMap data using trajectory mining. *Geo-Spatial Information Science*, *19*(1), 56–68. https://doi.org/10.1080/10095020.2016.1151213

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.-Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K. T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., … de Hoogh, K. (2013). Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmospheric Environment*, *72*, 10–23. https://doi.org/10.1016/j.atmosenv.2013.02.037

Blue-bike. (2019). *Blue-bike*. https://www.blue-bike.be/

Boniardi, L., Dons, E., Campo, L., Van Poppel, M., Int Panis, L., & Fustinoni, S. (2019). Is a Land Use Regression Model Capable of Predicting the Cleanest Route to School? *Environments*, *6*(8), 90. https://doi.org/10.3390/environments6080090

Briem, L., Heilig, M., Klinkhardt, C., & Vortisch, P. (2019). Analyzing OpenStreetMap as data source for travel demand models A case study in Karlsruhe. *Transportation Research Procedia*, *41*, 104–112. https://doi.org/10.1016/j.trpro.2019.09.021

Briggs, D. J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., Pryl, K., Reeuwijk, H. van, Smallbone, K., & Veen, A. van de. (1997). Mapping urban air pollution using GIS: A regression-based approach. *International Journal of Geographical Information Science*, *11*(7), 699–718. https://doi.org/10.1080/136588197242158

Broberg, A., & Sarjala, S. (2015). School travel mode choice and the characteristics of the urban built environment: The case of Helsinki, Finland. *Transport Policy*, *37*, 1–10. https://doi.org/10.1016/j.tranpol.2014.10.011

Buehler, R., & Dill, J. (2016). Bikeway Networks: A Review of Effects on Cycling. *Transport Reviews*, *36*(1), 9–27. https://doi.org/10.1080/01441647.2015.1069908

Buehler, R., & Pucher, J. (2012). Cycling to work in 90 large American cities: New evidence on the role of bike paths and lanes. *Transportation*, *39*(2), 409–432. https://doi.org/10.1007/s11116-011-9355-8

Cervero, R., Denman, S., & Jin, Y. (2019). Network design, built and natural environments, and bicycle commuting: Evidence from British cities and towns. *Transport Policy*, *74*, 153–164. https://doi.org/10.1016/j.tranpol.2018.09.007

Chen, P., Zhou, J., & Sun, F. (2017). Built environment determinants of bicycle volume: A longitudinal analysis. *Journal of Transport and Land Use*, *10*(1), 655–674. JSTOR.

Clifton, K. J., Singleton, P. A., Muhs, C. D., & Schneider, R. J. (2016). Representing pedestrian activity in travel demand models: Framework and application. *Journal of Transport Geography*, *52*, 111–122. https://doi.org/10.1016/j.jtrangeo.2016.03.009

Cui, Y., Mishra, S., & Welch, T. F. (2014). Land use effects on bicycle ridership: A framework for state planning agencies. *Journal of Transport Geography*, *41*, 220–228. https://doi.org/10.1016/j.jtrangeo.2014.10.004

Dataplatform fiets. (2019). *Welkom—Dataplatform fiets*. https://dataplatformfiets.be/

Dons, E., Van Poppel, M., Kochan, B., Wets, G., & Int Panis, L. (2013a). Modeling temporal and spatial variability of traffic-related air pollution: Hourly land use regression models for black carbon. *Atmospheric Environment*, *74*. https://doi.org/10.1016/j.atmosenv.2013.03.050

Dons, E., Van Poppel, M., Kochan, B., Wets, G., & Int Panis, L. (2013b). Modeling temporal and spatial variability of traffic-related air pollution: Hourly land use regression models for black carbon. *Atmospheric Environment*, *74*. https://doi.org/10.1016/j.atmosenv.2013.03.050

El Esawey, M., Lim, C., & Sayed, T. (2015). Development of a cycling data model: City of Vancouver case study. *Canadian Journal of Civil Engineering*, *42*(12), 1000–1010. https://doi.org/10.1139/cjce-2015-0065

Fietsberaad Vlaanderen. (2019). Conclusies. In *Fietstelweek 2018 Eindrapport versie 2* (p. 34). Fietsberaad Vlaanderen. https://www.fietstelweek.be/public/docs/eindrapport-2018.pdf?v=20190506

Fietssnelwegen. (2019). *Fietssnelwegen*. https://fietssnelwegen.be

Gao, J., Kamphuis, C. B. M., Dijst, M., & Helbich, M. (2018). The role of the natural and built environment in cycling duration in the Netherlands. *International Journal of Behavioral Nutrition and Physical Activity; London*, *15*. http://dx.doi.org/10.1186/s12966-018-0715-z

Gilbert, N. L., Goldberg, M. S., Beckerman, B., Brook, J. R., & Jerett, M. (2012). *Assessing Spatial Variability of Ambient Nitrogen Dioxide in Montréal, Canada, with a Land-Use Regression Model: Journal of the Air & Waste Management Association: Vol 55, No 8*. https://www.tandfonline.com/doi/abs/10.1080/10473289.2005.10464708

Hankey, S., Lu, T., Mondschein, A., & Buehler, R. (2017). Spatial models of active travel in small communities: Merging the goals of traffic monitoring and direct-demand modeling. *Journal of Transport & Health*, *7*, 149–159. https://doi.org/10.1016/j.jth.2017.08.009

Heesch, K. C., Giles-Corti, B., & Turrell, G. (2015). Cycling for transport and recreation: Associations with the socio-economic, natural and built environment. *Health & Place*, *36*, 152–161. https://doi.org/10.1016/j.healthplace.2015.10.004

Heinen, E., van Wee, B., & Maat, K. (2010). Commuting by Bicycle: An Overview of the Literature. *Transport Reviews*, *30*(1), 59–96. https://doi.org/10.1080/01441640903187001

Henderson, S. B., Beckerman, B., Jerett, M., & Brauer, M. (2007). *Application of Land Use Regression to Estimate Long-Term Concentrations of Traffic-Related Nitrogen Oxides and Fine Particulate Matter | Environmental Science & Technology*. https://pubs.acs.org/doi/abs/10.1021/es0606780#

Hochadel, M., Heinrich, J., Gehring, U., Morgenstern, V., Kuhlbusch, T., Link, E., Wichmann, H.-E., & Krämer, U. (2006). Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information. *Atmospheric Environment*, *40*(3), 542–553. https://doi.org/10.1016/j.atmosenv.2005.09.067

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, *42*, 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057

Hunt, J. D., & Abraham, J. E. (2007). Influences on bicycle use. *Transportation*, *34*(4), 453–470. https://doi.org/10.1007/s11116-006-9109-1

Hyde-Wright, A., Graham, B., & Nordback, K. (2014). *Counting Bicyclists with Pneumatic Tube Counters on Shared Roadways—ProQuest*. http://search.proquest.com/docview/1498090371?pq-origsite=summon

Int Panis, L. (2018). *Environmental economic evaluation of transport policy HC2: Air Quality & Land Use Regression*.

Kirner Providelo, J., & da Penha Sanches, S. (2011). Roadway and traffic characteristics for bicycling. *Transportation*, *38*(5), 765. https://doi.org/10.1007/s11116-011-9353-x

Kloog, I., Kaufman, L. I., & Hoogh, K. de. (2018). Using Open Street Map Data in Environmental Exposure Assessment Studies: Eastern Massachusetts, Bern Region, and South Israel as a Case Study. *International Journal of Environmental Research and Public Health*, *15*(11), 2443. https://doi.org/10.3390/ijerph15112443

Le, H. T., Buehler, R., & Hankey, S. (2018). Correlates of the Built Environment and Active Travel: Evidence from 20 US Metropolitan Areas. *Environmental Health Perspectives (Online); Research Triangle Park*, *126*(7). http://dx.doi.org/10.1289/EHP3389

Mertens, L., Compernolle, S., Deforche, B., Mackenbach, J. D., Lakerveld, J., Brug, J., Roda, C., Feuillet, T., Oppert, J.-M., Glonti, K., Rutter, H., Bardos, H., De Bourdeaudhuij, I., & Van Dyck, D. (2017). Built environmental correlates of cycling for transport across Europe. *Health & Place*, *44*, 35–42. https://doi.org/10.1016/j.healthplace.2017.01.007

Mertens, L., Van Dyck, D., Ghekiere, A., De Bourdeaudhuij, I., Deforche, B., Van de Weghe, N., & Van Cauwenberg, J. (2016). Which environmental factors most strongly influence a street's appeal for bicycle transport among adults? A conjoint study using manipulated photographs. *International Journal of Health Geographics*, *15*(1). https://doi.org/10.1186/s12942-016-0058-4

Mobit. (2019). *Mobit – Slim en flexibel fietsdelen*. http://www.mobit.eu

Moudon, A. V., Lee, C., Cheadle, A. D., Collier, C. W., Johnson, D., Schmid, T. L., & Weather, R. D. (2005). Cycling and the built environment, a US perspective. *Transportation Research Part D: Transport and Environment*, *10*(3), 245–261. https://doi.org/10.1016/j.trd.2005.04.001

OpenStreetMap. (2019). *OpenStreetMap*. OpenStreetMap. https://www.openstreetmap.org/

Osama, A., Sayed, T., & Bigazzi, A. Y. (2017). Models for estimating zone-level bike kilometers traveled using bike network, land use, and road facility variables. *Transportation Research Part A: Policy and Practice*, *96*, 14–28. https://doi.org/10.1016/j.tra.2016.11.016

Owen, N., De De Bourdeaudhuij, I., Sugiyama, T., Leslie, E., Cerin, E., Van Van Dyck, D., & Bauman, A. (2010). Bicycle Use for Transport in an Australian and a Belgian City: Associations with Built-Environment Attributes. *Journal* *of Urban Health*, *87*(2), 189–198. https://doi.org/10.1007/s11524-009-9424-x

Park, K., Sabouri, S., Lyons, T., Tian, G., & Ewing, R. (2019). Intrazonal or interzonal? Improving intrazonal travel forecast in a four-step travel demand model. *Transportation*. https://doi.org/10.1007/s11116-019-10002-0

Parkin, J., Wardman, M., & Page, M. (2008). Estimation of the determinants of bicycle mode share for the journey to work using census data. *Transportation*, *35*(1), 93–109. https://doi.org/10.1007/s11116-007-9137-5

Piatkowski, D. P., & Marshall, W. E. (2015). Not all prospective bicyclists are created equal: The role of attitudes, socio-demographics, and the built environment in bicycle commuting. *Travel Behaviour and Society*, *2*(3), 166–173. https://doi.org/10.1016/j.tbs.2015.02.001

Provincie Antwerpen. (2014). *Provinciale fietsbarometer—Fietstellingen in een integraal fietsbeleid*.

Provincie Antwerpen. (2019a). *Fietsbarometer*. Fietsbarometer. http://fietsbarometer.provincieantwerpen.be/geoloketten/?viewer=fietsbarometer

Provincie Antwerpen. (2019b, March 27). *Op 1 jaar tijd 14% meer fietsers in Antwerpen Fietsprovincie—Provincie Antwerpen*. https://www.provincieantwerpen.be/nieuws.masterdetail.html/p_detail_url/nl/drem/dienst-mobiliteit/nieuws/fietsbeleid/20190327_fietsbarometer.html

Pucher, J., Dill, J., & Handy, S. (2010). Infrastructure, programs, and policies to increase bicycling: An international review. *Preventive Medicine*, *50*, S106–S125. https://doi.org/10.1016/j.ypmed.2009.07.028

Rietveld, P., & Daniel, V. (2004). Determinants of bicycle use: Do municipal policies matter? *Transportation Research Part A: Policy and Practice*, *38*(7), 531–550. https://doi.org/10.1016/j.tra.2004.05.003

Ross, Z., English, P. B., Scalf, R., Gunier, R., Smorodinsky, S., Wall, S., & Jerrett, M. (2006). Nitrogen dioxide prediction in Southern California using land use regression modeling: Potential for environmental health analyses. *Journal of Exposure Science & Environmental Epidemiology*, *16*(2), 106–114. https://doi.org/10.1038/sj.jea.7500442

Ryan, P. H., & LeMasters, G. K. (2006). *A Review of Land-use Regression Models for Characterizing Intraurban Air Pollution Exposure: Inhalation Toxicology: Vol 19, No sup1*. https://www.tandfonline.com/doi/full/10.1080/08958370701495998?scroll=top&needAccess=true

Sallis, J. F., Conway, T. L., Dillon, L. I., Frank, L. D., Adams, M. A., Cain, K. L., & Saelens, B. E. (2013). Environmental and demographic correlates of bicycling. *Preventive Medicine*, *57*(5), 456–460. https://doi.org/10.1016/j.ypmed.2013.06.014

Sarjala, S. (2019). Built environment determinants of pedestrians’ and bicyclists’ route choices on commute trips: Applying a new grid-based method for measuring the built environment along the route. *Journal of Transport Geography*, *78*, 56–69. https://doi.org/10.1016/j.jtrangeo.2019.05.004

Schoner, J. E., Cao, J., & Levinson, D. M. (2015). Catalysts and magnets: Built environment and bicycle commuting. *Journal of Transport Geography*, *47*, 100–108. https://doi.org/10.1016/j.jtrangeo.2015.07.007

Statbel. (2020a). *About Statbel | Statbel*. https://statbel.fgov.be/en/about-statbel

Statbel. (2020b). *Census 2011*. https://www.census2011.be/data/fresult/highdegree_nl.html

Straatvinken. (2020). *Hoe gezond is het verkeer in jouw straat?* Straatvinken. https://straatvinken.be/

Sun, Y., Du, Y., Wang, Y., & Zhuang, L. (2017). Examining Associations of Environmental Characteristics with Recreational Cycling Behaviour by Street-Level Strava Data. *International Journal of Environmental Research and Public Health; Basel*, *14*(6), 644. http://dx.doi.org/10.3390/ijerph14060644

Telraam. (2019). *Telraam*. Telraam. https://telraam.net/nl/what-is-telraam

Transportation Research Board of the National Academies. (2014). *Estimating Bicycling and Walking for Planning and Project Development: A Guidebook* (p. 22330). Transportation Research Board. https://doi.org/10.17226/22330

VAB. (2019, February 15). *We fietsen vaker naar de winkel en het werk | VAB-Magazine*. https://magazine.vab.be/op-weg/mobiliteit/we-fietsen-vaker-naar-de-winkel-en-het-werk/

Vandenbulcke, G., Dujardin, C., Thomas, I., Geus, B. de, Degraeuwe, B., Meeusen, R., & Panis, L. I. (2011). Cycle commuting in Belgium: Spatial determinants and 're-cycling' strategies. *Transportation Research Part A:*

Policy and Practice, *45*(2), 118–137.
https://doi.org/10.1016/j.tra.2010.11.004

Vlaamse Regering. (2019). *Regeerakkoord van de Vlaamse Regering*.
https://www.vlaanderen.be/vlaamse-regering/regeerakkoord-van-de-vlaamse-regering

Wang, M., Li, Q., Hu, Q., & Zhou, M. (2013). QUALITY ANALYSIS OF OPEN STREET MAP DATA. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-2/W1*, 155–158.
https://doi.org/10.5194/isprsarchives-XL-2-W1-155-2013

Wang, Meng, Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., & Brunekreef, B. (2012). *Systematic Evaluation of Land Use Regression Models for NO2 | Environmental Science & Technology*.
https://pubs.acs.org/doi/abs/10.1021/es204183v

Weliwitiya, H., Rose, G., & Johnson, M. (2019). Bicycle train intermodality: Effects of demography, station characteristics and the built environment. *Journal of Transport Geography*, *74*, 395–404.
https://doi.org/10.1016/j.jtrangeo.2018.12.016

Winters, M., Brauer, M., Setton, E. M., & Teschke, K. (2010). Built Environment Influences on Healthy Transportation Choices: Bicycling versus Driving. *Journal of Urban Health*, *87*(6), 969–993.
https://doi.org/10.1007/s11524-010-9509-6

Yang, Y., Wu, X., Zhou, P., Gou, Z., & Lu, Y. (2019). Towards a cycling-friendly city: An updated review of the associations between built environment and cycling behaviors (2007–2017). *Journal of Transport & Health*, *14*, 100613. https://doi.org/10.1016/j.jth.2019.100613

Zahabi, S. A. H., Chang, A., Miranda-Moreno, L. F., & Patterson, Z. (2016). Exploring the link between the neighborhood typologies, bicycle infrastructure and commuting cycling over time and the potential impact on commuter GHG emissions. *Transportation Research Part D: Transport and Environment*, *47*, 89–103. https://doi.org/10.1016/j.trd.2016.05.008

Zhao, P. (2013). The Impact of the Built Environment on Bicycle Commuting: Evidence from Beijing. *Urban Studies*, *51*, 1019–1037. https://doi.org/10.1177/0042098013494423

Appendix

Report feedback meeting February 17^{th}

Feedback part 1

A few remarks were made concerning the first part of this master's thesis. First, the final report should contain an evaluation of the model's performance compared to other methods to predict bicycle counts. This evaluation could be integrated in the Discussion part. It could concern the model's R^2, but it is possible that not all prediction methods use an R^2. Consequently, an elaboration of how other method's performances are evaluated is possible as well. The fact that R^2 depends on the validation of the model has to be taken into account.

The second remark is that the study in this master's thesis uses an already-existing method, namely the LUR-method. This could mean that the study is not as innovative as other studies. The innovative aspect of this study can be found in the fact that this study applies the LUR method to bicycle counts instead of air pollution. Nevertheless, close attention will have to be paid when analyzing and interpreting the results to make sure the paper adds value to the literature.

Approach part 2

The time schedule for the second part of this master's thesis is achievable. Certain points of attention were mentioned.

The bicycle count data were collected from both temporal and permanent counting points. Because the temporal counters only registered cyclists during a limited period of the year certain irregularities may exist in the data set. Therefore, the temporal bicycle count data have to be rescaled. After rescaling, the average number of cyclists per day will be as representative for the temporal counting points as it is for the permanent ones.

For the development of the model, R will be used. Variables will be calculated using QGIS. Two counting points are located near the Dutch border. This could cause some problems developing the model, as variables are usually only available for Belgium or Flanders. Two solutions are possible: either these counting points will not be used for the development of the model, or additional data layers that cover the Netherlands will be collected to be able to properly calculate the variables for these counting points.

The model will be based on a linear regression equation. To develop this equation, certain conditions have to be met, for example both the predictor and the bicycle count variables have to be checked for multicollinearity. Every choice made during the study should be motivated, for example based on literature.

When an initial model is developed, alternative forms of the model could be looked at. A distinction could be made between bicycle counts on week and weekend days, for example.

Finally, the link with the real world could be made by meeting with the province and/or the city of Antwerp. They provided the bicycle data and are very curious as to what the results of the study will be. They can also give useful feedback concerning the model and the applied methods.

Used variables and their meanings

Table 18. Used variables.

Variable	Meaning	Buffer sizes (radii - m)	Source	Note
N	Average number of cyclists per day			Dependent variable.
X	X-coordinate			
Y	Y-coordinate			
L_cycleway	Length of cycleways	100, 300, 500, 1000	OSM	
brdg_	Presence of bike bridges	100, 300, 500, 1000	OSM	Value 0 or 1.
tun	Presence of bike tunnels	100, 300, 500, 1000	OSM	Idem.
zp	Number of *Zwarte punten*	100, 300, 500, 1000	AWV	*Zwarte punten* (black locations) are locations with the most severe accident rate.
park	Area of land use defined as 'park'	100, 300, 500, 1000, 2000, 4000, 6000	OSM	
ind	Area of industrial land use	100, 300, 500, 1000, 2000, 4000, 6000	OSM	
resi	Area of residential land use	100, 300, 500, 1000	OSM	
station	Number of railway stations	100, 300, 500, 1000, 2000, 4000, 6000	OSM	
avg_popdens	Weighted average population density	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Weighted by area of each municipality within the buffer.
agri	Area of agricultural land	100, 300, 500, 1000	Corine	
for	Area of forest	100, 300, 500, 1000	Corine	
lowveg	Area of land with low vegetation	100, 300, 500, 1000	Corine	
water	Area of water bodies	100, 300, 500, 1000	Corine	
parking	Area of parking space	100, 300, 500, 1000	OSM	
parking_b	Number of parking buildings	300, 500, 1000	OSM	Includes underground parking.
traff_l	Number of traffic lights	100, 300, 500, 1000	OSM	
ped	Length of pedestrian-only streets	100, 300, 500, 1000	OSM	
sch	Number of schools	300, 500, 1000	OSM	
h_edu	Number of colleges and universities	1000, 2000, 4000, 6000	OSM	
poi	Number of points of interest	100, 300, 500, 1000	OSM	

Table 18 (continued).

Variable	Meaning	Buffer distances (m)	Source	Note
jobs	Number of points of interest that create jobs	100, 300, 500, 1000	OSM	
int	Number of intersections	100, 300, 500, 1000, 2000, 4000	OSM	Except motorway intersections.
avg_v	Average speed limit	100, 300, 500, 1000, 2000, 4000, 6000	AWV	
l_hwy	Length of highways	100, 300, 500, 1000, 2000, 4000, 6000	AWV	
prim	Length of primary roads	100, 300, 500, 1000, 2000, 4000, 6000	AWV	
sec	Length of secondary roads	100, 300, 500, 1000, 2000, 4000, 6000	AWV	
velo	Number of *Velo* bike sharing stations	100, 300, 500, 1000, 2000, 4000, 6000	Stad Antwerpen	
avg_inc	Weighted average yearly income	500, 1000	Statbel	2017 data, on municipal level. Weighted by area of each municipality within the buffer.
empl_r	Weighted average employment rate (%)	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	2011 data, on municipal level. Weighted by area of each municipality within in the buffer.
avg_hh	Weighted average household size	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
reg_partner	Weighted average percentage of registered partnerships	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem. People engaged in an act of legally cohabiting (Statbel, 2020b).
married	Weighted average marriage rate (%)	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
20hi_edu	Weighted average rate of inhabitants (20y or older) with a higher educational degree	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
r_femalemale	Weighted average female/male ratio	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
avg_age	Weighted average age	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.

Table 18 (continued).

Variable	Meaning	Buffer distances (m)	Source	Note
u15_	Weighted average percentage of population under 15 years	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	2011 data, on municipal level. Weighted by area of each municipality within the buffer.
15-65_	Weighted average percentage of population between 15 and 65	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
65plus_	Weighted average percentage of population of 65 years and older	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
born_abr	Weigthed average percentage of inhabitants born abroad	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
foreign	Weighted average percentage of foreign inhabitants	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
imm	Weighted average percentage of inhabitants immigrated after 1980	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
before1991_	Weighted average percentage of buildings built before 1991	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
after1991	Weighted average percentage of buildings built after 1991	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
after2001_	Weighted average percentage of buildings built after 1991	100, 300, 500, 1000, 2000, 4000, 6000	Statbel	Idem.
nbuildings	Number of buildings	100, 300, 500, 1000, 2000	OSM	
distnear_st	Distance to the nearest railway station		OSM	
awv_aanl	Length of adjacent cycle lanes	100, 300, 500, 1000, 2000, 4000, 6000	AWV	Only for AWV roads.
awv_aanlh	Length of adjacent elevated cycle lanes	100, 300, 500, 1000, 2000, 4000, 6000	AWV	Idem.
awv_vrijl	Length of separated cycle tracks	100, 300, 500, 1000, 2000, 4000, 6000	AWV	Idem.
awv_sugg	Length of shared lane markings	1000, 2000, 4000, 6000	AWV	Idem.
lu	Number of land use types	100, 300, 500, 1000, 2000, 4000, 6000	OSM	
roadlength	Combined length of all roads within the buffer	100, 300, 500, 1000, 2000, 4000, 6000	OSM	

Location counting points

Table 19. Location of counting points that were used to develop the model. They are all situated in the province of Antwerp. 'Area' translates as follows: U = Urban, R = Rural, F = bike highway. (Fietssnelwegen, 2019).

Nr	ID	Location	Area	X	Y
1	ANT01	Steenplein, Antwerp	U	51.219230	4.395599
2	ANT02	Mercatorstraat, Antwerp	U	51.209022	4.423149
3	ANT05	Gerard Le Grellelaan, Antwerp	U	51.190523	4.406125
4	ANT09	Londenbrug, Eilandje, Antwerp	U	51.231394	4.405468
5	ANT10	IJzerlaanbrug, Antwerp	F	51.235806	4.431231
6	ANT12	Statielei, Antwerp	U	51.216018	4.445766
7	AREGV07	Hertevelden, Arendonk	R	51.311713	5.092592
8	BEEGV10	Antwerpseweg, Beerse	R	51.301820	4.794931
9	BERGV333	Cycle bridge Berchem Stn., Berchem	U	51.197689	4.433259
10	BORGV04A	N16, Scheldedijk, Bornem	R	51.118442	4.218056
11	BORGV04B	N16, Temsebrug, Bornem	F	51.118442	4.218056
12	BRB01	De Robianostraat, Borsbeek	U	51.194450	4.483680
13	BRB02	Jozef Reussenslei, Borsbeek	U	51.197460	4.490080
14	DUF03	Fietsostrade F1, Duffel	F	51.083520	4.495640
15	DUF04	Oude Liersebaan, Duffel	R	51.107950	4.550640
16	GROGV09	Jaagpad Albertkanaal, Grobbendonk	F	51.183144	4.720932
17	HOOGV06	Lodewijk de Koninckln, Hoogstraten	R	51.394930	4.756819
18	HOVGV03	Fietsostrade F1, Hove	F	51.155999	4.463873
19	HRS01	Westerlosesteenweg, Herselt	R	51.081120	4.913730
20	HRS02	Bergstraat, Herselt	R	51.014928	4.831088
21	HRS03	Diestsebaan, Herselt	R	51.060750	4.925090
22	KAPGV02	Georges Spelierlaan, Kapellen	F	51.321115	4.437013
23	KAS01	Herentalsesteenweg, Kasterlee	R	51.215700	4.908620
24	KAS02	Zevendonkseweg, Kasterlee	R	51.260080	4.914620
25	KASGV18	Oudemolsedijk, Kasterlee	R	51.222595	4.991111
26	LIEGV05	Netedijk, Lier	F	51.108683	4.526550
27	LIN01	Kontichsesteenweg, Lint	R	51.129940	4.479220
28	MECGV17	Fietsostrade F1, Mechelen	F	51.036077	4.490377
29	MEEGV17	Markdijk, Meersel-Dreef	R	51.498274	4.779328
30	OELGV20	Zandhovensesteenweg, Oelegem	R	51.212234	4.615297
31	RAMGV11	Bergstraat, Ramsel	F	51.031400	4.826072
32	RIJ02	Merksplassesteenweg, Rijkevorsel	R	51.349240	4.810780
33	RUMGV13	Francis van den Eedebrug, Rumst	R	51.071958	4.424261
34	TONGV19	Oevelse Dreef, Tongerlo	R	51.105392	4.903961
35	TURGV08A	Bels lijntje, Kanaalbrug, Turnhout	F	51.331328	4.941526
36	TURGV08B	Bels lijntje, Jaagpad, Turnhout	F	51.331328	4.941526
37	WILGV222	N177, Willebroek	F	51.081785	4.358405

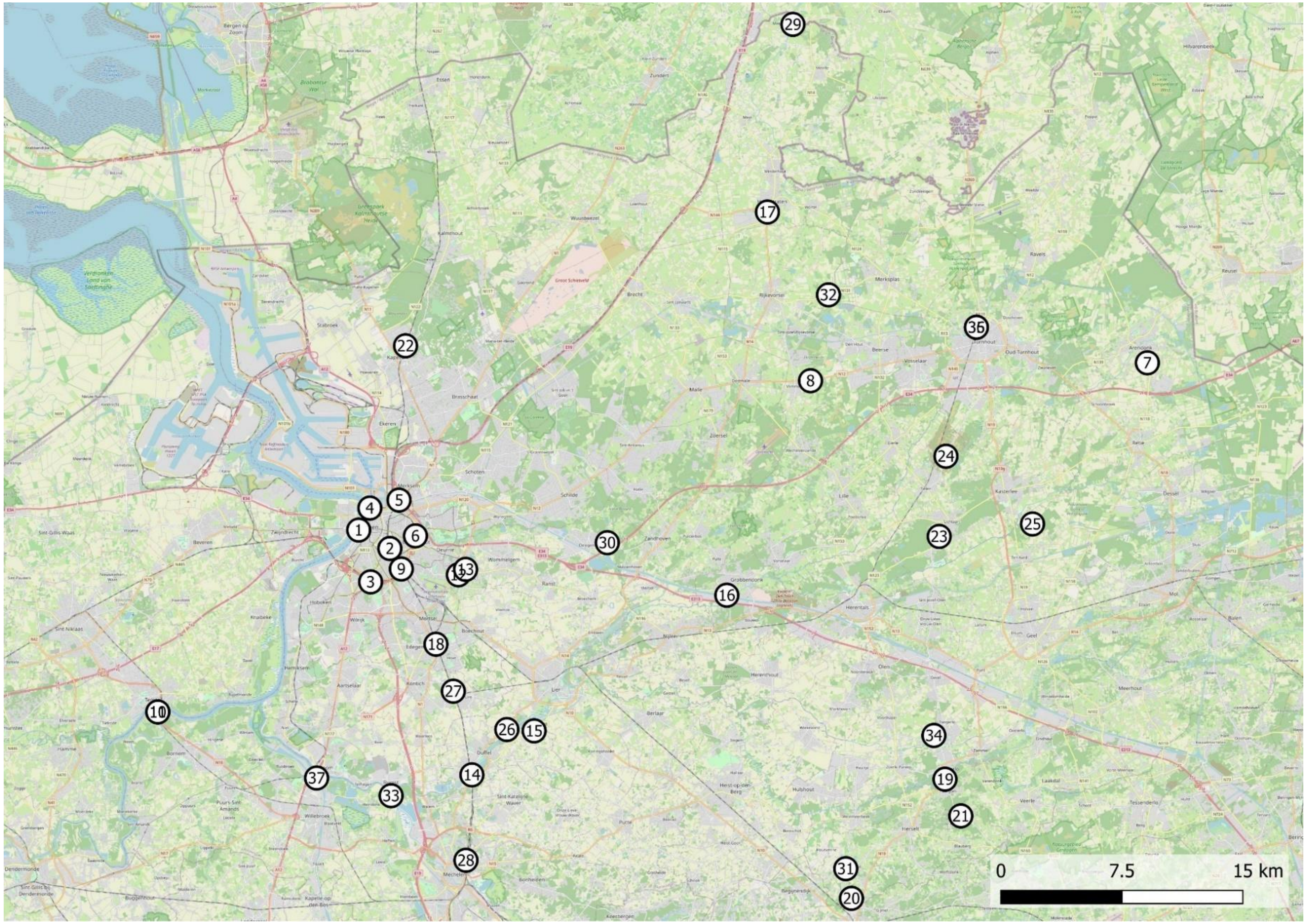


Figure 20. Locations of the counting points used to develop the model.

Location counting points (external validation)

Table 20. Counting points used for external validation and their locations. Points on a bicycle freeway are marked in bold (Fietssnelwegen, 2019). '*' indicates locations that are located in Flemish Brabant and were used to test transferability.

Point	ID	Location	X	Y
1	AAR01	Kerkeneinde, Aartselaar	51.1364	4.35732
2*	AFF01	Bellestraat, Affligem	50.89135	4.11866
3	ANT01	Boomsesteenweg, Wilrijk	51.17006	4.38643
4	ANT02	Borsbeeksebrug, Berchem	51.200752	4.436997
5	ANT03	D'Herbouvillekaai, Antwerp	51.20332	4.37388
6	ANT16	Justitiestraat, Antwerp	51.20877	4.40541
7	ANT17	Kasteelpleinstraat, Antwerp	51.21001	4.40194
8	ANT20	Brug N184, Borgerhout	51.20806	4.44277
9	ANT21	Scheldelaan, Berendrecht-Zandvliet-Lillo	51.368813	4.296445
10	ANT23	Posthofbrug, Berchem	51.19576	4.43365
11	**ANT26**	**Ringfietspad, Borgerhout**	51.20864	4.44469
12	ANT28	Desguinlei, Antwerp	51.194203	4.405439
13	ANT29	Van Eycklei, Antwerp	51.21111	4.41082
14	ANT30	Stenenbrug, Borgerhout	51.21304	4.44736
15	ANT31	Binnensingel, Berchem	51.19711	4.43067
16*	**ASS01**	**Fietsostrade F212, Asse**	50.90535	4.20876
17*	BEK01	Wissenbeemd, Bekkevoort	50.94929	5.00007
18*	BEK02	Staatsbaan, Bekkevoort	50.94507	4.97602
19	BER01	Tulpenstraat, Berlaar	51.120425	4.64803
20	BON01	Mechelsesteenweg, Bonheiden	51.03223	4.51636
21*	**DIE01**	**Fabriekstraat, Diest**	50.99361	5.03914
22*	DIE02	Fietspad Diest-Schaffen, Diest	50.99176	5.06546
23	**DUF01**	**Fietsostrade F1, Duffel**	51.109507	4.487998
24	EDE01	Kontichstraat, Edegem	51.150279	4.446689
25	EDE02	Prins Boudewijnlaan, Edegem	51.16133	4.42901
26	GEE02	Diestseweg, Geel	51.15149	4.99413
27	GEE15	Snelwegstraat, Geel	51.13914	4.94106
28*	GOO01	Bettestraat, Gooik	50.80851	4.07713
29*	GRB01	Robert Dansaertlaan, Groot-Bijgaarden	50.8664	4.26737
30	GRO01	Hofeinde, Grobbendonk	51.195493	4.754421
31*	HAA01	Haachtsebaan, Haacht	50.98531	4.64328
32*	HAA02	Schorisgat, Haacht	50.96839	4.63343
33	HEM01	Voetweg, Hemiksem	51.14716	4.34486
34*	**HER01**	**Vaartdijk, Herent**	50.92807	4.6784
35*	HER02	Mechelsesteenweg, Herent	50.89639	4.68536
36*	HER03	Terbankstraat, Herent	50.87904	4.66141
37*	**HOE02**	**Terhulpensesteenweg, Hoeilaart**	50.77034	4.44117
38*	HOE04	Tumulidreef, Hoeilaart	50.77015	4.44035
39*	HOL01	Horststraat, Holsbeek	50.932004	4.82949
40	**HRT01**	**Fietsostrade F106, Herentals**	51.149482	4.821079
41	HRT02	Geelseweg, Olen	51.159627	4.883728
42	HUL01	Vennekensstraat, Hulshout	51.04953	4.806299
43*	ITT01	Lenniksebaan, Sint-Pieters-Leeuw	50.81709	4.24949
44*	KAM01	Geilroedeweg, Kampenhout	50.94522	4.5527
45*	KAM02	Lauterweg, Kampenhout	50.9527	4.54797
46	**KON01**	**Fietsostrade F1, Kontich**	51.134678	4.476007
47*	**KOR01**	**Fietsostrade F3, Kortenberg**	50.89631	4.58047
48*	LEN01	Losgatstraat, Lennik	50.80224	4.18443
49*	**LEU12**	**Fietsostrade F3, Herent**	50.902734	4.697651

Table 20 (continued).

Point	ID	Location	X	Y
50*	LIED01	Affligemsestraat, Liedekerke	50.88196	4.09488
51*	**LIED03**	**Stationsstraat, Liedekerke**	50.8819	4.09332
52	LIL01	Gierlebaan, Lille	51.25615	4.84928
53*	**LON01**	**Leireken, Londerzeel**	50.99695	4.26248
54*	**LTR01**	**Fietsostrade F21/22, Linter**	50.84481	5.5928
55	**MEC02**	**Fietsostrade F1, Mechelen**	51.045027	4.494299
56*	NOS01	Zaventemsebaan, Zaventem	50.86765	4.2633
57*	NOS02	Mabtinusweg, Zaventem	50.88346	4.49114
58	OUT01	Steenweg op Turnhout, Oud-Turnhout	51.3209882	4.9688066
59	OUT02	Steenweg op Oosthoven, Turnhout	51.33574	4.96741
60	RIJ01	Sint-Lenaartsesteenweg, Rijkevorsel	51.347847	4.73953
61*	ROT01	Dijledijk, Rotselaar	50.96888	4.692876
62*	ROT02	Heirbaan, Rotselaar	50.96159	4.75198
63*	ROT03	Walenstraat, Rotselaar	50.4421	4.69221
64*	**SGR01**	**Waterloosestwg., St-Genesius-Rode**	50.76845	4.38269
65*	**SOZ01**	**Fietsostrade F214, Steenokkerzeel**	50.9113	4.46078
66*	**SPL01**	**Fietsostrade F20, St-Pieters-Leeuw**	50.78784	4.29191
67	STA01	Torense Weg, Stabroek	51.3324138	4.346626
68*	**STB01**	**Antwerpselaan, Grimbergen**	50.91284	4.34523
69*	**STB02**	**Boechoutlaan, Grimbergen**	50.90834	4.34326
70*	TER01	Brusselstraat, Ternat	50.8725	4.18333
71*	TER02	Nattestraat, Ternat	50.87828	4.17042
72*	TIE02	Molenbergweg, Tienen	50.80734	4.9108
73*	TRE01	Balenbergstraatje, Tremelo	51.00218	4.76581
74*	TRV01	Oud-Heverleestraat, Tervuren	50.83233	4.53896
75*	VIL01	Schoewever, Vilvoorde	50.94413	4.4483
76*	VIL02	Kleine Steenstraat, Vilvoorde	50.92793	4.4511
77*	**WEM01**	**Tentoonstellingslaan, Wemmel**	50.89503	4.31282
78*	WEO01	Astridlaan, Wezembeek-Oppem	50.831	4.5043
79*	WEO02	Lange Eikstraat, Wezembeek-Oppem	50.8523	4.4916
80*	**ZAV01**	**Tramlaan, Zaventem**	50.86222	4.48028
81*	**ZEM01**	**Fietsostrade F1, Zemst**	50.96004	4.45544
82*	**ZEM02**	**Westvaartdijk, Zemst**	50.99137	4.37859
83*	**ZOU02**	**Fietsostrade F21, Zoutleew**	50.2759	5.10412

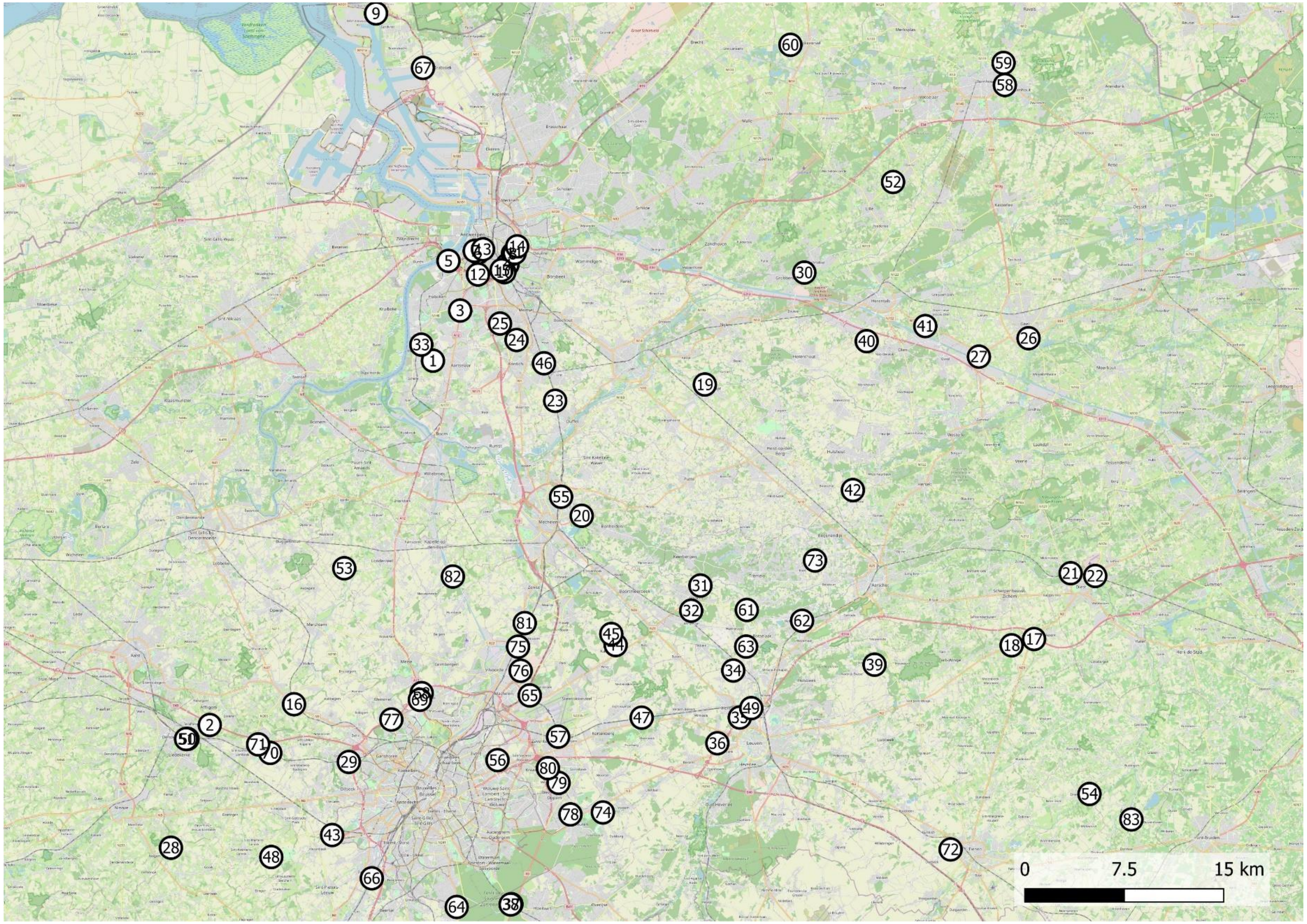


Figure 21. Counting points used for external validation and their locations.

Data registration Province of Antwerp

Provincie Antwerpen

data registratie

Ondertekend terug te sturen naar:
Provincie Antwerpen
Departement Ruimtelijke Ordening en Mobiliteit, Dienst Mobiliteit
Koningin Elisabethlei 22
2018 Antwerpen

tel :03 240 52 20
e-mail: fietsen@provincieantwerpen.be

Ondergetekende,

Naam: Tim Vervoort
Firma/Instelling: Uhasselt
Adres: Campus Hasselt | Martelarenlaan 42 | BE3500 Hasselt

verbindt zich ertoe om uitsluitend voor het project

Bicycle counts and built environment characteristics

in opdracht van

promotor Prof. Luc Int Panis / copromotor Evi Dons

volgende ter beschikking gestelde digitale bestanden te gebruiken volgens de hierna beschreven bepalingen:

Bestanden	***Bronvermelding***
Provinciale_Fietsbarometer_fietstelgegevens	Provincie Antwerpen (Fietsbarometer)

algemene voorwaarden gebruik GIS-data

1 Het studiebureau, vzw of vereniging is niet gemachtigd de bestanden, geheel of gedeeltelijk, over te dragen aan - of ter beschikking te stellen van - derden.
2 Het studiebureau, vzw of vereniging kan de bestanden of direct afgeleide producten (o.a. papieren kaarten, maar ook
3 Het studiebureau, vzw of vereniging zal in alle publicaties, rapporten, verslagen en kaartmateriaal waarin melding wordt gemaakt van de gegevens, refereren naar de bronvermeldingen zoals hierboven omschreven.
4 De provincie Antwerpen kan niet verantwoordelijk gesteld worden voor eventuele schade die mocht resulteren door of naar aanleiding van het gebruik van de bestanden.
5 Het studiebureau, vzw of vereniging zal de digitale bestanden, die aan de hand van de bestanden opgesteld worden, aan de provincie Antwerpen overmaken in GIS-formaat (shapefile (.SHP) of geodatabank (.GDB)). Op die manier worden afgeleide bestanden ingebracht in de provinciale GIS-databank.
6 Het studiebureau, vzw of vereniging zal voor alle digitale afgeleide bestanden een metadatafiche invullen met de editor van de metadatabank van het AGIV (http://metadata.agiv.be/). Een export van de metadatafiche moet in XML-formaat worden overgemaakt aan de provincie Antwerpen.

aanvullende voorwaarden voor het gebruik van GIS-data die eigendom zijn van de provincie Antwerpen

1 In publicaties waarin de provinciale data gebruikt wordt, wordt de provincie Antwerpen als eigenaar van de gegevens vermeld samen met het logo en een verwijzing naar de website www.provincieantwerpen.be
2 Op kaarten die de provinciale data of afgeleiden gebruikt, wordt steeds het logo van de provincie Antwerpen opgenomen samen met een verwijzing naar de website www.provincieantwerpen.be
3 Bij de aanvraag van gegevens zal de aanvrager alle gemeenten, waarvan er gegevens worden opgevraagd, inlichten over de aanvraag van en het gebruik van de provinciale GIS-data.

Ondertekend voor akkoord op: 29-01-2020

Naam en handtekening verantwoordelijke:

TIM VERVOORT

Agreements Master's thesis part 1

Afspraken 3085 Masterproef deel 1 - Academiejaar 2019-2020

Student:	Start MP1: sep 2019 / feb 2020
Promotor:	Begeleider: ....

Op dit formulier wordt door de promotor (en begeleider) aangeduid welke onderdelen van de masterproef in de context van het opleidingsonderdeel 3085 Masterproef deel 1 (12 studiepunten) dienen te worden afgewerkt voor de deadline. Op de volgende pagina vind je onderaan een overzicht van de belangrijke data wanneer het schriftelijke rapport ingediend moet worden voor de eerste en tweede zittijd. Het schriftelijk rapport dient alle onderstaand aangeduide zaken te integreren.

Dit formulier vormt een leidraad om de afspraken omtrent de verwachtingen tussen de student en de begeleiders vast te leggen. Dit formulier is een overeenkomst tussen de student en het team van begeleiders (i.e., promotor, co-promotor en begeleider). Het contract moet uiterlijk 31 oktober 2019 (deel 1 in sem 1) of 27 maart 2020 (deel 1 in sem 2) worden ingevuld en ondertekend. Dit contract wordt in tweevoud opgemaakt (één exemplaar voor de student en één exemplaar voor de promotor).

Minimale basisvereisten vastgelegd binnen studiefiche 3085 Masterproef deel 1
CONTEXT & PROBLEEMSTELLING
Beschrijving van de context van het onderzoek (Aanleiding, situatie, theoretisch of praktijkprobleem, ...).
Beschrijving van een constructief opgebouwde doelstelling (Duidelijke afbakening van wat men wel en niet in de masterproef zal meenemen).
THEORETISCHE & CONCEPTUELE BASIS VAN DE STUDIE
Literatuuronderzoek dat de centrale begrippen en de te onderzoeken relaties behandelt vanuit een theoretische en empirische invalshoek.
ONDERZOEKSVRAGEN
Duidelijke formulering van onderzoeksvragen en/of hypothesen die (1) op een logische wijze zijn afgeleid uit theorieën of een probleemstelling en (2) in verband zijn gebracht met vroeger verricht onderzoek.
ONDERZOEKSDESIGN
Motivering voor de keuze (incl. veronderstellingen en beperkingen) van een bepaalde onderzoeksbenadering (bv. een (quasi-)experimenteel design, een ontwerponderzoek, een case study, een survey onderzoek, ontwikkelen van een methode, implementatie van een gespecifieerde methode, toepassen van een tool/methode op een specifieke case, sensitiviteits analyse, ...).
Algemene beschrijving van een eerste concept van het onderzoeksdesign (aard van het kwantitatieve en/of kwalitatieve onderzoek) en van het ontwikkelingsdesign.
Beschrijving van de mogelijke respondenten / onderzoekseenheden / sampling aanpak (wie, omvang, eventuele subgroepen, ...). de te gebruiken datasets en de tools (voor simulaties en predicties).
Beschrijving van de mogelijk te gebruiken onderzoeksinstrumenten (bv. vragenlijst, interviewleidraad, observatie instrument, ...). simulators en analyzers voor problemen in mobiliteit.
PLANNING
Tijdsplanning voor de uitvoering van Masterproef deel 1.
Tijdsplanning voor het verdere verloop van het onderzoek (vooruitblik op Masterproef deel 2)

Specifieke (bijkomende) verwachtingen te bespreken met promotor	x = af te werken in MP1
Ontwikkeling van de te gebruiken onderzoeks- en ontwikkelings instrumenten.	
Voorstel hoe de data geanalyseerd zullen worden of gebruikt worden in validatie, simulatie en sensitiviteitesanalyses.	X
Schematisch overzicht van alle variabelen die centraal staan in de masterproef (onafhankelijke, afhankelijke, mediërende, covariabelen). Bij voorkeur in de vorm van een schematisch model.	X
Uitvoering van een pilot onderzoek (eveneens uittesten van de onderzoeksinstrumenten).	
~~Uitvoering van het hoofdonderzoek inclusief~~ een gedetailleerde planning voor implementatie.	X
Rapportage van de resultaten, geordend volgens de onderzoeksvragen/hypothesen. VAN DE LITERATUURSTUDIE	X
Beschrijving van de aanpak voor ~~"informed consent"; eventueel brieven en formulieren voor betrokkenen~~; NDA voor gebruik van data;...	X

Ik verklaar hierbij dat ik in samenspraak met mijn promotor / begeleider de verwachtingen m.b.t. 3085 Masterproef deel 1 heb afgetoetst.

Datum: 8 / 11 / 2019

Handtekening student: ....

Handtekening promotor / begeleider: ....

Overzicht belangrijke data

Deel 1 in sem 1	Deel 1 in sem 2	Activiteit
Week 1 van sem 1	Week 1 van sem 2	Infosessie: voorstelling onderwerpen + toelichting kernzaken leidraad
22/09/2019	17/02/2020	Top 3 met motivaties + CV
Uiterlijk 01/10/2019	Uiterlijk 28/02/2020	Toewijzing en bevestiging van onderwerp en promotor
22/10/2019	18/03/2020	Plan van aanpak indienen
Uiterlijk 31/10/2019	Uiterlijk 27/03/2020	Verplichte bespreking van plan van aanpak + Masterproef contract: bespreken en bepalen van doelstellingen in masterproef deel 1 en deel 2
17/01/2020	05/06/2020	Schriftelijk rapport masterproef deel 1 indienen voor 16.00u
30/01/2020 31/01/2020	18/06/2020 19/06/2020	Mondeling verdediging masterproef deel 1 + Uploaden presentatie
14/08/2020		Tweede zittijd: Rapport masterproef deel 1 indienen voor 16.00u
27/08/2020 28/08/2020		Tweede zittijd: Mondeling verdediging masterproef deel 1 + Uploaden presentatie