Title
Prediction Framework for Searching for Parking Average Vehicle Cruising Time and Emissions in Urban Areas

Author
Xiao, Runhua

Publication Date
2021

Peer reviewed|Thesis/dissertation

Prediction Framework for Searching for Parking
Average Vehicle Cruising Time and Emissions in Urban Areas

By

RUNHUA (IVAN) XIAO
THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Civil and Environmental Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Miguel Jaller, Chair

Michael Zhang

Alan Jenn

Committee in Charge

2021

ACKNOWLEDGMENTS

I would like to thank my major professor, Miguel Jaller, for his continued guidance on the thesis and an endless supply of fascinating projects. My sincere thank also goes to Prof. Michael Zhang, my area adviser and one of the members of my graduate committee. A special thanks also to Dr. Xiaodong Qian who worked with and guided me on my projects throughout the year. Another special thanks to the help provided by the technical and support staff at the Institute of Transport Studies at the University of California, Davis.

ABSTRACT

This thesis proposes and implements a novel framework to identify the factors influencing average cruising time (ACT) related to searching for parking and uses multiple machine learning (ML) models to predict ACT and average vehicle emissions (AEM) in New York City (NYC) and Los Angeles (LA). Specifically, the author 1) analyzes NYC datasets by spatial lag models to explore the factors affecting ACT; 2) uses the K-Means method to cluster land use type and vehicle composition, and perform a comparative analysis for different groups; 3) utilizes ML models to predict the land use and vehicle composition in NYC and LA, and use them into a Bayesian Ridge regression model to predict the ACT; and 4) uses a Gaussian-Link generalized linear model to calculate average cruising distance (ACD) based on ACT and other variables to estimate AEM generated by parking events. The major findings include 1) Residential and commercial areas have a significant positive correlation with ACT. The parking hotspots roughly coincide with the distribution of density, land use, and job opportunities, which may be associated to growing residential freight demand and sluggish parking supply. 2) A parking hotspot with high ACT is not necessarily a location with high AEM. Compared to locations with high ACT in the central areas, the heacy-duty-truck (HDT)-dominated areas at the periphery of the city can generate higher vehicle emissions. 3) AEMs are more affected by the type of dominating vehicles in the grid, and HDT-dominated grids generate four times of pollutants than light-duty-truck (LDT)-dominated ones.

Keywords: searching for parking; spatial lag model; comparative analysis; machine learning prediction; vehicle emission; commercial vehicles

INTRODUCTION

Searching for parking is a challenging task for many drivers when driving in cities, because sometimes the parking spaces closer to the destination may be unavailable or expensive, which forces them to cruise for a longer time and distance until they find available on-street spaces. This is even more important for the efficiency of delivery routes and operations. Studies reveal that drivers searching for parking in metropolitan areas not only cause mental frustration and longer cruising time, but contribute to traffic congestion and additional greenhouse gas (GHG) emissions as vehicles travel at lower speeds for drivers to look for parking spaces and signs, and also contributing to more fues consumed due to idling (Ng, 2016). On average, drivers spend about 17 hours a year searching for parking, and the estimated cost for paying the wasted time, fuel, and emissions is $345 per driver ("Drivers spend an average of 17 hours a year searching for parking spots," n.d.). Studies also suggest that drivers typically spend between 3.5 to 14 minutes to search for parking (Shoup, 2006), and this can lead to considerable economic, environmental and social impacts on cities (Arnott and Williams, 2017).

In terms of factors affecting the average crusing time (ACT) when searching for parking, studies have focused on parking demand, which is strongly associated with population and commercial establishment density, reflected on the number of buildings and establishments. Studies have analyzed parking demand for passenger and freight vehicles, though, in most cases, this has been done separately. Evidence shows that a mismatched parking supply in denser regions will exacerbate traffic congestion (Liu et al., 2018). Concentrations of freight activity in urban areas can evidence locations in which the demand for parking for commercial vehicles far exceeds local parking availability (Jaller et al., 2013). There are also views that the demand for on-street parking is related to the type of land use (Jaller et al., 2021; Simons, 2020). In the case of equal parking demand, land use types with a large number of off-street parking spaces (e.g., retail, storage) will have lower on-street parking demand. A study found that the influence of household properties, such as income, on automobile travel demand is remarkable (Schimek, 1996). In addition to the above factors, vehicle composition, population, point of interest (POI) density, the number of establishments and the number of employees of all industries which are coded by the North American

Industry Classification System (NAICS), also have an impact on ACT (U.S. Census Bureau). Due to differences in vehicle types and vehicles fleet compositions (e.g., share of passenger and freight vehicles), different emissions may occur in blocks with the same ACT, and due to the size limitations of parking spaces, it may take longer for large vehicles to find parking spaces. However, studies are lacking that consider all these factors to quantify the ACT and the associated impacts (e.g., emissions) for different land uses, and vehicle compositions.

Therefore, the aim of this paper is to propose and implement a novel framework to identify the factors affecting ACT in urban areas and quantify ACT and average vehicle emissions (AEM) when searching for parking. The author identifies the factors by implementing a spatial lag model and then perform comparative analyses for different factors under different groups of land use types and vehicle compositions. The study uses searching for parking data from the GeoTab data platform and complements with other datasets. The author uses New York City (NYC) and Los Angeles (LA) as case studies. The study contributes to searching for parking time and emission prediction using a model-based prediction framework. Several machine learning (ML) modeling strategies are employed ranging from Artificial Neural Network (ANN) to Bayesian Ridge regression models, using urban environmental conditions, socio-economic characteristics, and other network related attributes for classification and forecasting.

The remainder of the paper is organized as follows: The next section briefly presents the datasets used in this paper and the methodology. Next, the overall spatial lag model to identify factors affecting ACT in NYC and a comparative analysis across different land use types and vehicle compositions are presented. Then, the ACT and AEM are estimated for both NYC and LA. The paper ends with a discussion on the main findings and contributions of this work.

LITERATURE REVIEW

Although searching for parking is a paramount issue for urban planning and transportation, little related research in different forms of analysis and modeling exists. The previous literatures contributed to this issue and related to our work include the analysis of search for parking, cruising time prediction, and vehicle emission calculation.

The data-driven analysis of searching for parking is an emerging component of parking and curbside management related research, while there exist some studies that model parking and driver behavior and their impact on congestion and the environment. According to Box (2004), accidents related to parking and parking maneuvers and how these factors impact congestion are examined, and the study found that comparable streets with angle parking have two to three times the number of accidents per mile as those with parallel parking, which may also cause congestions that impede the vehicles searching for parking. There is a study that examines parking from a physical design perspective in relating parking occupancy and load to number of lanes and capacity (Valley, 1997). Recently, a case study for Seattle was done where the city-level traffic and parking data were explored to determine how much cruising for curbside parking contributes to overall traffic congestion. Results show that while percentage increase in travel time to through traffic vehicles depends on time of day, it does not appear to depend on high volumes of through traffic (Dowling et al., 2017). Another case study of Seattle utilizes a quantitative method to explore the parking cruising behavior of commercial vehicle drivers in urban areas using GPS data, suggesting that searching for parking decreased as more curb-space was allocated to commercial vehicles load zones and paid parking and as more off-street parking areas were available at trip destinations (Dalla Chiara and Goodchild, 2020). A study conducted surveys with drivers in Australia, to understand potential factors that influence drivers' cruising behavior, and revealed some factors affecting searching for parking at the individual level (Lee et al., 2017).

Although many studies have successfully applied different ML models for prediction, few studies have applied them to searching for parking. A multilayer network assignment model was proposed to simulate parking choices; and numerical tests of the model highlighted that it is able to simulate user parking choice and the impact of cruising for parking on congestion (Gallo et al., 2011). Sharma et al. (2021) explored machine learning to predict short-term wait time at a US-Mexico border crossing using Gradient Boosting Regression and Random Forest method. The model-based method encourages combining more sophisticated predictive algorithms and prediction methods on datasets, while the high data variability is challenging leading to non-reliable predictions. In terms of vehicle emissions, Čuljković (2018) evaluated the potential contribution of parking price to reducing energy consumption and CO_2 emissions and suggested that adequate parking pricing can eliminate cruising for parking and provide significant effects on emissions. However, the emissions related to cruising for parking were not quantified or modelled. Perugu et al. (2018) calculates emissions from grid-level to county, and quantifies the likely effects of the approach on a gridded on-road emission inventory, suggesting that results from the two methods SMOKE and EMFAC were within ±1%.

In general, most of the previous studies focused on the use of survey data to find factors that affect searching for parking, as well as individual data. A gap needs to be filled that addresses searching for parking for different vehicle compositions and land uses, by implementing a comprehensive and quantitative method for evaluating the factors affecting the cruising time and predicting ACT as well as vehicle emissions related to searching for parking.

METHODOLOGY

The section describres the modeling framework with a discussion of 1) descriptive and comparative models on factors influencing cruising time of searching for parking, 2) a model-based prediction for the average vehicle cruising time (ACT) related to searching for parking , and 3) average vehicle emissions (AEM) predictions. The first section mainly uses a spatial lag model to explore the influencing factors on cruising time. This is followed by a comparison and use of various ML models to predict cruising time and emission related to searching for parking. FIGURE 1 shows a schematic of the framework with the main modeling components, and the various datasets used (these are discussed next).

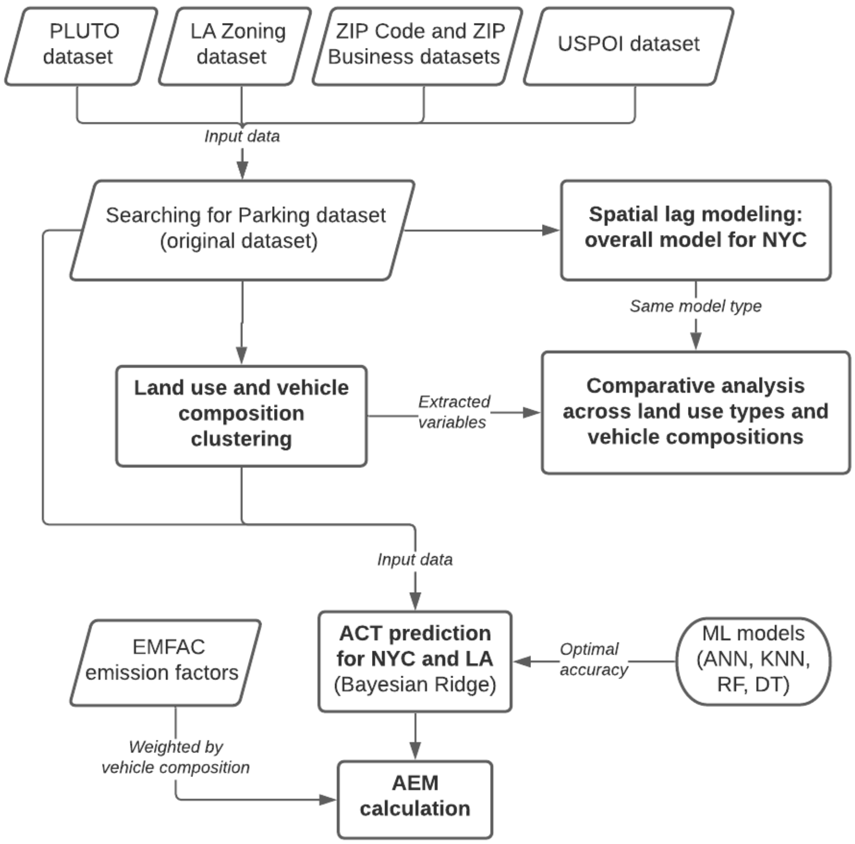


FIGURE 1 Technical scheme of the paper

Datasets

This study focuses on modeling ACT and AEM in NYC and LA. The Searching for Parking dataset is from the GeoTab data platform, and the data availability ranges from April 1, 2020 to October 1, 2020. The measurements of parking data are aggregated to the 7-character geohash level (153m × 153m) over a range of 6 months. Geotab tracks over 900,000 vehicles and generates Searching for Parking behavior data from over 250,000 vehicles per hour throughout North America. This allows Geotab to collect millions of vehicle location data points per second in near real-time and generate a comprehensive, aggregated updating average time to park value for each geohash of interest.

In addition, the author uses census socio-demographic and economic information at the ZIP code level, from the Maptitude Mapping Software. For example, the ZIP Code and ZIP Business datasets include data with the number of establishments, number of employees by NAICS codes, population and household income. The author gathers aggregate point of interest data for multiple types of establisments and locations. Another dataset useful from the company is USPOI that provides locations of points of interest (POI), which are aggregated as POI density. The analyses also use the Primary Land Use Tax Lot Output (PLUTO) dataset and the LA Zoning dataset from City of Los Angeles Hub for land use area information.

Spatial Lag Model

To model the ACT of a certain street block in a city and explore the factors affecting ACT, requires defining the best modeling approach. With a neighbor structure defined by the non-zero elements of the spatial weights matrix W, a spatially lagged variable is a weighted sum or a weighted average of the neighboring values for that variable (Fotheringham et al., 2003). In most commonly used notation, the spatial lag of dependent variable y is expressed as W_y. Based on the traditional Ordinary Least Square (OLS) model, the spatial lag model introduces a spatially lagged dependent variable vector, which assumes that spatial dependencies exist directly among the observations of a dependent variable:

$$y = X\beta + \rho W_y + \varepsilon \tag{1}$$

where X is the observation matrix of independent variables, β and ρ are estimated coefficients, and ε is the error term. With the assumption of spatial dependency, a spatial lag model is suitable for modeling under the condition that the dependent variable at one location is affected by the variable at the nearby locations. Spatial lag models are useful when dealing with spatial spillover effects, as spatial dependencies exist in many cases. If the ACT of each block is independent of each other, the traditional OLS model is merely a good choice. However, it cannot be ignored that the ACT between adjacent blocks affects each other. For example, on-street parking spaces in front of a famous restaurant always have a tight parking demand. The ACT of this block will be high, causing drivers to search in the nearby block to find parking spaces, because they can hardly find available spaces in that block, thereby increasing the ACT of the adjacent block.

ACT Prediction Framework

The ACT prediction framework is a citywide grid-based prediction modeling process, including a land use and vehicle composition prediction module based on various ML models and an ACT prediction module based on Bayesian Ridge regression model. In this study, the hierarchical data structure exist as land use and vehicle composition are nested within a grid and sharing the same data for predicting ACT.

This prediction framework is basically divided into two steps: model training and prediction. First, it implements the K-Means clustering method to extract the two variables of land use (NYC only) and vehicle composition, using the PLUTO dataset from NYC Open Data and the Searching for Parking dataset from Geotab. Inputting land use type, vehicle composition, census data including NAICS employment data (the analyses aggregate industry categories at the 1-digit NAICS), and POI data to the observations available in the Searching for Parking dataset, the framework uses a variety of ML models to train the independent variables and predict the ACT of each observation in the dataset. The purpose of the model training process is to ensure the predictive models get the highest prediction accuracy, which can justify the use of the ML models in the ongoing grid-based predictions. Secondly, the process divides the case study cities (NYC and LA) into grids of the same size, and weights the input data (both the aforementioned independent variables and ACT) equally to each Grid. For each grid, it then uses the specified models to predict land use type and vehicle composition, then ACT. The second step of model prediction uses different models to predict land use type, vehicle composition and ACT respectively, and models are selected based on optimal prediction accuracy. The land use type and vehicle composition extraction and prediction processes are stated in the following sub-sections.

Land Use and Vehicle Composition Extraction and Imputation Methods

The methods for extracting land use variables for NYC and LA are different. For NYC, the team used the PLUTO dataset, which contains the area of various types of land, so the K-Means method is used for clustering. The selection of the number of clusters is based on optimal Silouette score. The data available for LA comes from the LA City Planning Department, which can be used directly or modified to incorporate its land use type variables.

The extraction process of land use type and vehicle composition is as follows: 1) For PLUTO data, use the K-Means clustering method to classify the various types of land areas in the data set, and select the number of clusters according to the principle of the highest gap between classes. For LA's existing land use type data, merge and modify the type to correspond to NYC data; 2) Use K-Means clustering method to classify the five types of vehicle percentages in the Searching for Parking data set (cars & MPV, light-duty truck (LDT), medium-duty truck (MDT), heavy-duty truck (HDT) and other vehicles) to extract different vehicle components and label them; and 3) attach the extracted variables to the original dataset, and use the original dataset for model training.

The imputation process of the two variables for each grid incorporate various ML models. After dividing the case study city into grids and weighting other independent variables, the author used Artificial Neural Network (ANN), K-nearst neighbor (KNN), Random Forests (RF) and Decision Tree (DT) models to predict the land use and vehicle composition variables respectively. Using the confusion matrix, the author selected the final model specification according to the optimal accuracy.

ACT Prediction Methods: Bayesian Ridge Regression Model

Based on the weighted independent variable data of the grid and imputed land use and vehicle composition, the framework uses Bayesian Ridge regression model to predict the ACT of each grid. Bayesian Ridge regression estimates a probabilistic model of ACT prediction. To obtain a fully probabilistic model for prediction, ACT (denoted as y) is assumed to be Gaussian distributed around $X\beta$:

$$p(y|X,\beta,\alpha) = N(y|X\beta,\alpha) \tag{2}$$

where α is the complexity parameter that controls the amount of shrinkage. The larger the value of α, the greater the amount of shrinkage and thus the coefficients β become more robust to collinearity. In Bayesian Ridge, α is treated as a random variable that is to be estimated from the data. Also, the prior for the coefficient β is also given by a spherical Gaussian function. The parameters β, α and λ are estimated jointly during the fit of the model, the regularization parameters α and λ being estimated by maximizing the log marginal likelihood (MacKay, 1992).

The reasons for using Bayesian Ridge regression model mainly come from its adaptability and robustness. It adapts to the data at hand, and allows a natural mechanism to survive insufficient data or poorly distributed data by formulating linear regression using probability distributors rather than point estimates (Pedregosa et al., 2011). It can also be used to include regularization parameters in the estimation procedure.

Emission Calculation

The emission calculation process incorporates average vehicle cruising distance and emission factors for different vehicle composition within a grid. Bascially, the AEM for a single grid is calculated as follows:

$$AEM_{i,p} = ACD_i \cdot F_{i,p} \tag{3}$$

where $AEM_{i,p}$ is the average vehicle emissions of contanminant p for the specific grid i, ACD_i is the average vehicle cruising distance for grid i, and $F_{i,p}$ is the weighted vehicle cruising emission factor of contanminant p for grid i with its vehicle composition.

For vehicle cruising distance, imputation is required due to lack of direct data measurement. Due to the positive correlation between ACT and average vehicle cruising distance (ACD), the author used a generalized linear model based on the Gamma function to model average total geohashes (the dataset does not provide the cruising distance; it provides the number of geohashes traveled by the vehicle when searching for parking) and ACT. Then the process uses the average Manhattan distance, which is 204 m, to aproximate the relation between the average total geohashes and ACD.

The weighted vehicle cruising emission factors are provided by the 2021 version of the EMission FACtor (EMFAC) model developed by the California Air Resources Board, a model that estimates the official emission inventories of on-road mobile sources in California from 2000 to 2050 (EMFAC2021

Volume III). Vehicle cruising emissions related to searching for parking behavior are calculated separately for different vehicle compositions. These compositions are estimated using the K-Means clustering method described in the previous section. This study mainly focuses on the three types of contanminants in vehicle emissions generated by searching for parking, namely nitrogen oxides (NO_x), PM2.5 and carbon dioxide (CO_2). The weighted emission factors are calculated as follows:

$$F_{i,p} = f_{p,k} \cdot VP_{i,k} \tag{4}$$

where $f_{p,k}$ is the EMFAC emission factor of contanminant p for single vehicle type k, and $VP_{i,k}$ is the percentage of the vehicle type k in grid i. Each EMFAC emission factor for each vehicle composition is then calculated before calculating the final average vehicle emissions for a specific grid.

EMPIRICAL RESULTS

Spatial Lag Modeling of Factors Affecting ACT

Compared to the categorical land use variable of the LA Zoning data, the quantitative attributes of PLUTO data can be used to find factors that affect ACT. The dataset in use includes the NYC Searching for Parking data, the NYC census ZIP code data, and the PLUTO data, with 4,535 observations. The model on ACT is estimated via maximum likelihood spatial lag method, which is based on Gaussian kernel weights on ACT lagged variable (Rey and Anselin, 2010).

TABLE 1 shows the variable coefficients, t-values and the diagnostics of the spatial lag model. The independent variables include total number of establishments, number of employees across different NAICS codes, population, household income, POI density, land use area across different types (residential, commercial, office, retail, etc.), and the lagged variable of ACT. After preliminary analyses, the variables were carefully selected to decide the model specification. Only the variables with significant coefficients were included (i.e., the confidence interval ranges do not include a value with sign different from the mean).

The model that has the best goodness-of-fit was considered as the final form. Finally, 17 variables including the lagged ACT were found to be significantly related to ACT.

TABLE 1 Estimates of the Overall Spatial Lag Model Using NYC Dataset

Number of observarions	4535		**Mean dependent variable**		4.8268
Number of variables	18		**S.D. dependent variable**		1.1404
Degrees of freedom	4517		**Log likelihood**		-6426.829
Pseudo R-squared	0.2513		**Akaike info criterion**		12889.658
Spatial Pseudo R-squared	0.1963		**Sigma-square ML**		0.975
			S.E of regression		0.987
Variable name	**Coefficient**	**Std.Error**	**t-value**	**p-value**	**Elasticity**
(Intercept)	3.777	0.071	53.366	0.000***	/
Number of total establishments (in 1,000)	0.098	0.034	2.901	0.004***	5.16%
Employees in NAICS 2 (in 1,000)	-0.051	0.022	-2.317	0.021*	0.00%
Employees in NAICS 3 (in 1,000)	0.084	0.039	2.159	0.031*	0.00%
Employees in NAICS 5 (in 1,000)	-0.013	0.002	-7.790	0.000***	9.64%
Employees in NAICS 6-9 (in 1,000)	0.010	0.004	2.375	0.018*	14.80%
Median household income (in $1,000)	-0.005	0.001	-9.312	0.000***	9.64%
POI density (per 1,000 POI in the geohash)	0.218	0.116	-1.877	0.061`	8.54%
Commercial area (in 1M sq ft)	1.153	0.166	6.939	0.000***	0.00%
Residential area (in 1M sq ft)	0.874	0.086	10.216	0.000***	0.00%
Office area (in 1M sq ft)	-0.942	0.174	-5.420	0.000***	0.00%
Retail area (in 1M sq ft)	-0.736	0.304	-2.423	0.015*	0.00%
Garage area (in 1M sq ft)	-0.809	0.448	-1.806	0.071`	0.00%
Storage area (in 1M sq ft)	-0.824	0.536	-1.537	0.124	0.00%
Healthcare area (in 1M sq ft)	-0.426	0.180	-2.368	0.018*	1.08%
Number of buildings	-0.008	0.001	-6.485	0.000***	0.00%
Number of floors	0.004	0.000	7.970	0.000***	0.00%
Lagged ACT (min)	0.187	0.011	16.923	0.000***	/

Note: “ ` ”: $p < 0.1$; “ *”: $p < 0.05$; “ ** ” $p < 0.01$; “ *** ”: $p < 0.001$.

The spatial lag model results show that ACT of a geohash is significantly positively correlated with number of total establishments, number of employees from NAICS industries of manufacturing (code 3), educational services and healthcare/social assistance (code 6), recreations, accomodations and food services (code 7), public administration (code 9) and other services (code 8), POI density, sum of residential and commercial area, and number of floors. The results also show that number of employees from NAICS industries of utilities and construction (code 2) and professional services (code 5), median household income, sum of office, retail, garage, storage, health area, and number of buildings are significant negative factors affecting ACT. Most importantly, the results show that the coefficient of the spatially lagged variable of ACT is 0.187 with the highest magnitude among all the significant positive coefficients. The

ACT of neighboring geohashes is significantly positively correlated with this ACT, suggesting that a geohash with a high ACT value will increase the ACT of the adjacent geohash.

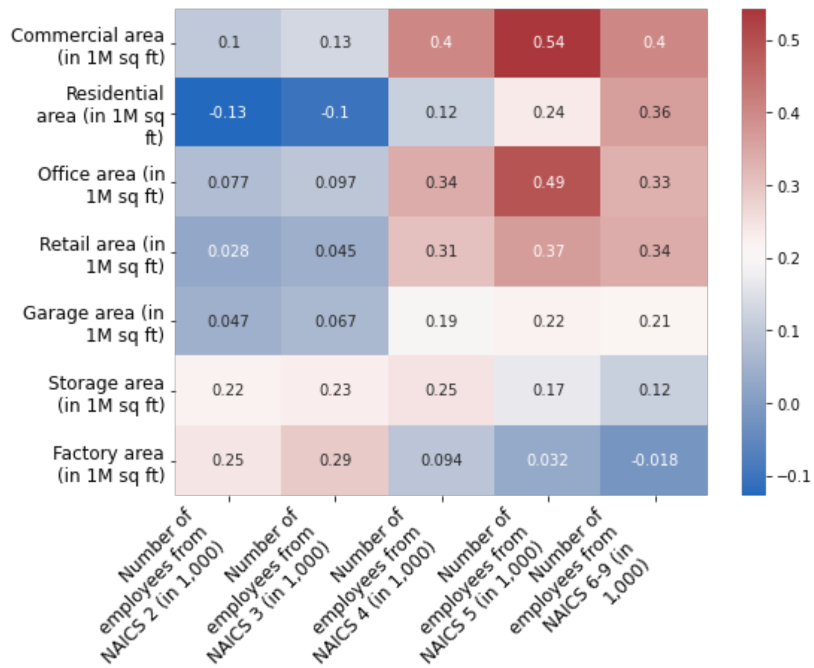
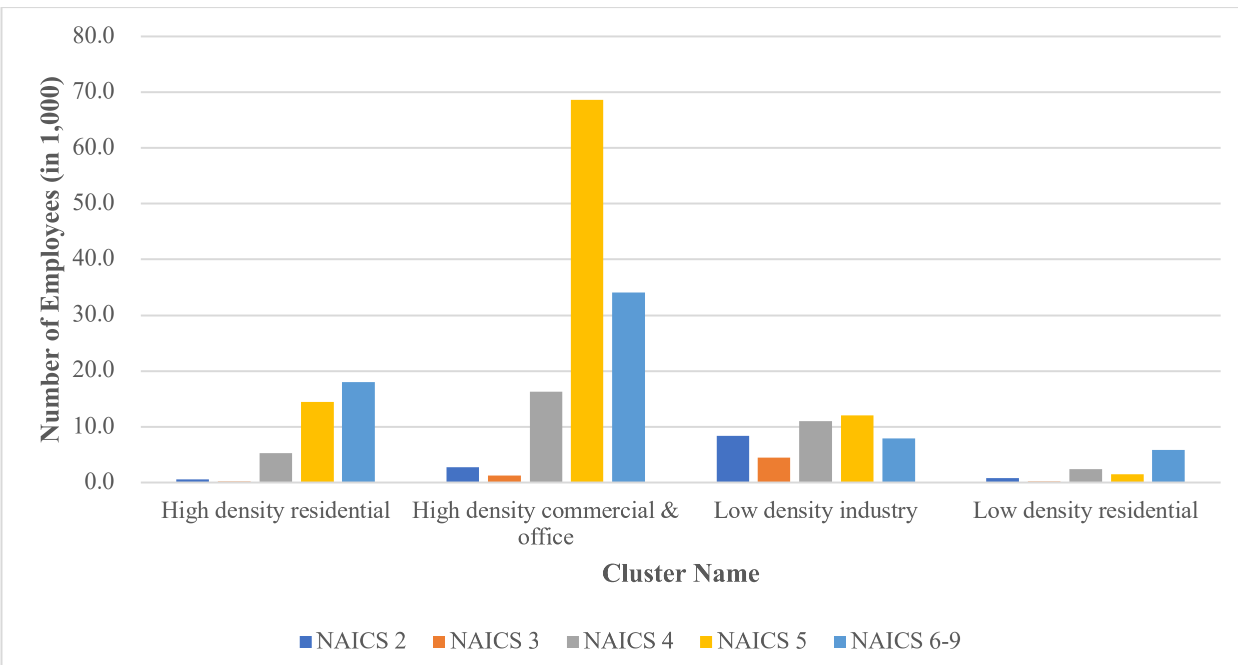


FIGURE 2 Partial Correlation Matrix between Number of Employees and Land Use Areas

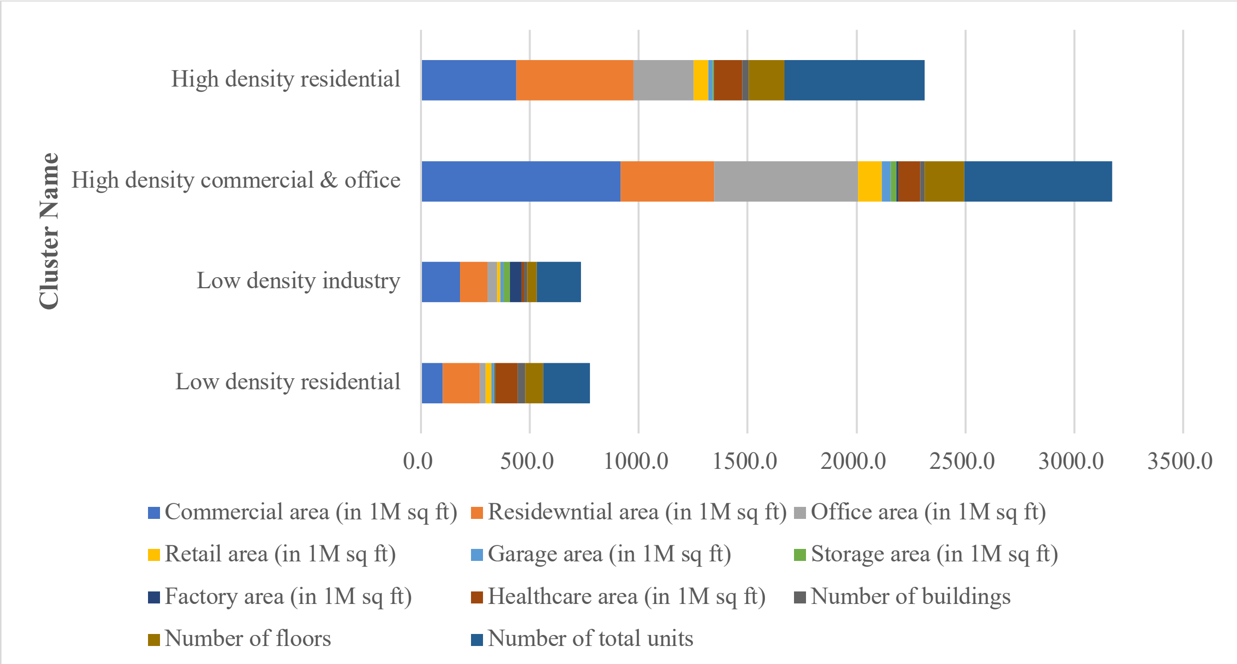
In the model, unexpected signs of coefficients were found for variables such as number of employees from utilities & construction (code 2) and professional services (code 5), which were expected to be positively correlated if significant. Observing the partial correlation matrix in FIGURE 2 shows that the number of employees from utilities & construction (code 2) has a high positive correlation with the area of storages and factories, and the number of employees from professional services (code 5) has a strong positive correlation with the area of commercial, office, retail and garage. These variables have shown their negative correlation with ACT. Therefore, the negative coefficients may be theorectically interesting, expanded in the discussion section.

The author also obtained the elasticities via log-log OLS model (see also TABLE 1). The elasticity calculation results show that the number of employees from the general service industries (code 6-9) have the highest elasticity at 14.8%. It shows that the same independent variable increase has the greatest impact on ACT on this variable. If a large number of general service industry employees gather, the value of ACT may be very high.

Before implementing the prediction framework, the author clustered NYC's land use types using PLUTO data. FIGURE 3 and FIGURE 4 summarize the categories of these two variables, and the mean value of each clustering index corresponding to each category. For land use types, we found four clusters: cluster 1 is in high-density areas with a large proportion of residential land use, cluster 2 is high-density commercial and office areas; clusters 3 and 4 are in low-density areas, but cluster 3 has more industry area, and cluster 4 is mostly residential. For vehicle composition, we found five clusters. According to the names of the categories, each cluster has a dominant vehicle type.



a) Employees from different industries



b) Sum of area for different land use types

FIGURE 3 Mean Values of Clusters on Land Use Types

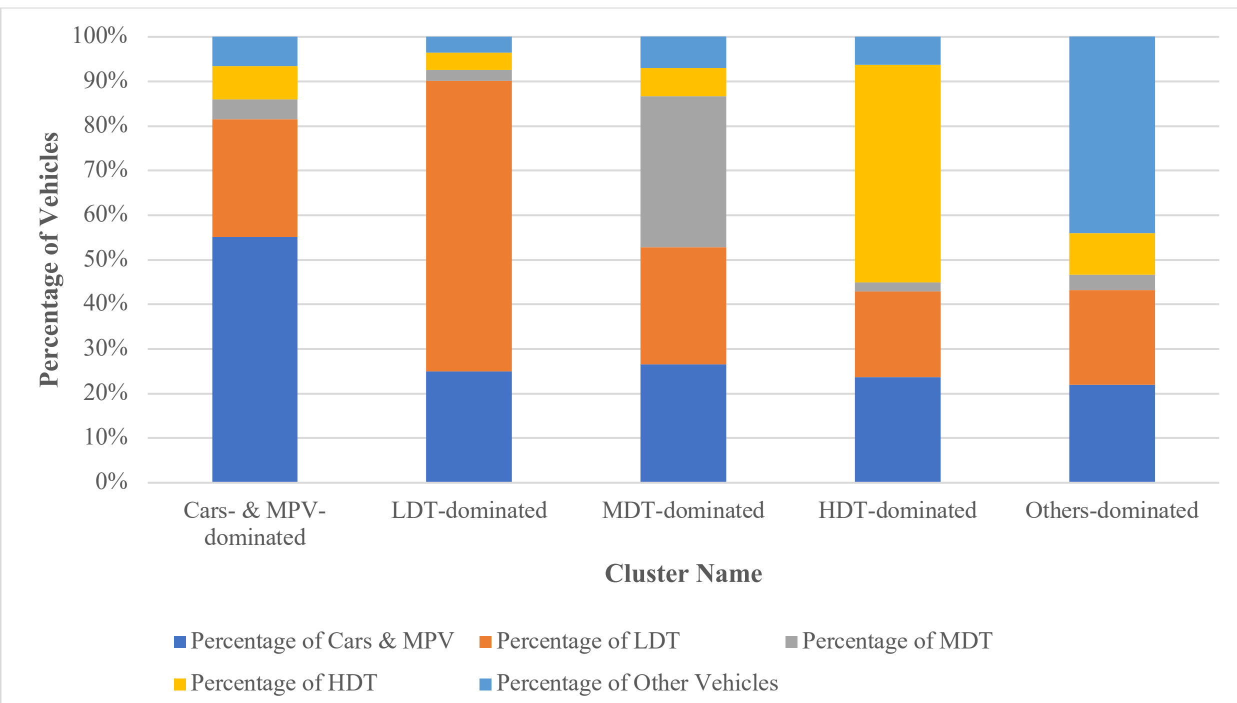


FIGURE 4 Mean Values of Clusters on Vehicle Compositions

The author also fit different spatial lag models for different land use and vehicle compositions. The selected variables and estimated coefficients for comparative spatial lag models are shown in TABLE 2. The result shows the coefficients are similar to the overall spatial model in TABLE 1, but each group has unique ones. We found that the area of residential and commercial land use has positive correlations and significant impacts on most groups, while other land use types have negative impacts on ACT. Moreover, the number of employees in trade, transport and warehousing industry (code 4) has a significant negative impact on the high-density commercial office group.

TABLE 2 Estimates of the Comparative Models across Land Use Types and Vehicle Compositions

	Land use types				**Vehicle compositions**			
Category	High density residential	High density commercial and office	Low density industry	Low density residential	Cars- and MPV-dominated	LDT-dominated	MDT-dominated,	HDT-dominated
Number of observations	916	304	118	3197	3428	667	112	191
Variable name	**Coefficients & Significances**				**Coefficients & Significances**			
(Intercept)	3.303***	6.246***	7.261*	2.528***	2.234***	3.648***	5.042***	5.450***
Total establishments (in 1,000)	/	0.476***	/	/	0.087***	/	/	/
Employees in NAICS 2 (in 1,000)	/	/	-0.369	/	/	/	-0.126*	/
Employees in NAICS 3 (in 1,000)	/	/	0.475	0.118*	/	/	/	/
Employees in NAICS 4 (in 1,000)	/	-0.105***	/	/	/	-0.020	/	0.049`
Employees in NAICS 5 (in 1,000)	-0.006	-0.024***	/	0.022`	-0.008***	-0.010*	0.023**	-0.016`
Employees in NAICS 6-9 (in 1,000)	-0.010	0.062***	/	0.014***	/	0.027***	/	/
Populatiuon (in 1,000)	/	-0.060***	/	/	/	/	-0.007*	/
Median household income (in $1,000)	/	/	/	-0.005***	-0.003***	-0.005***	-0.009	-0.008**
POI density (per 1,000 POI in the geohash)	/	-0.200	/	/	-0.302*	/	/	-1.347
Commercial area (in 1M sq ft)	0.933***	0.868***	/	0.814***	0.679***	0.761`	/	/
Residential area (in 1M sq ft)	0.622***	0.391	/	1.014***	0.579***	0.686***	1.105*	1.557*
Office area (in 1M sq ft)	-0.936***	-0.605**	/	/	-0.587***	-0.613	/	1.613*
Retail area (in 1M sq ft)	/	/	/	-0.841`	/	-0.987	-7.984**	/

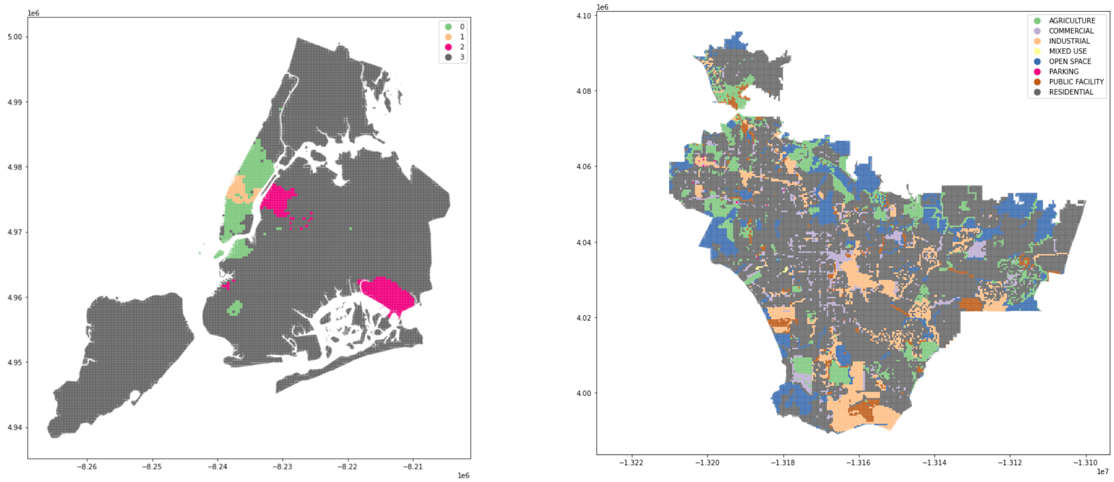
Garage area (in 1M sq ft)	-1.791	/	/	/	-0.660	/	/	/
Storage area (in 1M sq ft)	/	-1.150	/	/	/	/	/	/
Factory area (in 1M sq ft)	/	/	/	/	/	/	/	4.549`
Healthcare area (in 1M sq ft)	/	5.508	-7.480	-0.497***	-0.298	/	/	/
Number of buildings	0.005***	-0.025	/	/	-0.006***	-0.009**	/	-0.015*
Number of floors	/	/	/	/	0.003***	0.004***	/	0.007*
Number of total units (in 1,000)	0.060	-0.328	/	0.544**	/	/	/	/
Lagged ACT (min)	0.318*	-0.211	-0.341	0.421***	0.519***	0.245	0.161	-0.175

Note: “ ` ”: $p < 0.1$; “ *”: $p < 0.05$; “ ** ” $p < 0.01$; “ *** ”: $p < 0.001$.

ACT Prediction

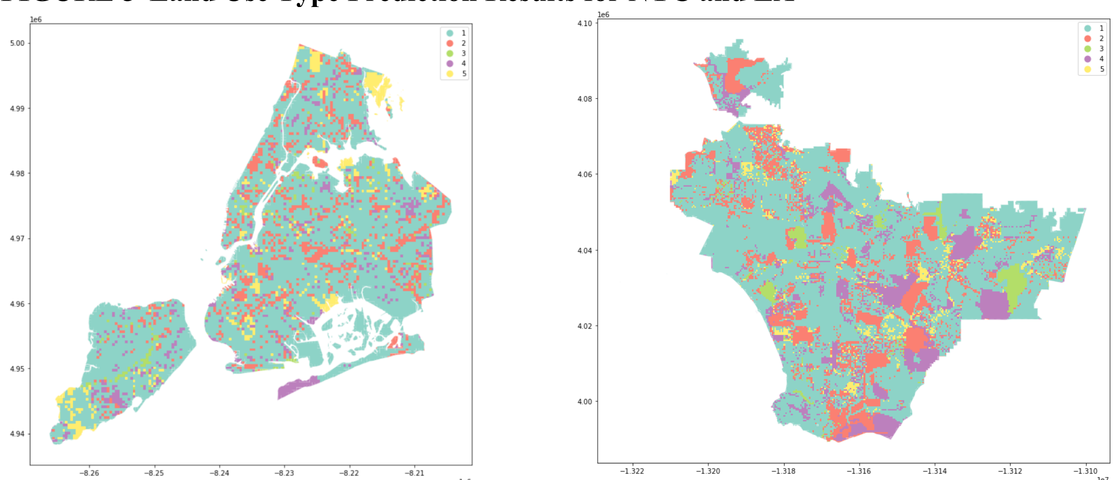
Basically, the territory of NYC and LA is divided into 400 m*400 m grids, incorporating census data, POI and other information of the grid. The process used the data from the original dataset (Searching for Parking) as the training dataset, and used various ML models (ANN, RF, DT, KNN, etc.) to predict the land use type and vehicle composition, and then predict the ACT of each grid.

The two variables, land use type and vehicle composition, are input into the Baysian Ridge regression model for predicting ACT. The author evaluated the confusion matrices and selected the most adequate model optimal on prodiction accuracy. For land use, due to the higher quantitative attributes provided by the PLUTO model, the NYC data models obtained higher accuracy than the LA data models’, which are above 0.9. According to the accuracy results, the author selected the RF and KNN models to predict the land use types of NYC and LA respectively. For vehicle composition, although ANN and KNN models can have higher prediction accuracy, the overall prediction results are severely biased, therefore the author selected the DT model for NYC data and the RF model for LA data. The land use type and vehicle composition prediction results using the selected ML models can be found in FIGURE 5 and FIGURE 6.



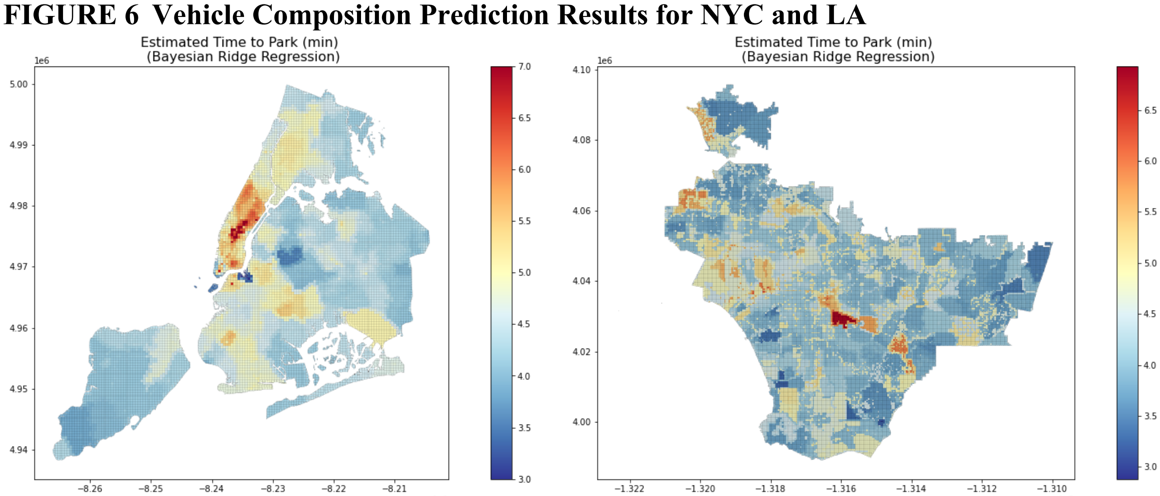
a) NYC b) LA

FIGURE 5 Land Use Type Prediction Results for NYC and LA



a) NYC b) LA

FIGURE 6 Vehicle Composition Prediction Results for NYC and LA



a) NYC b) LA

FIGURE 7 ACT Prediction Results for NYC and LA

FIGURE 7 shows the final prediction results of ACT using the Bayesian Ridge regression model for NYC and LA. For NYC, there are a total of 9,602 grids. The average ACT is 4.51 minutes and the standard deviation is 0.44 minutes. The minimum and maximum values of ACT are 1.79 minutes and 7.99 minutes, respectively. For LA, there are a total of 32,825 grids, with an average ACT of 4.27 minutes and a standard deviation of 0.52 minutes, and the minimum and maximum values are 2.88 minutes and 6.93 minutes, respectively. In general, NYC has a wider ACT range compared to LA, and as expected, hotspots where it is difficult to find on-street parking spaces have longer ACT.

Emission Calculation

As mentioned, the estimation of AEM is based on ACD and emission factors. For ACD, the author constructs a link between itself and the average total geohashes traveled before parking by establishing an average traveling distance for two adjacent geohashes. The simulation trial shows that the expected value of Manhattan distance between two random points in two adjacent geohashes is 204.01 m, by which the author can connect the average total geohashes to ACD. Although the author considered using various ML models to predict the average total geohashes with ACT, land use and vehicle composition as independent variables, the R^2 returned by the models did not exceed 0.05. Using RF regression also found that the importance of the latter two variables is almost zero. Therefore, the author considered using Gaussian-Link GLM regressions to predict the average total geohashes.

TABLE 3 Gaussian-Link GLM Model Results

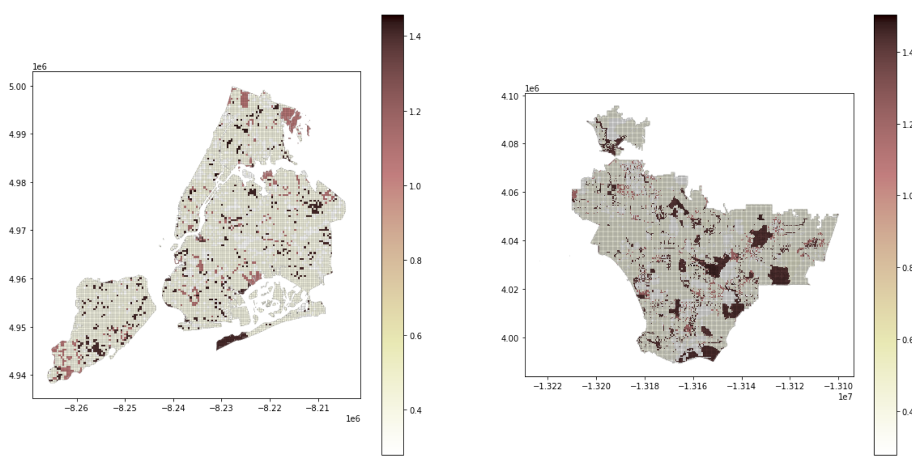
Dependent variable		Average total geohashes		**No. Observations**	4586
Model type		GLM		**Df Residuals**	4584
Model family		Gamma		**Df Model**	1
Link function		identity		**Scale**	0.0130
Method		IRLS		**Log-Likelihood**	-6286.1
No. Iterations		6		**Deviance**	57.233
Covariance Type		nonrobust		**Pearson chi^2**	59.6 (significant)
Variable	**Coefficient**	**Std.Error**	**t-value**	**p-value**	**C.I. [0.025, 0.975]**
Intercept	7.4056	0.062	120.198	0.000	[7.285, 7.526]
ACT (min)	0.2425	0.013	19.226	0.000	[0.218, 0.267]

TABLE 4 Weighted EMFAC Emission Factors for Different Vehicle Compositions

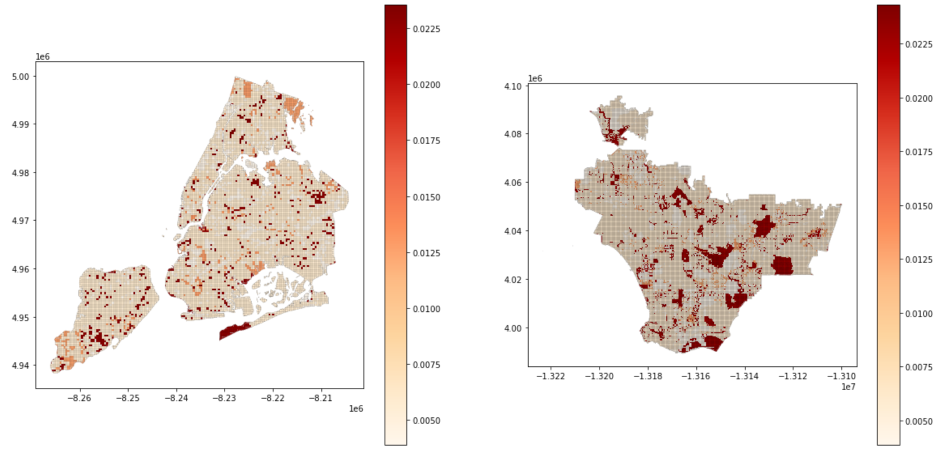
Vehicle Composition	Weighted EMFAC Factor (g/vehicle)		
	No_x	PM2.5	CO_2
Cars- & MPV-dominated	0.000229	0.000003	0.276442
LDT-dominated	0.000167	0.000002	0.259596
MDT-dominated	0.000242	0.000003	0.303469
HDT-dominated	0.000811	0.000013	0.506068
Others-dominated	0.000614	0.000007	0.466829

TABLE 3 shows the Gaussian-Link GLM model results with model diagnostics. The results show that the coefficients of ACT and the intercept are significant, suggesting that ACT can be connected to the average total geohashes, and thus linkng ACT to ACD by applying the formula in the GLM model and multiply by the expected value of Manhattan distance.

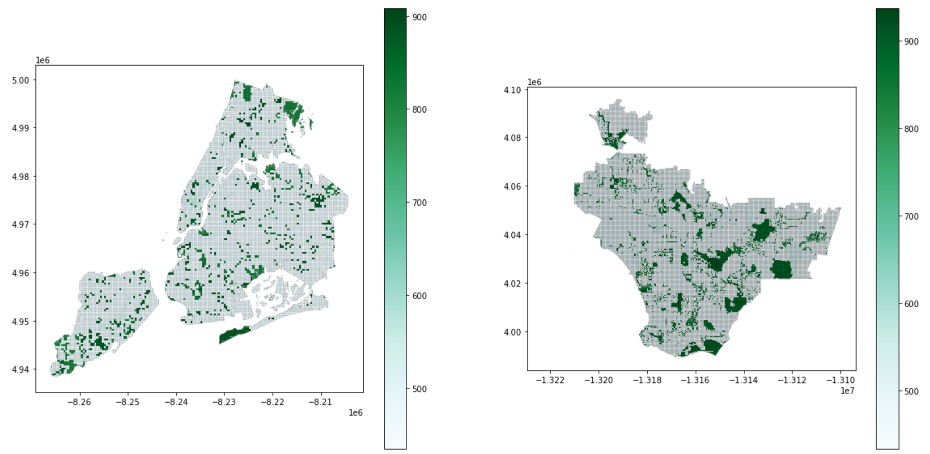
The author followed to estimate the emissions, using the emission factors in TABLE 4, and the results show that under the same ACD, the HDT-dominated vehicle composition produces the highest emissions of various contanminants, while the LDT-dominated ones are the lowest, even lower than the cars- and MPV-dominated ones. The results also show that for NO_x and PM2.5 pollutants, the emissions generated by HDT-dominated vehicles are more than four times that of LDT-dominated, while the CO_2 generated from cars- and MPV-dominated areas is only 0.027g/vehicle lower than that from MDT-dominated areas. Finally, the calculation results of emissions are shown in FIGURE 8.



a) NO_x Calculation Results



b) PM2.5 Calculation Results



c) CO_2 Calculation Results

FIGURE 8 AEM Calculation Results for NYC and LA

DISCUSSION

As a result of the spatial lag models including the overall model and the comparative models, the author found that besides some socio-demographic variables, ACT is positively affected by residential area and commercial area, and the influence of commercial land use area is slightly greater than residential. The variable differences of different comparative spatial lag models are analyzed by performing a comparative analysis. The results of comparative analysis and AEM weighted factors suggest that in areas with the highest AEM, which are HDT-donminated, ACT is not only positively correlated with residential and commercial area, but also has a more significant positive correlation with factory area.

High ACT tends to appear in high-density commercial and residential areas around urban downtown or sub-centers, which means that vehicles in these areas have a longer idle time. This is consistent with the overall spatial lag model and the original dataset. However, the prediction results show that the areas with high ACT are not necessarily areas with high AEM. Compared to the areas with high ACT in Manhattan, the HDT-dominated grids at the periphery of the city can produce higher AEM, which illustrates the two aspects of policy guidance. For one thing, policies may need to be adjusted, including parking priing, law enforcement, and infrastructure construction. For another, reducing the chance of searching for parking for heavy trucks with high emissions is very important, because cruising the same distance, heavy vehicles produce far higher emissions than passenger cars. Feasible improvement strategies include the construction and reservation of more heavy truck parking spaces, including exclusive loading and unloading curb spaces and loading and unloading zones for off-street parking lots for heavy trucks.

There are also some major findings from the study. First, the number of employees in the service industries (codes 6-9) has a positive impact on ACT in high-density commercial areas, as more concerntrated job opportunities attract higher trips and thus parking demand. This effect is particularly significant in NYC Midtown and LA Korea Town, which are dominated by commercial land uses. Moreover, in residential grids and those dominated by cars and MPVs, significant positive spatial lag effects of ACT are found, which means that ACT is affected by the positive correlation of surrounding grids. Still, the commercial and residential land uses have significant positive correlations with ACT in a majority of comparative models, while other land uses have negative ones. Different types of land uses can indirectly affect it by varying induced parking demand and supply. As negative significant coefficients are estimated by the overall spatial lag model, more office/retail/garage/storage/health area may mean lower parking and freight delivery demand and higher off-street parking space supply causing higher ACT, which are supported by the following paragraphs.

1) On-street parking spaces usually charge lower prices than off-street ones, which makes drivers more likely to search for on-street parking spaces. However, for land uses of office, retail, and healthcare that serve specific trip purposes, off-street parking spaces usually charge low prices. For example, shopping malls and supermarkets usually provide customers with free off-street parking spaces. This makes the price

advantage of on-street parking spaces non-existent near these types of land uses, so drivers are unlikely to spend a long time searching for on-street parking spaces.

2) Commercial vehicle demand continues to grow for off-street or on-street loading space in residential land use areas, and residential freight demand is substantial in mixed-use and residential census tracts (Chen et al., 2017). In areas where delivery demand is large, which usually tend to be residential and commercial, commercial vehicles may occupy excessive curb spaces and even have a higher chance of parking violations such as double parking or blocking entrances, compared to other land uses that usually have reserved off-street parking spaces or loading/unloading zones. Kim and Wang (2021) also suggests that double parking violations for commercial vehicles are likely to occur more frequently and repetitively than those for passenger cars.

3) Parking supply of off-street parking spaces are limited for some specific land uses in NYC. The parking and loading regulations require only 70% or even less than 50% of the parking ratio for high-density residential areas with accessible public transit. However, off-street parking spaces are not required in commercial areas in Lower and Mid-Manhattan and in Downtown Brooklyn. In these congested areas, there are low parking requirements or exemptions that may cause insufficient parking supply and thus the on-street parking spaces may face higher demand.

CONCLUSIONS

This paper uses a model-based comprehensive method to explore the factors that affect ACT related to searching for parking in NYC, and uses a combination of multiple ML models to predict ACT and vehicle emissions in NYC and LA. Specifically, the author 1) analyzes NYC datasets by spatial lag models to explore the influencing factors on ACT; 2) uses the K-Means method to cluster the two variables of land use type and vehicle composition, and compare different land use types and vehicle composition Analysis; 3) utilizes various ML models to predict the land use and vehicle composition variables of the NYC and LA city-wide grids, and inputs them into the Bayesian Ridge regression model to predict the ACT of each grid; and 4) simulates the average Manhattan distance for two random points selected in two adjacent geohashes and a Gaussian-Link GLM model that link between the average total geohashes and ACD, and uses the EMFAC emission factor to weight the average vehicle emissions produced by each grid. The major findings of this paper include: 1) Residential and commercial areas have a significant positive correlation with ACT. The parking hotspots roughly coincide with the distribution of density, land use, and job opportunities, which may be associated to growing residential freight demand and sluggish parking supply. 2) A parking hotspot with high ACT is not necessarily a grid with high AEM. Compared with the grid with high ACT in the central areas, the HDT-dominated grid at the periphery of the city can produce higher vehicle emissions. 3) AEMs are more affected by the type of dominating vehicles in the block, and HDT-dominated areas generate four times of pollutants than LDT-dominated areas. Compared to passenger cars, reducing the cruise time of searching for parking of heavy trucks is more effective at reducing emissions due to engine idling. Therefore, providing parking and loading/unloading areas could help minimize the ACT, and the associated emissions.

Given the results of this analysis, there are some limitations for the method and several ways in which future research could proceed. First, the quality of datasets regarding the data sources and methods of collection could be better. This paper uses the Searching for Parking data, which are aggregated without individual vehicle information such as route, speed and travel purpose. This does not allow the author to perform disaggregate level analysis to explore the factors influencing ACT. In addition, when performing the comparative analysis, the sample size of the group is likely to impact the final model specification. While it is possible that ACT for HDT-dominated geohashes have significantly different factors, the limited number of observations may have been the cause of the variables corresponding to these characteristics being deemed insignificant. These issues should be studied and taken into consideration for future studies.

REFERENCES

Arnott, R., Williams, P., 2017. Cruising for parking around a circle. Transportation Research Part B: Methodological 104, 357–375. https://doi.org/10.1016/j.trb.2017.07.009

Box, P.C., n.d. Curb-Parking Problems: Overview 5.

Chen, Q., Conway, A., Cheng, J., 2017. Parking for residential delivery in New York City: Regulations and behavior. Transport Policy 54, 53–60. https://doi.org/10.1016/j.tranpol.2016.12.005

Čuljković, V., 2018. Influence of parking price on reducing energy consumption and co2 emissions. Sustainable Cities and Society 41, 706–710. https://doi.org/10.1016/j.scs.2018.06.015

Dalla Chiara, G., Goodchild, A., 2020. Do commercial vehicles cruise for parking? Empirical evidence from Seattle. Transport Policy 97, 26–36. https://doi.org/10.1016/j.tranpol.2020.06.013

Dowling, C., Fiez, T., Ratliff, L., Zhang, B., 2017. How Much Urban Traffic is Searching for Parking?

Drivers spend an average of 17 hours a year searching for parking spots [WWW Document], n.d. URL https://www.usatoday.com/story/money/2017/07/12/parking-pain-causes-financial-and-personal-strain/467637001/ (accessed 7.28.21).

EMFAC2021 Volume III Technical Document Version 1.0.1 April, 2021, 2021. 271.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons.

Gallo, M., D'Acierno, L., Montella, B., 2011. A multilayer model to simulate cruising for parking in urban areas. Transport Policy 18, 735–744. https://doi.org/10.1016/j.tranpol.2011.01.009

Jaller, M., Holguín-Veras, J., Hodge, S.D., 2013. Parking in the City: Challenges for Freight Traffic. Transportation Research Record 2379, 46–56. https://doi.org/10.3141/2379-06

Jaller, M., Rodier, C., Zhang, M., Lin, H., Lewis, K., 2021. Fighting for Curb Space: Parking, Ride-Hailing, Urban Freight Deliveries, and Other Users. https://doi.org/10.7922/G22N50JJ

Kim, W., Wang, X. (Cara), 2021. Double parking in New York city: a comparison between commercial vehicles and passenger vehicles. Transportation. https://doi.org/10.1007/s11116-021-10212-5

Lee, J. (Brian), Agdas, D., Baker, D., 2017. Cruising for parking: New empirical evidence and influential factors on cruising time. Journal of Transport and Land Use 10, 931–943.

Liu, Q., Chen, P., Sun, F., 2018. Parking Policies in China's Metropolises: Rationales, Consequences, and Implications. Urban Policy and Research 36, 186–200. https://doi.org/10.1080/08111146.2017.1328353

MacKay, D.J.C., 1992. Bayesian Interpolation. Neural Computation 4, 415–447. https://doi.org/10.1162/neco.1992.4.3.415

Ng, W., 2016. Circling for parking is terrible for cities. Let's put an end to it. Sidewalk Talk. URL https://medium.com/sidewalk-talk/circling-for-parking-is-terrible-for-cities-lets-put-an-end-to-it-48c51921b776 (accessed 7.28.21).

North American Industry Classification System (NAICS) U.S. Census Bureau [WWW Document], n.d. URL https://www.census.gov/naics/?58967?yearbck=2017 (accessed 7.29.21).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Perugu, H., Ramirez, L., DaMassa, J., 2018. Incorporating temperature effects in California's on-road emission gridding process for air quality model inputs. Environmental Pollution 239, 1–12. https://doi.org/10.1016/j.envpol.2018.03.094

Rey, S.J., Anselin, L., 2010. PySAL: A Python Library of Spatial Analytical Methods, in: Fischer, M.M., Getis, A. (Eds.), Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Springer, Berlin, Heidelberg, pp. 175–193. https://doi.org/10.1007/978-3-642-03647-7_11

Schimek, P., 1996. Household Motor Vehicle Ownership and Use: How Much Does Residential Density Matter? Transportation Research Record 1552, 120–125. https://doi.org/10.1177/0361198196155200117

Sharma, S., Kang, D.H., Montes de Oca, J.R., Mudgal, A., 2021. Machine learning methods for commercial vehicle wait time prediction at a border crossing. Research in Transportation Economics 101034. https://doi.org/10.1016/j.retrec.2021.101034

Shoup, D.C., 2006. Cruising for parking. Transport Policy, Parking 13, 479–486. https://doi.org/10.1016/j.tranpol.2006.05.005

Simons, R.A., 2020. Driverless Cars, Urban Parking and Land Use. Routledge.

Valley, M., 1997. PARKING PERSPECTIVES. A SOURCEBOOK FOR THE DEVELOPMENT OF PARKING POLICY.