Transportation Research Part D 75 (2019) 1–12

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Examining threshold effects of built environment elements on travel-related carbon-dioxide emissions

Xinyi Wu, Tao Tao, Jason Cao*, Yingling Fan, Anu Ramaswami

Humphrey School of Public Affairs, University of Minnesota, Twin Cities, 301 19th Avenue South, Minneapolis, MN 55455, United States

ARTICLE INFO

Keywords:
Urban form
Travel behavior
Climate change
Greenhouse gas emissions
Machine learning

ABSTRACT

Understanding how built environment features are associated with travel-related carbon-dioxide (CO_2) emissions is essential for planners to encourage environmentally sustainable travel through transportation and land use policies. Applying gradient boosting decision trees to the data from the Minneapolis-St. Paul metropolitan area, this study addresses two gaps in the literature by identifying critical built environment determinants of CO_2 emissions, and more importantly, illustrating threshold effects of built environment elements. The results show that three neighborhood-level built environment factors have the strongest influences on CO_2 emissions: distance to the nearest transit stop, job density, and land use diversity. The distance to downtowns also has a substantial impact. This study further confirms that built environment variables are effective only within a certain range. These threshold effects offer valuable implications for planners to achieve desirable environmental benefits efficiently.

1. Introduction

Transportation accounts for a substantial share of greenhouse gas (GHG) emissions. About one-third of the emissions in the US are from transportation (Brandes et al., 2010). According to the European Environment Agency (2018), the transport sector contributes 27% of the total emissions in Europe. To better facilitate the reduction of carbon emissions caused by transportation, it is important to decrease auto-dependency and promote sustainable modes of transport. A growing body of literature examines how to alter individual travel behavior and related carbon emissions through transportation and land use policies.

Using aggregate analyses at the city or neighborhood level, many studies substantiate the association between built environment elements and travel-related carbon emissions. These studies provide encouraging evidence to create low-carbon cities through land use and transportation planning (Hankey and Marshall, 2010). However, aggregate studies are unable to reveal why and how built environment characteristics affect individual carbon emissions. Some studies address the limitation using disaggregate analyses at the individual/household level. They show that many built environment characteristics (such as density, diversity, and design) are significantly associated with carbon emissions (e.g., Cao and Yang, 2017). Furthermore, travel decisions (such as mode choice, travel distance, and travel speed) mediate the relationship between the built environment and carbon emissions (Choi and Zhang, 2017, Ma et al., 2015).

Many studies have touched on the relative importance of different built environment characteristics on travel behavior, but the influence on carbon emissions is in need of more exploration. The relative contribution is important to urban planners because they often need to decide which built environment dimension may be compromised given various constraints. Although some studies

* Corresponding author.
E-mail address: cao@umn.edu (J. Cao).

https://doi.org/10.1016/j.trd.2019.08.018

(Ewing and Cervero, 2010, Stevens, 2017, Salon et al., 2012) assess their relative influences on travel behavior such as vehicle miles traveled (VMT), the current study shows that the key determinants of carbon emissions and VMT somewhat differ. More critically, scholars often assume that the built environment has a pre-specified (mostly, linear and log-linear) relationship with carbon emissions (Kim et al., 2007, Zahabi et al., 2016). However, the land use-travel literature suggests that the relationship may deviate from any assumptions and also vary among built environment variables (Ding et al., 2018a, Ding et al., 2018b). This study relaxes the assumption of pre-defined relationships and reveals the true associations between built environment characteristics and carbon emissions, shedding further light on lowering carbon emissions through effective urban planning.

Using the data collected in the Minneapolis-St. Paul metropolitan area (Twin Cities), this study employs the gradient boosting decision tree approach to scrutinize built environment effects on travel-related carbon-dioxide (CO_2) emissions, while controlling for demographic characteristics. It addresses the aforementioned two research gaps in the literature by answering two central questions: (1) Which dimensions of the built environment are more important to CO_2 emission reduction? (2) Does the built environment have a non-linear effect on CO_2 emissions?

The paper is organized as follows. Section 2 reviews the literature on the relationship between the built environment and carbon emissions and justifies the research gaps. Section 3 introduces the data and the modeling approach. Section 4 presents the results. The final section summarizes key findings and discusses the implications for low-carbon planning.

2. Literature review

Transport-related carbon emissions are a combined outcome of modal split, distance traveled by different modes of transport, and mode-specific emission factors. Access to transit facilitates individual transit use and hence reduces carbon emissions (American Public Transportation Association, 2008). Higher employment density and mixed land use enable residents to internalize their trips within the neighborhoods and promote low-carbon travel behavior (Frank et al., 2007, Ewing and Cervero, 2010). Therefore, built environment variables that affect mode choice and VMT are expected to contribute to variation in carbon emissions. In the literature, both aggregate studies and disaggregate studies investigate the empirical connection between the built environment and carbon emissions.

Aggregate studies measure the built environment and carbon emissions at an "aggregate" level, such as city, census tract or neighborhood, to explore their association. For instance, Gallivan et al. (2015) analyzed the effect of transit on carbon emissions, using macro-scale and neighborhood-level data of more than 300 urbanized areas from nine metropolitan regions in the US. By changing density and accessibility, transit systems could lead to an 8% decrease in VMT, energy consumption, and carbon emissions. Some studies use aggregate data to predict carbon emissions. Hankey and Marshall (2010) developed six scenarios for different levels of urban sprawling, using the data from 142 urban areas in the US. They found that compact development is associated with carbon emission reduction by 15–20% by lowering VMT and encouraging efficient vehicles as well as low-carbon fuels. Overall, these aggregate studies suggest that in terms of both statistical significance and practical importance, land use and transportation policies have the great potential for carbon mitigation. However, because of the aggregate nature of the data, these studies can illustrate only the correlation between built environment elements and carbon emissions, but cannot scrutinize the mechanisms by which they affect individual travel decisions (Handy, 1996). This limitation calls for examining disaggregate data, which is employed in our study.

Disaggregate studies use individual- or household-level data. Some studies employ structural equation models to depict the causal pathways between built environment characteristics and carbon emissions. Based on the analyses of activity/travel diary surveys in Guangzhou, Beijing, and Austin, previous studies show that the built environment affects carbon emissions through its influences on vehicle ownership (Cao and Yang, 2017), travel distance (Cao and Yang, 2017, Ma et al., 2015, Choi and Zhang, 2017), mode choice (Ma et al., 2015), trip frequency (Choi and Zhang, 2017), and travel speed (Choi and Zhang, 2017). Furthermore, the literature on the built environment and fuel consumption suggests that carbon emissions may be affected by other travel behavior variables such as acceleration/deceleration and vehicle type choice (Wang et al., 2014, Brownstone and Golob, 2009, Zhu et al., 2019). In terms of specific built environment dimensions, previous studies conclude that travel-related carbon emissions are associated with the five "Ds" of the built environment. Specifically, distance to city center is negatively associated with carbon emissions (e.g., Cao and Yang, 2017, Wang et al., 2017). Density has a negative association with carbon emissions (e.g., Cao and Yang, 2017, Choi and Zhang, 2017). Land use diversity also has a negative effect (e.g., Wang et al., 2013, Choi and Zhang, 2017). Street connectivity, an indicator of road network design, is negatively correlated to carbon emissions (e.g., Wang et al., 2013, Xu et al., 2018). Distance to transit has a positive association with carbon emissions (e.g., Barla et al., 2011, Boarnet et al., 2017). Collectively, these disaggregate studies empirically elaborate on why and how the built environment contributes to carbon emissions.

Since the literature suggests that travel-related carbon emissions are influenced by the multi-dimensional built environment, planners wonder which dimensions have a stronger influence than others. This question is important for at least two reasons. First, the built environment contains measures at both regional (such as regional accessibility) and local (such as land use mix) levels. Changing the built environment at these two levels requires different approaches since land use planning is fragmented in the US. If the regional location of residences has a trivial effect on carbon emissions, local planning itself is sufficient to achieve the goal of carbon reduction. Otherwise, it is desirable to employ a regional approach, namely, coordinated effort among municipalities by the Metropolitan Planning Organization (MPO). Urban growth boundary in Portland, OR, and metropolitan urban service area in the Twin Cities are examples of regional planning. Second, if local planners cannot implement all planning instruments for some reason, a tradeoff among the instruments becomes necessary. Then, which characteristics of the built environment should they emphasize? The literature on land use and travel behavior offers hints. A recent meta-analysis shows that distance to downtown has the largest effect (elasticity) on VMT, followed by residential density, job accessibility by auto, street network design, and land use diversity (Stevens, 2017). Another meta-analysis concludes that transit use is affected primarily by transit proximity and road network design and secondarily by land use diversity (Ewing and Cervero, 2010). Because both transit and personal vehicles contribute to carbon emissions, the answer to the tradeoff question is unclear.

Furthermore, the literature shows that the built environment has a non-linear influence on travel behavior, consistent with the speculation of Van Wee and Handy (2016). The non-linear influence indicates that the marginal effect of an independent variable on the dependent variable depends on the value of that variable (Galster, 2018). Log-linear and quadratic functions are two common examples of non-linear influences. Threshold effect is a less studied non-linear influence. In this context, it means that the marginal effect of a built environment variable on travel behavior changes (either increases or decreases) dramatically after passing certain thresholds (Galster, 2018). For example, using data from Oslo, Norway, Ding et al. (2018a) found threshold effects of built environment elements on driving distance. Within 12 km from the city center, individual driving distance grows slowly as distance to the city center increases. Its slope becomes much steeper beyond the 12 km threshold. Its marginal effect diminishes after exceeding 20 km. Moreover, population density has a negative influence on driving distance, but the influence is within a narrow range of population density. They suggested that "a substantial part of the future population growth should take place within 12 km from the city center" (p. 116). A similar question emerges: where should planners direct future population and employment growth for the sake of carbon emission mitigation? Furthermore, mixed-use and dense development are often opposed by local residents because of Nimbyism (not in my back yard). Planners and public officials are cautious about up-zoning (Cervero et al., 2004). Under this circumstance, the following questions become intriguing: Should planners emphasize employment densification or population densification? How dense is enough to reduce carbon emissions efficiently? How about other dimensions of the built environment? This study addresses these research questions.

3. Data and method

3.1. Data and variables

The data came from the Neighborhood Environment, Daily Activities, and Well-Being Study conducted in the Minneapolis-St Paul Twin Cities area by the Sustainable Healthy Cities project from October 17, 2016 to October 25, 2017. Six neighborhoods were selected based on urban location versus suburban location, proximity to light rail transit, and income level (Fig. 1). In particular, for urban areas, we selected one low-income and one middle-income neighborhood without light-rail access and one low-income and one middle-income neighborhood with light-rail access. For suburban areas, we chose one low-income and one middle-income neighborhood. Urban areas have better transit services than suburban areas.

We sent out survey invitations to about 1700 households who lived in 921 randomly selected census blocks (out of the 2443 census blocks in the study neighborhoods) and our goal was to recruit 400 respondents. Although the size of different neighborhoods varies (Fig. 1), we aim to recruit a similar number of respondents from each neighborhood. In particular, all households were first sent a postcard with a brief study description and the contact information of the research team. Once contacted, the survey team explained the study in detail to the participants and those who chose to participate set an appointment with the survey team. Next, a survey team member visited the participants at home to collect their demographic information and provide instructions on how to use a smart phone application called Daynamica™ (previously called SmarTrAC (Fan et al., 2015, Fan et al., 2017)). The application

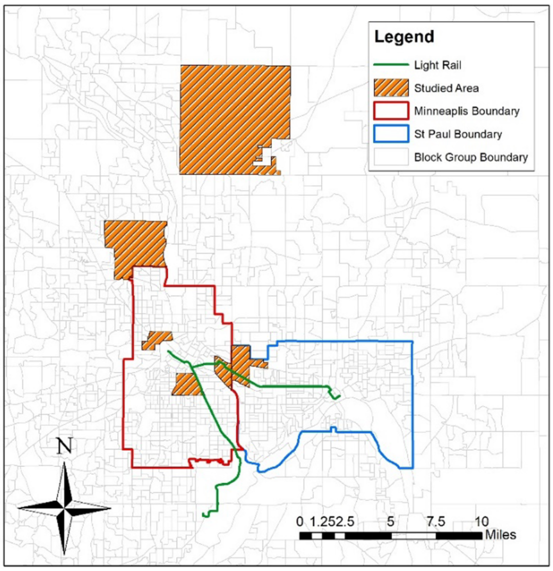


Fig. 1. Study area.

Table 1
Variable definition and descriptive statistics.

Variables			Description	Mean	Std. Dev.	Min	Max
CO2 emissions		Daily CO_2 emission amount	The daily CO_2 emission amount in pounds per day of a respondent	23.11	16.20	0.69	91.00
Built environment characteristics	Regional location	Distance to Downtown Minneapolis	The distance in mile from the respondent’s home to downtown Minneapolis	5.39	4.38	1.08	22.36
		Distance to Downtown St. Paul	The distance in mile from the respondent’s home to downtown St. Paul	10.56	4.94	5.26	28.69
	Transit	Distance to nearest stop	The distance in mile from the respondent’s home to the nearest regular bus stop	0.24	0.34	0	2.59
		Stop number	Number of transit stops within a half-mile buffer of the respondent’s home	24.34	15.42	0	60
		Presence of LRT	A dummy variable indicating whether or not the respondent’s home is within a half-mile buffer of the light rail station	0.14	0.35	0	1
	Density	Population density	Population density within a half-mile buffer of the respondent’s home	8.98	6.00	0.53	25.49
		Job density	Employment density within a half-mile buffer of the respondent’s home	7.84	8.01	0.10	31.65
	Design	Intersection number	Number of four or more way intersections within a half-mile buffer of the respondent’s home	25.51	17.81	0	67
	Diversity	Land use entropy index	Land use entropy within a half-mile buffer of the respondent’s home	0.53	0.15	0.03	0.95
Demographic characteristics		Age	Age	49.81	16.63	18	91
		Male	Male	0.34	0.47	0	1
		Employment condition	A dummy variable indicating whether the respondent is a full- or part-time employee	0.68	0.47	0	1
		Student	A dummy variable indicating whether the respondent is a full-time student	0.09	0.28	0	1
		Household size	Number of household members in the household	2.76	1.77	1	21
		Hispanic	A dummy variable indicating whether the respondent is Hispanic	0.03	0.18	0	1
		White	A dummy variable indicating whether the respondent is white	0.79	0.41	0	1
		Asian	A dummy variable indicating whether the respondent is Asian	0.04	0.19	0	1
		Black	A dummy variable indicating whether the respondent is black	0.10	0.29	0	1
		Less Bachelor	A dummy variable indicating whether the respondent has a degree less than bachelor, including 'less than a high school diploma', 'high school diploma or general educational development', 'associates or technical school degree', and 'some college, no degree'	0.28	0.45	0	1
		Bachelor	A dummy variable indicating whether the respondent has a degree of bachelor	0.33	0.47	0	1
		Low income	A dummy variable indicating whether the respondent’s income is less than $25,000	0.19	0.39	0	1
		Medium income	A dummy variable indicating whether the respondent’s income is between $25,000 and $75,000	0.34	0.48	0	1
		Having a driver's license	A dummy variable indicating whether the respondent has a driver’s license	0.90	0.30	0	1
		Home owner	A dummy variable indicating whether the respondent own current home	0.72	0.45	0	1

Table 2
CO_2 emission factors for modes of transport.

Mode	Emission rate (pounds/mile)
Car (or in vehicle)	0.96
Light Rail	0.36
Bus	0.64

Source: U.S. Department of Transportation (2010).

records individual daily trips and automatically remind the respondent to provide additional travel information (such as companion information and trip purpose). We tracked respondent trip data for seven consecutive days. The recorded information includes start and end locations, start and end times, and travel mode. After data processing, this study used 360 observations for further analysis.

Table 1 defines CO_2 emissions, built environment variables, and demographics used in this study and presents their descriptive statistics. Travel-related average daily CO_2 emissions C was calculated using the following equation:

$$C = \sum_{i=1}^{p} \sum_{j=1}^{q_i} \frac{D_{ij} R_m}{p}, \tag{1}$$

where D_{ij} is the distance of the jth trip on the ith day travelled by a respondent, R_m is CO_2 emission rate (pound/mile) for the corresponding mode of the trip, q_i is the number of trips on the ith day, and p is the number of valid days in a week. In this study, CO_2 emission rates (Table 2) are based on the estimates of the USDOT (U.S. Department of Transportation, 2010). In this sample, about 97% of CO_2 emissions are associated with personal vehicles (see Fig. A2 in the Appendix). This proportion is not surprising given the modal share in the U.S. in general and the mode-specific travel distance in this sample (see Fig. A1).

We assembled built environment characteristics around respondents' home locations, using publicly accessible datasets from Minnesota Geospatial Commons (https://gisdata.mn.gov/). It is worth noting that the land use entropy index is based on five types of land uses including agricultural, green area, industrial, official, and commercial uses. Land use data were derived from the 2016 Generalized Land Use Inventory dataset in the seven counties of the Minneapolis-St. Paul Metropolitan Area. Demographic variables were from the survey.

3.2. Modeling method

This study employed the gradient boosting decision tree (GBDT) model to explore the association between built environment characteristics and CO_2 emissions. The basis of performing GBDT is to build decision trees. The algorithm of decision trees is easy to understand, convenient to perform, and accurate in predicting, making it popular in dealing with many classification and regression problems. GBDT starts with classifying the sample into sub-groups through decision trees. The predicted values for the observations in each sub-group is the average of these observations. This classification generates prediction errors, which help the GBDT algorithm to adjust the weight of each independent variable in the next round of classification and prediction. During this iteration, a set of decision trees is built. GBDT will calculate the prediction error at each stage of the iteration and select the number of trees that minimizes the prediction error. The model with the optimal number of trees would be the basis for further analyses. The mathematical notation of GBDT can be found in the appendix.

This method is superior to traditional regression models in the following ways (Ding et al., 2018a). First, it produces predictions that are more accurate. Second, it can accommodate variables with missing values. Third, because the response to a predictor in decision trees depends on other predictors at the higher levels of trees, it can address interaction effects among predictors, easing the concern of multicollinearity (Elith et al., 2008). On the other hand, the GBDT approach cannot produce statistical inference; that is, we are unable to tell whether a variable has a statistically significant influence on the outcome variable.

4. Results

To obtain robust model results, a five-fold cross-validation procedure is employed to develop the GBDT model using the R programming language. Specifically, the sample is divided into five subsets, at each iteration, the model is fitted using four different subsets (80% of the data) and validated by the remaining subset (20% of the data). Overall, three important parameters are determined: the number of trees, shrinkage, and tree complexity. Following Ding et al. (2018a), this study set a maximum of 3000 trees, kept the shrinkage parameter at 0.001, and chose five-way interaction. After 2295 boosting iterations, the model obtained its best results. The pseudo-R^2, "the fraction of variation explained by the model" (Schonlau, 2005), is 0.30 (the R^2 of the linear regression model with the same set of independent variables is 0.22). Then, the relative importance of independent variables and partial dependence plots are derived for further analysis.

4.1. The relative influence of independent variables

Table 3 presents the relative influence of built environment characteristics and demographic variables. The relative influence of a predictor measures its relative empirical improvement in reducing prediction error. The total relative influence of all predictors adds

Table 3
The relative influence of independent variables.

Variables			Relative importance (%)	Sum (%)
Built environment characteristics	Regional location	Distance to Downtown Minneapolis	6.9	58.3
		Distance to Downtown St. Paul	5.9	
	Transit	Distance to the nearest stop	5.2	
		Stop number	4.5	
		Presence of LRT	0.1	
	Density	Population density	5.7	
		Job density	7.5	
	Design	Intersection number	4.1	
	Diversity	Land use entropy index	18.4	
Demographic characteristics		Bachelor	10.1	41.7
		Age	14.4	
		Low income	0.5	
		Employment condition	5.3	
		Household size	2.2	
		Male	0.5	
		Medium income	0.5	
		Number of vehicles	1.3	
		White	2.0	
		Student	0.3	
		Having a driver's license	1.3	
		Less Bachelor	2.7	
		Black	0.5	
		Hispanic	0.0	
		Asian	0.0	
Total Relative Importance				100

up to 100%. In terms of individual built environment characteristics, land use entropy index, a measure of land use diversity, has the largest contribution (18.4%), followed by job density (7.5%), distance to Downtown Minneapolis (6.9%) and distance to Downtown St. Paul (5.9%). The importance of land use diversity and job density makes sense. Denser and more mixed developments facilitate the use of transit and non-motorized modes (Ewing and Cervero, 2010). Even if individuals choose to drive, their driving distance tends to be shorter compared to those living in areas with low-density and homogenous land uses. As a group of variables, distance to the two downtowns has an important contribution as well. The regional location of a residence largely determines land use characteristics and transportation infrastructure surrounding the residence (Cao et al., 2019): urban areas tend to have denser development and more transit supply than suburban areas. Among demographics, having a Bachelor's degree and age have a relatively stronger predictive power for CO_2 emissions than others, contributing 10.1% and 14.4%, respectively.

Furthermore, the built environment is more important in predicting travel-related CO_2 emissions than demographics. Specifically, the nine built environment characteristics collectively contribute to almost sixty percent of the predictive power, whereas the 15 demographics account for about forty percent. This finding is consistent with two other studies that applied the GBDT approach to examine built environment effects on driving behavior (Ding et al., 2018a, Ding et al., 2018b). However, it is different from studies that employed parametric approaches (Singh et al., 2018, Kitamura et al., 1997, Stead, 2001): they often conclude that demographics have a more important influence on travel behavior than the built environment.

The divergence makes sense for a few reasons. First, different from linear regression, the GBDT approach does not assume any predefined relationship between predictors and the response. When built environment variables have threshold effects on CO_2 emissions, linear regression ignores these threshold effects, because of the linearity assumption, and produces biased results. This model specification underestimates built environment effects. Furthermore, many built environment variables are correlated with each other to some extent. For example, high density tends to be associated with mixed land use. Because of the correlation, the inclusion of one variable may make the other statistically insignificant. Researchers sometimes manually drop insignificant variables from the model. This practice understates built environment effects. Furthermore, built environment variables may have an interaction effect on travel behavior. For instance, the influence of transit supply on transit use may depend on density intensity. In the land use-travel literature, the possible interactions among built environment variables are often ignored (Boarnet, 2011). Accordingly, the impacts of built environment variables are not fully captured. By contrast, the model used in this study incorporates important advantages of tree-based methods: it can automatically capture interaction effects between variables and help handle the multicollinearity issue. Moreover, built environment variables could have a more significant impact on certain aspects of travel behavior than demographics. For example, after reviewing the literature, Ewing and Cervero (2001, pp. 106-107) concluded that "Trip frequencies appear to be primarily a function of the socioeconomic characteristics of travelers and secondarily a function of the built environment, trip lengths [of single trips] are primarily a function of the built environment and secondarily a function of socioeconomic characteristics, and mode choices depend on both … Studies of overall VMT or VHT find the built environment to be much more significant…" CO_2 emissions used in this study are closely related to VMT. Therefore, it makes sense that the built environment has a greater influence on CO_2 emissions than demographic variables.

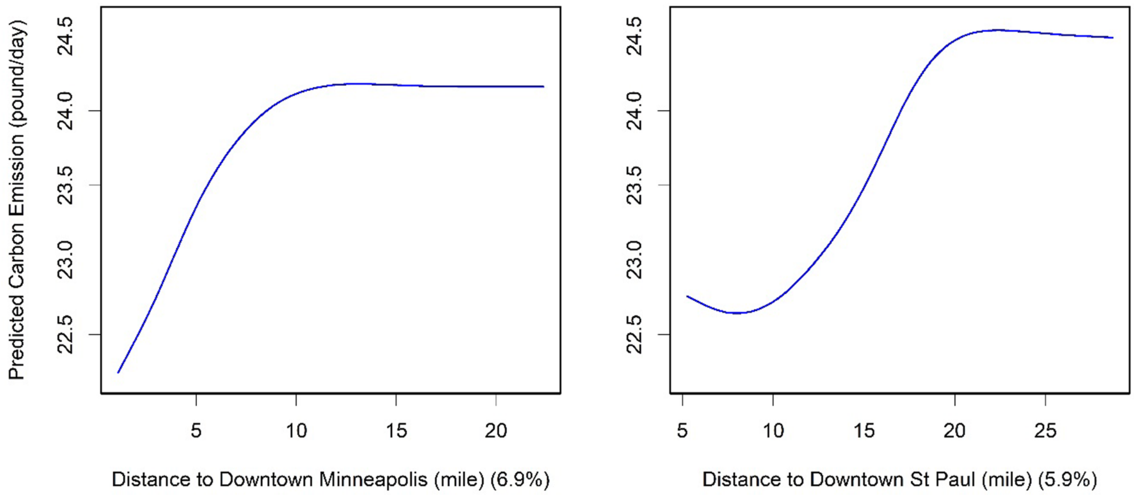


Fig. 2. The relationship between personal daily travel-related CO_2 emissions and distance to downtown Minneapolis (left) and downtown St. Paul (right).

4.2. The non-linear impact of built environment variables on CO_2 emissions

Using the gradient boosting decision tree algorithm, we produced partial dependence plots to illustrate the relationships between built environment variables and travel-related CO_2 emissions. A partial dependence plot demonstrates the marginal effect of an independent variable on the predicted response while controlling for all other variables in the model.

Fig. 2 illustrates the impact of distance to downtowns, with the relative importance presented along with the label for the horizontal axis. Distance to downtown Minneapolis has positive threshold effects on CO_2 emissions. Ten miles appear to be the elbow point. When the distance is smaller than ten miles, its slope is relatively big. However, when it exceeds ten miles, its slope decreases substantially. Living close to the primary center offers people higher accessibility and more mode choice alternatives to driving, thus lowering CO_2 emissions. By contrast, when people live farther from the primary center, they tend to reduce transit use and drive longer distances to reach their destinations. In other words, as home locations get closer to the city center, accessibility becomes larger. Accordingly, the distance required to access daily activities tends to be lower, resulting in decreased VMT and CO_2 emissions. Alternative modes such as walking, cycling, and riding transit also become more attractive, so CO_2 emissions of trips would be reduced. In terms of distance to downtown St. Paul, its effect on CO_2 emissions is the lowest when it reaches about nine miles. This pattern is due to the unique urban form of the Twin Cities area: it has two primary urban centers and Downtown Minneapolis and Downtown St. Paul are about nine miles apart. In our sample, those who live nine miles away from Downtown St. Paul reside close to Downtown Minneapolis. Based on the left figure in Fig. 2, it is reasonable that they have the lowest CO_2 emissions.

The two plots in Fig. 3 illustrate the associations between CO_2 emissions and transit-related variables: distance to the nearest transit stop and number of stops within half a mile of a residence. Overall, accessible transit service has a negative association with

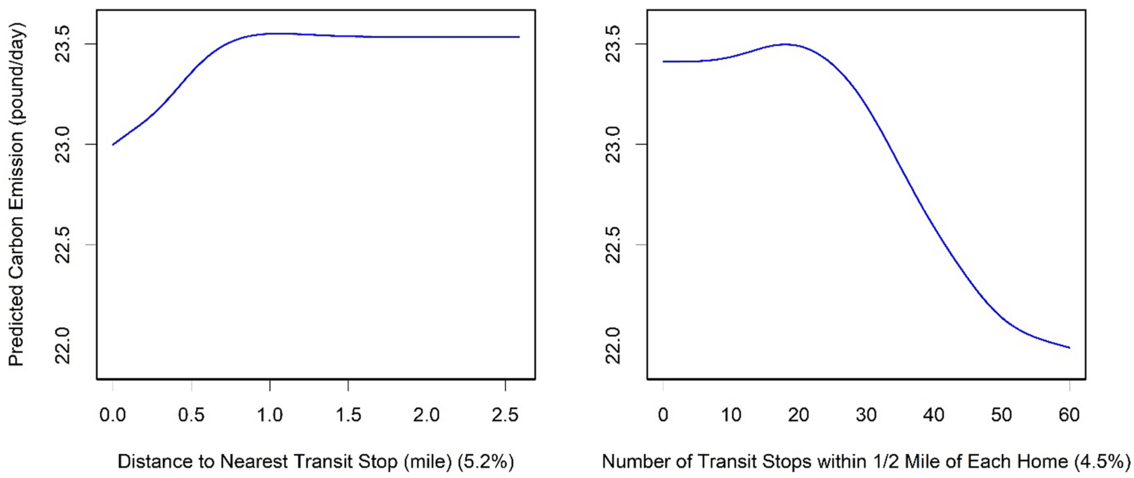


Fig. 3. The relationships between personal daily travel-related CO_2 emissions and transit features.

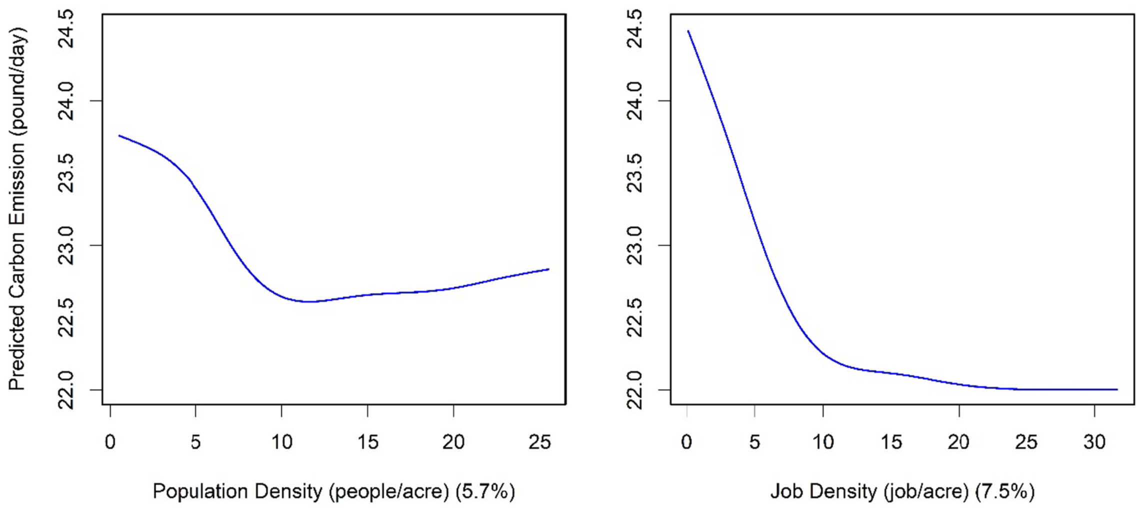


Fig. 4. The relationship between CO_2 emissions and density.

CO_2 emissions. This pattern is consistent with Gallivan et al. (2015) and Haas et al. (2010). Within a narrow range, distance to the nearest transit stop is positively associated with CO_2 emissions. In particular, as the distance grows from zero to approximately one mile, CO_2 emissions increase linearly from 23 to 23.5 lb per day. Beyond one mile, the impact on CO_2 emissions stabilizes and ceases to increase. Moreover, when there are fewer than 20 stops within a half-mile buffer of a residence, the number of stops shows a weak association with CO_2 emissions. Beyond this range, CO_2 emissions begin to have a substantial decrease as the number of stops increases. It is worth noting that the presence of LRT has a limited impact on CO_2 emissions, presumably because its effect has been captured by the other two transit variables.

The two plots in Fig. 4 show the influences of density variables: population density and job density. Overall, density has a negative relationship with CO_2 emissions, congruent with the literature on the density-VMT connection (Stevens, 2017). Nevertheless, the effect of population density reaches a low point at about ten people/acre and then slowly increases. This increase could be because of the concentration of human activities due to high density. Within the range of ten jobs per acre, employment density is negatively and linearly associated with CO_2 emissions. Then its marginal effect becomes negligible. Both the relative importance and partial dependence plots suggest that job density has a larger effect on CO_2 emissions than population density. The threshold effect pattern of job density is consistent with the findings of Ding et al. (2018a).

As a diversity indicator, land use entropy index is negatively correlated with CO_2 emissions (Fig. 5). This negative relationship is consistent with the literature (Frank et al., 2000, Zhang et al., 2012, Wang et al., 2013). However, the relationship appears only in a certain range of the entropy index. When land use diversity is relatively homogeneous (entropy < 0.4), it has a trivial influence on CO_2 emissions. When there are one relatively dominating land use type and several other land use types ($0.4 \le$ entropy ≤ 0.7), CO_2 emissions decrease by about five pounds per day. Land use diversity has no additional effect once all types of land use are relatively

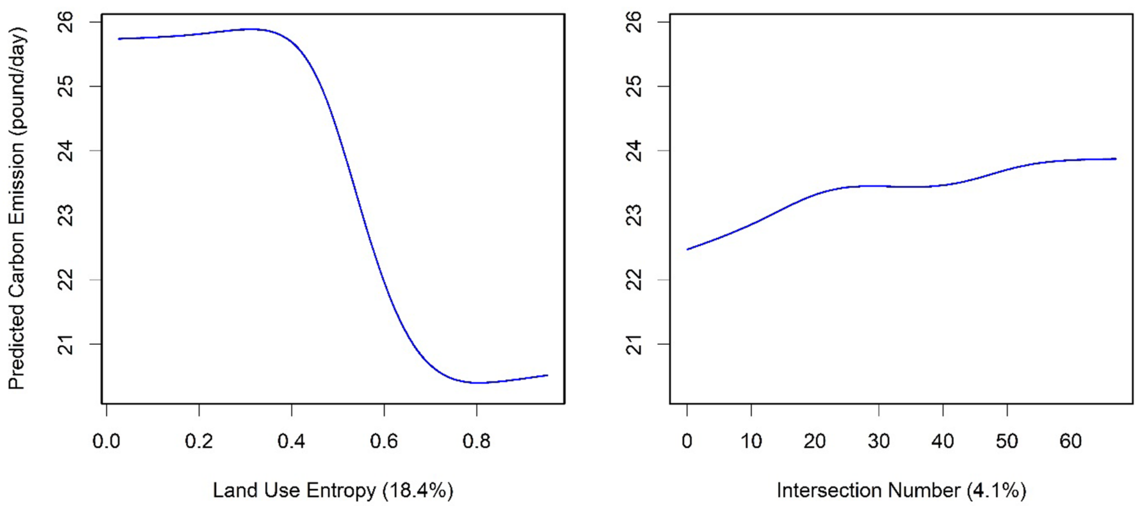


Fig. 5. The influences of diversity and network design on CO_2 emissions.

evenly distributed (entropy > 0.7). Overall, land use mix should reach a certain level to be effective and it has a diminishing return once it reaches another threshold.

The number of intersections within a half-mile buffer of residence shows a positive correlation with CO_2 emissions (Fig. 5). This positive relationship appears counterintuitive. Intersection density is a measure of street connectivity. The literature suggests that it is negatively associated with driving distance and positively associated with transit use (Ewing and Cervero, 2010). However, a grid street network could increase the number of car trips. Theoretically, it may increase or decrease mode split and driving distance "depending on how sensitive trips by each mode are to trip length" (Crane, 1996, p.124). Another possibility is that because a grid street network is conducive to transit operation and transit use (Ewing and Cervero, 2010), the growth in transit use increases CO_2 emissions offsetting the carbon benefits of the grid network.

5. Discussion and conclusions

This study adopted the gradient boosting decision tree approach to examine the nuanced relationship between built environment characteristics and travel-related CO_2 emissions. It makes two important contributions to the literature. First, it assesses the relative importance of different built environment characteristics in reducing CO_2 emissions and the collective contribution of built environment variables relative to demographic characteristics. Second, it depicts the non-linear relationships between built environment variables and CO_2 emissions, challenging the commonly used linearity assumption in the literature. Because the patterns of non-linear relationships vary among built environment variables, this renders parametric specification of non-linear relationships (such as log-linear regression, piecewise regression, and polynomial regression) inefficient and inaccurate.

This study has several limitations. First, the study area does not include neighborhoods outside of the beltway and far-flung suburban communities. Therefore, we may understate the impacts of built environment variables on CO_2 emissions. Second, the computation of CO_2 emissions was based on existing emission factors and we did not consider vehicle or transit occupancy, which affects the true emissions associated with travel. However, our approach is within mainstream literature. Third, although a large sample is desirable, we were unable to recruit more respondents because the survey was expensive and our budget was constrained. On the other hand, the literature has substantiated that the gradient boosting decision tree approach is a powerful tool to deal with small sample size: it can produce reliable results when the sample size is smaller than 100 (Isıkhan et al., 2016, Yang et al., 2010). Fourth, because the data are cross-sectional, the influence found in this study is more of an association than causality, similar to most studies in the literature. Fifth, the built environment variables used in this study are measured with 0.5 mile of the home location. Built environment measures at different spatial scales (0.5-, 1-, 2-, 3-, 5-mile radius buffers) may have differing impacts on CO_2 emissions. Future studies should examine the sensitivity of the built environment impacts using different spatial scales.

Nevertheless, this study offers important insights into the literature and planning practices. First, among individual built environment characteristics, these three variables have the largest predictive power: distance to the closest transit stop, job density, and land use diversity. Distance to downtowns, a proxy for destination accessibility (Ewing and Cervero, 2010), also has strong predictive power. More importantly, built environment variables collectively are more important in predicting CO_2 emissions than demographic variables, differing from most studies using parametric models. Furthermore, this study found that built environment variables have threshold effects on CO_2 emissions.

The empirical results have important implications for local planning. The city of Minneapolis aims to reduce CO_2 emissions by 80% by 2050 while attracting more residents and jobs. In the scenario of business as usual, population and employment will grow substantially in suburban and exurban areas in the Twin Cities. This study shows that if planners direct future population and employment growth to urban and inner-ring areas (or up to 12 miles from downtown Minneapolis), the amount of increase in transport-related CO_2 emissions associated with population and employment growth can be minimized. Minneapolis plans to adopt a few strategies to increase housing supply (and hence population density). It will allow "high-density housing in and near downtown" and "multifamily housing on select public transit routes," and increase housing density throughout the city (Minneapolis, 2018, pp. 105-106). The former two strategies are corresponding to the variables of distance to downtown and distance to transit, respectively, in this study. The city expects a 5% growth in employment opportunities by 2040 and establishes a policy to direct employment growth toward downtown and the areas well-served by transit. The latter action helps improve job density throughout the city. Furthermore, the city allows and facilitates "a dense mix of housing, employment, and commercial goods and services" near current and future light rail transit and bus rapid transit (Minneapolis, 2018, p.228). This action is related to three "Ds" of the built environment: increasing density and diversity near transit. Moreover, the city aims to produce "complete neighborhoods," which enable residents to live their daily lives by transit and active travel to reduce travel-related CO_2 emissions. Our research findings provide significant evidence that if implemented successfully, these policies can help the City of Minneapolis to achieve the goal of carbon mitigation. The threshold effects found in this study can also offer nuanced guidance on specific planning effort.

Furthermore, the results of this study can be generalized to other regions in the US and Europe. Specifically, land use mix, employment density, and distance to city centers are important for CO_2 emission mitigation. In general, cities should direct future development to existing urban areas, instead of sprawling to far-flung suburban areas, and promoting dense and mixed-use development around transit networks. Moreover, this study, as well as others (Ding et al., 2018a, Ding et al., 2018b), shows the prevalence of threshold effects of built environment variables on travel behavior and associated environmental outcomes, and the thresholds vary among different regions. The latter finding is not surprising given different urban forms and travel patterns in different regions. To better inform local planning, we encourage planning scholars to apply the GBDT approach to identify the thresholds in other regions.

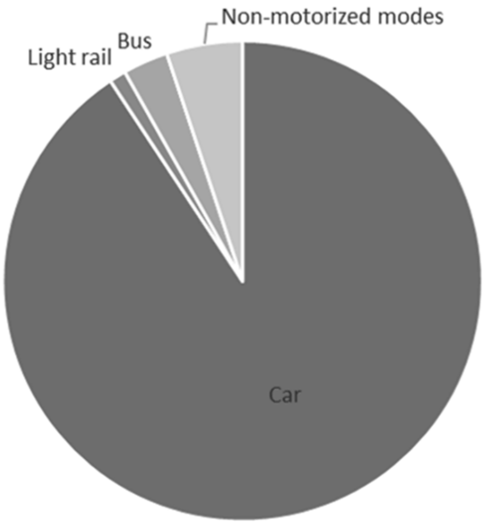
Acknowledgements

This study was funded by the Sustainable Research Network project of the National Science Foundation of USA: Integrated Urban Infrastructure Solutions for Environmentally Sustainable, Healthy and Livable Cities**.** (Award #: 1444745). Mr. Kirti Das led the survey administration.

Appendix A

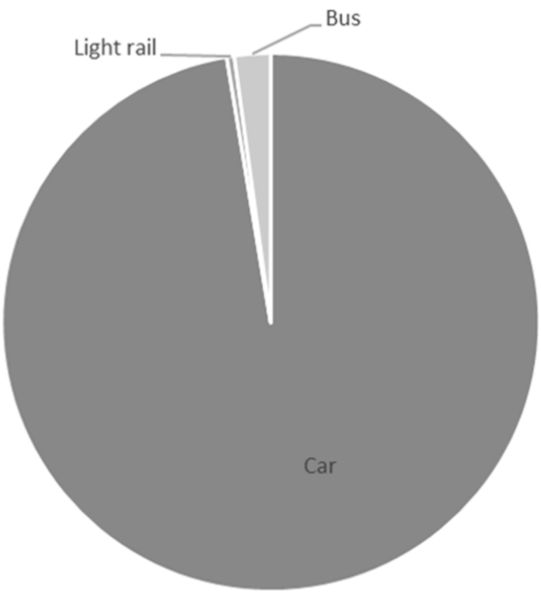
1. Travel distance and CO_2 emissions by mode
2. Gradient boosting decision tree

Gradient boosting assembles *M* decision trees to produce a strong prediction model. For a sample of ***(y, x)***, the goal of GBDT is to fit a function of $\widehat{F(x)}$ to minimize the loss function $\psi[y, F(x)]$. We developed the model using the R-based “gbm” package (Ridgeway, 2007). According to Friedman (2002), the output of a gradient boosting model in Step $m(0 < m \leq M)$ is:



Distance travelled by each mode (mile)	
Car	58,200
Light rail	720
Bus	1,951
Non-motorized modes	3,289

Fig. A1. Travel distance by mode.



Carbon Emission by modes (Pound)	
Car	55,872
Light rail	259
Bus	1,653

Fig. A2. CO_2 emissions by mode.

Table A1
The Relative Contribution of Built Envionment and Demographic Chracteristics in Predicting Daily VMT.

Independent variables			Relative importance (%)	Sum (%)
Built environment characteristics	Regional location	Distance to Downtown Minneapolis	9.3	56.7
		Distance to Downtown St. Paul	5.9	
	Transit	Distance to nearest stop	5.4	
		Stop number	2.9	
		Presence of LRT	0.1	
	Density	Population density	5.5	
		Job density	5.1	
	Design	Intersection number	4.1	
	Diversity	Land use entropy index	18.3	
Demographic characteristics		Bachelor	11.5	43.3
		Age	15.8	
		Low income	0.9	
		Full-time worker	4.5	
		Household size	2.0	
		Male	0.4	
		Medium income	0.4	
		Homeownership	1.5	
		White	3.7	
		Student	0.0	
		Having a driver's license	0.3	
		Less Bachelor	2.1	
		Black	0.3	
		Hispanic	0.0	
		Asian	0.0	

$$\widehat{F_m(x)} = \widehat{F_{m-1}(x)} + \xi \sum_{j=1}^{J} \gamma_{jm} I(x \in R_{jm}), \quad where\ 0 < \xi \leq 1 \tag{2}$$

where J is the number of regions partitioned by a decision tree, γ_{jm} is the value of optimal gradient for the region R_{jm}, which could make the current function $\widehat{F_m(x)}$ obtain the smallest loss; $I = 1$ if × falls into R_{jm} and $I = 0$ otherwise; ξ is the shrinkage parameter, also called learning rate. After M iteration of the steps above, we will have the final model.

Breiman et al. (1984) developed a measure to approximate the relative influence of an independent variable x_κ in predicting the dependent variable in a decision tree T:

$$I_\kappa^2(T) = \sum_{t=1}^{J-1} \hat{\tau}_t^2 I(v(t) = \kappa) \tag{3}$$

where $\hat{\tau}_t^2$is the reduction in squared error when predictor x_κ is used as the splitting variable, which is $v(t)$ in the equation above, and J is the number of partitioned regions by the tree T. For a series of decision trees$\{T_m\}_1^M$, we can generalize Eq. (3) by averaging all trees obtained through the GBDT approach:

$$I_\kappa^2 = \frac{1}{M} \sum_{m=1}^{M} I_\kappa^2(T_m) \tag{4}$$

3. Modeling results for VMT

When using VMT as the dependent variable, built environment characteristics show slightly less predictive power than them in the model using CO_2 emission (58.3% v. 56.7%). The most influential built environment features are relatively consistent between these two models. Land use mix is the most important in both models. Distance to Downtown Minneapolis and to Downtown St. Paul also have a great relative influence. But job density tend to be less influential when using VMT as the dependent variable (see Table A1).

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trd.2019.08.018.

References

American Public Transportation Association, 2008. Public Transportation Reduces Greenhouse Gases and Conserves Energy. American Public Transportation Association (APTA), Washington DC, US.

Barla, P., Miranda-Moreno, L.F., Lee-Gosselin, M., 2011. Urban travel CO2 emissions and land use: a case study for Quebec City. Transport. Res. Part D: Transp. Environ. 16, 423–428.

Boarnet, M.G., 2011. A broader context for land use and travel behavior, and a research agenda. J. Am. Plan. Assoc. 77, 197–213.

Boarnet, M.G., Wang, X., Houston, D., 2017. Can new light rail reduce personal vehicle carbon emissions? A before-after, experimental-control evaluation in Los Angeles. J. Region. Sci. 57, 523–539.

Brandes, U., Maccleery, R., Peterson, J., Johnston, M., 2010. Land Use and Driving: The Role Compact Development Can Play in Reducing Greenhouse Gas Emissions. Urban Land Institute, Washington, DC.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. CRC Press.

Brownstone, D., Golob, T.F., 2009. The impact of residential density on vehicle usage and energy consumption. J. Urban Econ. 65, 91–98.

Cao, X., Næss, P., Wolday, F., 2019. Examining the effects of the built environment on auto ownership in two Norwegian urban regions. Transport. Res. Part D: Transp. Environ. 67, 464–474.

Cao, X., Yang, W., 2017. Examining the effects of the built environment and residential self-selection on commuting trips and the related CO2 emissions: an empirical study in Guangzhou, China. Transport. Res. Part D: Transp. Environ. 52B, 480–494.

Cervero, R., Murphy, S., Ferrell, C., Goguts, N., Tsai, Y.-H., Arrington, G.B., Boroski, J., Smith-Heimer, J., Golem, R., Peninger, P., Nakajima, E., Chui, E., Dunphy, R., Myers, M., McKay, S., 2004. Transit-Oriented Development in the United States: Experiences, Challenges, and Prospects. Transportation Research Board, Washington, DC.

Choi, K., Zhang, M., 2017. The net effects of the built environment on household vehicle emissions: a case study of Austin, TX. Transport. Res. Part D: Transp. Environ. 50, 254–268.

Crane, R., 1996. On form versus function: will the new urbanism reduce traffic, or increase it? J. Plan. Educ. Res. 15, 117–126.

Ding, C., Cao, X., Næss, P., 2018a. Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. Transport. Res.. Part A 110, 107–117.

Ding, C., Cao, X., Wang, Y., 2018b. Synergistic effects of the built environment and commuting programs on commute mode choice. Transport. Res. Part A: Policy Practice 118, 104–118.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813.

European Environment Agency, 2018. Greenhouse gas emissions from transport.

Ewing, R., Cervero, R., 2001. Travel and the built environment: a synthesis. Transport. Res. Rec.: J. Transport. Res. Board 1780, 87–114.

Ewing, R., Cervero, R., 2010. Travel and the built environment – a meta-analysis. J. Am. Plan. Assoc. 76, 265–294.

Fan, Y., Wolfson, J., Adomavicius, G., 2017. Travel and Activity Capturing. Google Patents.

Fan, Y., Wolfson, J., Adomavicius, G., Vardhan Das, K., Khandelwal, Y., Kang, J. 2015. SmarTrAC: A smartphone solution for context-aware travel and activity capturing. Center for Transportation Studies, University of Minnesota.

Frank, L.D., Saelens, B.E., Powell, K.E., Chapman, J.E., 2007. Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? Soc. Sci. Med. 65, 1898–1914.

Frank, L.D., Stone Jr, B., Bachman, W., 2000. Linking land use with household vehicle emissions in the central Puget Sound: methodological framework and findings. Transport. Res. Part D: Transp. Environ. 5, 173–196.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38, 367–378.

Gallivan, F., Rose, E., Ewing, R., Hamidi, S., Brown, T. 2015. Quantifying Transit's Impact on GHG Emissions and Energy Use—The Land Use Component.

Galster, G.C., 2018. Nonlinear and threshold effects related to neighborhood: implications for planning and policy. J. Plan. Literature 33, 492–508.

Haas, P., Miknaitis, G., Cooper, H., Young, L., Benedict, A. 2010. Transit oriented development and the potential for VMT-related greenhouse gas emissions growth reduction. Report of the Center for Neighborhood Technology for the Center for Transit Oriented Development, pp. 1–64.

Handy, S., 1996. Methodologies for exploring the link between urban form and travel behavior. Transport. Res. Part D: Transp. Environ. 1, 151–165.

Hankey, S., Marshall, J.D., 2010. Impacts of urban form on future US passenger-vehicle greenhouse gas emissions. Energy Policy 38, 4880–4887.

Isikhan, S.Y., Karabulut, E., Alpar, C.R., 2016. Determining cutoff point of ensemble trees based on sample size in predicting clinical dose with DNA microarray data. Comput. Math. Methods Med. 2016, 9.

Kim, S., Ulfarsson, G.F., Hennessy, J.T., 2007. Analysis of light rail rider travel behavior: impacts of individual, built environment, and crime characteristics on transit access. Transport. Res. Part A: Policy Practice 41, 511–522.

Kitamura, R., Mokhtarian, P.L., Daidet, L., 1997. A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. Transportation 24, 125–158.

Ma, J., Liu, Z., Chai, Y., 2015. The impact of urban form on CO2 emission from work and non-work trips: the case of Beijing, China. Habitat Int. 47, 1–10.

Minneapolis. 2018. Minneapolis 2040 – The Citiy's Comprehensive Plan [Online]. Available: https://minneapolis2040.com/media/1429/minneapolis2040plan.pdf (Accessed Aug. 19, 2019).

Ridgeway, G., 2007. Generalized Boosted Models: A Guide to the gbm Package.

Salon, D., Boarnet, M.G., Handy, S., Spears, S., Tal, G., 2012. How do local actions affect VMT? A critical review of the empirical evidence. Transport. Res. Part D: Transp. Environ. 17, 495–508.

Schonlau, M., 2005. Boosted regression (boosting): an introductory tutorial and a Stata plugin. The Stata J. 5, 230–254.

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. Transport. Res. Part D: Transp. Environ.. 63, 23–36.

Stead, D., 2001. Relationships between land use, socioeconomic factors, and travel patterns in Britain. Environ. Plan. B: Plan. Des. 28, 499–528.

Stevens, M.R., 2017. Does compact development make people drive less? J. Am. Plan. Assoc. 83, 7–18.

U.S. Department of Transportation, 2010. Public Transportation's Role in Responding to Climate Change [Online]. Available: https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/PublicTransportationsRoleInRespondingToClimateChange2010.pdf (Accessed Aug. 20, 2019).

van Wee, B., Handy, S., 2016. Key research themes on urban space, scale, and sustainable urban mobility. Int. J. Sustain. Transport. 10, 18–24.

Wang, X., Khattak, A., Zhang, Y., 2013. Is smart growth associated with reductions in carbon dioxide emissions? Transport. Res. Rec.: J. Transport. Res. Board 2375, 62–70.

Wang, X., Liu, C., Kostyniuk, L., Shen, Q., Bao, S., 2014. The influence of street environments on fuel efficiency: insights from naturalistic driving. Int. J. Environ. Sci. Technol. 11, 2291–2306.

Wang, Y., Yang, L., Han, S., Li, C., Ramachandra, T.V., 2017. Urban CO2 emissions in Xi'an and Bangalore by commuters: implications for controlling urban transportation carbon dioxide emissions in developing countries. Mitig. Adapt. Strat. Glob. Change 22, 993–1019.

Xu, L., Cui, S., Tang, J., Yan, X., Huang, W., Lv, H., 2018. Investigating the comparative roles of multi-source factors influencing urban residents' transportation greenhouse gas emissions. Sci. Total Environ. 644, 1336–1345.

Yang, P., Yang, Y.H., Zhou, B.B., Zomaya, A.Y., 2010. A review of ensemble methods in bioinformatics. Curr. Bioinform. 5, 296–308.

Zahabi, S.A.H., Chang, A., Miranda-Moreno, L.F., Patterson, Z., 2016. Exploring the link between the neighborhood typologies, bicycle infrastructure and commuting cycling over time and the potential impact on commuter GHG emissions. Transport. Res. Part D: Transp. Environ. 47, 89–103.

Zhang, L., Hong, J., Nasri, A., Shen, Q., 2012. How built environment affects travel behavior: a comparative analysis of the connections between land use and vehicle miles traveled in US cities. J. Transp. Land Use 5, 40–52.

Zhu, W., Ding, C., Cao, X., 2019. Built environment effects on fuel consumption of driving to work: Insights from on-board diagnostics data of personal vehicles. Transport. Res. Part D: Transp. Environ. 67, 565–575.