Urban Planning (ISSN: 2183–7635)
2023, Volume 8, Issue 3, Pages 41–55
https://doi.org/10.17645/up.v8i3.6293

Article

Investigating the Nonlinear Relationship Between Car Dependency and the Built Environment

Jun Cao [1], Tanhua Jin [2,3], Tao Shou [1,*], Long Cheng [2,4], Zhicheng Liu [5], and Frank Witlox [2,6]

[1] School of Architecture, Southeast University, China
[2] Department of Geography, Ghent University, Belgium
[3] Institute of Cartography and Geoinformation, ETH Zurich, Switzerland
[4] Jiangsu Key Laboratory of Urban ITS, Southeast University, China
[5] Department of Electronic Engineering, Tsinghua University, China
[6] Department of Geography, University of Tartu, Estonia

* Corresponding author (beyondshou@seu.edu.cn)

Submitted: 19 October 2022 | Accepted: 14 February 2023 | Published: 25 July 2023

Abstract
Car-dominated daily travel has caused many severe and urgent urban problems across the world, and such travel patterns have been found to be related to the built environment. However, few existing studies have uncovered the nonlinear relationship between the built environment and car dependency using a machine learning method, thus failing to provide policymakers with nuanced evidence-based guidance on reducing car dependency. Using data from Puget Sound regional household travel surveys, this study analyzes the complicated relationship between car dependency and the built environment using the gradient boost decision tree method. The results show that people living in high-density areas are less likely to rely on private cars than those living in low-density neighborhoods. Both threshold and nonlinear effects are observed in the relationships between the built environment and car dependency. Increasing road density promotes car usage when the road density is below 6 km/km^2. However, the positive association between road density and car use is not observed in areas with high road density. Increasing pedestrian-oriented road density decreases the likelihood of using cars as the main mode. Such a negative effect is most effective when the pedestrian-oriented road density is over 14.5 km/km^2. More diverse land use also discourages people's car use, probably because those areas are more likely to promote active modes. Destination accessibility has an overall negative effect and a significant threshold effect on car dependency. These findings can help urban planners formulate tailored land-use interventions to reduce car dependency.

Keywords
built environment; car dependency; machine learning; nonlinearity; Puget Sound; threshold effects

Issue
This article is part of the issue "Car Dependency and Urban Form" edited by Kobe Boussauw (Vrije Universiteit Brussel), Koos Fransen (Vrije Universiteit Brussel / Ghent University), and Enrica Papa (University of Westminster).

1. Introduction

During the past several decades, car use has become a severe problem across the world. For example, almost half of the trips in European countries (e.g., Germany, Switzerland, and Austria) are made by private car (Buehler et al., 2017). The growth rate of car ownership in China has also been dramatic, which is similar to the historical process of developed countries (International Monetary Fund, 2005). Car-dependent issues in the U.S. are even worse. The rate of car ownership in the U.S. ranked first in the world, significantly higher than that in other countries (Pucher & Lefevre, 1996). Low density and urban sprawl in the U.S. have led to severe car dependency issues (Gilbert & Perl, 2011) since facilities and services (e.g., healthcare, and shopping centers) are sparsely distributed and cannot be reached and served efficiently by public transit and/or active modes. The extensive use of cars across the world has resulted in severe problems, such as traffic congestion, air pollution, and noise pollution (McIntosh et al., 2014). Understanding what contributes to the decline in car dependency can help planners reduce the detrimental effects of car use.

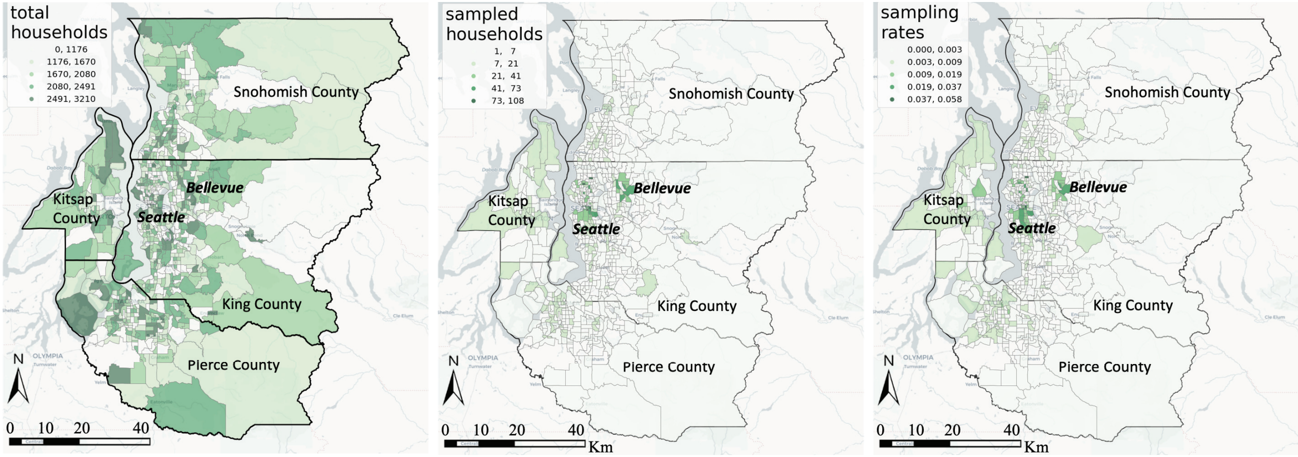
After "car dependency" was first introduced by Newman and Kenworthy (1989a, 1989b) in a study analyzing the relationship between travel patterns and land use factors from 32 global cities, extensive studies have discussed the influencing factors on car dependency. Socio-demographic characteristics can influence people's car use, such as age, gender, income level, education, and employment status (Naess, 2014). Manaugh et al. (2010) found that the number of automobile trips is positively associated with people's income level in Montreal, Canada. Another research in Detroit reached a similar finding and further found that education and employment status also has a positive effect on car use. People who have full-time jobs are more likely to use cars compared to those unemployed. How the built environment affects car use has also been extensively discussed in previous studies (Ding & Cao, 2019; Pinjari et al., 2011). Most existing studies concluded that built environments such as density, design, and destination accessibility have significant effects on car use. High density can lead to less car dependency (Van Acker & Witlox, 2010). Housing density has a negative impact on car dependency (Hong, 2017). Evidence from California witnessed that a decrease in density below 1,000 housing units per square mile is associated with a 5.5% increase in fuel consumption per household and a 4.8% increase in vehicle kilometers traveled (VKT) per capita (Zegras, 2010). Another study in Flanders, Belgium supported this finding that higher density increases the use of other modes, such as walking, cycling, and public transit (De Vos & Witlox, 2013). Car ownership is negatively associated with both residential density and employment density (Cervero & Arrington, 2008; Holtzclaw et al., 2002; Li et al., 2010). People living in areas with more diverse land uses are less likely to own a car (Potoglou & Kanaroglou, 2008). Those living in neighborhoods with pedestrian-friendly streets have fewer cars since these streets promote the use of non-motorized travel modes (e.g., walking, cycling; Zuo et al., 2018). Good access to transit services may encourage people to travel by public transit and thus decreases the possibility to use cars (Mavoa et al., 2012; McIntosh et al., 2014).

While most of the existing studies assumed a linear association between the built environment and car dependency (Van Acker & Witlox, 2010; Zegras, 2010), some researchers tried to uncover the nonlinear relationships between car use and urban form using exponential functions. Theoretical reasons for such nonlinear effects can be related to location theory and threshold theory for goods and services (Eldridge & Jones, 1991). For example, Newman and Kenworthy (1989a, 1989b, 1991, 2006, 2011a, 2011b) found that car use decreases exponentially with population density increasing by analyzing a group of global cities. The exponential function used by Newman and Kenworthy is one of the first attempts to uncover the nonlinear effects between car use and urban density. Exponential functions have been used in many previous studies to introduce nonlinearity (Holtzclaw et al., 2002), with the advantage of being smooth and differentiable and being able to derive the backpropagation algorithm. Unlike exponential function as a traditional statistical method that follows a constrained statistical assumption and is usually pre-defined, machine learning methods, such as the XGBoost model used in this research, are data-driven and are not statistically constrained, which will provide more sophisticated results. Many other researchers have also attempted to uncover the nonlinear built effects on travel patterns using machine learning methods, including driving distance (Ding et al., 2018), metro ridership (Ding et al., 2019), usage of shared mobility services (Cheng et al., 2023; Cheng, Wang, et al., 2022; Jin, Cheng, Zhang, et al., 2022), and public transit ridership (Chen et al., 2021). Relaxing the assumption of linearity using a machine learning method has several advantages in travel behavior analysis (Cheng et al., 2019; Liu et al., 2021; Xu et al., 2021; Zhang et al., 2020). First, former studies that assume linear relationships can only uncover a negative or positive effect of the influencing factors on travel behavior (Boarnet et al., 2008; Van Acker & Witlox, 2010; Zegras, 2010). The nonlinear relationships can illustrate a more complex relationship instead of a monotonous trend or effect. Moreover, the nonlinear relationships captured by machine learning methods can present more accurate estimates of the effects of influencing factors within different intervals of associated factors on travel behavior, which can help policymakers make targeted policies. This study, taking the Puget Sound Region, U.S, uses a machine learning method to explore the nonlinear associations between the built environment and car dependency.

The rest of this article is as follows. Section 2 introduces the data and variables. Section 3 explains how the gradient boost decision tree (GBDT) can be used to analyze nonlinear relationships. Section 4 discusses the nonlinear effects of the built environment on car dependency. Section 5 summarizes this research and proposes future research avenues.

2. Data and Variables

This study is based on the Puget Sound Region Travel Surveys from 2017 to 2021. The Puget Sound region (Figure 1) is in the U.S. state of Washington and consists of King, Kitsap, Pierce, and Snohomish counties, with the city of Seattle located in the region. The region includes 82 cities and towns with a total of over four million people and 1.5 million households (Figure 1a). As illustrated



(a) Total households **(b)** Sampled households **(c)** Sampling rates

Figure 1. Spatial distribution of total households, sampled households, and sampling rates in the Puget Sound Region.

in Figure 2, this region has multiple types of neighborhoods, such as high-density neighborhoods in downtown areas of Seattle, and low-density neighborhoods in Parkwood, Kitsap County. The surveys collected socio-demographic and geographic information about individuals and households, as well as detailed travel information. There are 136,079 trips involved in this research, which contains 8,287 households and 14,112 individuals. Travel information includes the number of trips, travel time, and travel mode. The travel surveys aim to help local and regional planning agencies prioritize transportation and land-use improvements. It should be noted that the Puget Sound Region Travel Survey uses a stratified address-based random sampling method, which combines proportional geographic sampling and compensatory sampling based on predicted response rates and targeted oversampling. Low-income households, those with no vehicles, and non-auto commuters are more targeted for policy goals (Puget Sound Regional Council, 2021). As illustrated in Figure 1c, those census tracts that have high sampling rates are located in the city of Seattle and Bellevue, two of the largest cities in the region. Since this research focuses on explaining the relationships between built environment variables and car use rather than on describing car use per se, these differences are not expected to materially affect the results (Babbie, 2009).

The dependent variable is whether a car is used as the main mode during one trip. It is a dummy variable, with one indicating that a car is used as the main mode, while zero otherwise. Among all trips surveyed, 63.47% of the trips use a car as the main mode while 36.53% use other modes. The explanatory variable is built environment attributes while the control variables include individuals' socioeconomic and demographic characteristics, household characteristics, and trip purposes (Table 1). While characteristics of individuals and households, as well as trip purposes, are sourced from the travel survey, built environment characteristics are collected from the U.S. Environmental Protection Agency's Smart Location Dataset (SLD), OpenStreetMap, and GTFS dataset. The SLD can be downloaded using the following link: https://www.epa.gov/smartgrowth/smart-location-database-technical-documentation-and-user-guide. The SLD data are all aggregated at the census block level with United States customary units (i.e., miles). Since the geographic information of the Puget Travel Survey is based on the census tract level, the SLD-sourced variables are conversed to the census tract level using the weighted average values with SI units (i.e., kilometers).

(a) High-density area (Downtown area of Seattle) **(b)** Median-density area (Everett, Snohomish County) **(c)** Low-density area (Parkwood, Kitsap County)

Figure 2. Representative photos of high-, median-, and low-density neighborhoods in the Puget Sound Region.

Table 1. Variable definition and descriptive statistics.

Variable	Frequency	Percentage
Dependent variables		
Whether or not a car is used as the main mode during one trip		
Yes (= 1)	86,376	63.47%
No (= 0)	49,703	36.53%
Independent variables		
Individual's socioeconomic and demographic characteristics (N = 14,112)		
Age		
16–34	4,440	31.46%
35–54	4,400	31.18%
55+	5,272	37.36%
Gender		
Male	6,808	48.24%
Not Male	7,304	51.76%
Education (Bachelor's degree or higher)		
Yes	8,684	61.54%
No	6,808	48.24%
License (Valid driver's license ownership)		
Yes	11,668	82.68%
No	2,444	17.32%
Household characteristics (N = 8,287)		
Household size		
1	3,048	36.78%
2	3,336	40.26%
3	957	11.55%
4+	946	11.42%
Household income		
Under $49,999	2,066	24.93%
$50,000–$99,999	2,359	28.47%
$100,000 or more	3,336	40.26%
Prefer not to answer	526	6.35%
Vehicle ownership		
0	1,222	14.75%
1 vehicle	3,796	45.81%
2 or more vehicles	3,269	39.45%
Residential type		
Single-family house	3,234	39.02%
Apartment/condo/others	5,053	60.98%
House ownership		
Yes	4,170	50.32%
No	4,117	49.68%

Table 1. (Cont.) Variable definition and descriptive statistics.

Variable	Frequency		Percentage	
Trip purpose (N = 136,079)				
Trip purpose of origins *				
Home	44,234		32.51%	
Work	15,156		11.14%	
Work-related	5,338		3.92%	
School	2,981		2.19%	
Meal	11,244		8.26%	
Shop	14,744		10.83%	
Social/recreation	19,708		14.48%	
Escort	7,323		5.38%	
Change mode	773		0.57%	
Errand/other	13,892		10.21%	
Non-response	686		0.50%	
Trip purpose of destinations *				
Home	43,879		32.25%	
Work	15,106		11.10%	
Work-related	5,383		3.96%	
School	2,991		2.20%	
Meal	11,265		8.28%	
Shop	14,775		10.86%	
Social/recreation	20,278		14.90%	
Escort	7,345		5.40%	
Change mode	780		0.57%	
Errand/other	13,783		10.13%	
Non-response	494		0.36%	
Built environment variables (Census Tract level)	Mean	Min	Max	Std
Density				
Residential density (10^3 housing units/km^2) **	0.97	0.001	15.82	1.49
Employment density (10^3 jobs/km^2) ***	1.50	0.00	124.36	6.90
Design				
Road density (km/km^2) ****	0.65	0.00	12.36	1.37
Intersection density (counts/km^2) ****	1.07	0.00	68.97	4.37
Pedestrian-oriented road density (km/km^2) *****	9.27	0.32	24.82	5.02
Building density (km^2/km^2) ****	0.16	0.00	0.49	0.08
Diversity				
Land use mix ******	0.70	0.23	0.96	0.11
Destination accessibility				
Transit service frequency (counts/km^2) *******	82	0.00	4567	372
10^3 jobs reached by public transit within 45 minutes ********	130.19	0.00	1121.77	194.23

Notes: * Trip purpose of origins (home) and destinations (work) means the respondent leaves home for the workplace. ** Pedestrian-oriented road density is network density in terms of facility kilometers of pedestrian-oriented links per square kilometer. It is sourced from D3apo in the SLD. Pedestrian-oriented facilities refer to any link having a low speed and pedestrian is permitted. *** Residential density is sourced from D1a in the SLD, which is the gross residential density (Housing Units/km^2) on unprotected land. **** Road density, intersection density, and building density variables are sourced from OpenStreetMap. ***** Employment density is sourced from D1c in the SLD, which is the gross employment density (jobs/km^2) on unprotected land. ****** Since land use data are not accessible for this region, employment entropy is used to represent the land use mix (Ozbilen et al., 2021; U.S. Department of Transportation, 2015). Employment entropy is sourced from D2b_E8Mix based on eight employment categories, including retail, office, service, industrial, entertainment, education, healthcare, and public administration. The entropy mixture of employment types can be calculated as: $H = -(\sum_{i=1}^{n} p_i * \ln(p_i)) / \ln(n)$, where p_i represents the share of each employment category i; and n is the number of employment types in each census tract. The value ranges from 0 to 1. The larger the value, the more mixed the job types are. A higher employment entropy can be assumed to represent more diverse land uses. ******* Transit service frequency per square kilometer is sourced from D4d in the SLD, which calculates the frequency of public transit services for each transit route during the weekday evening peak hour (from 4:00 PM to 7:00 PM). Transit stops within 0.4 kilometers of crow-fly distance from the boundary of the census block group were identified. ******** Jobs reached by public transit within 45 minutes variable is sourced from D5br in the SLD. It is distance decay weighted, which considers walking network travel time and GTFS schedules simulation.

3. Methodology

3.1. XGBoost Model

The GBDT method is a tree-based ensemble machine learning method (Friedman, 2001). As illustrated in Figure 3, multiple decision trees are built iteratively, and the outcomes of all trees are then combined to construct the final model. Each single tree aims to minimize a loss function, with more weights assigned to cases with a wrong prediction. The GBDT method has the advantage of excellent prediction power, making it one of the most widely recognized machine learning methods. The XGBoost is an advanced tree learning algorithm (Chen & Guestrin, 2016), which is able to deal with sparse and parallel data with a high computation speed. These improvements have made XGBoost a reputational and popular machine learning method in data science. While XGBRegressor is used for continuous outcome variables, XGBClassifier is used for categorical outcome variables.

For each tree, an outcome (i.e., whether a car is used as the main mode during a trip) y_i exists. The XGBoost model is built based on the features and K additive functions:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(X_i), \quad f_k \in \mathrm{F} \tag{1}$$

where f_k is a tree with leaf weights, and F indicates the space of decision trees. For each tree, the aim is to minimize the following:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{2}$$

where l is the difference between $\hat{y}_i$ and y_i. Ω is a term that penalizes the complexity of the model.

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \left\| \omega_i \right\| \tag{3}$$

$$\omega_i = -\frac{\sum_{i \in I_j} \partial^2_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right)}{\sum_{i \in I_j} \partial^2_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right) + \lambda} \tag{4}$$

where T indicates how many leave nodes in the tree, and ω_i represents the score of the i^{th} leaf, and γ and λ represent regularization parameters.

3.2. Interpretation of Results of the XGBoost

Explanatory variables are iteratively chosen randomly to construct a single decision tree in XGBoost. Relative importance is related to how many times a variable is selected to construct the model (Friedman, 2001). Relative importance is rescaled, the sum of which is one. Higher relative importance means a greater contribution of the variables. The relative importance of variable x_i can be obtained as follows:

$$I^2_{x_i} = \frac{1}{t} \sum_{k=1}^{t} I^2_{x_i}(T_k) \tag{5}$$

$$I^2_{x_i}(T_k) = \sum_{j=1}^{J} d_j \tag{6}$$

where J is the number of leaves in each tree; k is the number of additive trees; t is the number of iterations; T_k is the k^{th} tree function; d_j indicates the improvement in the square error term by making the j^{th} split based on the variable x_i.

How the outcome is influenced by independent variables can be illustrated by partial dependence plots (Tu et al., 2021). The x-axis represents the data distribution of the independent variable (Cheng et al., 2020). The partial dependence of $F(x)$ on x_S can be defined as follows:

$$F_{x_S}(x_S) = E_{x_C}\left[F(x_S, x_C)\right] = \int F(x_S, x_C) P(x_C)\, dx_C \tag{7}$$

$$F_{x_S}(x_S) = \frac{1}{n} \sum_{i=1}^{n} F(x_S, x_C^i) \tag{8}$$

where x_S are the features of which we want to estimate specific effects on car dependency and x_C are other variables; $P(x_C)$ is the probability density of x_C; n represents the number of samples.

Python "XGBoost" package is used to model the data. Model parameters are important for XGBoost, including

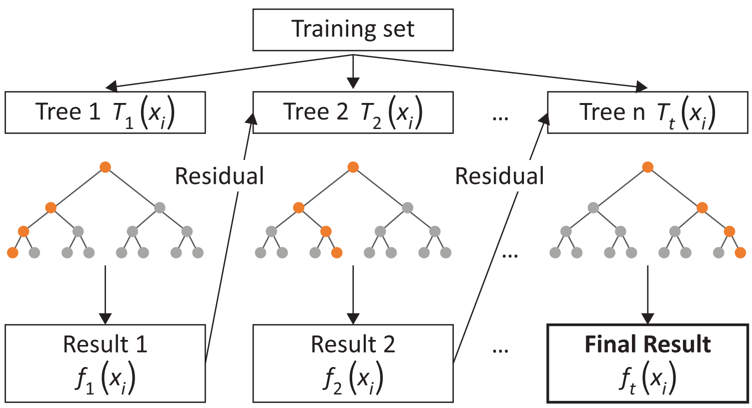


Figure 3. Schematic diagram of the GBDT method. Source: Authors based on Jin, Cheng, Liu, et al. (2022, p. 54).

the number of trees (n_estimators), shrinkage coefficient of each tree (learning_rate), and tree complexity (max_depth). Five-fold cross-validation was applied to search for optimum parameter values until the smallest F1 score occurs. Finally, the n_estimators, learning_rate, and max_depth was set as 200, 0.3, and 5 respectively for the model. To further help readers have a better understanding of the model performances of both the machine learning method and the binary regression model, we provide a table that illustrates more sophisticated performance metrics (i.e., precision, recall, F1_score, accuracy) for classification results. How these metrics can be calculated is illustrated from Equations 9 to 12, where TP indicates the correctly predicted positive class outcome of the model, TN demonstrates the correctly predicted negative class outcome, FP represents the incorrectly predicted positive class outcome, FN showing the incorrectly predicted negative class outcome. Precision is the rate of total correctly predicted instances of a class over total instances predicted as that class. Recall is the rate of total correctly predicted instances of a class over the total actual number of instances of that class. Accuracy is the rate of correctly predicted instances over the total number of instances. Accuracy represents a biased tendency towards the majority class in the imbalanced dataset as most of the data are from that class. Precision and recall can only illustrate the performance of each class. F1 score considers both values of precision and recall, and thus is regarded as a better representative model performance metric for the classification model. As illustrated in Table 2, all four model performance metrics of the XGBoost model are better than those of binary regression models.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

$$\text{F1\_score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN} \tag{12}$$

4. Results

4.1. Relative Importance of Independent Variables

Higher relative importance means a greater contribution of the variables. Regarding the average relative importance of different factor categories, household characteristics have the highest average relative importance, followed by destination accessibility and trip purpose (Table 3). In terms of the relative contribution of single variables, vehicle ownership is the most important variable, accounting for 33.27%. This is reasonable since people are more likely to use cars as their main travel mode when they have more cars in households (Buehler, 2011; Van Eenoo et al., 2022). Except for the highest contribution of vehicle ownership, the built environment has a higher average relative importance than individuals' and household socioeconomic and demographic characteristics. Some researchers have generally acknowledged the importance of socio-demographic characteristics in people's travel choices (e.g., Lanzendorf, 2010; Singh et al., 2018; Stead, 2001), such as the formulation of households and life domains. They claimed that individuals' travel behaviors are significantly influenced by their age, gender, and employment status. Others reached different findings that urban design and transportation infrastructure have a highly significant influence on car use, even after the correction for socio-economic effects (Holtzclaw et al., 2002; Lewis, 2018). This research aligns with the latter conclusion, which provides new insight into understanding the importance of the built environment on car dependency.

For built environment factors, destination accessibility variables have the highest relative importance, followed by design variables. The diversity variable has the lowest relative importance. In terms of single built environment variables, transit service frequency has the highest relative importance (12.03%), followed by pedestrian-oriented road density (7.74%). This is not surprising since transit service frequency may play a more important role in promoting people to use public transit while pedestrian-oriented road density also encourages people to take more active modes.

Table 2. Performance metrics of XGBoost and binary regression models.

		Recall	Precision	F1 score
Using car as a main mode	XGBoost	0.81	0.88	0.84
	Binary regression	0.74	0.85	0.79
Not using a car as a main mode	XGBoost	0.76	0.64	0.69
	Binary regression	0.64	0.48	0.55
Accuracy	XGBoost			0.79
	Binary regression			0.71

Table 3. Relative importance of independent variables.

Variable	Relative importance (%)	Average relative importance (%)
Individual's socioeconomic and demographic characteristics		
Age	1.74	2.32
Gender	1.77	
Education	3.47	
Employment	2.15	
License	2.49	
Household characteristics		
Household size	2.27	8.46
Household income	2.21	
Vehicle ownership	**33.27**	
Residential type	2.38	
House ownership	2.15	
Trip purpose		
Trip purpose of origins	4.37	4.55
Trip purpose of destinations	4.72	
Built environment variables		
Density		
Residential density (10^3 housing units/km^2)	2.70	2.40
Employment density (10^3 jobs/km)	2.51	
Building density (km^2/km^2)	2.00	
Design		
Road density (km/km^2)	2.31	4.12
Intersection density (counts/km^2)	2.30	
Pedestrian-oriented road density (km/km^2)	**7.74**	
Diversity		
Land use mix	2.07	2.07
Destination accessibility		
Transit service frequency	**12.03**	7.69
Jobs reached by public transit within 45 minutes	3.35	

4.2. Nonlinear Effects of the Built Environment Factors on Car Dependency

Partial dependence plots (Figures 4, 5, 6, and 7) are used to visualize the marginal effects of the built environment factors on car dependency (Tu et al., 2021). The x-axis presents the distributions of the built environment variables, and the x-axis presents the probability of using a car as the main mode. As illustrated in Figure 4a, car dependency is positively associated with residential density when the residential density is low. Such a positive effect turns into a negative one when the population density is high. Both high residential density and employment density will decrease people's car dependency, which aligns with previous research (Chatman, 2013; Newman & Kenworthy, 1989a, 1989b; Zegras, 2010). Increasing building density can also decrease people's car dependency. Cervero and Arrington (2008) found that there is a decline in car ownership as residential density increases. These neighborhoods may be equipped with more public services (i.e., healthcare, shopping center, and educational institutes), so that people may not need to drive a long distance to reach these public facilities. Moreover, densely populated neighborhoods are more likely to have more transportation facilities (e.g., buses, rails, shared services) so that people may have other travel options instead of car use. A significant threshold effect is observed when the population density is over 7,000 housing units per square kilometer. A significant decline is observed in the curve for employment density when the value is below 1,000 jobs. A similar threshold effect is also witnessed for the curve of building density. Newman and Kenworthy (2006) also found a threshold of the urban intensity (residents and jobs) at around 3,500 per square kilometer where car use significantly decreased. They further explained that below the threshold density of residents and jobs, the physical constraints of distance and time enforce car use as the norm.

Design variables have nonlinear effects on car dependency (Figure 5). The probability of using a car as the main mode increase continuously when the road density is below 6 km/km^2, afterward, the curve remains

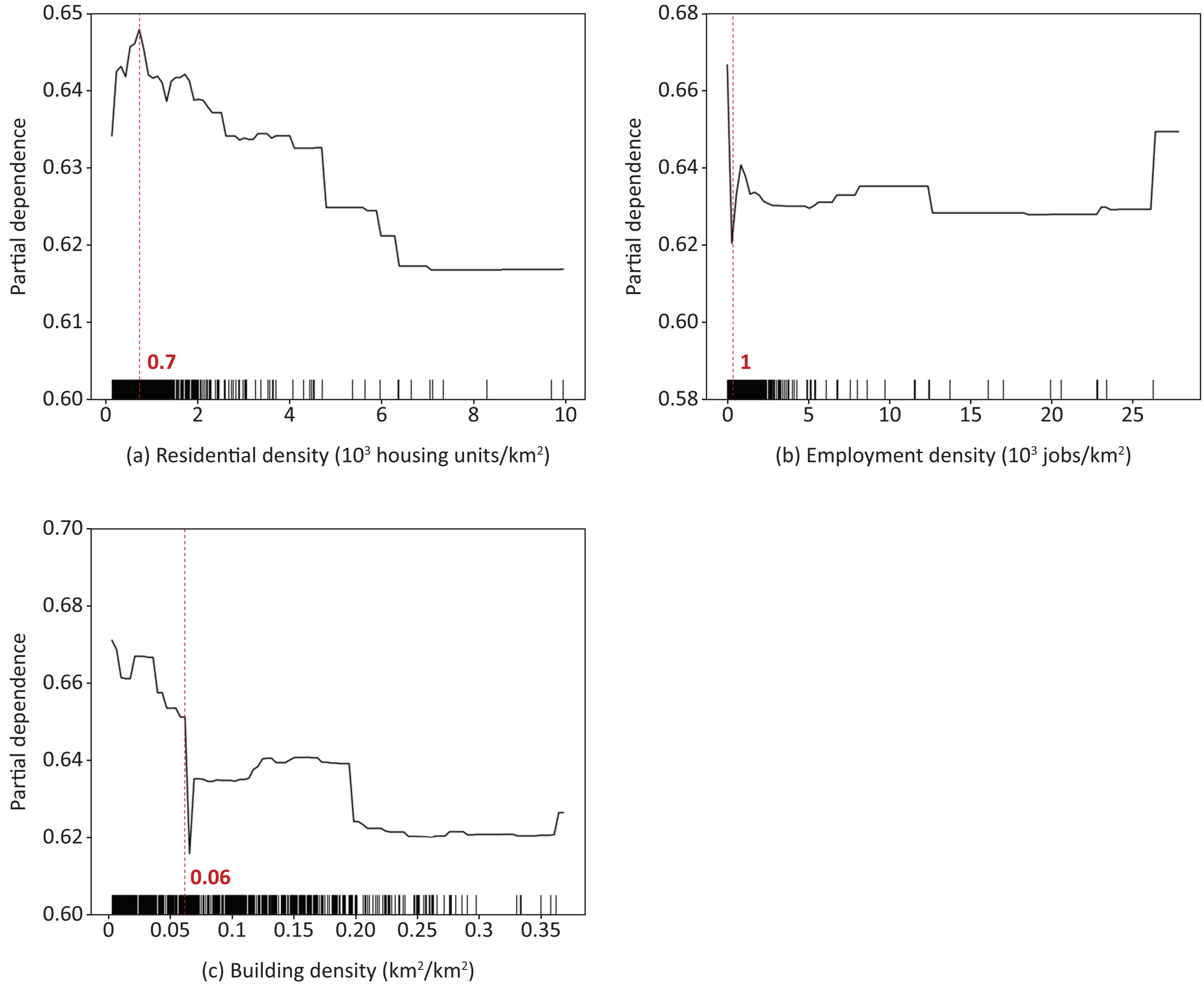


Figure 4. Nonlinear relationship between density variables and car dependency. Note: Y-axes represent the probability of using a car as the main mode.

unaffected. An efficient road network will promote car use. The positive association between road density and car use does not exist in areas with high road density, probably because these areas are more likely to be equipped with sufficient transportation infrastructures, such as public transit services (e.g., bus stops and metro stations) and shared mobility services (e.g., bike-sharing, ridesourcing). Car dependency is positively associated with intersection density when the intersection density is below 18. After the intersection density exceeds 38, the curve drops slightly and remains unchanged. Increased intersection density when the value is low means a good road network may facilitate car use. However, extremely high intersection density may be often accompanied by heavy traffic lights that may reduce people’s willingness to drive. Increasing pedestrian-oriented road density decreases car dependency. This is not surprising since a high pedestrian-oriented road density can promote active travel modes, which are alternative options for car use. This research further found that the most effective range of pedestrian-oriented road density to decrease car use is 14.5 km/km^2, which can provide an evidence-based policy for local government and urban planners.

Car dependency has a positive association with land use mix in neighborhoods with a relatively low land use mix and a negative association in areas with highly mixed land use (Figure 6). Our finding indicates that areas with highly mixed land use are less likely to use cars as the main mode. This is probably because diverse land use promotes the use of active modes (e.g., walking, cycling; Cheng, Jin, et al., 2022), which, in turn, will decrease the use of private cars. Such restraint is not observed in areas with relatively lower land use mix. A similar finding was reached by Cervero (1996), who found that people are more likely to travel by transit, foot, or bicycle when mixed land development within several blocks. Beyond this distance, mixed-use activities appear to induce auto use since automobiles can efficiently link work and shopping activities.

Destination accessibility has an overall negative effect on car dependency (Figure 7). This is consistent with previous studies (Wiersma et al., 2017) that higher

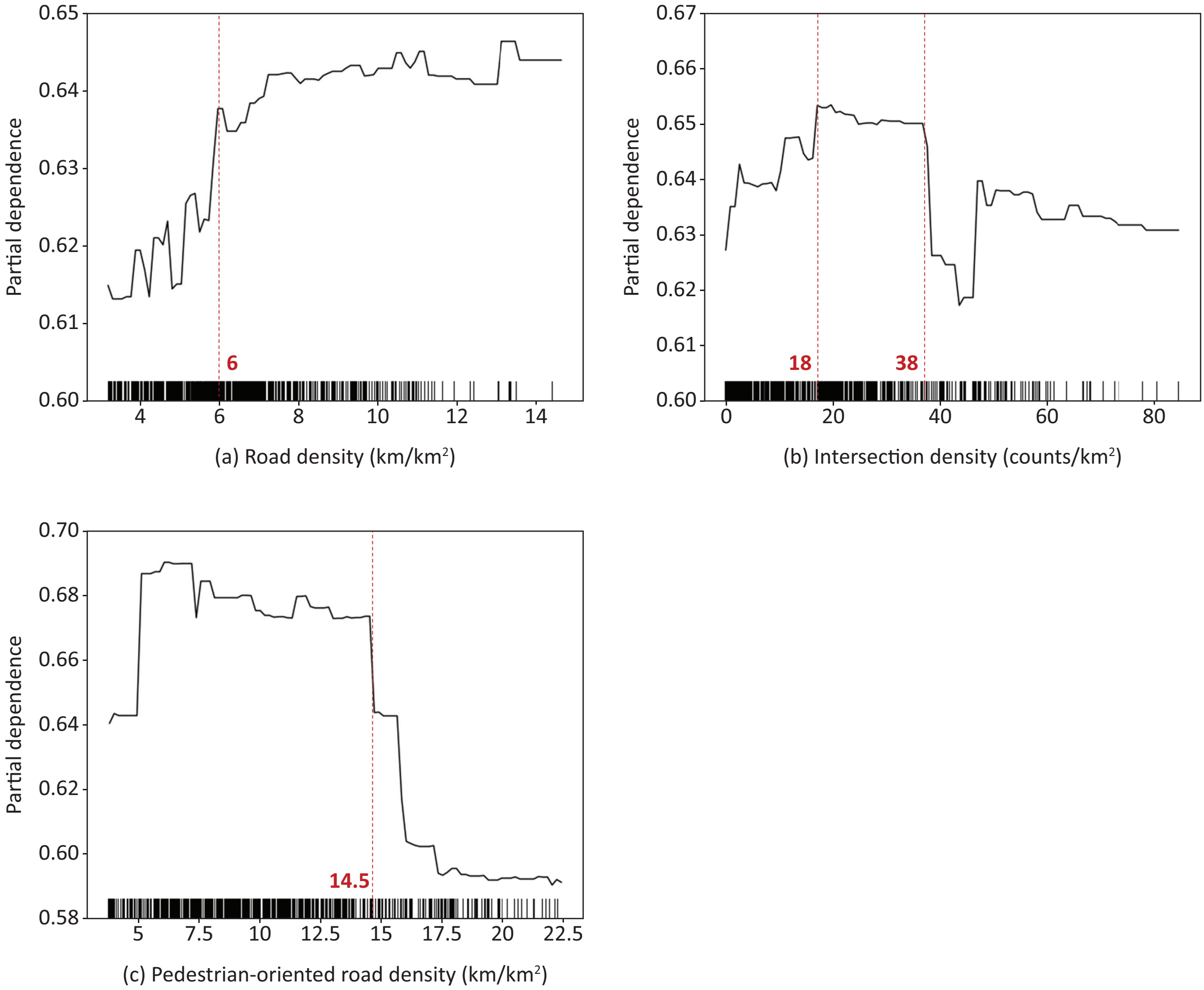


Figure 5. Nonlinear relationship between design variables and car dependency. Note: Y-axes represent the probability of using a car as the main mode.

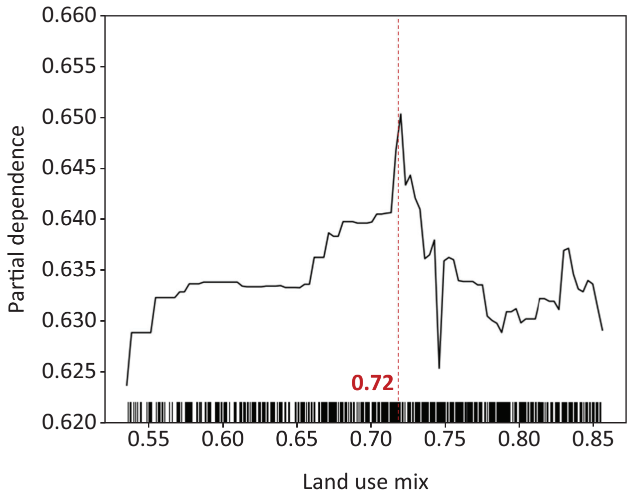


Figure 6. Nonlinear relationship between diversity variable and car dependency. Note: Y-axes represent the probability of using a car as the main mode.

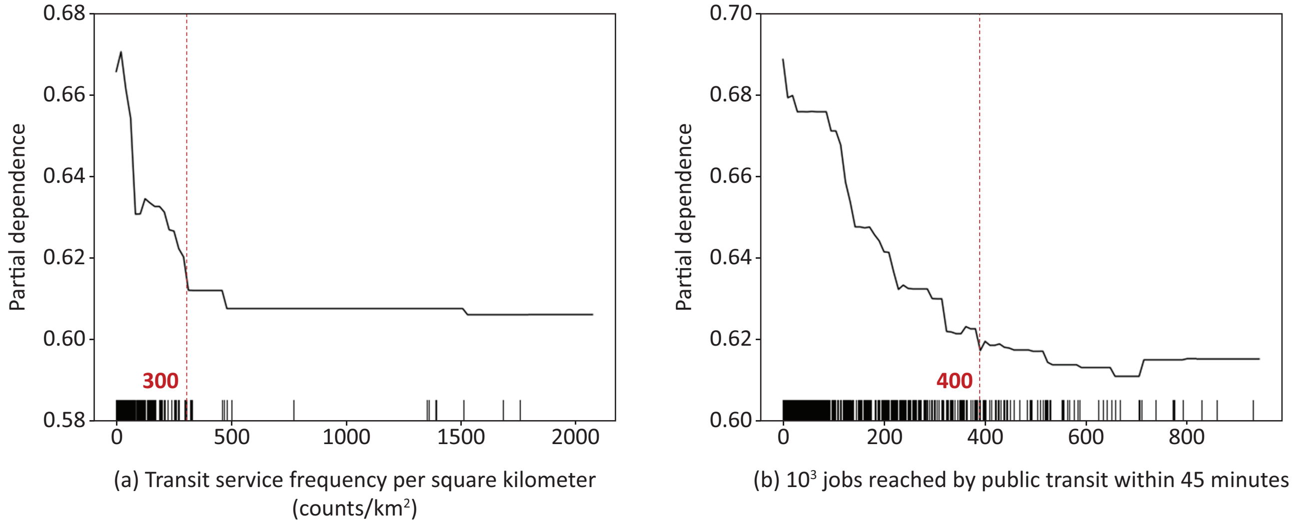


Figure 7. Nonlinear relationship between destination accessibility variables and car dependency. Note: Y-axes represent the probability of using a car as the main mode.

public transit accessibility increases the possibility of public transit use, and in turn, reduces car use. A significant threshold effect is observed for both transit service frequency and jobs reached by public transit. Car dependency witnesses a dramatic decline when the transit service frequency is below 300 per kilometer, afterward, the curve remains unaffected. This may suggest that people's demand for public transit services is satisfied when the transit frequency per square kilometer is 300. Further increasing transit frequency may not be able to decrease car dependency significantly. A similar pattern is also observed for job accessibility. Areas with high job accessibility are favored by public transit more. One explanation is that these areas can provide enough demand that can well maintain the efficiency of public transit systems. Moreover, these areas have high commuting demand, and car use is normally restricted to avoid severe congestion, such as high parking costs.

5. Conclusions

The extensive use of private cars has caused many problems for society. Reducing car dependency and thus relieving the severe issues caused by car dependency has become one of the key objectives of transportation development and land use interventions. Many previous studies have confirmed that compact development and transit-oriented development could be effective strategies to reduce car use and lower the externalities of car dependency (Saeidizand et al., 2022). How to implement efficient planning policies is vital for policymakers and transportation planners. This study analyzed the nonlinear relationship between the built environment and car dependency using a machine learning method, taking Puget Sound Region as the case area. Results show that except for the highest contribution of vehicle ownership, the built environment has a higher average relative importance than individuals' and household socioeconomic and demographic characteristics. This differs from some previous studies, suggesting that the way people travel is strongly affected by individuals' age, gender, income, and employment status (Boussauw & Witlox, 2011). The finding also provides new evidence to further support that built environment factors have more significant impacts on car use (Holtzclaw et al., 2002; Lewis, 2018). For built environment factors, destination accessibility variables have the highest relative importance, followed by design variables. The overall effects of the built environment factors on car dependency are consistent with previous studies (Ding & Cao, 2019; Newman & Kenworthy, 1989a, 1989b; Pinjari et al., 2011). For example, high density leads to low car dependency. Sufficient public transit services and high public transit accessibility can decrease the possibility of using a car as the main mode of a trip. This study further found that built environment factors have significant nonlinear and threshold effects on car dependency, which also provides new insight into the previous nonlinearity studies using the exponential function method. Moreover, the nonlinear relationship captured using a machine learning method releases the pre-defined statistical assumptions that will gain more sophisticated results. This research uncovered that the effect of a built environment variable is only effective within specific intervals of this attribute, which also provides evidence-based guidance for nuanced land use interventions, at least for the government of the Puget Sound Region.

Our results will be useful to provide policy implications for Puget Sound Region to reduce car dependency. First, both high residential density and employment density can lead to low car dependency, which comes with no surprise to further support population densification and increasing employment opportunities near the neighborhoods can reduce people's car use. Second, an efficient road network and pedestrian-friendly street design are helpful to reduce car dependency. An effective road network can encourage people to use shared mobility services more (e.g., bike-sharing, ridesourcing) based on previous studies (e.g., Cheng, Jin, et al., 2022; Jin, Cheng, Zhang, et al., 2022). High pedestrian-oriented road density can also encourage active travel modes, which in turn, reduce car use. Third, good access to public transit services can increase people's use of public transit services and decrease people's car use. Increasing density (i.e., population density, employment density, and building density) can reduce people's car use, which is a valuable strategy for urban planning. However, we should also acknowledge that it is not easy to implement densification since it is faced with challenges for some cities to increase density. Promoting road design and increasing public transit services can be much more operational ways to reduce car dependency. This research further found that the negative association between design and destination accessibility variables and car dependency is effective when the values of built environment variables are within a specific range. These can provide evidence-based guidelines to help policymakers to use limited resources to reduce car dependency through targeted strategies.

The study has several limitations, which promote future research agendas. First, the built environment may have not only a direct impact on travel behavior but also an indirect influence through residential self-selection, which was not considered in this research. Second, the nonlinear relationships between the built environment and car dependency are analyzed only in the Puget Sound Region, validated evidence from other case areas should be provided to test the generalizability of our findings. Nonetheless, this study examines how the built environment affects car dependency, which would help to support targeted and nuanced planning policies to encourage sustainable transportation systems.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No. 52208009 and No. 52278011) and the China Scholarship Council (No. 202008320289). Tanhua Jin would like to acknowledge that most of the work was completed during her stay at her first affiliation, Ghent University.

Conflict of Interests

The authors declare no conflict of interests.

References

Babbie, E. (2009). The practice of social research (12th ed.). Cengage Learning.

Boarnet, M. G., Greenwald, M., & McMillan, T. E. (2008). Walking, urban design, and health: Toward a cost-benefit analysis framework. Journal of Planning Education and Research, 27(3), 341–358. https://doi.org/10.1177/0739456X07311073

Boussauw, K., & Witlox, F. (2011). Linking expected mobility production to sustainable residential location planning: Some evidence from Flanders. Journal of Transport Geography, 19(4), 936–942. https://doi.org/10.1016/j.jtrangeo.2010.12.002

Buehler, R. (2011). Determinants of transport mode choice: A comparison of Germany and the USA. Journal of Transport Geography, 19(4), 644–657. https://doi.org/10.1016/j.jtrangeo.2010.07.005

Buehler, R., Pucher, J., Gerike, R., & Götschi, T. (2017). Reducing car dependence in the heart of Europe: Lessons from Germany, Austria, and Switzerland. Transport Reviews, 37(1), 4–28. https://doi.org/10.1080/01441647.2016.1177799

Cervero, R. (1996). Mixed land-uses and commuting: Evidence from the American Housing Survey. Transportation Research Part A: Policy and Practice, 30(5), 361–377.

Cervero, R., & Arrington, G. B. (2008). Vehicle trip reduction impacts of transit-oriented housing. Journal of Public Transportation, 11(3), 1–17. http://doi.org/10.5038/2375-0901.11.3.1

Chatman, D. G. (2013). Does TOD need the T? On the importance of factors other than rail access. Journal of the American Planning Association, 79(1), 17–31. https://doi.org/10.1080/01944363.2013.791008

Chen, E., Ye, Z., & Wu, H. (2021). Nonlinear effects of built environment on intermodal transit trips considering spatial heterogeneity. Transportation Research Part D: Transport and Environment, 90, Article 102677. https://doi.org/10.1016/j.trd.2020.102677

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Cheng, L., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2019). Applying a random forest method approach to model travel mode choice behavior. Travel Behaviour and Society, 14, 1–10. https://doi.org/10.1016/j.tbs.2018.09.002

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, Article 102552. https://doi.org/10.1016/j.trd.2020.102552

Cheng, L., Huang, J., Jin, T., Chen, W., Li, A., & Witlox, F. (2023). Comparison of station-based and free-floating bikeshare systems as feeder modes to the metro. Journal of Transport Geography, 107, Article 103545.

Cheng, L., Jin, T., Wang, K., Lee, Y., & Witlox, F. (2022). Promoting the integrated use of bikeshare and metro: A focus on the nonlinearity of built environment effects. Multimodal Transportation, 1(1), Article 100004. https://doi.org/10.1016/j.multra.2022.100004

Cheng, L., Wang, K., De Vos, J., Huang, J., & Witlox, F. (2022). Exploring non-linear built environment effects on the integration of free-floating bikeshare and urban rail transport: A quantile regression approach. Transportation Research Part A: Policy and Practice, 162, 175–187.

De Vos, J., & Witlox, F. (2013). Transportation policy as spatial planning tool; reducing urban sprawl by increasing travel costs and clustering infrastructure and public transportation. Journal of Transport Geography, 33, 117–125. https://doi.org/10.1016/j.jtrangeo.2013.09.014

Ding, C., & Cao, X. (2019). How does the built environment at residential and work locations affect car ownership? An application of cross-classified multilevel model. Journal of Transport Geography, 75, 37–45. https://doi.org/10.1016/J.JTRANGEO.2019.01.012

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.

Ding, C., Cao, X., & Liu, C. (2019). How does the station-area built environment influence Metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds. Journal of Transport Geography, 77, 70–78.

Eldridge, J. D., & Jones, J. P., III. (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

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.

Gilbert, R., & Perl, A. (2011). Transportation in the post-carbon world. Post Carbon Institute.

Holtzclaw, J., Clear, R., Dittmar, H., Goldstein, D., & Haas, P. (2002). Location efficiency: Neighborhood and socio-economic characteristics determine auto ownership and use-studies in Chicago, Los Angeles and San Francisco. Transportation Planning and Technology, 25(1), 1–27. https://doi.org/10.1080/03081060290032033

Hong, J. (2017). Non-linear influences of the built environment on transportation emissions: Focusing on densities. Journal of Transport and Land Use, 10(1), 229–240. https://doi.org/10.5198/jtlu.2015.815

International Monetary Fund. (2005). World economic outlook. https://www.imf.org/en/Publications/WEO/Issues/2016/12/31/Globalization-and-External-Imbalances

Jin, T., Cheng, L., Liu, Z., Cao, J., Huang, H., & Witlox, F. (2022). Nonlinear public transit accessibility effects on housing prices: Heterogeneity across price segments. Transport Policy, 117, 48–59. https://doi.org/10.1016/j.tranpol.2022.01.004

Jin, T., Cheng, L., Zhang, X., Cao, J., Qian, X., & Witlox, F. (2022). Nonlinear effects of the built environment on metro-integrated ridesourcing usage. Transportation Research Part D: Transport and Environment, 110, Article 103426.

Lanzendorf, M. (2010). Key events and their effect on mobility biographies: The case of childbirth. International Journal of Sustainable Transportation, 4(5), 272–292.

Lewis, S. (2018). Neighborhood density and travel mode: New survey findings for high densities. International Journal of Sustainable Development & World Ecology, 25(2), 152–165. https://doi.org/10.1080/13504509.2017.1321052

Li, J., Walker, J. L., Srinivasan, S., & Anderson, W. P. (2010). Modeling private car ownership in China: Investigation of urban form impact across megacities. Transportation Research Record, 2193(1), 76–84. https://doi.org/10.3141/2193-10

Liu, J., Wang, B., & Xiao, L. (2021). Non-linear associations between built environment and active travel for working and shopping: An extreme gradient boosting approach. Journal of Transport Geography, 92, Article 103034. https://doi.org/10.1016/j.jtrangeo.2021.103034

Manaugh, K., Miranda-Moreno, L. F., & El-Geneidy, A. M. (2010). The effect of neighbourhood characteristics, accessibility, home–work location, and demographics on commuting distances. Transportation, 37, 627–646. https://doi.org/10.1007/s11116-010-9275-z

Mavoa, S., Witten, K., McCreanor, T., & O'sullivan, D. (2012). GIS based destination accessibility via public transit and walking in Auckland, New Zealand. Journal of Transport Geography, 20(1), 15–22.

McIntosh, J., Trubka, R., Kenworthy, J., & Newman, P. (2014). The role of urban form and transit in city car dependence: Analysis of 26 global cities from 1960 to 2000. Transportation Research Part D: Transport and Environment, 33, 95–110. https://doi.org/10.1016/j.trd.2014.08.013

Naess, P. (2014). Tempest in a teapot: The exaggerated problem of transport-related residential self-selection as a source of error in empirical studies. Journal of Transport and Land Use, 7(3), 57–79. https://doi.org/10.5198/jtlu.v7i3.491

Newman, P., & Kenworthy, J. (1989a). Cities and automobile dependence: An international sourcebook. Gower Publishing.

Newman, P., & Kenworthy, J. (1989b). Gasoline consumption and cities: A comparison of US cities with a global survey. Journal of the American Planning Association, 55(1), 24–37.

Newman, P., & Kenworthy, J. (1991). Transport and urban form in thirty-two of the world's principal cities. Transport Reviews, 11(3), 249–272. https://doi.org/10.1080/01441649108716787

Newman, P., & Kenworthy, J. (2006). Urban design to reduce automobile dependence. Opolis: An International Journal of Suburban and Metropolitan Studies, 2(1), 35–52.

Newman, P., & Kenworthy, J. (2011a). “Peak car use”: Understanding the demise of automobile dependence. World Transport Policy & Practice, 17(2), 31–42. http://www.eco-logica.co.uk/pdf/wtpp17.2.pdf

Newman, P., & Kenworthy, J. (2011b). The density multiplier: A response to Mees. World Transport Policy & Practice, 17(3), 32–44.

Ozbilen, B., Wang, K., & Akar, G. (2021). Revisiting the impacts of virtual mobility on travel behavior: An exploration of daily travel time expenditures. Transportation Research Part A: Policy and Practice, 145, 49–62.

Pinjari, A. R., Pendyala, R. M., Bhat, C. R., & Waddell, P. A. (2011). Modeling the choice continuum: An integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. Transportation, 38, 933–958. https://doi.org/10.1007/s11116-011-9360-y

Potoglou, D., & Kanaroglou, P. S. (2008). Modelling car ownership in urban areas: A case study of Hamilton, Canada. Journal of Transport Geography, 16(1), 42–54. https://doi.org/10.1016/j.jtrangeo.2007.01.006

Pucher, J., & Lefevre, C. (1996). The urban transport crisis in Europe and North America. Springer.

Puget Sound Regional Council. (2021). 2021 Puget Sound regional travel study. https://www.psrc.org/media/6322

Saeidizand, P., Fransen, K., & Boussauw, K. (2022). Revisiting car dependency: A worldwide analysis of car travel in global metropolitan areas. Cities, 120, Article 103467. https://doi.org/10.1016/j.cities.2021.103467

Singh, A. C., Astroza, S., Garikapati, V. M., Pendyala, R. M., Bhat, C. R., & Mokhtarian, P. L. (2018). Quantifying the relative contribution of factors to household vehicle miles of travel. Transportation Research Part D: Transport and Environment, 63, 23–36. https://doi.org/10.1016/j.trd.2018.04.004

Stead, D. (2001). Relationships between land use, socioeconomic factors, and travel patterns in Britain. Environment and Planning B: Planning and Design, 28(4), 499–528. https://doi.org/10.1068/b2677

Tu, M., Li, W., Orfila, O., Li, Y., & Gruyer, D. (2021). Exploring nonlinear effects of the built environment on ridesplitting: Evidence from Chengdu. Transportation Research Part D: Transport and Environment, 93, Article 102776. https://doi.org/10.1016/j.trd.2021.102776

U.S. Department of Transportation. (2015). Land use mix. https://www.transportation.gov/mission/health/land-use-mix

Van Acker, V., & Witlox, F. (2010). Car ownership as a mediating variable in car travel behaviour research using a structural equation modelling approach to identify its dual relationship. Journal of Transport Geography, 18(1), 65–74. https://doi.org/10.1016/j.jtrangeo.2009.05.006

Van Eenoo, E., Fransen, K., & Boussauw, K. (2022). Car dependency beyond land use: Can a standardized built environment indicator predict car use? Journal of Transport and Land Use, 15(1), 117–136. https://doi.org/10.5198/jtlu.2022.2073

Wiersma, J., Bertolini, L., & Straatemeier, T. (2017). Adapting spatial conditions to reduce car dependency in mid-sized “post growth” European city regions: The case of South Limburg, Netherlands. Transport Policy, 55, 62–69. https://doi.org/10.1016/j.tranpol.2016.12.004

Xu, Y., Yan, X., Liu, X., & Zhao, X. (2021). Identifying key factors associated with ridesplitting adoption rate and modeling their nonlinear relationships. Transportation Research Part A: Policy and Practice, 144, 170–188. https://doi.org/10.1016/j.tra.2020.12.005

Zegras, C. (2010). The built environment and motor vehicle ownership and use: Evidence from Santiago de Chile. Urban Studies, 47(8), 1793–1817. https://doi.org/10.1177/0042098009356125

Zhang, W., Zhao, Y., Cao, X. J., 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, Article 102445. http://doi.org/10.1016/j.trd.2020.102445

Zuo, T., Wei, H., & Rohne, A. (2018). Determining transit service coverage by non-motorized accessibility to transit: Case study of applying GPS data in Cincinnati metropolitan area. Journal of Transport Geography, 67, 1–11. https://doi.org/10.1016/j.jtrangeo.2018.01.002

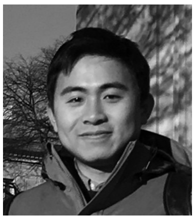
About the Authors

Jun Cao is currently a lecturer at the School of Architecture, Southeast University, China. He received his bachelor’s degree and master’s degree in urban planning from Southeast University, Nanjing, China, in 2014 and 2017. He received his PhD in urban planning from Southeast University in 2021. His research interests include urban morphology, urban design, and urban computing.

Tanhua Jin is currently a PhD student at the Department of Geography, Ghent University, and an academic guest at the Institute of Cartography and Geoinformation, ETH Zurich. She received her bachelor's degree and master's degree in urban planning from Southeast University, China, in 2016 and 2019, respectively. Her research interests include travel behaviour and land use, shared mobility, accessibility, and transportation equity, as well as big data and machine learning in transportation analytics.

Tao Shou is an assistant professor at the School of Architecture at Southeast University. His teaching and research interests relate to urban renewal, vernacular tectonics, and rural landscape architecture. He has published research in these areas and also directed a series of design projects in some typical regional sites of south China. He is a senior member of both the Architectural Society of China and the American Institute of Architects, recently focused on the research and practices in new low-carbon community construction.

Long Cheng received his BS degree in transport and traffic in 2011 and his PhD degree in transport engineering in 2016, both from Southeast University, Nanjing, China. He is currently a postdoctoral researcher at the Department of Geography of Ghent University. His research interests include multimodal transport, shared mobility, travel behavior analysis, and transport and land use integration.



Zhicheng Liu is a postdoctoral researcher at Alibaba Group and the Department of Electronic Engineering, Tsinghua University, Beijing, China. He received his BE and PhD degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2014 and 2021. His research interests include urban computing, graphical models, and data science. His research works have been published in IEEE Transactions on Knowledge and Data Engineering (TKDE) and AAAI (Association for the Advancement of Artificial Intelligence) conference, and others.

Frank Witlox holds a PhD in urban planning (Eindhoven University of Technology), a master's degree in Applied economics, and a master's degree in maritime sciences (both University of Antwerp). Currently, he is head of the department and senior full professor of economic geography at the Department of Geography of Ghent University. His research focuses on travel behavior analysis and modeling, travel and land use, and sustainable mobility issues.