Exploring nonlinear built environment effects on driving with a mixed-methods approach[1]

Tao Tao[1] and Petter Næss[2]

[1] University of Minnesota (taotao@umn.edu)

[2] Norwegian University of Life Sciences

Abstract

Recent studies have been exploring the complex nonlinear relationships between built environment attributes and driving using machine learning approaches. However, these nonlinear relationships lack causal explanations. This study applied a mixed-methods approach to data from a smaller European city, Stavanger, Norway. Our results showed that transport rationales for choosing activity locations and travel modes, along with configurations of the jobs and other facilities, provide causal explanations for the nonlinear and threshold effects of built environment attributes on people's driving-related behavior. Distance to city center plays the most important role and its nonlinear relationship reflects the influence of the polycentric city structure of Stavanger on driving. For Stavanger and similar cities, compact development around the city center helps to rein the auto dependence. Furthermore, the thresholds of nonlinear relationships provide planning guidelines to support compact development policies.

Keywords: land use, travel behavior, machine learning, mixed-methods approach, qualitative analysis

1 Introduction

Auto dependence has caused many problems for society, such as traffic crashes, air pollution, and obesity. This issue is more serious in small cities as studies have found that the share of driving in small cities is higher than that in large cities (Tennøy et al., 2022). Therefore, reducing car use is important for sustainable development in small cities. Planning scholars have been interested in studying the relationships between the built environment and auto use and how these relationships could support planning policies to reduce driving and encourage healthy travel modes such as transit and active travel. The purpose of this article is to illuminate nonlinear relationships of built environment characteristics with weekly car-driving distance, frequency of car commuting, and car ownership in a smaller European urban area. This study combines a machine learning approach (gradient boosting decision tree, GBDT) with an in-depth interpretation of qualitative interviews. The GBDT approach enables an exploration of nonlinear relationships between built environment attributes and driving-related behavior. The qualitative analysis sheds light on the causal mechanisms of these nonlinear relationships. This mixed-methods approach is applied to the Norwegian urban region of Stavanger. Despite being one of the larger Norwegian cities, Stavanger should be characterized as a smaller urban area in a wider European context.

Although most quantitative analyses on the associations between built environment characteristics and travel behavior are based on the assumption of linear relationships, there are theoretical reasons (notably from location theory on the thresholds and ranges for goods and services (Christaller, 1966; Eldridge & Jones, 1991)) that the influences of built environment characteristics on travel are not linear. For example, Figure 1 illustrates a relationship between distance from residence to an employment center and commuting distance. Initially, commuting distance increases moderately with distance to an employment center. The dense inner-city area usually covers a broader area than just the very center point. People living within the inner city therefore all have a quite high likelihood of finding suitable jobs within a short distance from home. Then, after the first threshold, commuting distance increases much faster than before. The density gradient is nonlinear, with a marked density difference between the inner city and the suburbs. Therefore, the curve could be expected to rise more steeply from the outskirts of the inner city and outward. Finally, because the attractiveness of jobs in the employment center will gradually diminish as the distance to the center exceeds a certain level (Ding, Cao, & Næss, 2018), the curve becomes less steep again after the second threshold.

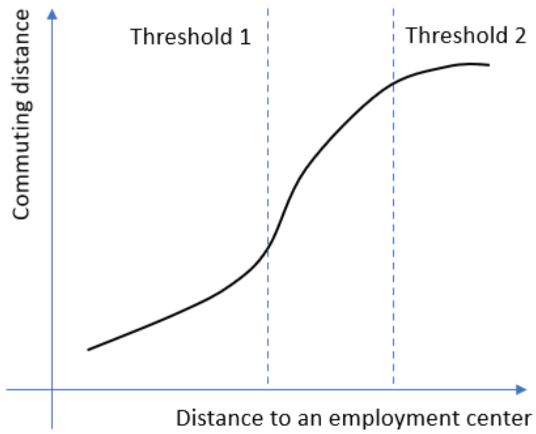

Figure 1. A conceptual relationship between distance to an employment center and commuting distance

Many studies have shown that relaxing the assumption of linearity helps improve theoretical understandings and practices. First, allowing nonlinear relationships produces more accurate estimates for the influences of built environment characteristics on travel behavior (Cheng et al., 2020; L. Yang et al., 2021). Second, the nonlinear relationships provide new insights into the influences and help planners design effective policies and programs. For example, Tao et al. (2020) examined the nonlinear association between the built environment and active travel in the Twin Cities area, the US. Their result suggested that compact development in the areas within four miles from downtown Minneapolis can effectively encourage people's amount of active travel.

One way of dealing with nonlinearity could be to identify the theoretical shape of the association of an independent variable with the dependent variable and then transform the variables by relevant mathematical functions (Næss et al., 2019). For example, the relationship between the distance from the dwelling to the city center and commuting distance could, for a monocentric city region, be depicted by transmuting the distance to the city center through a hyperbolic-tangential function along with a quadratic function (Næss et al., 2019). However, specifying the parameters of such a curve is a time-consuming iterative process, and doing so for each independent variable would be very laborious, especially when taking into consideration mutual influences between the various independent variables in a multivariate context. The

GBDT approach offers a unique opportunity for describing and comparing the effect sizes of nonlinear associations between built environment characteristics and travel behavior. The previous papers where GBDT or similar approaches have been applied to depict nonlinear relationships between built environment characteristics and travel have all been based solely on quantitative methods (Ding, Cao, & Næss, 2018; Zhang et al., 2020).

These studies, however, lack causal explanations of the nonlinear relationships. In the literature, several quantitative methods have been applied to study the causal effects of built environment attributes on travel behavior, such as longitudinal research design (Handy et al., 2005; Krizek, 2003; Van De Coevering et al., 2021) and instrumental variables (Boarnet & Sarmiento, 1998; Heres-Del-Valle & Niemeier, 2011; Vance & Hedel, 2007). Longitudinal research design compared the travel behavior before and after people's moving to new residential locations and examined how the changes in built environment attributes affect people's travel behavior. For example, Handy et al. (2005) applied a quasi-longitudinal design to explore how the changes in people's perceived built environment attributes influenced people's driving amount in California, US. Instrumental variables are those being correlated with built environment attributes (i.e., people's choice in residential locations) but are exogenous to travel behavior. Applications of instrumental variables can address the endogeneity issue of built environment attributes. For example, Vance and Hedel (2007) used percentages of buildings built in different periods and percentages of different populations at zip code levels as instrumental variables for built environment attributes and studied the impacts of built environment attributes on people's driving distance. Studies with these quantitative methods, although being better able to demonstrate the plausibility of causal relationships between built environment attributes and travel behavior, lack sufficient identification of causal mechanisms for explaining the relationships. Several scholars have argued that qualitative studies can supplement quantitative methods and provide rich causal explanations for the relationships between built environment attributes and travel behavior (Clifton & Handy, 2003; Næss, 2015). A qualitative analysis can better explain the causal mechanisms that produce the statistical patterns, especially since it can show how the rationales for activity location and travel mode choice encourage people to choose different travel modes and differ in their travel distances, depending on their residential locations.

The present study mainly contributes twofold to the literature. First, this study explores the nonlinear relationships between built environment attributes and driving-related travel behavior in Stavanger, which makes the study the first to uncover the threshold effects of the built environment in a small city. Second, this study combines the quantitative GBDT approach with qualitative interview interpretation. With the transport rationales, which indicate people's backgrounds, motivations, and justifications when making travel-related decisions, summarized from the interviews, this study illustrates the causal mechanisms underlying the nonlinear relationships.

The rest of the paper is organized as follows. We introduce the causal linkages between the built environment and driving-related travel behavior in Section 2. We introduce the data and methods in Section 3. We discuss the transport rationales extracted from the qualitative interviews in Section 4. We discuss the relative importance and nonlinear relationships generated from quantitative analysis in Section 5. We conclude our study in the final section.

2 Causal linkages

There are important interrelationships between the three dependent variables of our study, and explanations of how they are causally related to built environment characteristics must include people's reasons for choices of travel modes and of the activity locations that generate travel.

Weekly driving distance is logically determined by the total weekly travel distance and the share of this distance traveled as a car driver. On average for workforce participants, commuting makes up a large part of the total travel distance. The frequency of car commuting and commuting distance, therefore, exert substantial influence on the average weekly driving distance among workforce participants. Moreover, the choice between car travel and other travel modes is influenced by travel distance, since most people find non-motorized travel, particularly walking, acceptable only for relatively short trips. For both commuting and non-work travel, built environment characteristics that influence trip distances therefore also influence travel mode choices indirectly.

Whether or not people travel as car driver depends on whether they have a car at their disposal and possess a driver's license. Auto ownership depends on several socio-demographic and attitudinal factors (Scheiner, 2010). The built environment also plays a role since differently located and designed neighborhoods can, to varying degrees, facilitate car driving or other travel modes. The location of the neighborhood relative to daily-life trip destinations such as jobs, places of education, and stores is an important part of such facilitating. If long trip distances need to be overcome to reach daily activities and the provision of public transit is poor in the neighborhood, car travel may stand out as the only acceptable travel mode to reach the activities within a constrained time budget (Hägerstrand, 1970).

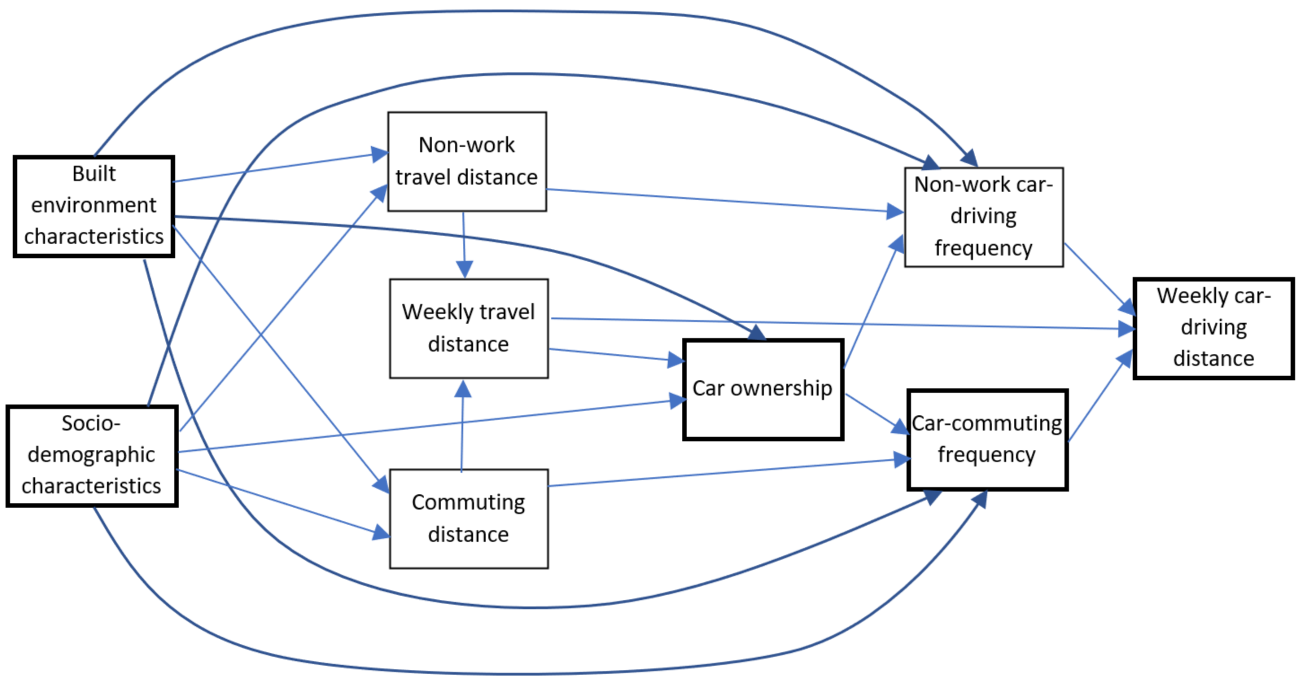

Figure 2. Theoretical relationships between different characteristics influencing weekly car-driving distance (Variables considered in the quantitative analysis of the present paper are shown with a bold outline. For the sake of simplicity, we only show the influences of different characteristics on driving-related travel behavior and ignored some effects in the opposite direction.)

Figure 2 shows how built environment characteristics, socio-demographic characteristics, and different aspects of travel behavior could theoretically be expected to influence the weekly car-driving distance. As can be seen, the main dependent variable of the present paper, i.e., the weekly car-driving distance, is influenced by the two other dependent variables of the paper (car-commuting frequency and car ownership), which are in their turn influenced by the built environment and socio-demographic characteristics.

Residents' *transport rationales*, i.e., their backgrounds, motivations, and justifications when making decisions on travel (Næss & Jensen, 2005), make up important links in the causal mechanisms through which residential location affects travel behavior. Rationales for location of activities and travel mode choice are particularly important. The former rationales influence whether people prefer to use local facilities or rather travel further to find a better facility (Næss & Jensen, 2005; Næss et al., 2018), whereas the latter rationales can explain why people living in the same neighborhood may prefer different travel modes for trips to the same destination (Næss & Jensen, 2005; Næss et al., 2018).

3 Data and method

3.1 Study area

The study area is Stavanger, which is a city on the southwestern coast of Norway. In 2020, Stavanger had 228,000 inhabitants. Its continuous urban area includes the historical city of Stavanger, the large suburban employment center of Forus, and the previously separate town of Sandnes (Figure 3). Therefore, Stavanger has a polycentric urban region. The area within the semi-ellipse defined by the city center, Forus, and Sandnes (Figure 3) is the main concentration of activities in Stavanger. Forus has the largest concentration of workplaces with more than 45,000 jobs, compared to around 27,000 and 10,000 in the central parts of Stavanger and Sandnes, respectively (Figure 3). In addition, the main city center is the major concentration of non-work activities, including cultural arenas, restaurants, bars, specialized stores, fitness centers, and civic buildings. Compared with other small cities with a similar population in the Nordic context[2] (Table 1), Stavanger has a similar population density but is rather unique in its polycentric city structure.

[2] Nordic countries include Denmark, Finland, Iceland, Norway, and Sweden, which are geographically close and culturally similar to each other.

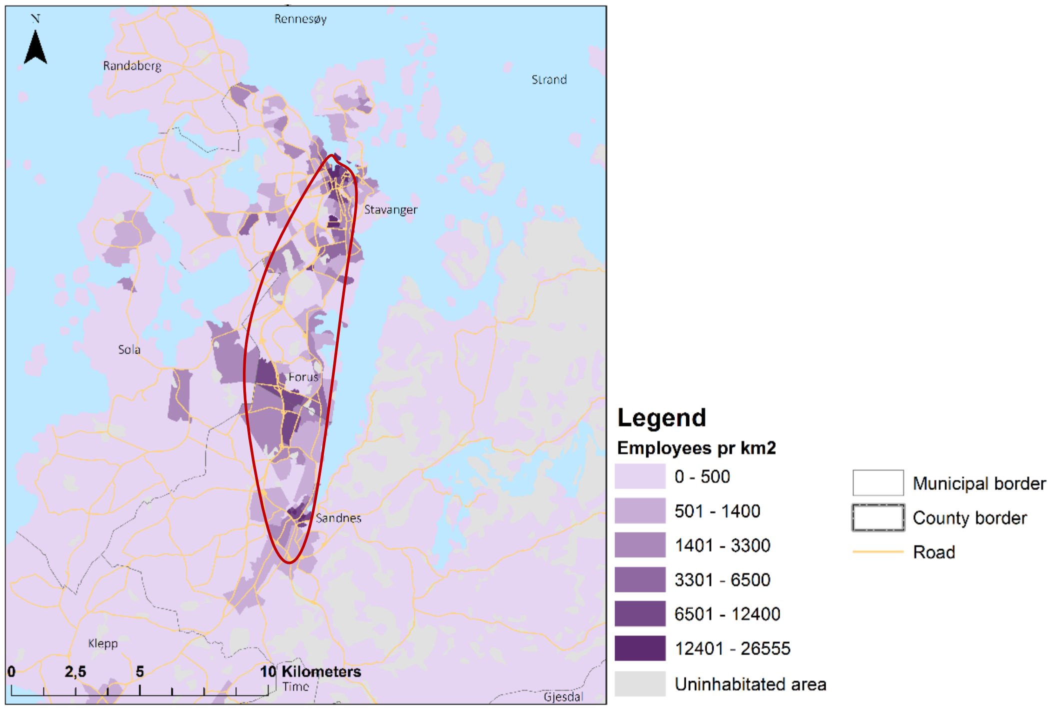

Figure 3. Employment densities within different parts of the Stavanger metropolitan area[3] (The red semi-ellipse generally demacrates corridor within which the main concentrations of jobs and places of education are located and includes the three activity centers.)

Table 1. Comparison among a selection of Nordic small cities

	Stavanger	**Bergen**	**Trondheim**	**Reykjavik**
Country	Norway	Norway	Norway	Iceland
Population in 2020	228,000	265,000	192,000	232,000
Population density (per hectare)	28.6	29.7	32.5	31.8
City structure	Polycentric	Monocentric	Monocentric	Monocentric

3.2 Data

We used the data from the Stavanger study, which is part of a larger study also comprising the Oslo metropolitan area. In each case city region, a questionnaire survey was combined with qualitative interviews. The questionnaire survey was performed from May to June in 2015 and was mainly web-based, with respondents drawn randomly by the national census authorities from a geographically stratified sample. In each metropolitan area, 12,500 persons, supplemented with 2,500 persons who had moved into newly constructed dwellings during the last two years, were selected. Moreover, those who wanted could receive a paper version of the

[3] Map by Anja Fleten Nielsen, Institute of Transport Economics.

questionnaire and a stamped envelope for submission. Nearly 3,400 people completed their responses. After deleting invalid observations, the final sample was reduced to 3,232. 1,328 of those were from Stavanger. Given that some respondents did not answer all questions, the sample sizes in the models were slightly smaller than 1,328.

We recruited 33 participants in the qualitative interviews from the survey respondents (more than 900) who had answered that they were willing to participate in this part of the study. Sixteen of the interviewees were from the Stavanger region. Interviewees were selected to include different household types, different socio-demographic characteristics, and different kinds of locations of their dwellings (in the inner city, in the proximity of a second-order center, and in non-central areas). Furthermore, we recruited interviewees who had typical or atypical travel patterns. Figure 4 shows approximately where the residences of the Stavanger interviewees were located. We conducted each interview for 60 to 90 minutes in a semi-structured form. A map of the Stavanger region (or Greater Oslo for the interviews in this part of the study) was placed on the table to enable interviewers and the interviewee to point at places on the map and identify locations talked about. The interviews included questions about different aspects of travel behavior and related topics, socio-demographic characteristics, residential preferences, residential mobility, physical activity, and health. The questionnaire design was informed from theoretical considerations as well as by results from our earlier qualitative studies (Næss, 2013; Naess, 2015). The purpose of the interviews was mainly explanatory. We also tried to be open to new aspects that had previously been overlooked. We recorded the audio of the interviews and transcribed it verbatim into text.

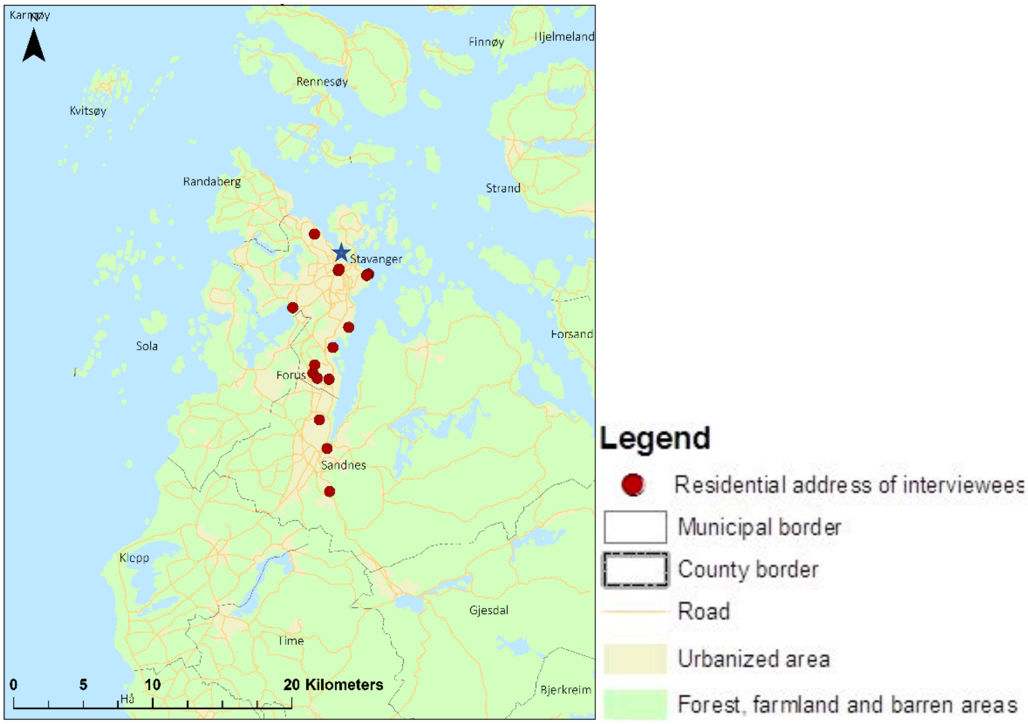

Figure 4. Locations of the dwellings of the Stavanger interviewees[4] (The blue asterisk shows the main city center. The locations of some interviewees are overlapped.)

Our dependent variables are three types of driving behavior: weekly driving distance, commuting frequency by car, and car ownership (Figure **5**). These three variables provide a more complete picture of people's driving in Stavanger compared with previous literature on nonlinear relationships between built environment characteristics and driving, which has only focused on one aspect of driving (Ding, Cao, & Wang, 2018; Ding, Cao, & Næss, 2018; Zhang et al., 2020). 67% of the respondents have their weekly driving distances below 100 km. 44% commute by car four days or more during a week. 50% have only one car and 44% have two or more cars in their households. We augmented the data with several GIS-measured built environment variables based on the addresses of respondents' residences. Such variables included in the present paper are residential distances to the main city center, to the closest second-order center (i.e., Forus and Sandes), and to the closest local center, as well as population density and job density in the neighborhood of the dwelling (Table 2). We also controlled for several socio-demographic characteristics (Table 2). The variance inflation factors of the

[4] Map by Anja Fleten Nielsen, Institute of Transport Economics.

independent variables are all smaller than 3, showing that multicollinearity is not an issue in this study.

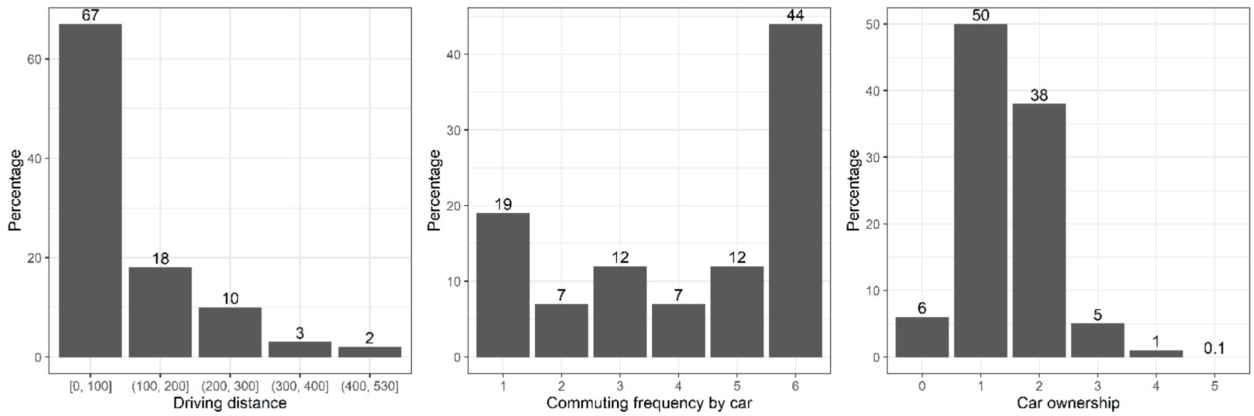

Figure 5. Distribution of driving distance, commuting frequency by car, and car ownership. (Commuting frequency by car: 1 = Not at all; 2 = Less than 1 day a month; 3 = 1-3 days a month; 4 = One day a week; 5 = 2-3 days a week; 6 = 4 or more days a week.)

It is worth noting that the analyses in the present paper do not include residential preferences as control variables. As shown in two separate articles from the Oslo and Stavanger study (Wolday et al., 2018; Wolday et al., 2019), the effects of the built environment variables are only to a very small extent affected by whether or not residential preferences are included as control variables. Moreover, other criteria are generally more important than travel attitudes to people's choices of their residences, and residential preferences do not only influence people's choices of where to live but are also influenced by people's experiences from living in a specific neighborhood (Lin et al., 2017; Wolday et al., 2019).

Table 2. Variable description and descriptive statistics

Variable	Description	Driving distance model (N=1,174)		Car commuting frequency model (N=973)		Car ownership model (N=1,208)	
		Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Driving distance	Driving distance over the whole week (km)	96.14	103.28				
Commuting frequency by car	Frequency of going from home to workplace/place of education by car in a typical month in spring 1 = Not at all 2 = Less than 1 day a month 3 = 1-3 days a month 4 = One day a week 5 = 2-3 days a week 6 = 4 or more days a week			4.19	1.98		
Car ownership	Number of cars at the household's disposal					1.45	0.72
Socio-demographic characteristics							
Male	A dummy variable indicating that the respondent is a male	0.48	0.5	0.48	0.5	0.48	0.5
Age	Age	46.4	16.69	42.31	14.15	46.37	16.6
Workforce	A dummy variable indicating that the respondent is employed	0.72	0.45	0.89	0.32	0.72	0.45
Personal income	Personal gross annual income 1 = Less than 100,000 Kroner 2 = 100,000 to 199,999 Kroner 3 = 200,000 to 299,999 Kroner 4 = 300,000 to 399,999 Kroner 5 = 400,000 to 499,999 Kroner 6 = 500,000 to 599,999 Kroner 7 = 600,000 to 699,999 Kroner 8 = 700,000 to 799,999 Kroner 9 = 800,000 Kroner and above	5.72	2.36	6.03	2.35	5.75	2.36
Household income	Household gross annual income 1 = Below 200,000 Kroner 2 = 200,000 to 399,999 Kroner 3 = 400,000 to 599,999 Kroner 4 = 600,000 to 799,999 Kroner 5 = 800,000 to 999,999 Kroner 6 = 1,000,000 to 1,199,999 Kroner 7 = 1,200,000 Kroner or more	5.25	1.68	5.48	1.61	5.26	1.68
Education	Highest completed education level	3.69	1.35	3.75	1.33	3.7	1.34
Household size	Number of household members	2.43	1.37	2.58	1.4	2.5	2.46
Teenager	Number of children between 7 and 17 years old	0.42	0.82	0.51	0.87	0.43	0.82

Children	Number of children younger than 7 years old	0.24	0.57	0.27	0.62	0.24	0.59
Built environment characteristics							
Distance to city center	Driving distance from the residence to the main city center of Stavanger (km)	10.29	6.43	10.27	6.44	10.32	6.5
Distance to second-order center	Driving distance from the residence to the closest second-order center (km)	6.83	4.88	6.87	4.89	6.89	4.96
Distance to local center	Driving distance from the residence to the closest local center (km)	2.71	2.08	2.73	2.04	2.75	2.15
Population density	Number of residents per hectare within the 1 km^2 grid square to which the residential address belongs	21.53	12.95	21.67	12.94	21.48	13.01
Employment density	Number of full-time jobs per hectare within the 1 km^2 grid square to which the residential address belongs	7.64	10.22	7.4	9.12	7.57	10.13

3.3 Method

We used a mixed-methods research design including both qualitative and quantitative analysis in this study. As described earlier, we assumed that the results of qualitative analysis can greatly support the findings from quantitative analysis.

In the qualitative part, we used a scheme to interpret the interviews. In our earlier studies, we developed an interpretation scheme as a tool for interview analysis, designed for explanatory qualitative research (Næss, 2005, 2013). We refined and developed this scheme further in the Oslo and Stavanger study. The scheme consisted of more than 40 research sub-questions to be answered from the information given in the interviews. Some of the questions were purely descriptive but necessary to enable us to answer other, analytic questions included in the interpretation scheme. We used only the written text as a base when interpreting the interviews, except for a few occasions when we were uncertain whether the transcriber had understood the interviewee correctly. We then also checked the audio files. Two members of the four-person research team separately interpreted each interview. For each of the 45 research sub-questions, each interpreter thus wrote what the particular interview could inform about. Thereupon, the two interpreters of each interview discussed their interpretations and made a common interpretation.

Extracting *transport rationales* from the interviews is an important and challenging task during the interpretation. The interviewees normally did not state their rationales explicitly in the interviews. The rationales were instead abstractions that we, the researchers, inferred from the interviewees' information about their routine activities and their traveling to reach these activities; their existing and earlier places of work; what they considered the acceptable length of a journey to work were they to find a new job, as well as journeys to other facilities; their reflections about the reasons for choosing to live where they live, and their thoughts about the dwelling and residential neighborhood; the locations of their preferred stores, cultural facilities, restaurants, outdoor recreation areas, places for physical exercise, etc.; their information about any changes in activity pattern, travel behavior or car ownership after moving; and their thoughts concerning whether and how their activity pattern and/or travel behavior would be different if they lived in a different part of the metropolitan area. Apart from the information given by each interviewee, our interpretations were also informed by the questionnaire answers of the interviewees as well as by the general respondents.

In the quantitative analysis, we applied the gradient boosting decision tree (GBDT) method (Friedman, 2001, 2002) to estimate the influence of socio-demographic characteristics and built environment variables on driving-related travel behavior. Specifically, we estimated three models: a driving distance model, a car commuting frequency model, and a car ownership model. This method has been applied in several recent studies about the influence of the built environment on travel behavior (Wang & Ozbilen, 2020; Wu et al., 2019). GBDT is a combination of two approaches: the decision tree approach and the gradient boosting approach. The decision tree approach uses a tree structure to divide a sample space into several subsamples based on specific criteria and then applies the average of the dependent variable to predict the response. One single decision tree usually performs poorly in the accuracy of its prediction. The gradient boosting approach addresses this issue by combining many decision trees in a sequential order to produce a stronger model. Please see Tao (2021) for more discussion about the mathematical algorithm of the GBDT method.

Compared with traditional statistical methods such as linear regression and generalized linear regression, the GBDT method has several advantages. First, GBDT is powerful in estimating the irregular nonlinear relationships between dependent and independent variables. Many studies have found that the effects of built environment variables on travel behavior are nonlinear and vary among different variables. GBDT is a useful tool to address these complex relationships. Traditional methods, however, rely on transformation functions (e.g., quadratic form and logarithmic form) and can only estimate regular nonlinear relationships. Second, GBDT can make a more accurate prediction. Third, GBDT can better handle missing values and outliers in independent variables, which can help reserve a larger size of sample for data analyses. GBDT also has two limitations. One is that it cannot produce p-values to evaluate the significance of independent variables. Instead, GBDT generates relative importance which can be used to evaluate the practical importance of independent variables. The other is that GBDT is subject to overfitting. In this study, we applied the cross-validation method to address the issue of overfitting.

To interpret the GBDT models, we used relative importance and accumulative local effect (ALE) plots. Relative importance stands for the contribution of independent variables to estimating the dependent variable. It is the proportion of the variance reduction by one independent variable among the total variance reduction. The relative importance of all independent variables adds up to 100%. ALE plots can visualize the relationships between dependent and independent variables (Apley & Zhu, 2020). Specifically, ALE plots show the marginal effect of one independent variable on the dependent variable when controlling for other independent variables. ALE plots can better handle multicollinearity among independent variables (Molnar, 2020).

We used the "gbm" package (Greenwell et al., 2020; Ridgeway, 2020) in R to estimate the GBDT models. Three parameters are required to estimate the models: tree depth, learning rate, and number of trees. Tree depth indicates the number of levels of the decision tree structure. It is a positive integer. Learning rate decides how large proportion of the results of decision trees will be combined into the final model. It is a value ranging from 0 to 1. Number of trees shows how many decision trees will be combined into the final model. We followed the practices in similar travel behavior studies (J. Yang et al., 2021; Zhang et al., 2020) and set tree depth as 10 and learning rate as 0.001 for the three models. We applied five-fold cross-validation to search for the optimal number of trees that generate the smallest root mean squared error (RMSE). After searching, driving distance model, car commuting frequency model, and car ownership model have 1,989, 1,639, and 3,133 trees, respectively (Figure 6). We then used the "gbm" package to generate the relative importance and the "ALEPlot" package (Apley, 2018) to produce ALE plots.

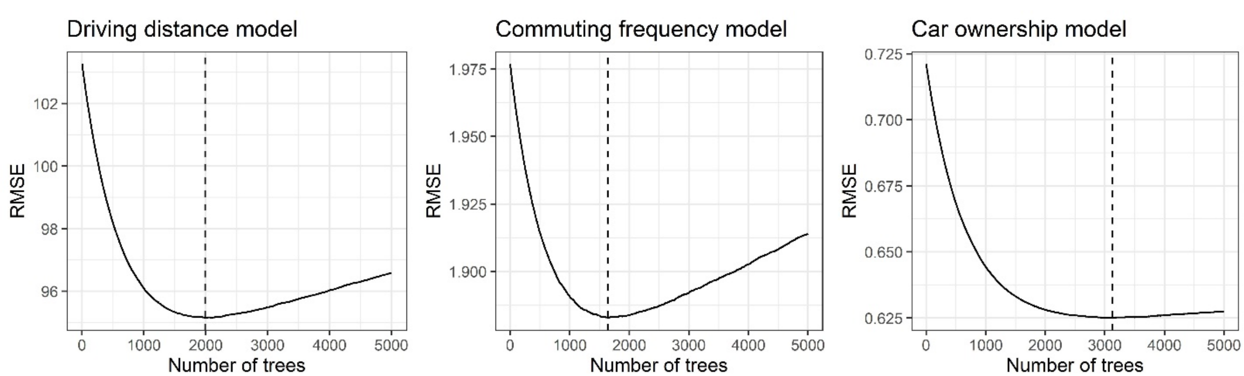

Figure 6. Results of cross validations

4 Transport rationales

We summarized transport rationales of people's activity location choice and travel mode choice from the qualitative interviews. These rationales play an important role in explaining the causal mechanisms through which built environment attributes affect driving distance, car commuting frequency, and car ownership. For a more extensive presentation of the rationales found among Stavanger interviewees, see Næss et al. (2018).

4.1 Rationales for activity location choice

The two main rationales encountered among the interviewees for activity location choices are choosing the best facility and minimizing the friction of distance. The latter is a concept that includes the time consumption, cost, effort, and inconvenience of going from one place to another (Lloyd & Dicken, 1972). Each of these two rationales plays a role for all interviewees, but their influence varies between the interviewees as well as activity types. Two other rationales also play a role in some interviewees' choices of locations for leisure activities: maintaining social contacts (encountered in about half of the interviews) and variety seeking (about a third of the interviews).

The rationale of choosing the best facility implies that people tend to be willing to travel beyond the closest facility of a category if they can then find a more suitable facility for their desired activity. For example, people may prefer a more interesting and well-paid job in a different part of the metropolitan area rather than a less interesting or lower-paid job opportunity in their local district. The interviews show several examples of such prioritizations. In addition, jobs matching a worker's qualifications may simply not exist in the local neighborhood, and if they do, an applicant from outside the neighborhood may be preferred by the employer instead of the local applicant. The likelihood of finding a job within a short commuting distance is, therefore, highest among residents living close to the main concentrations of workplaces. In Stavanger, the largest concentration of jobs is in a suburban second-order center, Forus, with more than one and a half times as many jobs as in the city center of Stavanger. The 'best facility' rationale is also important for several non-work activities, such as visits to cultural arenas and restaurants. The main city center of Stavanger has the highest concentration of such facilities.

The rationale of choosing the best facility is traded off against a rationale of minimizing the friction of distance. People are not willing to spend unlimited time, money, and effort on daily traveling. Work, higher education, and some kinds of outdoor recreation/sports activities are the kinds of daily-life activities our interviewees are willing to travel the longest distances to reach. For other less specialized activities, they tend to limit their choices of locations much more to the local district or neighborhood. This is particularly the case for grocery stores, primary schools, and kindergartens, but also areas for short recreational walks.

For example, two interviewees who had moved from less central locations to the inner city of Stavanger both stated that they had diminished their trip distances due to their moves, reflecting the combination of these two main rationales. One of these interviewees had reduced his travel distances for most trip purposes (journeys to work, political meetings, cultural events, sport, and other leisure activities) after moving from a single-family house 11 km from the city center to an apartment less than one km from the city center. For the other interviewee, a pensioner, the move from a suburb of the second-order center town Sandnes to downtown Stavanger had reduced his travel distances to cultural events matching his taste and interests, as well as for visits to friends. Conversely, a female interviewee had increased her travel distances for daily purposes after moving from the inner city to a suburban single-family house neighborhood. Another female interviewee had moved from downtown Stavanger to a new apartment in a transformed industrial area in a suburb and had increased her travel distances to most destinations. The move had also increased her son's distance to his temporary job. On the other hand, moving had reduced her husband's commuting distance to his job in the suburban second-order center of Forus.

4.2 Rationales for travel mode choice

The main rationales encountered for travel mode choice are convenience and comfort, aversion of frustration, and timesaving. The convenience and comfort rationale appears in all interviews and in many forms, such as avoiding physical efforts (e.g., carrying heavy items), avoidance of harsh weather, and traveling smoothly (e.g., not having to change between different transit lines or search for a parking space). What is considered convenient travel for daily-life trip purposes depends to a great extent on the kinds of the neighborhood in which the interviewees live. For residents living in the inner-city of Stavanger, where traffic lights, congested streets, and scarcity of parking opportunities make driving more cumbersome than in the suburbs, non-motorized travel may often be perceived as more convenient than driving. Walking distances from stores to home or transit stops are also on average shorter in the inner part of Stavanger than in the suburbs, where driving is often considered the only convenient mode when bringing groceries or other purchases home.

Frustration mainly plays a role when traveling by motorized modes, whereas bicycling and walking are not experienced by the interviewees as frustrating. Frustration aversion thus encourages interviewees to prefer non-motorized modes for some trips. Frustration caused by long waiting times, delayed departures, and buses sometimes not showing up at all can induce people to choose travel modes other than transit. Frustration when traveling by car is typically caused by congestion and demanding traffic situations (stressful traffic, especially in the central parts of the city).

The time-saving rationale is emphasized particularly among suburban interviewees, although it evidently plays some role for the remaining interviewees as well, often in combination with the convenience rationale. Most of the suburban interviewees need fast means of transportation since their travel distances to relevant activities are often long. Walking distances to transit stops and frequencies of transit services are also less favorable in the suburbs than in the areas close to downtown Stavanger, whereas driving speeds are generally higher in the suburbs. This also applies to the Forus second-order center, which has ample parking opportunities and wide access roads but is not very well served by transit. Transit is often perceived as a time-consuming mode among suburban residents living close to a transit stop. The time-saving rationale thus encourages suburbanites to travel by car. For interviewees living in the inner city, distances to relevant destinations are often so short that travel time consumption will be small no matter what travel mode is chosen, especially for non-work travel. The same applies to the commutes of those who live close to the Forus employment center and have their workplace there.

For example, one interviewee had moved to the city center from a suburban single-family house and, due to this move, changed from car to walking or biking for his trips to the workplace, political meetings, cultural events, sport, and other leisure activities. Similarly, for another interviewee, walking had become his most common travel mode to cultural events and other regular and non-work destinations after moving to downtown Stavanger. Previously, this interviewee made most such trips by car or, in some cases, by transit. Another interviewee had moved with his family to the inner city of Stavanger from a suburb of another Norwegian city, Trondheim. This interviewee told that moving had caused him to travel more by bike, partly due to a milder winter climate in Stavanger but also because more trip destinations were located close to his current residence. Accordingly, his driving on weekdays had been reduced. They had also stopped using one of their two cars and intended to store it away. When living in Trondheim, they used both cars, reflecting the location of key trip destinations beyond acceptable non-motorized travel distance when living in their previous suburban neighborhood. This interviewee also stated that he and his family would be more dependent on car travel if they lived in a suburb or in the second-order center Sandnes. On the other hand, a retired couple had replaced driving with walking when buying daily necessities as a result of moving from a suburb to the downtown of Sandnes. They had also changed from car to train for their regular visiting trips to relatives in Stavanger.

Conversely, a female interviewee had changed from walking to driving for trips to virtually all intra-metropolitan activities after moving from the inner city to a suburban neighborhood. She said that she did not have a car when she lived at her previous residential address, and tried to live without having her own car the first year after moving to her present dwelling. However, she bought a car after realizing how inconvenient it was to reach her relevant destinations by transit from her new residential location. Another female interviewee who had moved from the inner city to a densified suburban neighborhood used to walk to such destinations before the move. After moving, she first increased her travel by car but shifted to predominantly going by transit for journeys to work and cultural events after realizing that the transit connections from her new dwelling were quite good. Her shift to bus and train reflected the transit-oriented character of her new local neighborhood, which was located close to a train station and bus stop in the major transit corridor of the region. On the other hand, the move had made her son more dependent on car travel. Her husband continued to commute by car despite having moved closer to his job in the suburban second-order center Forus, where the transportation infrastructure is very car-oriented with ample parking, wide roads, and poor transit level of service.

In addition to the above-mentioned main rationales, several secondary rationales also play some role for the interviewees' travel mode choices, often in combination with one or more of the main rationales: a wish for physical exercise while traveling, limitation of travel expenses, safety, social contact and caretaking during the journey, long-term habits, environmental concerns, and esthetics. However, these rationales only very modestly affect the relationships between characteristics of the built environment and travel mode choices.

In summary, the transport rationales encountered in the interviews contribute to longer travel distances among those living far from the city center of Stavanger and the Forus employment center. Moreover, they encourage interviewees living in the inner city of Stavanger to choose non-motorized modes rather than car, especially for moderate-length trips, distinct from the suburban residents for whom the rationales tend to encourage car travel for most trip purposes. The rationales thus offer improved evidence of causal mechanisms underlying the statistical relationships between residential location and travel. The evidence of causality is further strengthened through quasi-longitudinal analyses of travel behavior among Stavanger respondents who had moved to their current dwelling less than two years ago (not shown in this paper but presented in other articles from the same study). These analyses show that moving further away from the city center tended to increase commuting distances and trip distances for most non-work purposes (Næss et al., 2019), driving distances (Næss et al., 2017) as well as auto ownership (Cao et al., 2019), whereas inward moving had the opposite effects.

5 Relative importance and nonlinear relationships

We estimated three models with the GBDT method: a weekly driving distance model, a car commuting frequency model, and a car ownership model. We included the same set of independent variables in these models: socio-demographics and built environment variables. We present the results of relative importance and nonlinear relationships in this section. Finally, we compare the results in Stavanger with those in a few other, similar-sized Nordic cities.

5.1 Relative importance

Table 3 presents the relative importance for all independent variables in the three models. The relative importance of socio-demographics and built environment variables varies by type of behavior.

In the driving distance and car commuting frequency models, the collective relative importance of the built environment variables is slightly higher than that of socio-demographics. In the car ownership model, the built environment totally has smaller relative importance than socio-demographic characteristics. These results show that built environment attributes contribute more to car usage, while socio-demographics contribute more to car ownership.

Among the built environment variables, residential distance to the city center is the most influential one on driving. This is because the driving conditions are worse, the transit provision is better, and the number of trip destinations within acceptable walking or cycling distance is higher (especially for non-work activities but also to a great extent for working) for residents living in the areas close to the city center of Stavanger. Car trips will also on average be shorter for these residents to non-work activities. Car trips for commuting will be even shorter on average for those living close to the Forus second-order center, but those who live at the outskirts of the metropolitan area have a long driving distance to Forus as well as to the city center of Stavanger. In the three models, distance to local center has a larger relative importance than distance to second-order center. This is because, compared to second-order center, people would be more likely to travel to local center for activities such as grocery stores, primary schools, and kindergartens based on the rationale of minimizing friction of distance. The relative importance of socio-demographics varies with the type of driving-related behavior. For example, personal income is influential on weekly driving distance and frequency of car commuting. However, its relative importance on car ownership is rather small. On the other hand, household size and income are the variables showing the strongest influence on car ownership.

Table 3. Relative importance of variables in the models

	Driving distance model (N = 1,174)			Car commuting frequency model (N = 973)			Car ownership model (N = 1,208)		
	Relative importance (%)	Rank	Sum (%)	Relative importance (%)	Rank	Sum (%)	Relative importance (%)	Rank	Sum (%)
Socio-demographics									
Personal income	13.9	2	49.8	9.9	4	46.4	3.2	10	58.6
Gender	11.8	3		1.6	12		0.5	13	
Age	11.6	4		15.3	2		9.8	4	
Household income	5.4	9		6.7	8		18.6	2	
Education	2.9	10		5.8	9		4.0	9	
Household size	2.4	11		2.8	11		18.6	1	
Workforce	0.8	12		2.9	10		0.5	14	
Teenager	0.8	13		1.1	13		1.7	11	
Children	0.3	14		0.3	14		1.6	12	
Built environment characteristics									
Distance to city center	19.7	1	50.2	15.9	1	53.6	10.8	3	41.4
Employment density	8.8	5		11.9	3		7.0	7	
Distance to local center	7.8	6		9.0	5		8.4	5	
Population density	7.6	7		8.3	7		8.3	6	
Distance to second-order center	6.4	8		8.5	6		6.9	8	

Note: The relative importance of all independent variables adds up to 100% in each model.

5.2 Relationships between the built environment and travel behavior

The figures below show the relationships between built environment variables and people's driving-related behavior. We applied the same scale on the y axes of the plots generated from the same model. Overall, driving-related behavior is positively correlated with the three distance variables and negatively correlated with the two density variables. These relationships are consistent with the literature (Ewing & Cervero, 2010). Besides, there exist clear nonlinear and threshold relationships, which can be explained by people's transport rationales and the geographical configuration of jobs and other facilities in Stavanger.

The relationships between distance and driving-related behavior are shown in Figure 7. There are two important thresholds in these three relationships: one is at 18 km and the other is between 8 and 10 km.

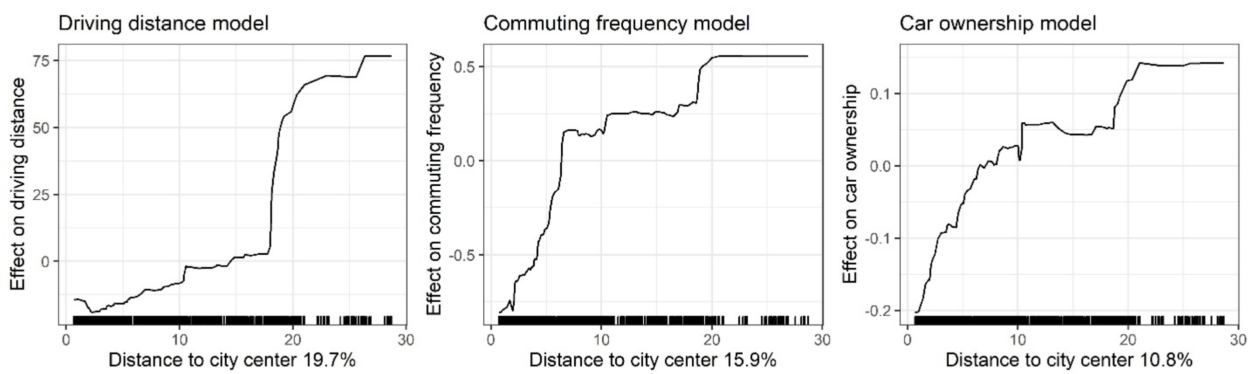

Figure 7. Relationships between distance to city center and travel behavior in Stavanger (The ticks on the x axes show the distribution of the corresponding built environment variable.)

The threshold of 18 km is shared by all three types of driving-related behavior. After 18 km, people's driving distance, car commuting frequency, and car ownership all have a substantial increase, especially for driving distance. The choice of workplace location is the activity location where the relative importance of the rationale of choosing the best facility is at its highest, compared to the rationale of minimization of the friction of distance. The implication of this is that most workers, especially those with specialized job qualifications, are willing to search for jobs within a large radius from their place of residence. As can be seen in Figure 8, the main concentrations of jobs are located within the semi-elliptic corridor including the inner districts of Stavanger (27,000 jobs), the Forus area (45,000 jobs), and the inner districts of Sandnes (10,000 jobs), stretching out to around 18 km from the city center of Stavanger. This semi-ellipse also includes the Ullandhaug area where the University of Stavanger is located. Residents living within this area will have a relatively equally large opportunity of finding a job within a moderate commuting distance even if they place more importance on the quality of the job (regarding job content, salary, terms of employment, etc.) than to proximity to the residence. For workers living outside this semi-elliptic corridor, the likelihood of finding a job satisfying the 'best facility' criteria within a moderate distance from home will be much smaller. In addition, transit provision is considerably poorer outside the corridor, and transit commuting will therefore be more time-consuming and less convenient than car travel for workers living in these parts of the metropolitan area. Roads are rarely congested in the outskirts, and car commuting will therefore also be less frustrating than in the more central parts of the metropolitan area. Given the longer average commuting distances when living outside the semi-elliptic corridor, the attractiveness of non-motorized travel modes for commuting will also be lower. Thus, the main rationales for job location (choosing the best facility) as well as for travel mode choice (convenience and comfort, timesaving, and frustration aversion) identified in the interviews contribute to higher demand of auto usage and ownership among residents living outside the semi-elliptic corridor shown in Figure 8. Most of these residents live more than 18 km away from Stavanger's city center. When distance to city center is longer than 18 km, the relationship with driving distance increases faster than those with car commuting frequency and car ownership. This is because the trip distances increase substantially as the respondents live far away from the city center of Stavanger as well as from the Forus area.

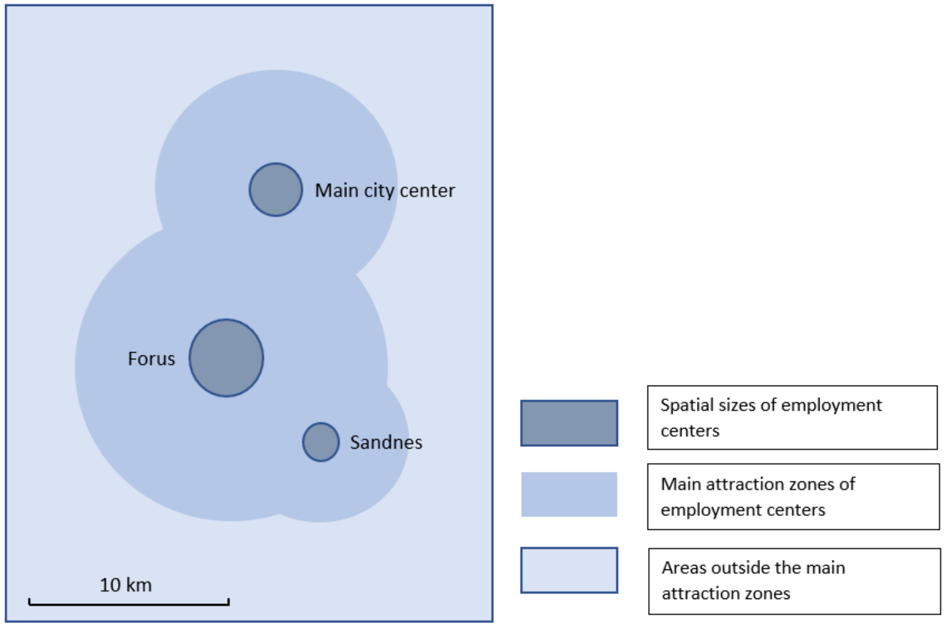

Figure 8. Conceptual illustration of the locations and spatial sizes of employment centers, main attraction zones of employment centers, and areas outside the main attraction zones

The other threshold occurs between 8 and 10 km from the city center, depending on the type of driving behavior. During the interval between the city center and this threshold, people's driving distance, car commuting frequency, and car ownership increase as distance to the city center increases. The relationship with driving distance increases more slowly than those with car commuting frequency and car ownership within a short distance from the city center. Although people's car commuting frequency and car ownership increase quickly as they live farther from the city center, their trip distances are still often short since many activity destinations are nearby in these areas. In particular, this applies to the commuting distances of those who live near the Forus employment center, which is located around 9 km from the city center of Stavanger. The increased propensity of choosing car as travel mode when the distance to the city center of Stavanger increases is counteracted by a tendency of reduced commuting distances among those who live close to the Forus area. Therefore, the corresponding driving distances increase slowly. Between the 8-10 km threshold and 18 km, the influence of distance to city center on driving-related behavior becomes moderate or even trivial. In the area between this threshold and 18 km from the city center, people live farther from the city center but closer to either Forus or Sandnes. In this case, based on the rationales of minimizing the friction of distance, people may choose the jobs especially in Forus but also in Sandnes, as these employment centers are closer. Thus, their driving demand for commuting may be lower. At the same time, based on the rationale of choosing the best facility, they have higher driving demand to the city center for other purposes as it has the highest concentration of attractive non-work destinations, such as cultural arenas, restaurants, bars, etc. People's driving demand is dynamically stable under the influence of both the city center and Forus or Sandnes in areas from somewhere between 8 and 10 km to 18 km.

The relationships of distance to second-order center and distance to local center are shown in Figure 9 and Figure 10, respectively. Generally, the threshold of distance to second-order center is located somewhere between 2.5 and 5 km. The two second-order centers (Forus and Sandnes), and particularly Forus, attract workers from far beyond the local neighborhoods. Since especially Forus but also Sandnes has few non-work facilities, they do not attract many non-work trips (and for Forus, the number of non-work trip destinations are very few). On average for all trips, then, the reduction in travel distance by living close to one of the two second-order centers is therefore more moderate. And for driving distance, the strongly car-oriented road and parking situation at Forus contributes to an additional diminishing of the reduction in driving distance resulting from living close to this second-order center. In terms of travel modes, the effect of living close to Forus applies mainly to the use of non-motorized modes, since transit accessibility to Forus is very poor. Therefore, the threshold plausibly reflects the average acceptable distance for walking and biking, which are driven by the convience and time-saving rationales for travel modes choice. The threshold of distance to local center is somewhere between 0.9 and 1.5 km. Note that a spike occurs between 1 km and 2 km in the relationship between distance to local center and car ownership for unknown reasons. The much shorter distance ranges within which any effects of proximity to local centers can be observed reflects that the facilities existing in these centers are mainly not very specialized ones, i.e., facilities for which minimizing the friction of distance normally plays a stronger role than the 'best facility' rationale. The attraction zone of such centers is therefore rather small, and this explains why the curves of travel behavior flatten already after a few km from these centers if any effect can at all be observed. The relatively large number of local centers also implies that most residents do not live very far from the closest center. To the extent that these centers include employment opportunities, the jobs normally attract workers from far beyond the closest surroundings of these centers and will therefore not affect the average travel distance and mode among the residents of these center's main (non-work) attraction zone much.

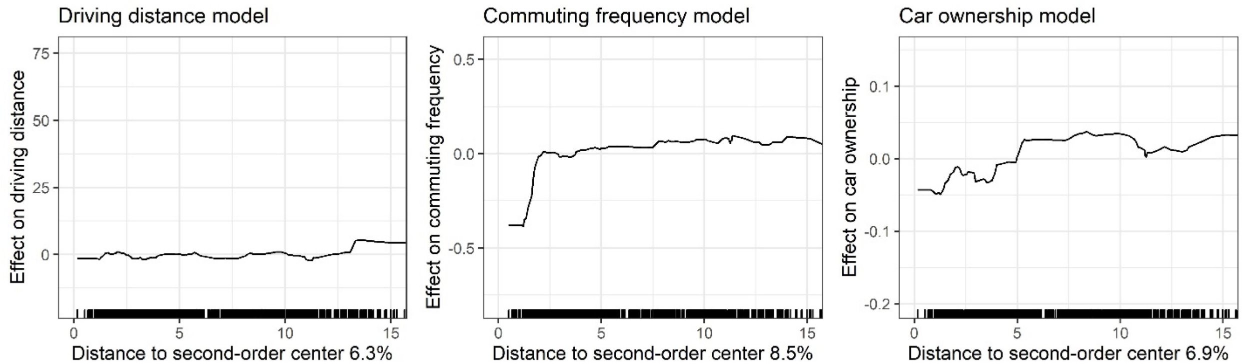

Figure 9. Relationships between distance to second-order center and travel behavior in Stavanger (The ticks on the x axes show the distribution of the corresponding built environment variable.)

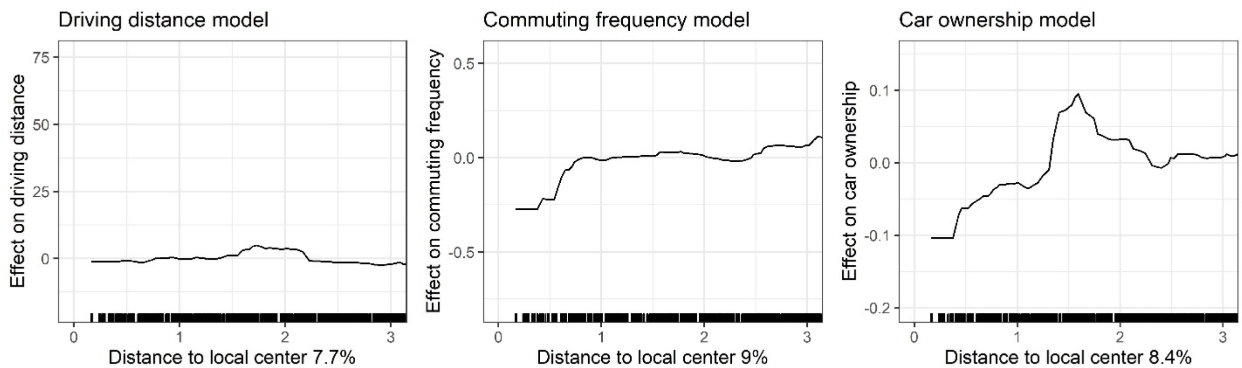

Figure 10. Relationships between distance to local center and travel behavior in Stavanger (The ticks on the x axes show the distribution of the corresponding built environment variable.)

We presented the relationships of two density variables in Figure 11 and Figure 12, respectively. Effects of job density and population density are moderate and occur mainly in the low-density part of the range. When population density is smaller than 15 people per hectare or employment density is smaller than 5 jobs per hectare, people's driving distance, car commuting frequency, and car ownership are generally higher. This reflects a tendency of the areas with the very lowest densities to have very poor transit service with few departures and long average distances to stops. For residents of these parts of the metropolitan area, transit is therefore normally perceived as inconvenient and very time-consuming, resulting in high car dependency.

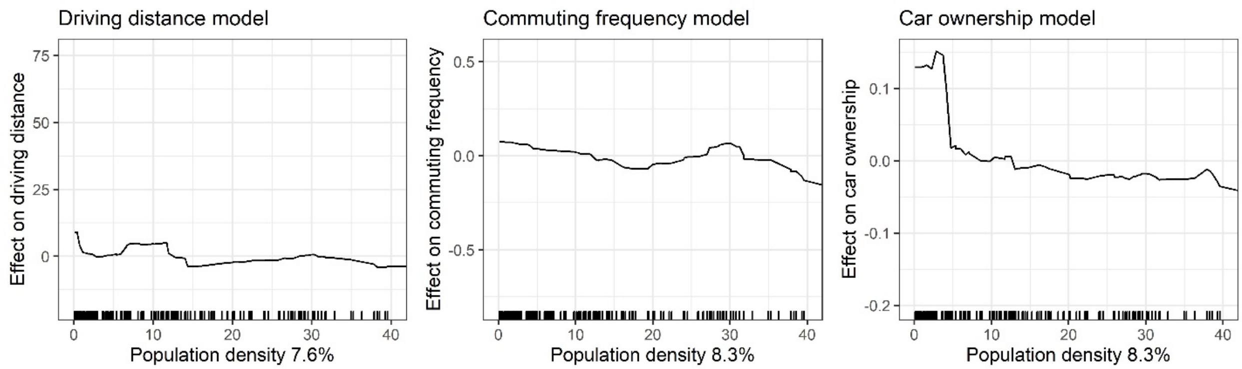

Figure 11. Relationships between population density and travel behavior in Stavanger (The ticks on the x axes show the distribution of the corresponding built environment variable.)

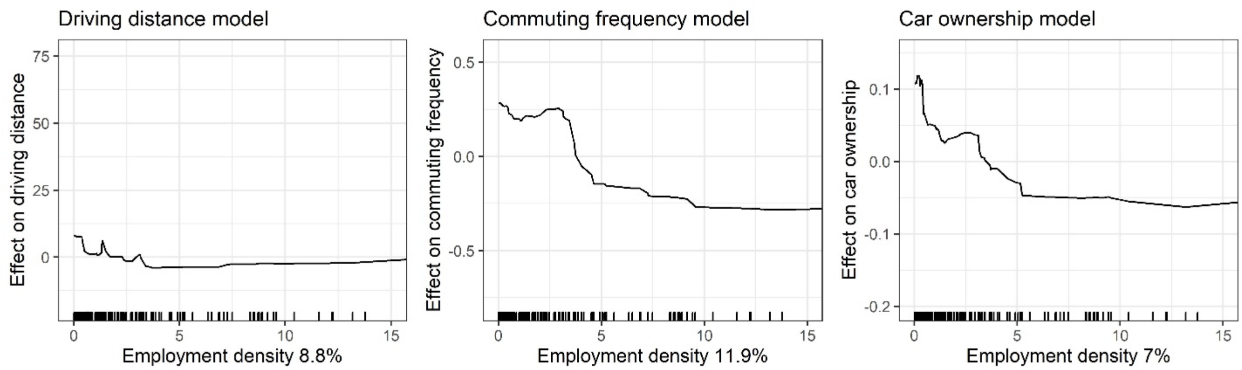

Figure 12. Relationships between employment density and travel behavior in Stavanger (The ticks on the x axes show the distribution of the corresponding built environment variable.)

5.3 Discussions

In this study, we found that residential distance to the main city center plays the most important role among the built environment variables in influencing people's driving-related travel behavior in Stavanger. This finding is consistent with several studies on Nordic cities with similar population size. In Reykjavik (Næss et al., 2021), which is an Icelandic city with 232,000 people, residential distance was found to be the dominant built environment characteristic influencing car-driving distance and especially the frequency of car commuting, both with a strong and highly significant effect ($p < 0.001$ for both variables). For frequency of car commuting, residential distance to the city center was the only built environment characteristic showing a significant effect. For car-driving distance, there was also a significant effect of distance to the local center ($p = 0.002$), but considerably weaker than that of distance to the main city center. Studies of residential built environment characteristics and travel in the similar-sized Norwegian cities of Bergen (population: 265,000) and Trondheim (population:

192,000) (Engebretsen et al., 2018) also show effects on car-driving distance mainly of residential distance to the city center.

With nonlinear relationships, this study found that there exist clear threshold effects of built environment attributes on driving behavior in Stavanger, which has not been uncovered in similar studies on small cities. For distance to city center, Stavanger has two clear thresholds. One is between 8 to 10 km, and the other is 18 km from the city center. These thresholds, as explained in the last section, are mainly due to the polycentric city structure. The other two distance variables and two density variables also present clear thresholds.

6 Conclusion

Using data from a small European city, Stavanger, this study applied a mixed-methods approach to study the nonlinear and threshold relationships between built environment attributes and three types of driving-related travel behavior, including weekly driving distance, car commuting frequency, and car ownership. To our best knowledge, this is the first study that uses both qualitative interview analysis and quantitative machine learning approach to examine the nonlinear influence of the built environment on driving-related travel behavior and explore the causal explanations of these nonlinear effects in a small-city context.

Transport rationales for choosing activity locations and travel modes, along with configurations of the jobs and other facilities, provide causal explanations of the nonlinear and threshold effects of the built environment attributes on people's driving-related travel behavior. The influence of distance to city center has thresholds at somewhere between 8 and 10 km and 18 km, which can be explained by the main transport rationales and the polycentric city structure of Stavanger. The influences of distance to second-order center and local center have thresholds at somewhere between 2.5 and 5 km and somewhere between 0.9 and 1.5 km, which can mainly be explained by minimization of the friction of distance, convenience, and time-saving. The influences of two density variables have thresholds at lower density values where the population base for public transit becomes very poor. These thresholds can thus be explained by the rationales of convenience and comfort and time-saving. Second, distance to city center plays the most important role among considered built environment variables, which is consistent with the findings of the limited number of other studies on small cities.

The results will be very useful to provide policy implications for Stavanger and similar cities to reduce the auto dependence of their residents. First, compact development around the main city center helps to rein the auto use as distance to city center contributes most to people's driving and people who live close to city center generally have a lower level of driving and car ownership. Such policies include population densification and increasing new employment opportunities around the main city center. Second, thresholds generated from the nonlinear relationships provide planning guidelines to carry out the policies of compact development. The nonlinear relationship of distance to main city center indicates that areas within 18 km from the city center and inside the semi-elliptic demarcation shown in Figure 3 should be prioritized to carry out compact development. The thresholds related to population density and employment density show that densities below 15 people and 5 jobs per hectare make people very dependent on car travel. For small cities, their financial capacities are weaker than those of large cities. These planning guidelines help them to use limited resources to reduce driving through efficient planning policies.

References

Apley, D. (2018). *ALEPlot: Accumulated Local Effects (ALE) Plots and Partial Dependence (PD) Plots*. https://cran.r-project.org/package=ALEPlot

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(4), 1059-1086. https://doi.org/10.1111/rssb.12377

Boarnet, M. G., & Sarmiento, S. (1998). Can Land-use Policy Really Affect Travel Behaviour? A Study of the Link between Non-work Travel and Land-use Characteristics. *Urban Studies*, *35*(7), 1155-1169. https://doi.org/10.1080/0042098984538

Cao, X., Næss, P., & Wolday, F. (2019). Examining the effects of the built environment on auto ownership in two Norwegian urban regions. *Transportation Research Part D: Transport and Environment*, *67*, 464-474. https://doi.org/10.1016/j.trd.2018.12.020

Cheng, L., De Vos, J., Zhao, P., Yang, M., & Witlox, F. (2020). Examining non-linear built environment effects on elderly's walking: A random forest approach. *Transportation Research Part D: Transport and Environment*, *88*, 102552. https://doi.org/10.1016/j.trd.2020.102552

Christaller, W. (1966). *Central places in southern Germany*. Englewood Cliffs.

Clifton, K. J., & Handy, S. L. (2003). Qualitative Methods in Travel Behaviour Research. In P. Jones & P. R. Stopher (Eds.), *Transport Survey Quality and Innovation* (pp. 283-302). Emerald Group Publishing Limited. https://doi.org/10.1108/9781786359551-016

Ding, C., Cao, X., & Wang, Y. (2018). Synergistic effects of the built environment and commuting programs on commute mode choice. *Transportation Research Part A: Policy and Practice*, *118*, 104-118. https://doi.org/10.1016/j.tra.2018.08.041

Ding, C., Cao, X. J., & Næss, P. (2018). Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transportation Research Part A: Policy and Practice*, *110*, 107-117. https://doi.org/10.1016/j.tra.2018.02.009

Eldridge, J. D., & Jones, J. P. (1991). Warped Space: A Geography of Distance Decay. *The Professional Geographer*, *43*(4), 500-511. https://doi.org/10.1111/j.0033-0124.1991.00500.x

Engebretsen, Ø., Næss, P., & Strand, A. (2018). Residential location, workplace location and car driving in four Norwegian cities. *European Planning Studies*, *26*(10), 2036-2057. https://doi.org/10.1080/09654313.2018.1505830

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, *76*(3), 265-294. https://doi.org/10.1080/01944361003766766

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189-1232. https://doi.org/10.1214/aos/1013203451

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367-378. https://doi.org/10.1016/s0167-9473(01)00065-2

Greenwell, B., Boehmke, B., & Cunningham, J. (2020). *gbm: Generalized Boosted Regression Models*. https://cran.r-project.org/package=gbm

Hägerstrand, T. (1970). What about people in Regional Science? *Papers of the Regional Science Association*, *24*(1), 6-21. https://doi.org/10.1007/bf01936872

Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment*, *10*(6), 427-444. https://doi.org/10.1016/j.trd.2005.05.002

Heres-Del-Valle, D., & Niemeier, D. (2011). CO2 emissions: Are land-use changes enough for California to reduce VMT? Specification of a two-part model with instrumental variables. *Transportation Research Part B: Methodological*, *45*(1), 150-161. https://doi.org/10.1016/j.trb.2010.04.001

Krizek, K. J. (2003). Residential Relocation and Changes in Urban Travel: <i>Does Neighborhood-Scale Urban Form Matter?</i>. *Journal of the American Planning Association*, *69*(3), 265-281. https://doi.org/10.1080/01944360308978019

Lin, T., Wang, D., & Guan, X. (2017). The built environment, travel attitude, and travel behavior: Residential self-selection or residential determination? *Journal of Transport Geography*, *65*, 111-122. https://doi.org/10.1016/j.jtrangeo.2017.10.004

Lloyd, P., & Dicken, P. (1972). Location in space: a theoretical approach to economic geography. In: Harper & Row Publishers.

Molnar, C. (2020). *Interpretable Machine Learning - A Guide for Making Black Box Models Explainable*. lulu.com. https://christophm.github.io/interpretable-ml-book/

Næss, P. (2005). Residential location affects travel behavior—but how and why? The case of Copenhagen metropolitan area. *Progress in Planning*, *63*(2), 167-257. https://doi.org/10.1016/j.progress.2004.07.004

Næss, P. (2013). Residential location, transport rationales and daily-life travel behaviour: The case of Hangzhou Metropolitan Area, China. *Progress in Planning*, *79*, 1-50. https://doi.org/10.1016/j.progress.2012.05.001

Næss, P. (2015). Built Environment, Causality and Travel. *Transport Reviews*, *35*(3), 275-291. https://doi.org/10.1080/01441647.2015.1017751

Naess, P. (2015). Residential location and travel behavior. In P. Pinho & C. Silva (Eds.), *Mobility Patterns and Urban Structure* (pp. 151-186). Ashgate Publishing.

Næss, P., Cao, X. J., & Strand, A. (2017). Which D's are the important ones? The effects of built environment characteristics on driving distance in Oslo and Stavanger. *Journal of Transport and Land Use*, *10*(1), 945-964.

Næss, P., & Jensen, O. B. (2005). *Bilringene og cykelnavet: Boliglokalisering, bilafhængighed og transportadfærd i Hovedstadsområdet*. Aalborg Universitetsforlag.

Næss, P., Peters, S., Stefansdottir, H., & Strand, A. (2018). Causality, not just correlation: Residential location, transport rationales and travel behavior across metropolitan contexts. *Journal of Transport Geography*, *69*, 181-195. https://doi.org/10.1016/j.jtrangeo.2018.04.003

Næss, P., Stefansdottir, H., Peters, S., Czepkiewicz, M., & Heinonen, J. (2021). Residential Location and Travel in the Reykjavik Capital Region. *Sustainability*, *13*(12), 6714. https://doi.org/10.3390/su13126714

Næss, P., Strand, A., Wolday, F., & Stefansdottir, H. (2019). Residential location, commuting and non-work travel in two urban areas of different size and with different center structures. *Progress in Planning*, *128*, 1-36. https://doi.org/10.1016/j.progress.2017.10.002

Ridgeway, G. (2020). *Generalized Boosted Models: A guide to the gbm package*. https://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf

Scheiner, J. (2010). Social inequalities in travel behaviour: trip distances in the context of residential self-selection and lifestyles. *Journal of Transport Geography*, *18*(6), 679-690. https://doi.org/10.1016/j.jtrangeo.2009.09.002

Tao, T. (2021). *The gradient boosting decision tree approach – Mathematical illustration*. https://vtao1989.github.io/blog/files/GBDT.pdf

Tao, T., Wu, X., Cao, J., Fan, Y., Das, K., & Ramaswami, A. (2020). Exploring the non-linear relationship between the built environment and active travel in the Twin Cities. *Journal of Planning Education and Research*, 1-16. https://doi.org/https://doi.org/10.1177/0739456X20915765

Tennøy, A., Knapskog, M., & Wolday, F. (2022). Walking distances to public transport in smaller and larger Norwegian cities. *Transportation Research Part D: Transport and Environment*, *103*. https://doi.org/10.1016/j.trd.2022.103169

Van De Coevering, P., Maat, K., & Van Wee, B. (2021). Causes and effects between attitudes, the built environment and car kilometres: A longitudinal analysis. *Journal of Transport Geography*, *91*, 102982. https://doi.org/10.1016/j.jtrangeo.2021.102982

Vance, C., & Hedel, R. (2007). The impact of urban form on automobile travel: disentangling causation from correlation. *Transportation*, *34*(5), 575-588. https://doi.org/10.1007/s11116-007-9128-6

Wang, K., & Ozbilen, O. (2020). Synergistic and threshold effects of telework and residential location choice on travel time allocation. *Sustain Cities Soc*, *63*, 102468. https://doi.org/10.1016/j.scs.2020.102468

Wolday, F., Cao, J., & Næss, P. (2018). Examining factors that keep residents with high transit preference away from transit-rich zones and associated behavior outcomes. *Journal of Transport Geography*, *66*, 224-234. https://doi.org/10.1016/j.jtrangeo.2017.12.009

Wolday, F., Næss, P., & Cao, X. (2019). Travel-based residential self-selection: A qualitatively improved understanding from Norway. *Cities*, *87*, 87-102. https://doi.org/10.1016/j.cities.2018.12.029

Wu, X., Tao, T., Cao, J., Fan, Y., & Ramaswami, A. (2019). Examining threshold effects of built environment elements on travel-related carbon-dioxide emissions. *Transportation Research Part D: Transport and Environment*, *75*, 1-12. https://doi.org/10.1016/j.trd.2019.08.018

Yang, J., Cao, J., & Zhou, Y. (2021). Elaborating non-linear associations and synergies of subway access and land uses with urban vitality in Shenzhen. *Transportation Research Part A: Policy and Practice*, *144*, 74-88. https://doi.org/10.1016/j.tra.2020.11.014

Yang, L., Ao, Y., Ke, J., Lu, Y., & Liang, Y. (2021). To walk or not to walk? Examining non-linear effects of streetscape greenery on walking propensity of older adults. *Journal of Transport Geography*, *94*, 103099. https://doi.org/10.1016/j.jtrangeo.2021.103099

Zhang, W., Zhao, Y., Cao, X., Lu, D., & Chai, Y. (2020). Nonlinear effect of accessibility on car ownership in Beijing: Pedestrian-scale neighborhood planning. *Transportation Research Part D: Transport and Environment*, *86*. https://doi.org/10.1016/j.trd.2020.102445