Ineffective built environment interventions: how to reduce driving in American suburbs?[1]

Tao Tao and Jason Cao

Humphrey School of Public Affairs, University of Minnesota

Abstract

Designing effective built environment policies to reduce auto use is key to promoting sustainable transportation in suburban areas. However, most studies on the association between the built environment and auto use focus on the entire region rather than the suburban areas. In addition, previous studies often ignore the nonlinear association. Applying Gradient Boosting Decision Trees to the data in the Twin Cities, USA, this study explores the nonlinear relationships between built environment attributes and driving distance in suburban areas and illustrates how the relationships differ from those in urban areas. The results show that suburban residents are less sensitive to the built environment than urban residents. More importantly, built environment policies that work in urban areas might not be useful in suburban areas. Although many studies have advocated population densification and mixed-use development for driving mitigation, this study suggests that these policies are ineffective in suburban areas. Instead, promoting job accessibility and densifying intersection density could reduce auto use in suburban areas. Densifying transit stops has a small but nontrivial contribution to mitigating auto use.

Keywords: sprawl, land use, VMT, machine learning, nonlinear relationships

1 Introduction

Limited developable lands and high development costs in urban areas have been pushing new residential development to suburban areas (Levitt and Eng 2021). Accordingly, American suburbs have experienced a higher average annual population growth than urban areas since 2014 (Frey 2019). This trend is likely to continue because policymakers have encountered significant challenges when implementing high-density development policies in urban areas. For example, the city of Minneapolis eliminated single-family zoning in its 2040 comprehensive plan in 2018 to promote compact development (Ibrahim 2018). However, this plan has to be suspended as the city was sued by three environmental organizations for its potential negative environmental impacts (Du and Navratil 2022). In addition, the COVID-19 pandemic enables many people to telecommute, encouraging them to live farther from the city center (Sisson 2022). Because suburbanites tend to drive a much longer distance than urban residents, designing effective built environment interventions to reduce their auto use is key to sustainable transportation.

The literature on the built environment and auto use has shed light on potential interventions. There are, however, at least two research gaps that hinder planners' understanding. First, previous studies often examine the relationships between built environment attributes and auto use for the entire metropolitan area. However, their findings and associated planning implications may not be applicable to suburban areas. Some empirical studies have shown heterogeneous effects of the built environment across their study areas (Salon 2015; Elldér, Haugen, and Vilhelmson 2020; Chen, Wang, and Akar 2017). Second, many recent studies have suggested that the relationships between built environment attributes and auto use are irregularly nonlinear (Ding, Cao, and Næss 2018; Wang and Ozbilen 2020). Although some studies consider suburban areas, they often assume a priori and apply (generalized) linear models. These traditional models may not fit the irregularity well and hence underestimate the efficacy of using built environment interventions to alter travel behavior (Wu et al. 2019). Therefore, empirical studies should accommodate the potential irregular nonlinearity.

To fill the gaps, we apply the gradient boosting decision tree (GBDT) method to household travel survey data from the Twin Cities area. This study addresses two research questions: 1) Are built environment attributes equally important to predicting driving distance in urban and suburban areas? And 2) how do the relationships between built environment attributes and driving distance in suburban areas differ from those in urban areas? The answers to the two inquiries will inform planners of built environment interventions specific to suburban areas. ~~Policies that work in urban areas, such as population densification and land use diversification, might not be effective at all in suburban areas. In fact, local context matters to policymaking. Ignoring the divergence between suburban areas and urban areas may misguide planning practices in suburban communities.~~

The rest of this paper is organized as follows. We review the studies on the relationships between built environment attributes and auto use in Section 2. We introduce the data and analysis method in Section 3. We illustrate and discuss the results in Section 4. In the final section, we summarize key results and provide policy implications for reducing driving in suburban areas.

2 Literature Review

During the past three decades, urban planners have been interested in implementing built environment interventions to reduce the use of personal vehicles. Many studies have recommended policies such as population densification, mixed-use development, and grid street pattern as supported by the empirical evidence on the connections between built environment attributes and auto use (Ewing and Cervero 2010; Stevens 2017). For example, Ewing et al. (2015) studied the influences of built environment attributes on driving distance using household travel data from 15 regions in the US. They found that job accessibility, land use mix, and intersection density are all negatively associated with driving distance. Næss, Cao, and Strand (2017) examined a similar research topic in the Greater Oslo and Greater Stavanger regions in Norway, respectively. Their quantitative and qualitative analyses showed that distance to the city center is positively correlated with driving distance in the two regions. In addition, population density and job density are negatively correlated with driving distance in Oslo. Yang and Cao (2018) explored the built environment impact on travel-related carbon dioxide emissions, a negative outcome closely associated with auto use, in Guangzhou, China. They found that land use mix is negatively correlated with carbon dioxide emissions from daily commuting, social, and shopping trips. These studies highlight planning policies such as compact development and urban growth boundary.

However, these interventions may not be applicable to suburban areas because the evidence is often based on data from entire metropolitan areas. Most studies quantify the average effects of built environment attributes on auto use for the entire region but fail to uncover the potential variation of these effects across different subareas in the regions (Boarnet 2017). If the variation exists, one-size-fits-all approaches will not be effective.

Conceptually, the relationships between some built environment attributes and driving distance should be different between urban areas and suburban areas. For example, the relationship between residential distance to city center and driving distance is positive in urban areas. However, in suburban areas, the increasing trend of this relationship may become slower or even disappear. The time geography theory suggests that people have limited time spent on their daily travel as they have to do other more important activities (Hägerstrand 1970; Ellegård 2018). When residential distance to city center exceeds a certain threshold, which usually occurs in suburban areas, people may drive to closer destinations instead of those in the city center to save time from traveling. Therefore, the relationships between residential distance to city center and driving distance could differ between urban and suburban areas. Another example is intersection density. Crane (1996) suggested that intersection densification may reduce the detours of driving and shorten trip distance, but it may also lead to more trips because of induced demand. In urban areas where trips are usually very short, the effect of shortening trip distance may be smaller than that of inducing more trips. Thus, intersection densification may increase driving distance. In suburban areas, trips are longer than those in urban areas. The effect of shortening trip distance may be larger than that of inducing more trips. In this case, intersection density may be negatively correlated with driving distance. Overall, intersection density may have different relationships with driving distance in urban and suburban areas.

Empirical studies have shown that the relationship between the built environment and auto use is heterogeneous within a region. Some develop separate models for different subareas in a study area and compare the relationships. Using data from California, Salon (2015) estimated the correlations between built environment attributes and driving distance in the central city, urban, suburban, rural-in-urban, and rural areas, and found that the correlations vary by area. For example, road density around residence has a larger elasticity in urban areas than in suburban and rural-in-urban areas and it is insignificant in the central city and rural areas. These results imply that increasing road density works best in urban areas, but is ineffective in the central city and rural areas. Using data from the Los Angeles metropolitan region, Boarnet et al. (2011) divided the region into five quintiles based on employment accessibility and estimated the relationships between driving distance and employment density for each quintile. They found that the coefficient of employment accessibility is only significant for the middle three quintiles but not for the top and bottom ones. In other words, employment accessibility has a trivial influence on driving distance when it is very low or very high in the region. Other studies apply geographically weighted regression (GWR) models, which can capture the variation in built environment effects across the entire region. For instance, Chen, Wang, and Akar (2017) explored the relationships between built environment attributes and people's activity space, a proxy for driving distance, by applying the GWR approach to data from the Greater Cleveland region, the US. They found that the coefficients of built environment variables and their corresponding significance level are determined by regional location. For example, distance to the nearest transit facility has a statistically significant and positive relationship with activity space only in some areas (e.g., the eastern part). This finding implies that improving transit access can promote individuals' activity space in limited locales. Overall, these studies suggest that it is critical to examine the effects of built environment attributes on travel in suburban areas. The evidence will offer targeted implications for mitigating auto use in the areas of future population growth.

Furthermore, prominent scholars have contended that the associations between the built environment and auto use should be characterized as nonlinear and threshold relationships although previous studies often assume them to be linear (Van Wee and Handy 2016; Boarnet 2017). For example, the correlation between distance to the city center and driving distance is generally positive because people residing farther away from the city center need to drive a longer distance to reach many of their destinations. However, when distance to the city center exceeds a threshold, its effect on driving distance may reach a plateau. According to the time-geography theory (Hägerstrand 1970), people have limited time spent on activities and traveling among activity destinations in a day. Residents in suburban and exurban areas are more likely to travel to closer destinations instead of those around the city center than their urban counterparts so that they can have sufficient time for their daily activities. If nonlinearity is overlooked, the estimated relationship between distance to the city center and driving distance is biased and the resulting planning implication is flawed.

Some recent studies employing machine learning approaches have shed light on the nonlinear relationships between built environment attributes and auto use. Ding, Cao, and Næss (2018) explored the correlation between the built environment and driving distance in Oslo, Norway. Using the gradient boosting decision trees method, they found that all the built environment variables included in the model have nonlinear and threshold relationships with driving distance. For example, within 12 km from the city center, driving distance grows slowly as the distance to the center increases. However, once exceeding 12 km, driving distance grows sharply. Applying the same approach to the US National Household Travel Survey, Wang and Wang (2021) studied the relationships between built environment attributes and driving distance of young adults. They found that the connection between population density and driving distance is nonlinear. When population density is within 6,000 people per square mile, the relationship is negative. Beyond the threshold, the corresponding relationship becomes flat. Other studies have also shown that nonlinear relationships are prevalent in the connections of the built environment with travel behavior and related outcomes (Yang, Ao, et al. 2021; Wang and Ozbilen 2020; Yang 2023; Yang and Zhou 2020).

3 Data and Method

3.1 Data

The data came from the most recent Travel Behavior Inventory (TBI) of the Twin Cities metro area (Figure 1). It is a household travel survey carried out by the Metropolitan Council (2019), the local metropolitan planning organization, from October 1st, 2018 to September 30th, 2019. To facilitate participation, the TBI offered participants three ways to report their travel diaries: smartphone application, online questionnaire, and call center. All participants were provided with the same questionnaire. In addition, the survey was available in several languages, including English, Spanish, Karen, Oromo, Somali, and Hmong.

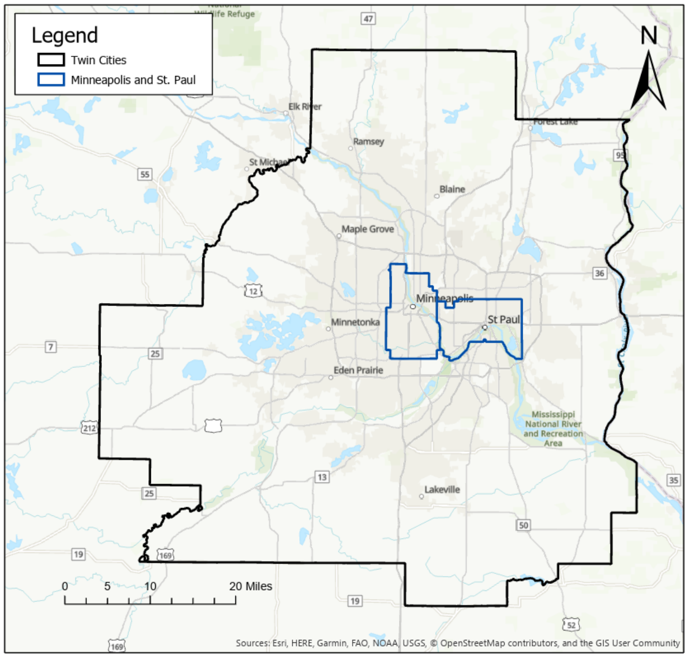

Figure 1. Study area

The survey recruited 13,215 residents from 6,558 households located in the region. In Minnesota, only people older than 16 years old can hold a driver's license. Therefore, we included 11,005 participants who were older than 16 years old in the analysis. Figure 2 presents the distribution of participants by census block group (CBG). The participants resided in approximately 90% of the CBGs in the Twin Cities. Therefore, the sample represents various geographical locations of the study area. The dependent variable is the average weekday driving distance, calculated based on the travel behavior reported in the TBI.

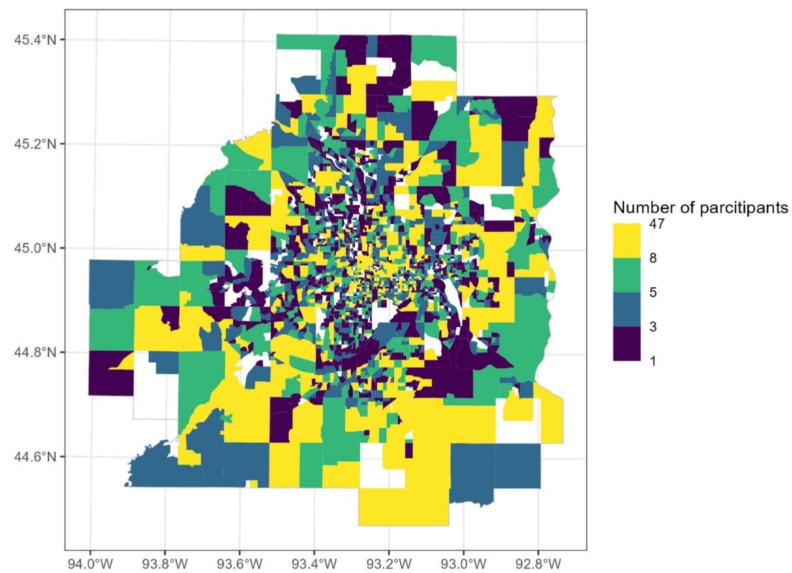

Figure 2. Distribution of the number of survey participants at CBG level
Note: CBGs with no color indicate no valid participants in those areas.

Besides travel diaries, the survey asked participants about their demographic characteristics and residential location at the CBG level. Based on the CBGs where participants live, we measured the built environment through 5Ds: density (population density), diversity (land use entropy), street design (intersection density), distance to transit (transit stop density), and destination accessibility (job accessibility, distance to downtown Minneapolis, distance to downtown St. Paul), as these variables influence driving distance (Ewing and Cervero 2010). In Table 1, we defined and described both demographics and built environment attributes, independent variables in this study.

Table 1. Descriptive statistics of variables

Variable	Description	Source	Suburban area (N = 7,390)		Urban area (N = 3,615)	
			Mean	SD	Mean	SD
Average weekday driving distance	Average daily driving distance in vehicle miles traveled on weekdays during the survey period	Travel survey	25.24	28.88	16.17	23.62
Demographics						
Male	A dummy variable indicating whether the respondent is male	Travel survey	0.46	0.5	0.45	0.5
Age	Age category of the respondent 3 = 16-17 4 = 18-24 5 = 25-34 6 = 35-44 7 = 45-54 8 = 55-64 9 = 65-74 10 = 75 or older	Travel survey	7.19	1.79	6.71	1.75
Worker	A dummy variable indicating whether the respondent is employed (full-time, part-time, self-employed)	Travel survey	0.65	0.48	0.71	0.45
Student	A dummy variable indicating whether the respondent is a student	Travel survey	0.07	0.25	0.1	0.3
Education	Educational background of the respondent 1 = Less than high school 2 = High school graduate/General educational development 3 = Some college 4 = Vocational/technical training 5 = Associate degree 6 = Bachelor's degree 7 = Graduate/post-graduate degree	Travel survey	5.12	1.8	5.44	1.75
White	A dummy variable indicating whether the respondent is Caucasian	Travel survey	0.88	0.33	0.83	0.38
Income	The respondent's household income 1 = Under $15,000 2 = $15,000-$24,999 3 = $25,000-$34,999 4 = $35,000-$49,999 5 = $50,000-$74,999 6 = $75,000-$99,999 7 = $100,000-$149,999 8 = $150,000-$199,999 9 = $200,000-$249,999 10 = $250,000 or more	Travel survey	6.15	1.95	5.53	2.2
Household size	Number of members in the respondent's household	Travel survey	2.48	1.29	2.15	1.27
Built environment variables measured at the census block group (CBG) where the respondent’s household is located						

Variable	Description	Source	Suburban area (N = 7,390)		Urban area (N = 3,615)	
			Mean	SD	Mean	SD
Population density	Number of people per acre	ACS[2]	4.43	3.43	16.3	13.88
Land use entropy	The entropy index of land use mix	MGC	0.59	0.19	0.53	0.24
Transit stop density	Number of transit stops per square mile as of Sep. 2018	Metro Transit	13.04	18.76	55.76	40.41
Job accessibility	Number of jobs in thousand that can be reached by auto in 20 minutes	Accessibility observatory[3]	353.57	233.83	829.37	187.09
Intersection density	Number of four- or more-way intersections per square mile	MGC[4]	19.73	21.32	88.46	47.67
Distance to downtown Minneapolis	The driving distance in miles from the centroid of the CBG to downtown Minneapolis	Google map[5]	18.19	7.85	6.12	3.89
Distance to downtown Saint Paul	The driving distance in miles from the centroid of the CBG to downtown Saint Paul	Google map	20.7	8.64	9.06	4.38

[2] American community survey 2014-2018 5-year estimates: https://www.census.gov/programs-surveys/acs/news/data-releases.html
[3] Accessibility observatory at the University of Minnesota: https://access.umn.edu/
[4] Minnesota geospatial commons: https://gisdata.mn.gov/
[5] Google maps platform (distance matrix API): https://developers.google.com/maps/documentation/distance-matrix/overview

3.2 Defining suburban areas

There is no consensus on how to define suburban areas in the US. Airgood-Obrycki, Hanlon, and Rieger (2020) presented three popular approaches: census-convenient, suburbanism, and typology. The census-convenient approach, which is widely used recently, defines the areas outside of the primary cities in a region as suburban areas. The suburbanism approach defines the areas with the suburban way of life (e.g., high levels of car ownership, car commuting rate, and single-family housing rate) as suburban areas. The typology approach defines the areas with the suburban urban form (e.g., low population density and the specific period when the majority of buildings were developed) as suburban areas. In addition, different definitions could further lead to different typologies. For example, based on the typology definition, suburban areas could be split into inner suburban areas and outer suburban areas.

While there are multiple definitions, scholars often define suburban areas based on the need of their studies (Airgood-Obrycki, Hanlon, and Rieger 2020). Following the practice of the Metropolitan Council (Metropolitan Council 2020), we defined the areas outside of the two central cities as suburban areas (Figure 1). This is consistent with the census-convenient definition. In this way, the results of our study could be directly applied to local planning decision making without any further modification. The descriptive statistics of the variables in urban and suburban areas are shown in Table 1.

3.3 Modeling approach

We constructed two models to estimate the relationships between built environment attributes and driving distance for suburban and urban areas, respectively, controlling for demographics. We applied the GBDT method to estimate their nonlinear relationships. In the past five years, the method has been used in many studies on the built environment and travel behavior (Ding, Cao, and Næss 2018; Wang and Ozbilen 2020; Yang, Liang, et al. 2021). It is a machine learning method proposed by Friedman (2001, 2002) and integrates two approaches: decision tree and gradient boosting. The decision tree approach is used to divide the sample into subsamples based on the values of independent variables. The average value of the independent variable in each subsample is used to predict the case that holds similar values of independent variables. However, a single decision tree is weak in prediction. The gradient boosting approach combines the results of hundreds of decision trees sequentially into a stronger model. See (Tao 2021) for a detailed mathematical explanation of the GBDT method.

We chose the method for several reasons. First, it is efficient in estimating the irregularly nonlinear relationships between dependent and independent variables. Traditional statistical models (i.e., ordinary least squares regression and generalized linear regression) assume that the relationships are linear or generalized linear. However, if this assumption does not hold, the results of traditional models may be biased (Breiman 2001). The GBDT approach does not assume the relationships and, thus, is capable of estimating the relationships that are closer to the true ones. Second, it has a better prediction performance than traditional models. Third, it can accommodate missing values and outliers in the sample. This minimizes the bias resulting from observation deletion. However, the method has a few limitations. First, overfitting can be a major issue. In this study, we applied cross validations to overcome this problem. Second, it cannot provide statistical inferences such as p values and confidence intervals. We used relative importance and scale of influences to measure the importance of independent variables.

We used the "gbm" package in R (Ridgeway 2007; Greenwell, Boehmke, and Cunningham 2020) to estimate the GBDT models. Three parameters need to be specified for a good prediction performance, including tree depth, learning rate, and number of trees. The tree depth parameter, a non-zero positive integer, is related to the complexity of one decision tree. A decision tree with a larger value of tree depth is more complex. Learning rate indicates how much proportion of the result of each decision tree will be integrated into the final model. It is a value from 0 to 1. A smaller learning rate can provide a better prediction, but the modeling process needs more computation resources. The number of trees is how many decision trees will be combined into the final model. More decision trees are better but are subject to overfitting at the same time. Following the common practice in previous built environment and travel behavior studies (Yang, Cao, and Zhou 2021; Zhang et al. 2020), we set tree depth to be 10 and learning rate to be 0.001. We applied a five-fold cross validation to search for the number of decision trees that can provide the least root mean squared error loss.

To interpret the results of the models, we used two approaches: relative importance and accumulated local effect (ALE) plot. Relative importance measures the contribution of an independent variable to predicting the dependent variable. Specifically, it is the proportion of the variance reduced by the independent variable among the variance reduced by all independent variables during the modeling process (Molnar 2020). It is expressed as percentage. Therefore, the relative importance of all independent variables in one model adds up to 100%. An ALE plot shows the marginal effect of one independent variable on the dependent variable (Molnar 2020). It helps to visualize their relationship while controlling for other independent variables (see Appendix for the mathematical notation of estimating ALE). We used the "ALEplot" package in R (Apley and Zhu 2020) to generate the ALE plots of the relationships between built environment attributes and driving distance.

4 Results

We present the results of two models: the suburban model and the urban model. Again, the former uses the respondents who lived in the areas outside the city boundaries of Minneapolis and St. Paul in the Twin Cities, and the latter uses those who resided in the areas within the city boundaries. To facilitate comparison, the two models contain the same set of independent variables: built environment attributes and demographics. The final models include 3,658 and 1,995 trees for the suburban and urban areas, respectively.

4.1 Relative importance

Table 2 presents the relative importance of independent variables to predicting driving distance in the two models. Built environment attributes jointly have a contribution of 42.2% in the suburban model, which is smaller than that in the urban model (57.1%). This suggests that the built environment is less important to driving distance of people living in suburban areas than of those residing in urban areas.

Table 2. Relative importance (RI) of independent variables

Variable	Suburban model			Urban model		
	RI (%)	Ranking	Sum (%)	RI (%)	Ranking	Sum (%)
Built environment characteristics						
Job accessibility	11.7	4	42.2	9.9	3	57.1
Distance to Minneapolis	6.3	6		9.6	4	
Land use entropy	6.1	8		9.4	5	
Intersection density	5.5	9		6.9	9	
Population density	4.9	10		7.5	7	
Distance to St. Paul	4.1	11		6.4	10	
Transit stop density	3.6	12		7.4	8	
Demographics						

Variable	Suburban model			Urban model		
	RI (%)	Ranking	Sum (%)	RI (%)	Ranking	Sum (%)
Worker	15.6	1	57.8	8.1	6	42.9
Male	12.8	2		3.5	12	
Education	12.7	3		12.3	1	
Age	7.6	5		5.5	11	
Income	6.2	7		10.8	2	
Household size	1.6	13		2.0	13	
White	1.2	14		0.5	14	
Student	0.1	15		0.3	15	
Pseudo R^2	0.19			0.14		
Sample size	7,390			3,615		

Note: The RI of all independent variables in a model adds up to 100%.

Among the built environment attributes, job accessibility and distance to Minneapolis are the two most important variables in both models. Land use entropy, population density, and transit stop density have smaller contributions in the suburban model than in the urban model. In addition, distance to Minneapolis and distance to St. Paul have smaller relative importance in suburban areas than in urban areas. This suggests that the two city centers are less attractive to people living outside the beltway. For job accessibility and intersection density, their relative contributions to predicting driving distance are similar between the two areas.

As to demographic characteristics, employment status, gender, and education level are the top three important predictors in the suburban model. In the urban model, education level, income, and employment status are the three most influential variables.

4.2 Relationships between driving distance and built environment attributes

Figure 3 to Figure 9 present and compare the relationships (i.e., accumulated local effect plots) between driving distance and built environment attributes in the two models. All the effects have been centered to zero. Specifically, we subtracted the average effect of the built environment attribute from its individual effect across all observations. Centering ALE to zero enables a more intuitive interpretation of the plot. A positive value indicates an effect higher than the average prediction, and a negative value indicates a lower-than-average effect. This standardization makes it easier to understand the influence of the feature on the predicted response. The standardization also facilitates the comparison of ALE plots for the same attribute between different models by providing a common baseline of zero, which is pivotal for this study. Furthermore, we use the same scale for the vertical axes in all the plots and for the horizontal axes for plots of the same variables. The ticks on the x axes show the distribution of the corresponding built environment variable.

The relationships between job accessibility and driving distance are shown in Figure 3. In general, driving distance is negatively correlated with job accessibility in both suburban and urban areas. The scales of the associations are similar. Given that job accessibility is important to predicting driving distance in both models (Table 2), this result suggests that increasing job accessibility is effective in reducing driving distance in both areas. However, job accessibility has a trivial influence on driving distance when it exceeds 250 thousand jobs in suburban areas. The corresponding threshold in urban areas is 750 thousand jobs, which is much higher than that in suburban areas.

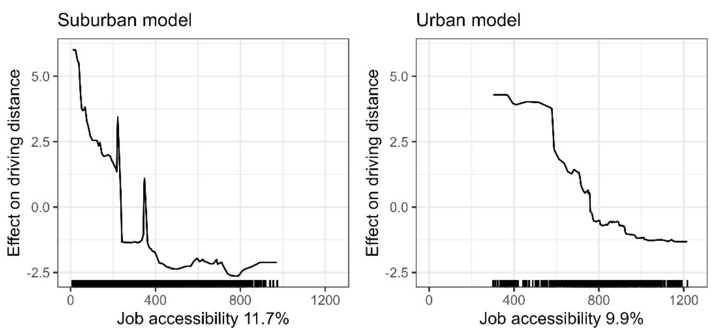

Figure 3. Relationships between job accessibility and driving distance in two areas

Figure 4 presents the relationships between population density and driving distance. Driving distance is roughly negatively correlated with population density in urban areas. However, their relationship in suburban areas fluctuates around zero and does not show a clear pattern. This suggests that population densification is not an effective approach for driving reduction in suburban areas.

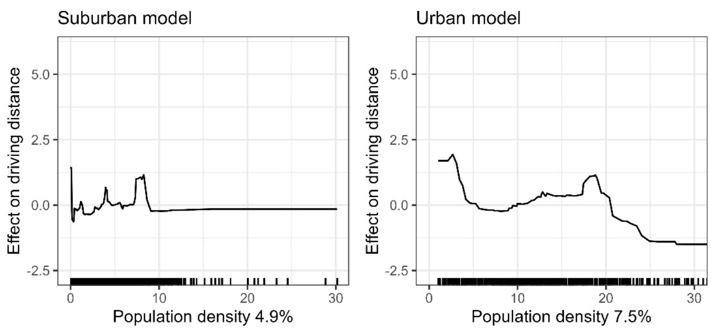

Figure 4. Relationships between population density and driving distance in two areas

Figure 5 shows the relationships between intersection density and driving distance. Their association in urban areas is positive. In the urban areas with a very high level of intersection density, better street connectivity may induce more travel (Crane 1996). In suburban areas, driving distance is negatively correlated with intersection density. With a lower level of intersection density, street connectivity is poorer. Therefore, driving distance is longer due to more detours in suburban areas (Crane 1996). When the number of intersections in suburban areas exceeds 40, intersection density has a trivial effect on driving distance. When intersection density is within the range of 50-~~100~~120, its relationship with driving distance is similar between the two areas. When intersection density exceeds 120, the corresponding interval contains too few sample points, and the relationship there is not reliable for interpretation. Their scales of influence are comparable in the two areas. This result suggests that intersection densification is helpful in lowering driving distance in suburban areas and the effect is maximized when there are at least 40 intersections of four or more ways per square mile.

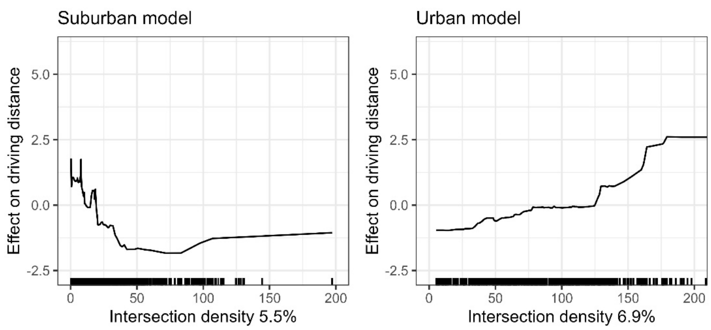

Figure 5. Relationships between intersection density and driving distance in two areas

The relationships between land use entropy and driving distance are shown in Figure 6. The relationship is negative in urban areas. In the areas with relatively homogenous land uses, people need to drive a longer distance to reach destinations. By contrast, in the areas with diverse land use types, people can reach their destinations by walking, biking, or transit and even if they drive, they have short driving distances. However, the relationship in suburban areas does not show a clear pattern. This result suggests that increasing land use mix in suburban areas is not effective for driving reduction. It is worth noting that in the suburban model, a sudden jump appears at 0.3 for unknown reasons.

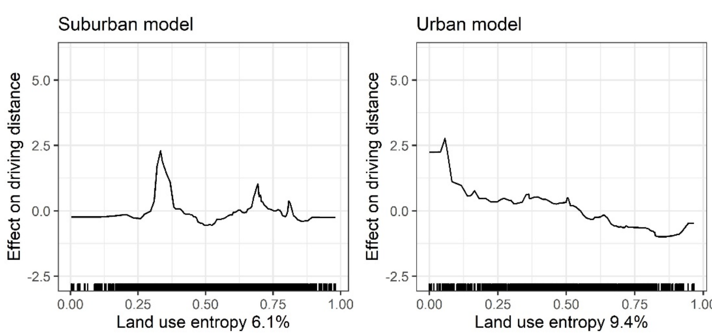

Figure 6. Relationships between land use entropy and driving distance in two areas

Figure 7 presents the relationships between transit stop density and driving distance. Generally, the relationships are negative in both areas. In urban areas, transit stop density has an influential effect within 25 stops per square mile, but its effect becomes trivial when exceeding this threshold. In suburban areas with lower transit stop density (i.e., smaller than 30 stops per square mile), there is a slight increase in driving distance as the density grows. This might be because, when transit stop density increases from 0 to a higher value, the driving demand in the corresponding areas also increases. When transit stop density is higher than 30 stops per square mile, there is a substantial decrease in driving distance as the ~~number~~ density of transit stops increases. After reaching 35 transit stops per square mile, the effect of transit supply becomes stable. Although the relative importance of transit stop density in suburban areas (3.6%) is smaller than that in urban areas (7.4%), their scales of influence are similar. The results suggest that increasing transit stop density has a moderate effect on reducing driving distance in suburban areas.

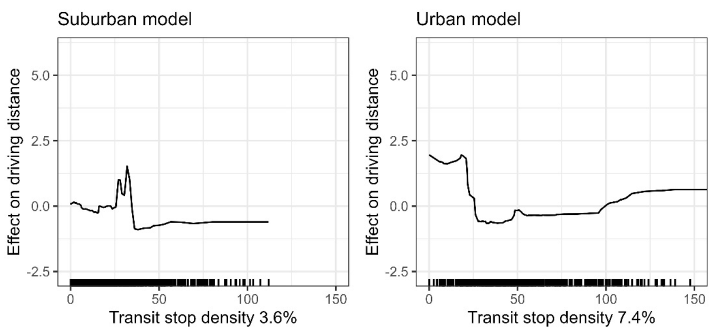

Figure 7. Relationships between transit stop density and driving distance in two areas

Figure 8 shows the relationships between distance to Minneapolis and driving distance. Driving distance is positively correlated with distance to Minneapolis in both suburban and urban areas. The scale of the corresponding influence in urban areas is approximately 1.6 miles larger than that of suburban areas. The range of influence in urban areas is 6.3 miles, changing from -2.5 to 3.8 miles. However, the range of influence in suburban areas is only 4.7 miles, changing from -1.2 to 3.5 miles. It is worth noting that 1.6 miles account for about 10% of the average driving distance of urban residents.

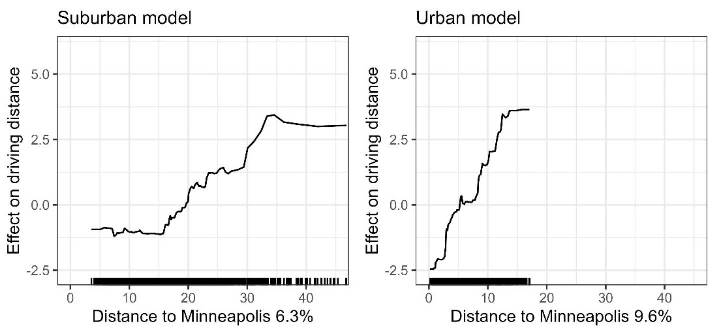

Figure 8. Relationships between distance to Minneapolis and driving distance in two areas

The relationships between distance to St. Paul and driving distance are shown in Figure 9. Distance to downtown St. Paul is positively correlated with driving distance in urban areas. However, the correlation is generally negative in suburban areas. A potential reason is that the city center of Minneapolis (the business center) has a larger influence on driving behavior than the city center of St. Paul (the political center). Some areas, although farther from the city center of St. Paul, are closer to the city center of Minneapolis. People living in these areas mostly drive to Minneapolis for their activities and, thus, have a lower driving distance. The scale of the influence in suburban areas is smaller than that in urban areas. The scale of the influence of distance to St. Paul in urban areas is about 4.7 miles, ranging from -1 to 3.7 miles. However, the corresponding scale in suburban areas is 3.6 miles, ranging from -1.2 to 2.4 miles.

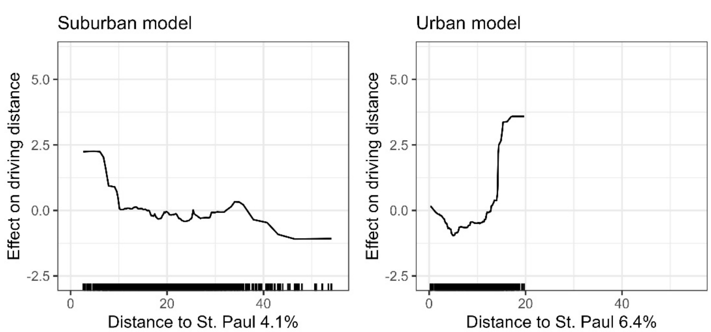

Figure 9. Relationships between distance to St. Paul and driving distance in two areas

4.3 Discussion

Commented [TT1]: Did not track change, but the main revision is move paragraphs from other places and some minor changes on the flow.

A comparison of the relative importance shows that the collective contribution of built environment attributes to the prediction of driving distance in suburban areas is less than that in urban areas, suggesting a less important role of built environment interventions in modifying driving behavior in suburban areas. This result is likely because most built environment attributes, except for job accessibility, have less variance in the suburban areas than in the urban areas (Table 1).

According to the relative importance and nonlinear relationships of individual built environment attributes, population density and land use entropy, important in urban areas, have trivial influences on driving distance in suburban areas. On the other hand, job accessibility and intersection density are effective for mitigating auto use in suburban areas, as well as urban areas. Transit stop density has a moderate effect in suburban areas. Furthermore, in suburban communities, the areas with at least 250 thousand jobs within the 20-minute driving distance tend to have the lowest driving distance. So do the areas with 40 or more intersections per square mile and 35 transit stops per square mile. It is worth noting that the thresholds may be specific to suburban communities in the Twin Cities and additional studies are needed to substantiate their generalizability to other regions or countries.

Suburbanization is an ongoing process in the US. Although the sprawl to suburban areas is slower in the 21 century than in the last century, large urbanized areas in the US became less compact from 2000 to 2010 (Hamidi and Ewing 2014). By examining the trend of urban sprawl in the US from 1910 to 2012, Barrington-Leigh and Millard-Ball (2015) also showed that urban sprawl continued in low-connectivity areas. They concluded that "sprawl begets sprawl" (Barrington-Leigh and Millard-Ball 2015, 8249). Due to the recent surge in real estate industry, American suburbs have continued to expand during the past decade. For example, Woodbury, a suburban city in the Twin Cities, has made substantial greenfield development for residential use (Johnson 2022). Accordingly, the number of residents in Woodbury has grown by around 27% from April 1, 2010, to July 1, 2022[6]. We expect this trend to persist, partly owing to the proliferation of telecommuting after the COVID-19 pandemic (Sisson 2022).

[6] https://www.census.gov/quickfacts/fact/table/woodburycityminnesota,MN,US/POP060210, accessed on September 26, 2023.

By implementing strategic planning policies, we could change the course and curtail excessive automobile dependency in American suburbs. This study provides guidance for strategically achieving this target. Built environment policies that work in urban areas might not be useful in suburban areas and policymakers need to design context-specific land use and transportation policies for driving reduction. Many studies have proposed population densification and mixed-use development to mitigate auto use. However, this study suggests that, in suburban areas, these policies are infertile. Instead, promoting job accessibility and densifying intersection density seems productive to reducing auto use in suburban areas. Increasing transit supply also has a nontrivial contribution to mitigating auto use.

Attracting job opportunities to suburban areas through policies such as cash grants, tax credits, and infrastructure investment can help increase job accessibility. The goal is to cultivate employment clusters in suburban cities, as opposed to scattered development or "edgeless cities". For example, the Livable Communities Act by the Metropolitan Council offers grants for job creation and infrastructure development and re-development in local communities. In addition, this study suggests that designing communities with small blocks and a dense road network and providing more transit services such as express transit routes can reduce car use in suburban areas. It is worth noting that these policies complement each other and should be implemented together. The application of only one policy without the other two may create more issues. For example, only increasing job accessibility but without providing transit services in suburban areas may induce more driving trips.

This study enhances the existing literature in two significant ways. Firstly, it delves into the influence of built environment attributes on driving in suburban areas. Understanding the distinct built environment impacts in suburban and urban areas, we can formulate more customized land use policies in reducing car reliance. Secondly, it considers nonlinearity when assessing the connections between the built environment and driving distance. The results provide more accurate estimates of the contributions and complex impact exerted by the built environment.

5 Conclusions

To shed light on planning policies for suburbs, this study examined the nonlinear relationships between built environment attributes and average weekday driving distance in suburban areas of the Twin Cities and compared the results with those in urban areas (i.e., Minneapolis and St. Paul). Specifically, we constructed two GBDT models to estimate the nonlinear relationships in suburban areas and urban areas, respectively. Then, we compared the relative importance of built environment attributes to driving distance and their ALE plots.

The results showed that built environment attributes in suburbs contribute less to predicting driving distance than those in urban areas. Moreover, although population density and land use mix are negatively associated with car use in urban areas, their associations with car use in suburbs do not show a clear pattern. Despite its critical role in urban areas, transit stop density moderately affects driving in suburbs. On the other hand, job accessibility and intersection density have negative associations with car use in suburbs, as well as in urban areas.

Future studies should try to measure built environment attributes at the individual level (i.e., using the buffers around respondents' residences). Because the data used in this study provides geographical information of home locations at the CBG level, we measured built environment variables at this level. These measures may not accurately capture the built environment around respondents' residences. The measurement errors may misidentify built environment thresholds. In addition, future studies should incorporate attitudinal variables to estimate the differing impacts of the built environment on driving behavior of urban residents and suburbanites, controlling for residential self-selection. The issue of residential self-selection arises when people choose where to live based on their travel needs and preferences. It results from two main sources: individuals' demographic characteristics and their predisposition towards travel and land use (Mokhtarian and Cao 2008). Because the TBI survey did not include attitudinal measures, we are unable to assess how they confound the differing impacts. With that said, we controlled for demographics and hence partly accounted for the residential self-selection effect. Finally, this study is one of the early ones that focus on the differing impacts of the built environment on driving distance between suburban and urban areas. We defined the areas outside of the two central cities as suburban areas. However, suburban areas are not the same and can be categorized into different types based on different definitions. Future studies should further explore the various impacts of the built environment in different types of suburban areas. Such studies might provide more nuanced land use policies to modify driving behavior of people residing in different suburban areas.

Acknowledgement

This research is funded by the University of Minnesota Doctoral Dissertation Fellowship.

Appendix

Accumulated local effect (ALE) is an instrument of interpretable machine learning that quantifies the impact of an individual feature on the predicted response. It assembles local effects of the feature while considering interactions with other features. The estimation of an uncentered ALE includes three main steps. First, we divide the values of the target feature into several intervals based on its percentiles. Second, we calculate the local effect of the target feature within each interval. A local effect refers to the average difference in predictions, which are estimated by replacing the value of the target feature with the upper and lower boundaries of the corresponding interval while the values of all other features are kept constant. Finally, we accumulate the local effects across all the intervals.

Following the notation in Molnar (2020), the uncentered ALE of an independent variable is calculated as follows by Equation (A1):

$$\hat{f}_{j,ALE}(x) = \sum_{k=1}^{k_j(x)} \frac{1}{n_j(k)} \sum_{i:x_j^{(i)} \in N_j(k)} \left[f\left(z_{k,j}, x_{-j}^{(i)}\right) - f\left(z_{k,j-1}, x_{-j}^{(i)}\right) \right], \quad \text{(A1)}$$

where $\hat{f}_{j,ALE}(x)$ is the uncentered ALE value of the jth independent variable; $N_j(k)$ indicates the kth interval; $z_{k,j}$ is the upper boundary and $z_{k,j-1}$ is the lower boundary of $N_j(k)$; $n_j(k)$ indicates the number of observations located in $N_j(k)$; $x_j^{(i)}\left(i \in (1, \dots, n)\right)$ indicates the value of the jth independent variable of the ith observation located in $N_j(k)$ and n indicates the number of observations in the sample; $x_{-j}^{(i)}$ indicates the values of other independent variables of the ith observation; $f(\cdot)$ denotes the estimation function; and $k_j(x)$ indicates the number of intervals to accumulate.

The centered ALE is computed by subtracting the average value of the uncentered ALE from the original value, as shown below by Equation (A2):

$$f_{j,ALE}(x) = \hat{f}_{j,ALE}(x) - \frac{1}{n}\sum_{i=1}^{n}\hat{f}_{j,ALE}\left(x_j^{(i)}\right). \quad \text{(A2)}$$

Reference

Airgood-Obrycki, Whitney, Bernadette Hanlon, and Shannon Rieger. 2020. "Delineate the U.S. suburb: An examination of how different definitions of the suburbs matter." *Journal of Urban Affairs* 43 (9):1263-1284. doi: 10.1080/07352166.2020.1727294.

Apley, Daniel W., and Jingyu Zhu. 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. doi: 10.1111/rssb.12377.

Barrington-Leigh, C., and A. Millard-Ball. 2015. "A century of sprawl in the United States." *Proc Natl Acad Sci U S A* 112 (27):8244-9. doi: 10.1073/pnas.1504033112.

Boarnet, Marlon. 2017. "Land use, travel behavior, and disaggregate travel data." In *The geography of urban transportation*, edited by Genevieve Giuliano and Susan Hanson. New York: The Guilford Press.

Boarnet, Marlon, Douglas Houston, Gavin Ferguson, and Steven Spears. 2011. "Land use and vehicle miles of travel in the climate change debate: Getting smarter than your average bear." In *Climate Change and Land Policies*, edited by Gregory Ingram and Yu-Hung Hong. Cambridge, Massachusetts: Lincoln Institute of Land Policy.

Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)." *Statistical Science* 16 (3):199-231. doi: 10.1214/ss/1009213726.

Chen, Na, Chih-Hao Wang, and Gulsah Akar. 2017. "Geographically Weighted Regression Approach to Investigate Spatial Variations in Activity Space." *Transportation Research Record: Journal of the Transportation Research Board* 2671 (1):40-50. doi: 10.3141/2671-05.

Crane, Randall. 1996. "On Form versus Function: Will the New Urbanism Reduce Traffic, or Increase It?" *Journal of Planning Education and Research* 15 (2):117-126. doi: 10.1177/0739456x9601500204.

Ding, Chuan, Xinyu (Jason) Cao, and Petter Næss. 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. doi: 10.1016/j.tra.2018.02.009.

Du, Susan, and Liz Navratil. 2022. "Court orders Minneapolis to cease implementation of 2040 plan." *Star Tribune*. https://www.startribune.com/court-minneapolis-ordered-to-cease-implementation-of-2040-plan/600182511/.

Elldér, Erik, Katarina Haugen, and Bertil Vilhelmson. 2020. "When local access matters: A detailed analysis of place, neighbourhood amenities and travel choice." *Urban Studies* 59 (1):120-139. doi: 10.1177/0042098020951001.

Ellegård, Kajsa. 2018. *Thinking Time Geography Concepts, Methods and Applications*. London: Routledge.

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

Ewing, Reid, Guang Tian, Jp Goates, Ming Zhang, Michael J Greenwald, Alex Joyce, John Kircher, and William Greene. 2015. "Varying influences of the built environment on household travel in 15 diverse regions of the United States." *Urban Studies* 52 (13):2330-2348. doi: 10.1177/0042098014560991.

Frey, William H. 2019. "Big city growth stalls further, as the suburbs make a comeback." accessed 04-07-2022. https://www.brookings.edu/blog/the-avenue/2019/05/24/big-city-growth-stalls-further-as-the-suburbs-make-a-comeback/.
Friedman, Jerome H. 2001. "Greedy function approximation: A gradient boosting machine." *The Annals of Statistics* 29 (5):1189-1232. doi: 10.1214/aos/1013203451.
Friedman, Jerome H. 2002. "Stochastic gradient boosting." *Computational Statistics & Data Analysis* 38 (4):367-378. doi: 10.1016/s0167-9473(01)00065-2.
Greenwell, Brandon, Bradley Boehmke, and Jay Cunningham. 2020. "gbm: Generalized Boosted Regression Models." https://cran.r-project.org/package=gbm.
Hägerstrand, Torsten. 1970. "What about people in Regional Science?" *Papers of the Regional Science Association* 24 (1):6-21. doi: 10.1007/bf01936872.
Hamidi, Shima, and Reid Ewing. 2014. "A longitudinal study of changes in urban sprawl between 2000 and 2010 in the United States." *Landscape and Urban Planning* 128:72-82. doi: 10.1016/j.landurbplan.2014.04.021.
Ibrahim, Mukhtar M. 2018. "Days before final vote, group sues Minneapolis over 2040 plan." *Star Tribune*. https://www.startribune.com/days-before-final-vote-group-sues-minneapolis-over-2040-plan/501840971/.
Johnson, Brian. 2022. "Developer proposes 446 new homes in Woodbury." https://finance-commerce.com/2022/10/developer-proposes-446-new-homes-in-woodbury/.
Levitt, Zach, and Jess Eng. 2021. "Where America's developed areas are growing: 'Way off into the horizon'." The Washington Post, accessed 02-24-2022. https://www.washingtonpost.com/nation/interactive/2021/land-development-urban-growth-maps/.
Metropolitan Council. 2019. 2019 Travel behavior inventory household survey results.
Metropolitan Council. 2020. A typology of suburban change.
Mokhtarian, Patricia L., and Xinyu Cao. 2008. "Examining the impacts of residential self-selection on travel behavior: A focus on methodologies." *Transportation Research Part B: Methodological* 42 (3):204-228. doi: 10.1016/j.trb.2007.07.006.
Molnar, Christoph. 2020. "Interpretable Machine Learning - A Guide for Making Black Box Models Explainable." In: lulu.com. https://christophm.github.io/interpretable-ml-book/.
Næss, Petter, Xinyu (Jason) Cao, and Arvid Strand. 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.
Ridgeway, Greg. 2007. "Generalized Boosted Models: A guide to the gbm package." https://pbil.univ-lyon1.fr/CRAN/web/packages/gbm/vignettes/gbm.pdf.
Salon, Deborah. 2015. "Heterogeneity in the relationship between the built environment and driving: Focus on neighborhood type and travel purpose." *Research in Transportation Economics* 52:34-45. doi: 10.1016/j.retrec.2015.10.008.
Sisson, Patrick. 2022. "How the Pandemic Supercharged Sprawl." accessed 03-06-2022. https://www.bloomberg.com/news/features/2022-01-05/a-supernova-of-suburban-sprawl-fueled-by-covid.
Stevens, Mark R. 2017. "Does Compact Development Make People Drive Less?" *Journal of the American Planning Association* 83 (1):7-18. doi: 10.1080/01944363.2016.1240044.
Tao, Tao. 2021. "The gradient boosting decision tree approach – Mathematical illustration." https://vtao1989.github.io/blog/files/GBDT.pdf.

Van Wee, Bert, and Susan Handy. 2016. "Key research themes on urban space, scale, and sustainable urban mobility." *International Journal of Sustainable Transportation* 10 (1):18-24. doi: 10.1080/15568318.2013.820998.

Wang, Kailai, and Ozbilen Ozbilen. 2020. "Synergistic and threshold effects of telework and residential location choice on travel time allocation." *Sustain Cities Soc* 63:102468. doi: 10.1016/j.scs.2020.102468.

Wang, Kailai, and Xize Wang. 2021. "Generational differences in automobility: Comparing America's Millennials and Gen Xers using gradient boosting decision trees." *Cities* 114. doi: 10.1016/j.cities.2021.103204.

Wu, Xinyi, Tao Tao, Jason Cao, Yingling Fan, and Anu Ramaswami. 2019. "Examining threshold effects of built environment elements on travel-related carbon-dioxide emissions." *Transportation Research Part D: Transport and Environment* 75:1-12. doi: 10.1016/j.trd.2019.08.018.

Yang, Jiawen, Jason Cao, and Yufei Zhou. 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. doi: 10.1016/j.tra.2020.11.014.

Yang, Linchuan, Yibin Ao, Jintao Ke, Yi Lu, and Yuan Liang. 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. doi: 10.1016/j.jtrangeo.2021.103099.

Yang, Linchuan, Yuan Liang, Qing Zhu, and Xiaoling Chu. 2021. "Machine learning for inference: using gradient boosting decision tree to assess non-linear effects of bus rapid transit on house prices." *Annals of GIS*:1-12. doi: 10.1080/19475683.2021.1906746.

Yang, Wenyue. 2023. "The nonlinear effects of multi-scale built environments on CO2 emissions from commuting." *Transportation Research Part D: Transport and Environment* 118. doi: 10.1016/j.trd.2023.103736.

Yang, Wenyue, and Xiaoshu Cao. 2018. "Examining the effects of the neighborhood built environment on CO2 emissions from different residential trip purposes: A case study in Guangzhou, China." *Cities* 81:24-34. doi: 10.1016/j.cities.2018.03.009.

Yang, Wenyue, and Suhong Zhou. 2020. "Using decision tree analysis to identify the determinants of residents' CO2 emissions from different types of trips: A case study of Guangzhou, China." *Journal of Cleaner Production* 277:124071. doi: https://doi.org/10.1016/j.jclepro.2020.124071.

Zhang, Wenjia, Yajing Zhao, Xinyu Cao, Daming Lu, and Yanwei Chai. 2020. "Nonlinear effect of accessibility on car ownership in Beijing: Pedestrian-scale neighborhood planning." *Transportation Research Part D: Transport and Environment* 86. doi: 10.1016/j.trd.2020.102445.