Impacts of transportation network companies on urban mobility

Citation: Diao, M., Kong, H. & Zhao, J. Impacts of transportation network companies on urban mobility. Nat Sustain 4, 494–500 (2021)

As Published: 10.1038/S41893-020-00678-Z

Publisher: Springer Science and Business Media LLC

Persistent URL: https://hdl.handle.net/1721.1/139853

Version: Original manuscript: author's manuscript prior to formal peer review

Impact of transportation network companies on urban congestion: Evidence from large-scale trajectory data

Xinwu Qian[1], Tian Lei[1], Jiawei Xue[1], Zengxiang Lei[1], Satish V. Ukkusuri[1]

[a] *Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, USA.*

Abstract

We collect vehicle trajectory data from major transportation network companies (TNCs) in New York City (NYC) in 2017 and 2019, and we use the trajectory data to understand how the growth of TNCs has impacted traffic congestion and emission in urban areas. By mining the large-scale trajectory data and conduct the case study in NYC, we confirm that the rise of TNC is the major contributing factor that makes urban traffic congestion worse. From 2017 to 2019, the number of for-hire vehicles (FHV) has increased by over 48% and served 90% more daily trips. This results in an average citywide speed reduction of 22.5% on weekdays, and the average speed in Manhattan decreased from 11.76 km/h in April 2017 to 9.56 km/h in March 2019. The heavier traffic congestion may have led to 136% more NOx, 152% more CO and 157% more HC emission per kilometer traveled by the FHV sector. Our results show that the traffic condition is consistently worse across the different times of day and at different locations in NYC. And we build the connection between the number of available FHVs and the reduction in travel speed between the two years of data and explain how the rise of TNC may impact traffic congestion in terms of moving speed and congestion time. The findings in our study provide valuable insights for different stakeholders and decision-makers in framing regulation and operation policies towards more effective and sustainable urban mobility.

Keywords: Transportation network companies, Trajectory data, Urban traffic congestion, Emission

1. Introduction

Transportation network companies (TNCs), which connect travelers with drivers through app-based platforms, have expanded rapidly in recent years. Based on a recent report, TNCs have more than doubled the overall size of the for-hire ride services sector since 2012, making the for-hire vehicle (FHV) sector a major provider of urban transportation services by the end of 2018 [1]. The popularity of TNCs is the result of numerous advantages including improved convenience, higher flexibility, shorter waiting time and lower trip fare as compared to traditional taxi services. However, the overgrowth of TNCs also brings new concerns and challenges for urban traffic management. Although TNCs claim that they help to reduce congestion, official reports and many studies have enumerated signs of road traffic getting worse after the emergence of TNCs. It is reported that private-ride TNC services (Uber, Lyft) have introduced an overall 180 percent more traffic to urban road networks and added billions of vehicle miles traveled (VMT) in the nation's largest metro areas [1]. Another recent study also asserted that TNCs are the biggest contributor to the growth of traffic congestion in San Francisco [2]. These researches depict the big picture of the influence of the overgrowing TNCs on traffic congestion and their findings largely agree with the impression among the general public. However, understanding the precise impact of TNCs on urban traffic is intrinsically difficult, as the change of traffic condition can be the result of the compounding effect of many other factors including population, employment, and change of road network capacity, letting alone the rise of TNCs. And TNCs barely release data that are of sufficient spatial resolution and temporal coverage to allow for tracing their service and evaluating their impacts.

Despite its difficulties, understanding the effects of TNCs on traffic conditions has become an increasingly important topic for transportation planners and policymakers especially in large cities. Our interpretation of the TNC effects will be directly reflected in the way we regulate TNCs and how we may integrate them into the existing transportation system [3]. And our decisions and policies will largely affect the mobility needs of millions of urban travelers and even the livings of hundreds of thousands of TNC drivers. The previous study suggested that TNCs have the potential for reducing road traffic by replacing individual trips with ride-sharing services [4]. But recent research indicated that rapidly increasing TNCs have a negative effect on traffic conditions by attracting transit riders [5]. In particular, the influence of the entry of TNCs on congestion was assessed based on historical area-level panel data. Erhardt et al. [2] studied the impact of TNCs' on traffic congestion through a before-and-after evaluation of the 2010 and 2016 traffic conditions. While they specifically took the change of population, employment and road network into consideration, their results may be largely affected by their counterfactual case in 2016 which was projected from the 2010 baseline with no TNC trips using San Francisco's travel demand model.

In this study, we design a control experiment for gaining accurate insights on the impact of TNC's on urban road traffic by scraping the data from TNC platforms in New York City (NYC) in 2017 and 2019. We limit our discussion to four major boroughs (Brooklyn, Bronx, Manhattan, and Queens) in NYC and argue that the rise of TNCs is the foremost contributing factor to the statistically significant changes, if any, of the road traffic condition based on the following facts:

1. We eliminate the impact of the population since NYC's total population declined from 8.623 million in 2017 to 8.399 million as of July 2018 [6].
2. We eliminate the impact due to employment changes as the labor force and employment are of the identical level in both years (4.13 million and 4.11 million) for NYC.
3. There are no major transportation projects reported in NYC since 2014 according to NYCDOT [7].
4. Registration of standard vehicles declined from 1,913,663 in 2017 to 1,912,468 in 2018 [8].
5. The number of TNC drivers increased from 58,900 in April 2017 to 87,600 in March 2019 (48.7% more) [9].
6. The number of daily TNC trips increased from 393,918 in April 2017 to 769,729 in March 2019 (95.4% more) [9].
7. The number of medallion taxis remains the same but the number of daily trips decreased from 334,865 in April 2017 to 252,634 in March 2019 (24.5% fewer) [9].
8. Transit usage in NYC experienced a drop from 2017 to 2018. It is reported that daily weekday subway ridership in NYC was 5.44 million in 2018, which declined by about 2.6% compared with

2017 (143,000 fewer riders per day). Also, weekday bus ridership in NYC also experienced a drop of 5.9% from 2017 to 2018 (1.81 million).

Table 1: Background facts in NYC

Item	2017	2019
Population (million)	8.623	8.399 (End of 2018)
Employment (million)	4.13	4.11 (End of 2018)
Standard vehicles registration	1,913,663	1,912,468 (End of 2018)
Daily weekday subway ridership (million)	5.58	5.44 (End of 2018)
Daily weekday bus ridership (million)	1.92	1.81 (End of 2018)
Number of TNC drivers	58,900	87,600
Number of daily TNC trips	393,918	769,729
Number of daily taxi trips	334,865	252,634

These facts help to narrow the only dominating contributing factor to the rise of TNC if we may observe any meaningful changes in road traffic conditions. To obtain the most precise understanding of road traffic conditions, we have scraped one month of FHV trajectory data in April 2017 and one month of FHV trajectory data in March 2019. And we use the trajectory data from Uber, the largest TNC in NYC, for further analysis. The scraped trajectory data contain the GPS record of the online Uber drivers every few seconds and can be used to visualize and quantify the spatiotemporal change of traffic conditions. And the large amount of data we collected help to obtain findings that are statistically meaningful. We then classify the trajectory data into moving activities and stationary activities for fine-level analysis of the time spent in congestion and speed during travel. We introduce macroscopic energy models to further calculate the change in fuel consumption and emission during the two years. Through comprehensive numerical experiments, we conclude that the increase of FHVs contributes to significant speed reduction in NYC with a daily average drop of 22.5% on weekdays. As for Manhattan, the average speed declines from 11.76 km/h to 9.56 km/h on weekdays and from 14.98 km/h to 13.51 km/h on weekend in less than two years. We report that the increased traffic congestion, along with the growing number of TNC trips, double the tailpipe emissions from the TNC sector since 2017.

The rest of the study is organized as follows. We briefly review related literature on trajectory analysis in the next section. Section 3 introduces the main methods used in this study, including the developed data collection method, the validation of data quality, activity identification from trajectory data and energy and emission calculation. Section 4 presents comprehensive results and discussion on understanding the FHVs' impact. Finally, we summarize key findings and future directions in section 5.

2. Literature

With the rapid development of data collection methods and availability of traffic-related big data in cities, estimating city-level fuel consumption using vehicle trajectory data has gained a lot of interest.

GPS trajectory data have been widely used to understand mobility patterns [10, 11] and travel behavior [12], discover flexible routes [13] and monitor real-time traffic situation (visualize traffic jam) [14] due to their advantages of large coverage, good continuity, low cost and rich information about vehicles' movements. In recent years, GPS trajectory data were used for large-scale fuel consumption estimation to provide a more accurate vision of national or regional level vehicular emissions. Shang et al. [15] calculated the gas consumption and emissions using GPS trajectories generated by over 32,000 taxis in Beijing over a period of two months based on the estimated travel speed of each road segment using a context-aware matrix factorization approach. Du et al. [16] explored the fuel consumption pattern and analyzed the temporal and spatial distribution characteristics of average fuel consumption in Beijing using large samples of historical floating vehicle trajectory data, where a fuel consumption forecasting model was established using the back-propagation neural network. Gately et al. [17] quantified the emissions from traffic congestion and identified local hotspots with highly elevated annual emissions at regional scales using a large database of hourly vehicle trajectory data CO_2 from road vehicles on 280,000 road segments in eastern Massachusetts. Luo et al. [18] analyzed the energy consumption and emissions and their spatial-temporal distribution in Shanghai using GPS trajectory data obtained from taxis.

Vehicular emission models can be summarized as two types: macroscopic models [19, 20] and microscopic models [21, 22], which focus on different aspects of vehicle emissions calculations and analysis. For large-scale fuel consumption estimation, macroscopic models are usually used where emissions factors are modeled as functions of the average speed of vehicles [23]. However, these estimations do not consider different driving modes or driving patterns which have been proved to have an obvious effect on vehicle fuel consumption [24]. For example, engine start [25] or idling speed [26] will increase vehicle exhaust emissions. Lack of consideration of these parameters may lead to erroneous estimations. For large-scale emissions estimation, such erroneous estimations may result in a misunderstanding of the overall traffic states and emission levels in the region. While GPS trajectory data can reveal detailed information about vehicle driving modes and traffic states, it therefore provides the possibility of identifying different driving activities that will influence vehicle fuel consumption [27, 28]. In this paper, a two-step integrated emission estimation method [29, 30] that incorporates driving activities (considered in microscopic models) into COPERT model (macroscopic model) is adopted to provide more accurate fuel consumption estimation of Manhattan using GPS trajectory data obtained from Uber. With this method, driving activities of each vehicle are first specified as moving activities and stationary activities. COPERT model is then applied to calculate the emissions of all trajectories considering both types of driving activities of each vehicle. The integrated estimation method ensures more accurate emissions and fuel consumption estimation in a city-level scheme and at the same time provides a more detailed sense of TNC's influence on traffic conditions.

3. Method

3.1. Data Collection

To gain insights on the impact of FHVs on urban traffic, we develop the data crawler, which simulates the ride requesting behavior on the mobile app, to fetch the trajectory data from major TNCs including

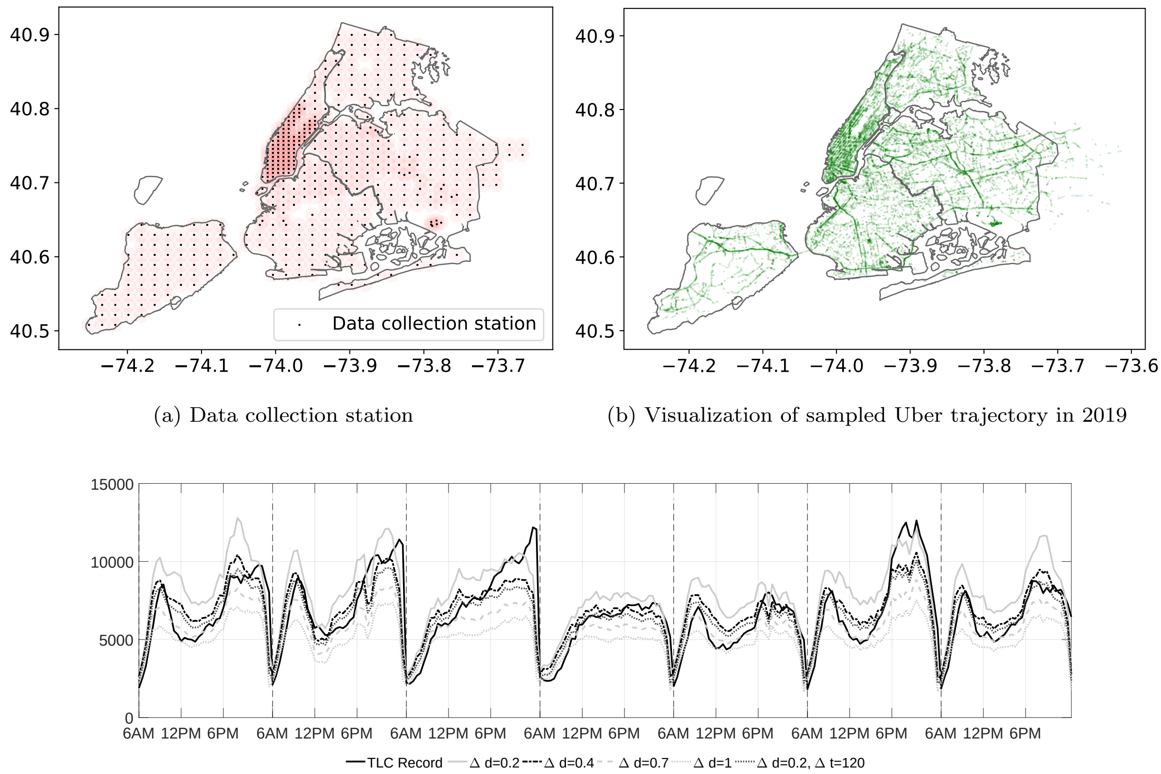
Uber and Lyft. Our data crawler sends the trip starting location as the pingClient message to TNCs' mobile API and receives back the sequences of coordinates of eight closest online FHV drivers as well as the *surge price* (SP) and *estimated time of arrival* (ETA). Online vehicles refer to those who are available for picking up passengers and the vehicles will no longer be recognized if they start a trip or if they go offline. The collected trajectories therefore capture the cruising behavior of FHVs. But different from taxis where street hailing is permitted, FHVs cruise to the next pick up location assigned by the platform and the data therefore well reflect the actual traffic condition. By placing a sufficient number of data collection stations with proper spacing and collection frequency, we are able to collect abundant vehicle trajectories to restore the citywide operation dynamics of FHVs. In this study, only the trajectory data from Uber are used as it is the dominant TNC in NYC with approximately 70% market share [31]. Trajectory records collected include the information of timestamp (in Unix), latitude, longitude, driver ID (only first 6 letters shown here), product ID (e.g. UberX and UberXL) and bearing. The sample of collected trajectory records can be seen in Table 2.

Table 2: Sample data records

Product ID	Driver ID	Epoch	Bearing	Latitude	Longitude
2083	b97fed	1491760511750	344	40.67387	-73.80141
39	657dbb	1491753163395	209	40.77918	-73.95079
694	6b25cd	1491748277252	299	40.78273	-73.9495
4	73c3f4	1491732814910	191	40.71448	-74.01372
39	5f486	1491733990716	299	40.75755	-73.96903

We conduct citywide data collection in NYC and the data analyzed in this study were collected from April 7 to May 3rd (6 AM to 11 PM) in 2017 and February 7 to March 13 (24 hours) in 2019 from Uber API. The data collection was performed at the frequency of 5 seconds for each data collection station in 2017, and this frequency was set to 1 minute in 2019 due to the change of functional mechanism of Uber API. As suggested in [32], Uber may dynamically alter the ID assigned to each driver and the data collected therefore do not contain privacy information related to any individual drivers. And the data collection stations only send pingClient messages to Uber server for obtaining nearby vehicles' trajectories without actually requesting a ride. Hence our data collection was conducted in an ethical manner that neither hacked any driver or passenger privacy information nor sent real ride requests which may disturb Uber operations.

We set the same data collection station configuration in 2017 and 2019 which consists of 470 stations. The amount and spatial placement of the stations are carefully calibrated to ensure sufficient coverage of the actual operation dynamics. In the beginning, we randomly placed a set of data collection stations spreading over the entire NYC area, with each location having two data collection stations, and sent pingClient message every 5 minutes for 12 consecutive hours. The test results suggested that over 99.99% of feedback messages between the two stations at the same location were exactly the same. And we therefore assigned one data collection station per location. Another set of experiments was conducted to identify appropriate spacing between two adjacent data collection stations. We used historical taxi demand distributions to divide the whole study area into three sub-regions based on the trip demand level. We varied the spacing from 100m to 1,500m between two adjacent stations in each sub-region and deployed 9 neighboring stations in each region to measure data repetition among the 9 stations for a 12-hour data collection. Finally, we chose the largest spacing that reached at least 40% repetition. The resulting distribution of the data collection stations and the sampled spatial trajectory coverage are shown in Figure 1.



(a) Data collection station

(b) Visualization of sampled Uber trajectory in 2019

(c) Comparison between inferred number of trips vs TLC reported Uber trips [33]

Figure 1: Study area and the configuration of data collection stations

The 470 data collection stations fetched around 100 GB data per day in 2017 and 5.17 GB data per day in 2019. To validate the quality of the data, we infer the number of Uber trips from 2017 data and compare this number with the FHV trips reported by NYCTLC [34] for every 15-minutes time interval of entire NYC. In particular, we track the trajectory of each unique driver ID and consider a trip was taken place if (1) the time gap ($\Delta t \geq 60$) and spatial displacement (Δd) between consecutive records exceeds certain threshold or (2) the record was the last trajectory identified for the driver ID. The validation results with various distance and time thresholds are presented in Figure 1c. While the collected data by no means capture complete FHV operation information, we observe that the inferred number of trips well resembles the reported trip level and the trip trend is closely aligned with the actual trip tendency with the proper choice of distance and time threshold. This demonstrates the quality of the data we collected and suggests that the data yield sufficient coverage of actual FHV dynamics.

Finally, we choose the data between Feb 27 to March 12 in 2019 and April 12 to April 25 in 2017 and compare the change of traffic states in two years. This time selection is to ensure the dates are comparable in the time vicinity given the availability and the quality of the data we collected. Moreover, we only focus on investigating the change in Manhattan as the case study which is the borough of the heaviest congestion and highest FHV trip level in NYC.

3.2. Activity identification

Based on the collected FHV trajectories, we next convert the trajectories into space-time path segments (STPS) following the method proposed in Kan et al. [29]. The main reason for STPS construction is to identify different vehicle activities during a sequence of GPS records for accurately estimating the trajectory speed and inferring energy consumption and emission. In particular, we focus on separating stationary activities (SA) from moving activities (MA) in the trajectories so that we may make the best use of the high-resolution trajectories to restore the stop-and-go traffic states. MA and SA will contribute to differentiating between the amount of time urban traffic caught in gridlock and the velocity of the moving traffic. In addition, the functionality of engines differs between idle state and when the vehicle is in motion. MA and SA will therefore result in the more accurate characterization of fuel consumption and emission for urban traffic, where MA can be used with emission models for vehicle in motion and SA can be used with emission models for idle engine state to obtain comprehensively evaluate the actual emissions and fuel consumptions. Studies have shown that this approach can achieve over 88% accuracy when using macroscopic emission model [29] and over 94% accuracy when using microscopic emission model [30] when compared to actual fuel consumption.

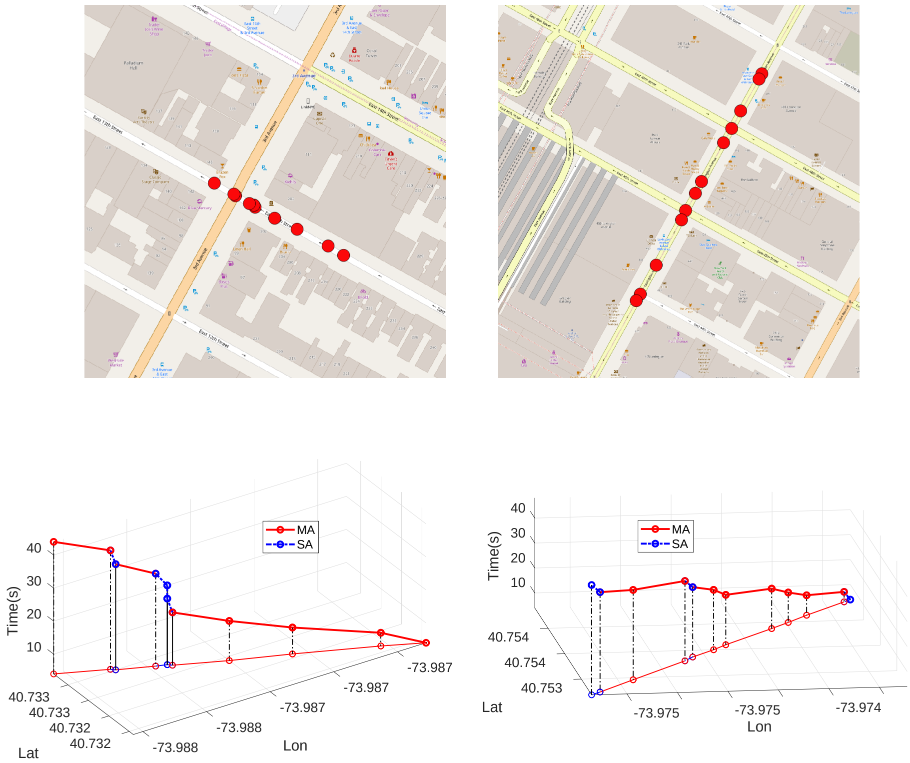
To best identify trajectory activities, we first preprocess the data to remove consecutive trajectory records of time gap that is shorter than 2 seconds or longer than 15 seconds. The removal of short time intervals helps to mitigate GPS errors. On the other hand, we may underestimate the number of SA for including records of longer time gaps as intermediate SA will be consolidated and reflected as MA if these short time intervals are included. The resulting time gaps between consecutive trajectory records mostly lie between 4 seconds and 6 seconds. The preprocessing eliminates around 15% of trip records in the data and we then identify SA and MA based on the velocity (km/h) of the trip segment:

$$V_{i,i+1} = \frac{\|coord_{i+1} - coord_i\|}{t_{i+1} - t_i} \tag{1}$$

where $\|\cdot\|$ measures the euclidean distance between consecutive trajectory records in kilometers. And we define the state of trajectory segment as:

$$S_{i,i+1} = \begin{cases} SA, & \text{if } V_{i,i+1} < 5 \\ MA, & \text{if } V_{i,i+1} \geq 5 \end{cases} \tag{2}$$

The threshold of $V_{i,i+1} < 5$ for separating SA and MA is selected to mitigate GPS errors that may lead to the false classification of actual identities. We present two sample trajectories and their constructed STPS and identified SA and MA in Figure 2. As seen in the figure, by using the velocity threshold, we are able to accurately identify the non-moving or near non-moving activities as SA and the actual moving trajectories as MA. After activity identification, we observe there are over 4.8 million daily activities for 2017 data and over 3.2 million daily activities for 2019 data between 7 AM to 11 PM. And these large number of activities will be sufficient to obtain statistically meaningful results in the following sections.



(c) Sample trajectory of length 103 meters

(d) Sample trajectory of length 149 meters

Figure 2: Example of collected FHV trajectory and the constructed STPS

3.3. Estimating fuel consumption and emission

Total vehicle emission is usually categorized into cold emission and hot emission. Hot emission entails the emission when the engine is operating at a normal temperature, and the cold emission denotes the emission at transient thermal operation. In this study, we only consider hot emission due to lack of data to classify cold start activities and also because hot emission usually dominates the total emission for long trips. As reviewed in the earlier section, both MOVES and COPERT are popular models for energy and emission calculation and MOVES are specifically tailored to emission standards in the US. Nevertheless, the MOVES model requires the calculation of vehicle specific power which needs the second by second acceleration and engine specification data. This calls for the need of trajectory interpolation and is better suited for long trajectories. Our data primarily contains trajectories over short segments (as shown in Figure 2) and interpolation may result in high estimation errors. As a consequence, we use COPERT model for fuel and emission calculation and assume all vehicles under Euro 3 standards with a capacity of 1.4-2.0L. Note that not all Uber vehicles may comply with the Euro 3 standards and there is no available data to understand the type of vehicles in the Uber fleet. In addition, Euro 3 standards may not fully comply with the US EPA standards and hence the exact value calculated for emission and fuel consumption may not be taken as an accurate measure for NYC. Nevertheless, the change of standard will only affect the model parameters but not the relationship between velocity and the corresponding fuel consumption and emission and the obtained results still capture the relative change between 2017 and 2019.

Based on the aforementioned specifications, for MA, fuel consumption (denoted as $FC(g/km)$) can be calculated based on trajectory segment speed $V(km/h)$ as:

$$FC_{MA} = \frac{217 + 0.253V + 0.00965V^2}{1 + 0.096V - 0.000421V^2} \tag{3}$$

As for SA state, we estimate the fuel consumption based on vehicle idle time T [35]as:

$$FC_{SA} = 0.361mL/s * 0.75g/mL * T = 0.27g/s * T \tag{4}$$

where the density of gasoline is taken as 0.75g/mL.

As for hot emission, following the Tier 3 method of COPERT model[36], the emission factor (EF (g/km)) during MA state are speed-dependent:

$$EF_{MA} = (1 - RF)\frac{aV^2 + bV + c + \frac{d}{V}}{eV^2 + fV + g} \tag{5}$$

where RF is the reduction factor. The corresponding parameters for measuring EF of CO, HC and NO_x are presented as follows:

Table 3: Emission parameter for small vehicles in COPERT model

item	**a**	**b**	**c**	**d**	**e**	**f**	**g**
CO	0	11.4	71.7	0	-0.248	35.4	1
NO_x	6.53e-6	-1.49e-3	9.29e-2	0	3.97e-5	-1.22e-2	1
HC	1.2e-5	-1.1e-3	5.57e-2	0	-1.88e-4	3.65e-2	1

Finally, the calculation of EF under the SA state takes the following form:

$$EF_{SA} = \alpha * T \tag{6}$$

where T is the idle time and the parameter $\alpha(mg/s)$ for CO, NO_x and HC are 13.889, 0.556 and 2.222 respectively [35].

3.4. FHV as probe vehicles

Based on the previous discussions, we are able to make FHV as the probe vehicle for characterizing the traffic condition in Manhattan with the large-scale trajectory data collected. Since Uber has a large fleet of vehicles roaming around NYC, the performance metrics calculated from Uber vehicles will serve as a close approximation of the actual metrics of all vehicles on road. If we consider $\mathcal{P}$ as the complete trajectory data generated by the entire Uber fleet, then our collected data $\mathcal{P}_C \subset \mathcal{P}$ which can be viewed as the sub-population randomly drawn from $\mathcal{P}$. As a consequence, the average performance metrics calculated from our collected data is the sample mean of the entire population. And the mean value of the metrics obtained in our data will be close to the expected value in $\mathcal{P}$ based on the law of large numbers. These suggest that the traffic condition mined from our data can well represent the actual traffic condition of the road network.

In this study, we are primarily interested in the spatiotemporal velocity metrics and the corresponding energy and emission level. In particular, we propose to measure the following velocity metrics:

$$V_{i,t} = \frac{\sum_k D_k^{MA}}{\sum_k T_k^{MA} + T_k^{SA}}, \text{if activity } k \text{ is at location } i \text{ within time } t \tag{7}$$

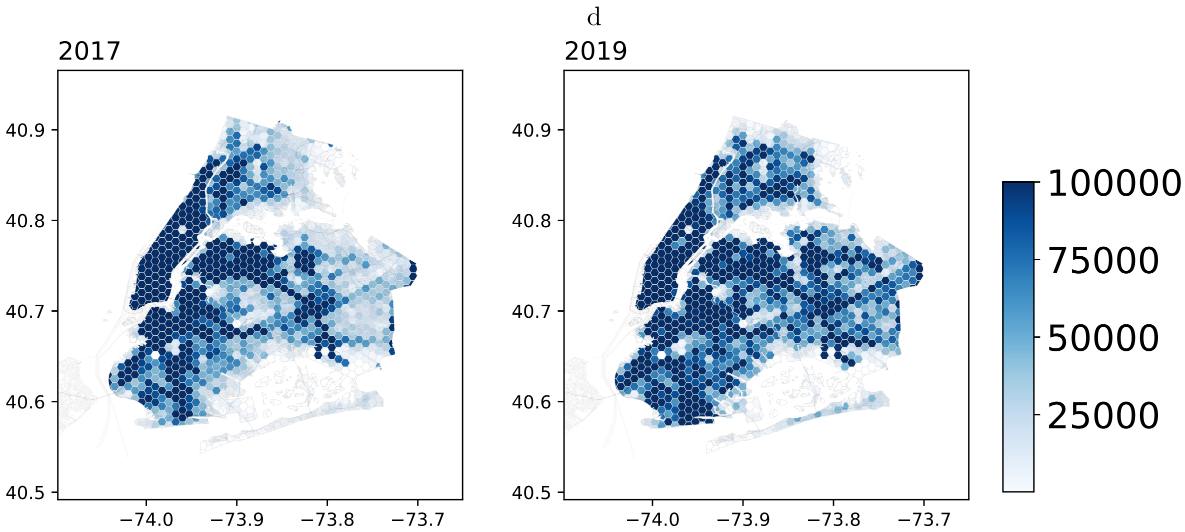
$$V_{i,t}^{MA} = \frac{\sum_k D_k^{MA}}{\sum_k T_k^{MA}}, \text{if activity } k \text{ is at location } i \text{ within time } t \tag{8}$$

$$R_{i,t}^{SA} = \frac{\sum_k T_k^{SA}}{\sum_k T_k^{MA} + T_k^{SA}}, \text{if activity } k \text{ is at location } i \text{ within time } t \tag{9}$$

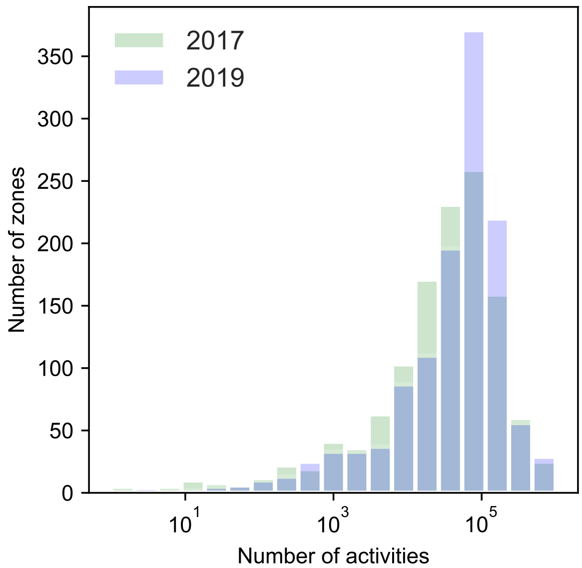
where $V_{i,t}, V_{i,t}^{MA}$ and $R_{i,t}^{SA}$ represents the mean velocity, mean MA speed (speed when the vehicle is in motion) and mean SA time (proportion of time spent in stationary traffic congestion) respectively.

4. Results

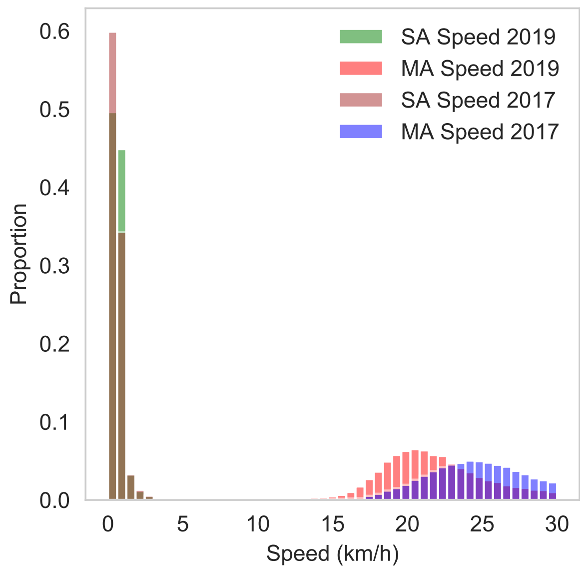
4.1. Overview of identified activities



(a) Spatial distribution of identified activities



(b) Histogram for zonal activity level

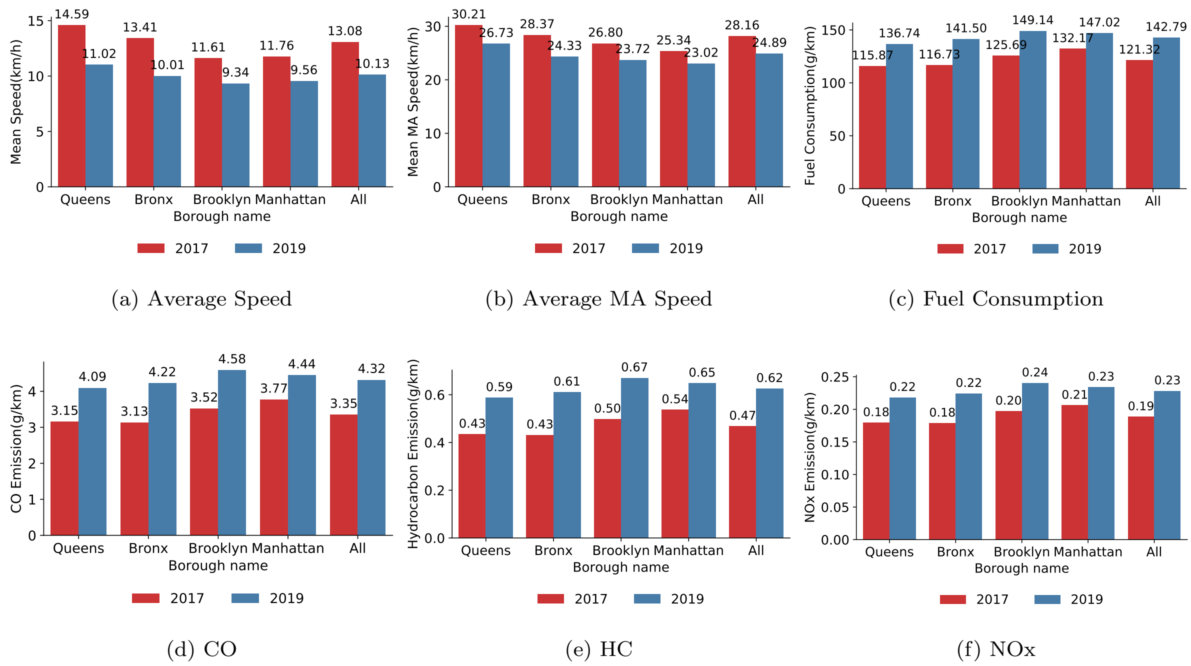


(c) MA and SA speed distribution

Figure 3: Distribution of number of identified activities in 2017 and 2019.

As mentioned earlier, due to the change in data collection frequency, there exists a significant difference in the amount of data collected in 2017 and 2019. To overcome this issue and deliver fair comparison, we perform sampling from the 2017 data and include 30% of the total identified activities. This results in similar number of total activities identified in 2017 and 2019, as shown in Figure 3a. There are over 123.8 million activities identified during our study period from the 2017 data and the corresponding value is 117.6 million for 2019, which suggests the same level of identified activities between the two years. These activities cover the entire study area and we also can verify the expansion of Uber's service coverage areas from 2017 to 2019 based on the spatial distribution of the identified activities. In particular, we report that 70% of the zones in 2017 and 74.5% of the zones have more than 10,000 identified activities (see Figure 3b), representing over 150 activities for each 15-minutes time interval at each location. This vast amount of activities delivers superior spatiotemporal coverage and ensures the obtained results are statistically meaningful. Finally, we present in Figure 3c the validity of the identified SA and MA based on equation 2, where we also measure the speed of SA from the spatial displacement and time gaps. We can verify that over 90% of the identified SA have the speed lower than 1 km/h and there exists a small fraction of SA with speed lower than 3 km/h which we suspect to be caused by GPS errors. On the other hand, we observe that MA is perfectly separated from the SA based on the speed metric and we can readily tell the differences between 2017 and 2019 data from the distributions of the corresponding MA speed. We next present detailed analyses of the changes in traffic condition and emission based on the identified activities.

4.2. Overall change in traffic condition

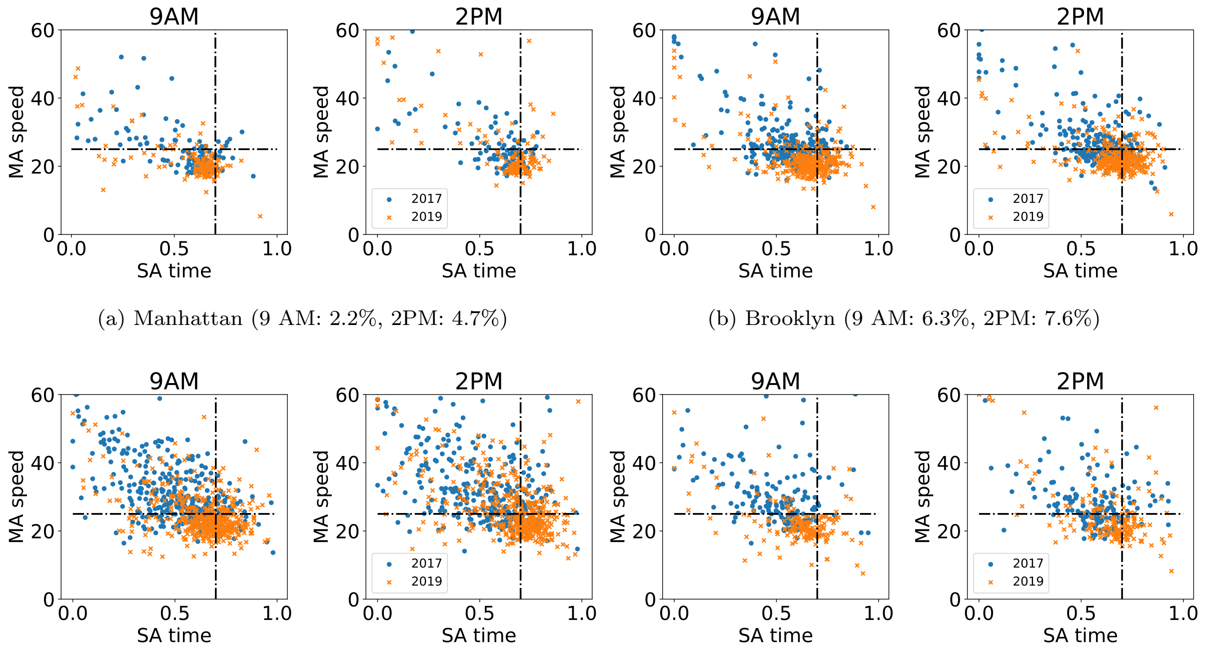


(a) Average Speed (b) Average MA Speed (c) Fuel Consumption

(d) CO (e) HC (f) NOx

Figure 4: Average borough-wide performance during weekdays

We first show the comparisons of daily average speed, energy consumption and emission across the study area and the results can be found in Figure 4. One immediate observation from the results is the deterioration of citywide traffic performances from 2017 to 2019, and such observation is consistent across the four major boroughs in our study area. We find that the citywide average daily speed reduced from 13.08 km/h in 2017 to 10.13 km/h in 2019, representing a significant drop of 22.6%. Among the four boroughs, Manhattan and Brooklyn are the areas with the worst traffic condition and we observe the average speed reduction around 19%. Meanwhile, we see a notable increase in energy consumption and emission due to the worse borough-wide traffic condition. For each kilometer traveled, the vehicles in NYC now consume 21 grams more gasoline and emit 1 more gram of CO, 0.15 more grams of HC and 0.04 more grams of NOx on average as compared to the 2017 state. If we assume the metrics calculated from cruising FHVs also apply to the full FHV sector, and project these values onto the increase of FHV trips while assuming the same average distance per trip, these translate into that FHVs have introduced 152% more CO, 157.7% more HC and 136.5% more NOx for every kilometer they traveled.



(a) Manhattan (9 AM: 2.2%, 2PM: 4.7%)

(b) Brooklyn (9 AM: 6.3%, 2PM: 7.6%)

(c) Queens (9 AM: 9.1%, 2PM: 13.3%)

(d) Bronx (9 AM: 3.5%, 2PM: 9.4%)

Figure 5: Relationship between standardized SA time and standardized MA speed. The percentage of areas that exceeds the predefined threshold is shown in the bracket.

There is, however, one particular drawback associated with the cruising trajectory data when it is used for probing citywide traffic conditions. Since FHV drivers may choose to park by side of the street and wait for future orders from the platform, this may lead to overestimation of the actual number of SA activities on the road and the calculated average zonal speed, therefore, constitutes the lower bound of actual travel speed. By inspecting the relationship between $V_{i,t}^{MA}$ and $R_{i,t}^{SA}$, we may gain additional insight on this particular issue. Specifically, when there is heavy congestion in certain areas, we should observe low MA speed and high SA time which captures the stop-and-go traffic pattern during congestion and gridlock. Similarly, when traffic is light, the trajectory data should reveal high MA speed and low SA time. However, when Uber drivers choose to park and wait rather than cruise, we are likely to encounter the anomalies with both high MA speed and high SA time. This motivates us to explore the percentage of observations that fall into the latter abnormal state and reveal the park-and-wait behavior of FHV drivers. By inspecting the 2017 and 2019 data, we find that the 75% percentile of MA speed across the data is approximately 25 km/h and that for the SA time is around 0.7. We then set the threshold of $\bar{V}^{MA} = 25$ and $\bar{R}^{SA} = 0.7$ and measure the proportion of areas in each borough with both MA speed and SA time being higher than the thresholds. The results are shown in Figure 5.

We report that during morning peak hours (9 AM) all boroughs present traces of the park-and-wait behavior. Manhattan has the lowest value of 2.2% while Queens has the highest percentage of park-and-wait observations (9.1%), followed by Brooklyn (7.6%) and the Bronx (3.5%). And the percentage of park-and-wait observations is increased for all boroughs during the off-peak time (2 PM), where Queens still has the highest value of 13.3% and we find a drastic jump in the Bronx to 9.4% and that of Brooklyn is 7.6%. These findings explain why the estimated speed in Brooklyn is lower than in Manhattan despite the fact that Manhattan has the highest number of FHVs and passenger demand. In this situation, the cruising trajectory data may slightly underestimate the average speed and overestimate the actual energy consumption and emission. And the change in MA speed serves as a more accurate metric for comparing the change of traffic conditions across different boroughs. On the other hand, the results also indicate that the obtained traffic condition and emission are relatively accurate in Manhattan as there are few park-and-wait observations. Moreover, the measured average speed of 11.76 km/h in 2017 is well aligned with the speed of 11.2 km/h reported in the NYC mobility report [37]. We next zoom into Manhattan borough and discuss how traffic conditions and emission change over time.

We summarize the time-varying performance metrics for both weekday and weekend in Manhattan in Figure 6 and 7. For measuring the changes in traffic conditions, we plot the average speed, average MA speed, as well as average SA time from 2017 and 2019 data and the corresponding changes between the two years, are visualized by the shaded area. While the average weekday speed in Manhattan decreased from 11.76 km/h to 9.56 km/h (Figure 4), this reduction can be further decomposed into two parts. On one hand, more FHVs result in slower-moving speed so that there is a reduction of 9.2% in average MA speed in Manhattan. On the other hand, there is also an increase in average SA time by 7.04%. These provide strong evidence showing that the traffic condition is worse in 2019 than in 2017, and such observation is consistent across different times of the day. As for the weekend, the mean speed is 14.98 km/h in 2017 and 13.51 km/h in 2019 respectively, suggesting a reduction of 9.8%. During weekdays, we find that the peak hours (especially morning peak during 7-9 AM) have the worst traffic condition and also suffer the greatest decline in average speed and MA speed. The changes during off-peak hours are relatively minor. For the weekend, we observe that notable changes in mean speed mainly take place during off-peak hours (7 AM to 12 PM on weekend) and during the nighttime period (7 PM to 10 PM). The decreases in mean speed during weekdays and weekends correspond to the largest drop in travel speed in Manhattan sine 2015 [37] and eventually lead to higher fuel consumption and more tailpipe emission across different times of the day. For Manhattan, we observe that, during weekdays, vehicles will consume 10.0% more gasoline and exhale 12.0% more NO_x, 16.1% more CO and 18.6% more HC for each kilometer they traveled in Manhattan in 2019 as compared to those in 2017. These results highlight the critical traffic congestion issues related to the rise of TNC in NYC, and possibly around the world: the fast expansion of TNCs quickly saturates the road network, resulting in the increase of fuel consumption and vehicle emission for all road traffic and even significant addition from the compound of increasing worse traffic condition and more FHV trips.

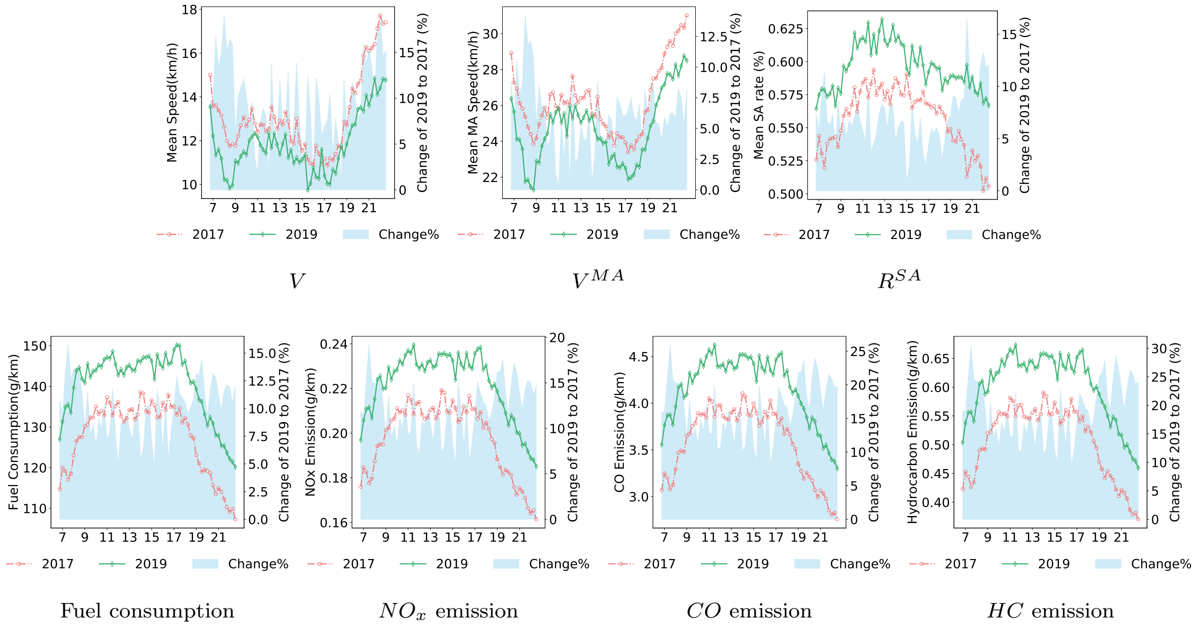


Figure 6: Change of city-wide metrics on weekdays between 2017 and 2019

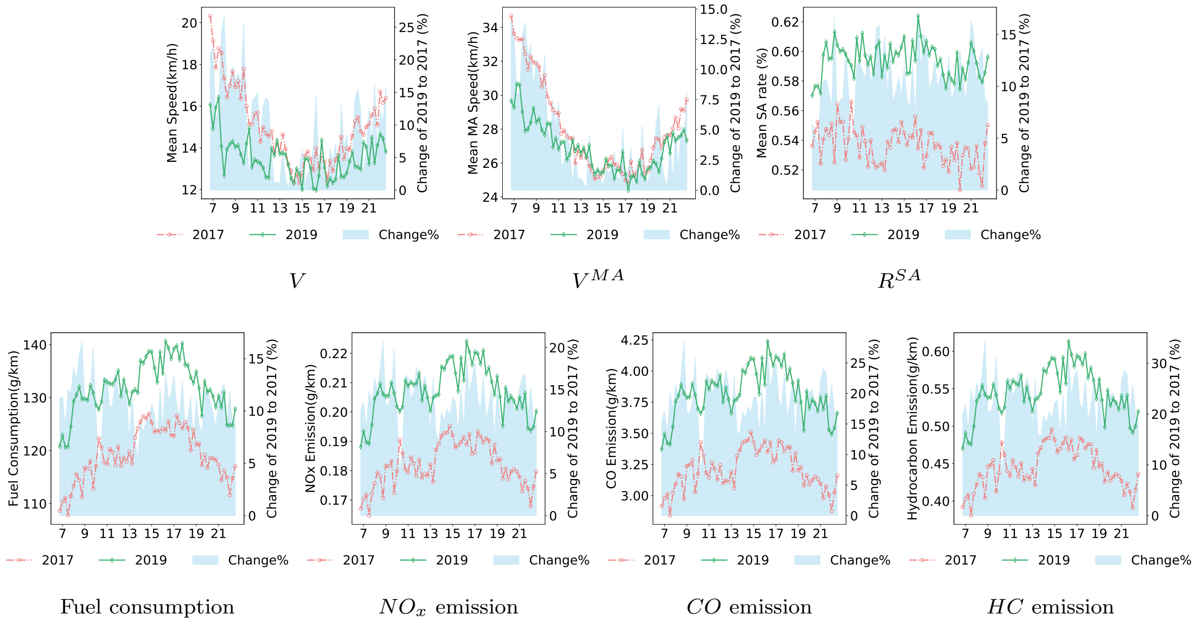
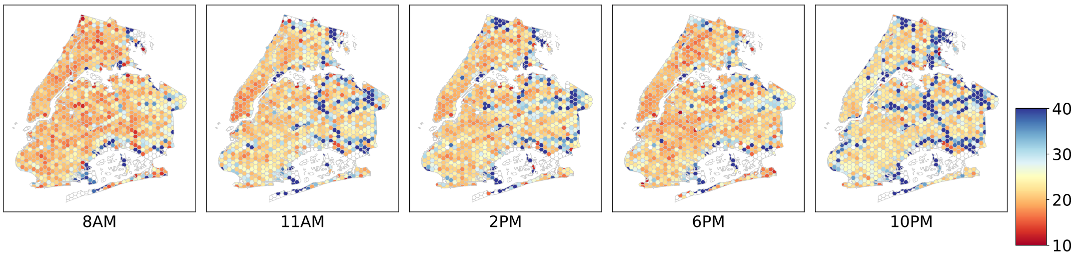
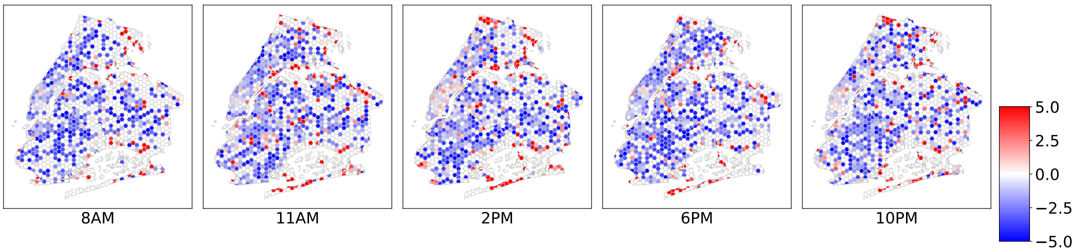


Figure 7: Change of city-wide metrics on weekend between 2017 and 2019

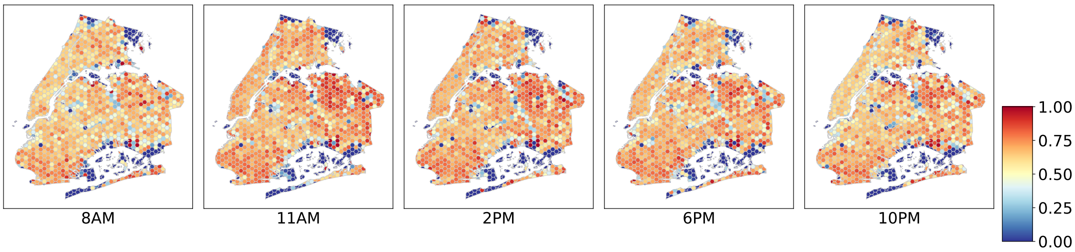
4.3. Spatial impact



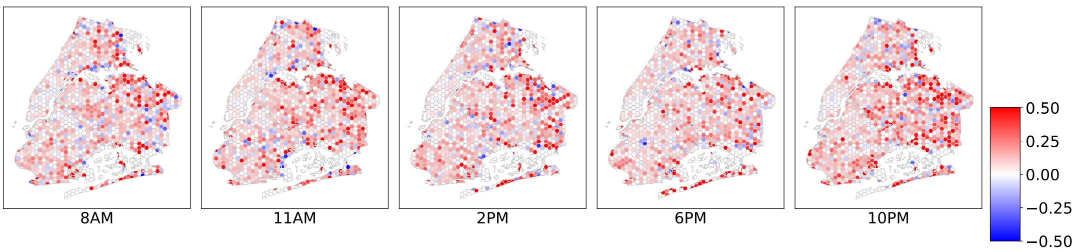
(a) Mean MA speed distribution in 2019 on weekdays



(b) Change of mean MA speed between 2017 and 2019 on weekdays



(c) Mean SA ratio distribution in 2019 on weekdays



(d) Change of mean SA ratio between 2017 and 2019 on weekdays

Figure 8: Spatial distribution of mean speed and SA ratio on weekdays

We next present the spatial distributions of the mean MA speed and SA time in 2019 and their changes as compared to 2017, and the results during weekdays are shown in Figure 8. Based on Figure 8a, we can clearly distinguish between the traffic peak hours (8 AM and 6 AM) and off-peak periods. In particular, we find that heavy traffic congestion in Manhattan persists across the day time, whereas there are notable differences in terms of the MA speed in other boroughs between peak and off-peak hours. Location-wise, we observe that lower and middle Manhattan, as well as the areas in other boroughs that are adjacent to Manhattan, are found to suffer the heaviest congestion with the average MA speed being less than 20 km/h. And with the rise of TNCs, we notice a city-wide reduction of MA speed in despite of the particular times of the day, where there are 88.2%, 83.1%, 78.9%, 85.1% and 81.8% of all 1371 areas that have slower travel speed for 8 AM, 11 AM, 2 PM, 6 PM, and 10 PM. On weekend, the corresponding values are 76.9%, 76.0%, 73.3%, 75.1% and 70.3% respectively. These numbers indicate that the traffic condition is more affected during weekdays than on weekend, and the congestion is worse during peak hours on both weekdays and weekends.

Note that with an excessive number of FHV vehicles on the road in cruising mode, these drivers are likely to present different driving behavior than other commuting drivers. Specifically, they need to pay close attention to their smartphones for incoming passenger orders which distracted them from the road. And they may have to frequently merge into or diverge from the slow traffic to pick up passengers or find temporary parking spots to save their cruising cost. These undoubtedly introduce significant disturbance to the already slow traffic and adding frequent stop-and-go activities and shockwaves into the traffic flow. These can be confirmed from the spatial distribution of the SA ratio, as shown in Figure 8c. Being different than the findings from the distributions of the MA speed, the morning peak hour has a lower level of SA ratio and the SA ratio is found to be higher at the time of more number of cruising drivers (see x-axis in Figure 9 for the number of identified cruising drivers). This is likely due to lower demand levels and more cruising FHVs in off-peak hours, resulting in greater disturbance to the traffic and more stop-and-go traffic. The average SA ratio is 0.503 for 8 AM, and 0.542,0.535,0.532, and 0.508 for 11 AM, 2 PM, 6 PM and 10 PM respectively. For every 10 minutes of driving in NYC, these numbers translate into over 5 minutes sitting in non-moving traffic and highlight the huge amount of wasted time for the large number of daily travelers who drive themselves or rely on taxis and FHVs. And when compared to the 2017 scenario, there are 66.4%, 71.5%, 71.0%,70.6% and 73.4% of the areas at 8 AM, 11 AM, 2 PM, 6 PM and 10 PM that experience increased SA ratio. Finally, following our previous discussion on the relationship between MA speed and SA ratio, we can also visually identify the places with dominant park-and-wait behavior. And such behavior is primarily found in peripheral areas of Brooklyn, the Bronx, and Queens with lower trip intensity and less traffic. It is easier for drivers to spot a parking place in these places and wait for future orders from the ride-hailing platforms.

4.4. Active drivers and speed change

Finally, we build the connection between the change of vehicle speed and the number of available Uber drivers on the road during weekdays. In particular, we track the active Uber drivers as the number of unique driver IDs identified across the study areas for every 15 minutes time interval. And we separate our daily observations and group similar time intervals to increase the number of individual observations for comparison between 2017 and 2019. For each year, we use the observations from 8 consecutive time intervals (every 2 hours) on each of the 10 weekdays. This gives 80 observations for each 2 hour time period for both 2017 and 2019, and each observation contains the number of active Uber drivers and the average speed in NYC. The relationship between active Uber drivers and the speed at different times of the day are presented in Figure 9. One immediate finding based on the results is that the traffic condition in 2017 and 2019 are in entirely different states and the differences between the two years are distinguishable even if we visualize the observations of all time periods in one plot (7:00-23:00). And

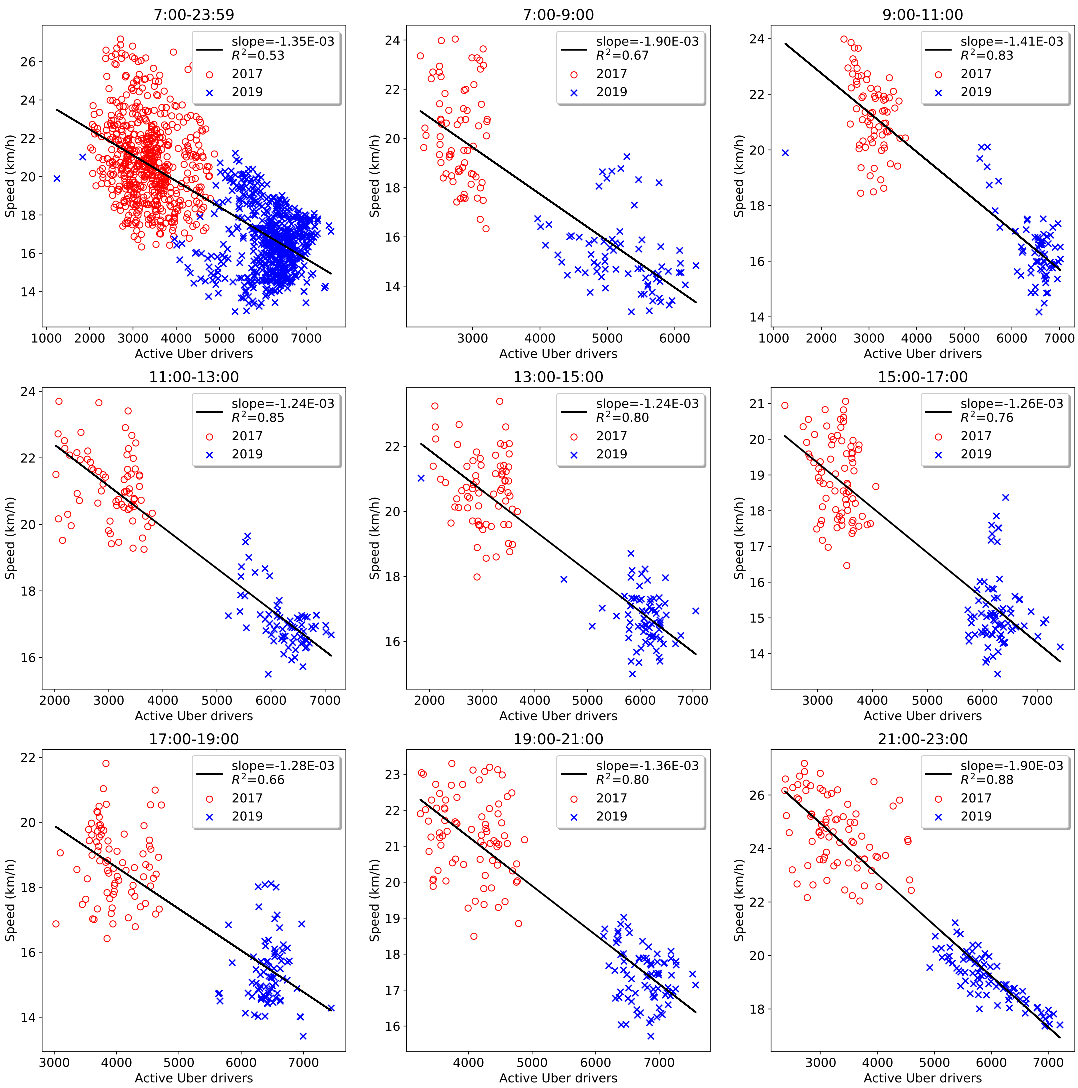


Figure 9: city-wide average speed comparison across different time of the day

the two years of data are also linearly separable in all of the individual time periods. We calculate the Pearson correlation coefficient between the number of active drivers and the average speed for each two-hour period of time, and the resulting coefficients are -0.82,-0.91,-0.92,-0.89,-0.87,-0.81,-0.89 and -0.94 for each of the time period. All these values are close to -1 and they suggest the significant negative correlation between the number of FHV drivers and the average travel speed in Manhattan. While correlation does not necessarily imply causation, if we may eliminate the impact from other contributing factors such as those shown in Table 1, the strong negative correlation likely hints that the increase in FHV drivers is the primary contributing factor to the citywide worse traffic congestion and emission. Note that the active Uber drivers we identified from the data may only serve as a proxy of the total number of Uber drivers in service. During the time with high passenger demand (e.g. 7:00-9:00, 17:00-19:00), we may identify fewer drivers than that during off-peak hours. This is because our data capture cruising Uber drivers and the drivers are less likely in cruising state when there is more passenger demand than the vehicle supply. Nevertheless, the results are still valid as we compare the same time of the day in two different years. And the number of identified drivers as shown by the x-axis also echo our finding on the impact of excessive drivers in cruising on travel speed and SA ratio (Figure 8). As the observation is based on the consistent results across all time periods over multiple days of data for two years, this confirms that the increase of FHV trips and TNC drivers are one significant contributing factor to urban traffic congestion. If we fit a simple linear regression model over the data, as shown by the lines in Figure 9, the number of active Uber drivers alone may explain up to 88% (21:00-23:00 with $R^2 = 0.88$) of the variability for the reduction of travel speed and such linear relationship may well fit the observations for most of the time periods. Based on the coefficients of active Uber drivers, we further notice that the impact of the number of FHVs on travel speed can be categorized into two cases depending on the number of cruising drivers in the city. In the first case (7:00-9:00, 21:00-23:00), the impact of this is reflected by both the reduction in MA speed and the increase in SA time. For the second state (9:00-21:00), the impact of FHV vehicles is primarily reflected by the increase in SA ratio, where the park-and-wait behavior from excessive cruising drivers as well as the disturbance to normal traffic from more number of FHV trips together lead to the worse traffic congestion and emission. In general, for the first state, the increase in FHV vehicles has a greater impact on the speed reduction than the second state with the fitted coefficients being 50% higher.

5. Conclusion

In this study, we collect and mine large-scale FHV data and provide comprehensive understandings of how the rise of TNCs impacts the traffic congestion and emissions in urban areas. We choose Manhattan in NYC as the study area and conduct analyses of the trajectory data in 2017 and 2019. We classify stationary and moving activities from the trajectory data and calculate the mean speed, energy consumption and fuel consumption based on the classified MA and SA. Our results suggest that the increase of FHV trips in NYC has resulted in an average citywide speed reduction of 22.5% on weekdays and the average speed in Manhattan has decreased from 11.76 km/h in April 2017 to 9.56 km/h in March 2019. And if we consider the increase of FHV trips over the two years. Our results confirm that the increase of TNC vehicles is one of the major contributing factors to the increase in traffic congestion. And we also articulate two different ways, which depend on the overall congestion level, that the increase in FHVs may affect traffic congestion with different magnitude of speed reduction.

As a major byproduct of the worse traffic conditions, our results highlight emerging energy consumption and emissions issues from the TNC sector. We have shown in our study that the increase in FHVs and the number of trips has led to 136% more NOx, 152% more CO and 157% more HC emissions per kilometer traveled by the FHV sector within two years. This finding is obtained under a conservative assumption where the duration and distance of each passenger trip stay the same. In reality, however, the revealed decrease in MA speed and increase in SA ratio are indicative of longer trip duration as well as longer cruising time before an FHV may reach the next passenger. We may speculate from this observation that the actual contribution of the TNC sector could be much higher than the reported values in this study. In this regard, immediate actions should be taken against the overgrowth of TNCs in urban areas. Based on our results, there are two practical directions that may help to mitigate the energy and emission issues. First, as a short-term measure, the entry of FHVs in heavily congested areas should be strictly regulated. We have shown that more FHVs contribute to not only slow-moving speed but also more congestion and stop-and-go traffic. The latter is the primary source of tailpipe emissions and regulating FHV service in congested areas helps to avoid the additive effect of more traffic and worse emissions per individual vehicle. But more importantly, considering a large number of trips served by the TNC sector, policies should be framed to encourage and facilitate the adoption of alternate fuel vehicles in the ride-hailing industry which can achieve significant long-term savings of the energy and emission costs.

We believe that the "failure" of TNCs in populated urban areas can be attributed to three primary reasons. One straightforward reason is the overgrowth of the number of TNCs that exceeds the already limited capacity of the urban road network. It is noted that the increase in the number of TNC drivers contributes differently to traffic congestion as compared to regular commuters. This can be reflected by much more frequent merges and diverges for picking up and dropping off passengers. And these introduce vital disturbances to regular traffic flow and result in more stop-and-go traffic. As a consequence, TNC vehicles not only add traffic, they also downgrade the capacity of the road network.

The second reason is due to the competitive nature of the TNC market. The market involves competition among different service provides and the traditional taxi sector, it also includes the competition among the drivers of the same TNC platform [38]. Such competition adds another layer of inefficiency if there is excessive supply than the actual demand, which is often the case during off-peak periods of passenger demand and corresponds to our analyses of the time periods with low average speed but a high number of active drivers. We should be aware that TNCs' prime time only accounts for approximately 6 hours (morning peak + evening peak) per day or 25% of the time daily. But for the rest of the day, there are more number of drivers competing for fewer number of passengers, resulting in excessive cruising miles and searching time. And we have pointed out in our analyses that TNC drivers will need to pay attention to their smartphones during cruising and such distracted driving is one notorious casual factor for traffic accidents.

Finally, we consider the lack of effective regulation and operation mode to be another reason. We assert that the observations that "TNC worsens urban traffic congestion and emissions" should not be viewed as contradicting the potential of TNCs for improving the efficiency and sustainability of our urban mobility. Indeed, several studies have pointed out that TNC could be a highly effective solution for efficient travel (e.g. 60% to 90% empty trips may be reduced if passengers and drivers are optimally matched [39]) and have validated the effectiveness of properly designed ridesharing mechanisms [40, 41]. But at present, there is no evidence showing how efficiently are TNC drivers and passengers being matched and the 'real' ridesharing which actually combines multiple single rides only accounts for a small amount of the total number of TNC trips [42]. Apparently the current TNC practice, which is primarily revenue driven, is still far from its optimal performance considering aspects of social benefits and overall sustainability. It is therefore necessary to frame regulations to strike the balance between the TNC's business model and social welfare. And the findings in our study provide important insights for evaluating the actual externalities from the TNC sector and will be valuable for decision and policymakers in framing effective regulations. As an example, NYC recently started the congestion surcharge for TNC and taxi trips entering Manhattan (south of 96th street) [43]. Our findings largely favor this regulation as the first step to mitigate the congestion impacts from the TNC sector, but also suggest the possibility for the surcharge to be varying spatially and temporally.

References

[1] Bruce Schaller. The new automobility: Lyft, uber and the future of american cities. 2018.

[2] Gregory D Erhardt, Sneha Roy, Drew Cooper, Bhargava Sana, Mei Chen, and Joe Castiglione. Do transportation network companies decrease or increase congestion? *Science advances*, 5(5):eaau2670, 2019.

[3] Maarit Moran and Philip Lasley. Legislating transportation network companies. *Transportation Research Record*, 2650(1):163–171, 2017.

[4] Susan Shaheen, Nelson Chan, Apaar Bansal, and Adam Cohen. Shared mobility: A sustainability & technologies workshop: definitions, industry developments, and early understanding. 2015.

[5] Eun Hye Grace CHOI. *The Effects of transportation network companies on traffic congestion*. PhD thesis, KDI School, 2017.

[6] NYC current and projected population., accessed June, 2019. Available online at `https://www1.nyc.gov/site/planning/planning-level/nyc-population/current-future-populations.page`.

[7] NYCDOT: major transportation projects, accessed Oct, 2019. Available online at `https://www1.nyc.gov/html/dot/html/about/major-transportation-proj.shtml`.

[8] New York Department of Motor Vehicles. Statistical data and summaries.

[9] New York City Taxi and Limousine Commission. Aggregated reports - data reports monthly indicators.

[10] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. Understanding mobility based on gps data. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 312–321. ACM, 2008.

[11] Yu Zheng, Yukun Chen, Quannan Li, Xing Xie, and Wei-Ying Ma. Understanding transportation modes based on gps data for web applications. *ACM Transactions on the Web (TWEB)*, 4(1):1, 2010.

[12] Jinjun Tang, Han Jiang, Zhibin Li, Meng Li, Fang Liu, and Yinhai Wang. A two-layer model for taxi customer searching behaviors using gps trajectory data. *IEEE Transactions on Intelligent Transportation Systems*, 17(11):3318–3324, 2016.

[13] Favyen Bastani, Yan Huang, Xing Xie, and Jason W Powell. A greener transportation mode: flexible routes discovery from gps trajectory data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 405–408. ACM, 2011.

[14] Zuchao Wang, Min Lu, Xiaoru Yuan, Junping Zhang, and Huub Van De Wetering. Visual traffic jam analysis based on trajectory data. *IEEE transactions on visualization and computer graphics*, 19(12):2159–2168, 2013.

[15] Jingbo Shang, Yu Zheng, Wenzhu Tong, Eric Chang, and Yong Yu. Inferring gas consumption and pollution emission of vehicles throughout a city. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1027–1036. ACM, 2014.

[16] Yiman Du, Jianping Wu, Senyan Yang, and Liutong Zhou. Predicting vehicle fuel consumption patterns using floating vehicle data. *Journal of Environmental Sciences*, 59:24–29, 2017.

[17] Conor K Gately, Lucy R Hutyra, Scott Peterson, and Ian Sue Wing. Urban emissions hotspots: Quantifying vehicle congestion and air pollution using mobile phone gps data. *Environmental pollution*, 229:496–504, 2017.

[18] Xiao Luo, Liang Dong, Yi Dou, Ning Zhang, Jingzheng Ren, Ye Li, Lu Sun, and Shengyong Yao. Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: a case of shanghai, china. *Journal of cleaner production*, 142:926–935, 2017.

[19] Kyoungho Ahn and Hesham Rakha. The effects of route choice decisions on vehicle energy consumption and emissions. *Transportation Research Part D: Transport and Environment*, 13(3):151–167, 2008.

[20] Mohammad Amin Pouresmaeili, Iman Aghayan, and Seyed Ali Taghizadeh. Development of mashhad driving cycle for passenger car to model vehicle exhaust emissions calibrated using on-board measurements. *Sustainable cities and society*, 36:12–20, 2018.

[21] Maryam Shekarrizfard, Ahmadreza Faghih-Imani, Louis-Francois Tétreault, Shamsunnahar Yasmin, Frederic Reynaud, Patrick Morency, Celine Plante, Louis Drouin, Audrey Smargiassi, Naveen Eluru, et al. Regional assessment of exposure to traffic-related air pollution: Impacts of individual mobility and transit investment scenarios. *Sustainable Cities and Society*, 29:68–76, 2017.

[22] Aarshabh Misra, Matthew J Roorda, and Heather L MacLean. An integrated modelling approach to estimate urban traffic emissions. *Atmospheric Environment*, 73:81–91, 2013.

[23] Jianlei Lang, Shuiyuan Cheng, Ying Zhou, Yonglin Zhang, and Gang Wang. Air pollutant emissions from on-road vehicles in china, 1999–2011. *Science of The Total Environment*, 496:1–10, 2014.

[24] Boski P Chauhan, GJ Joshi, and Purnima Parida. Driving cycle analysis to identify intersection influence zone for urban intersections under heterogeneous traffic condition. *Sustainable cities and society*, 41:180–185, 2018.

[25] Jean-Yves Favez, Martin Weilenmann, and Jan Stilli. Cold start extra emissions as a function of engine stop time: Evolution over the last 10 years. *Atmospheric Environment*, 43(5):996–1007, 2009.

[26] SM Ashrafur Rahman, HH Masjuki, MA Kalam, MJ Abedin, A Sanjid, and H Sajjad. Impact of idling on fuel consumption and exhaust emissions and available idle-reduction technologies for diesel vehicles–a review. *Energy Conversion and Management*, 74:171–182, 2013.

[27] Eric Jackson, Lisa Aultman-Hall, Britt A Holmén, and Jianhe Du. Evaluating the ability of global positioning system receivers to measure a real-world operating mode for emissions research. *Transportation research record*, 1941(1):43–50, 2005.

[28] Carolien Beckx, Luc Int Panis, Davy Janssens, and Geert Wets. Applying activity-travel data for the assessment of vehicle exhaust emissions: Application of a gps-enhanced data collection tool. *Transportation Research Part D: Transport and Environment*, 15(2):117–122, 2010.

[29] Zihan Kan, Luliang Tang, Mei-Po Kwan, and Xia Zhang. Estimating vehicle fuel consumption and emissions using gps big data. *International journal of environmental research and public health*, 15(4):566, 2018.

[30] Zihan Kan, Luliang Tang, Mei-Po Kwan, Chang Ren, Dong Liu, Tao Pei, Yu Liu, Min Deng, and Qingquan Li. Fine-grained analysis on fuel-consumption and emission from vehicles trace. *Journal of cleaner production*, 203:340–352, 2018.

[31] Taxi and ridehailing usage in New York City, accessed Sep, 2019. Available online at `https://toddwschneider.com/dashboards/nyc-taxi-ridehailing-uber-lyft-data/`.

[32] Le Chen, Alan Mislove, and Christo Wilson. Peeking beneath the hood of uber. In *Proceedings of the 2015 Internet Measurement Conference*, pages 495–508. ACM, 2015.

[33] Xinwu Qian, Dheeraj Kumar, Wenbo Zhang, and Satish Ukkusuri. Understanding the operational dynamics of mobility service providers: A case of Uber. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, 2019.

[34] TLC trip record data, accessed May, 2019. Available online at `https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page`.

[35] Rahmi Akçelik, Robin Smit, and Mark Besley. Calibrating fuel consumption and emission models for modern vehicles. In *IPENZ Transportation Group Conference*, 2012.

[36] Leon Ntziachristos and Zissis Samaras. Methodology for the calculation of exhaust emissions, 2018. Available online at `https://www.emisia.com/utilities/copert/documentation/`.

[37] NYC Department of Transportation. New York City mobility report, 2018.

[38] Xinwu Qian and Satish V Ukkusuri. Taxi market equilibrium with third-party hailing service. *Transportation Research Part B: Methodological*, 100:43–63, 2017.

[39] Xianyuan Zhan, Xinwu Qian, and Satish V Ukkusuri. A graph-based approach to measuring the efficiency of an urban taxi service system. *IEEE Transactions on Intelligent Transportation Systems*, 17(9):2479–2489, 2016.

[40] Paolo Santi, Giovanni Resta, Michael Szell, Stanislav Sobolevsky, Steven H Strogatz, and Carlo Ratti. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences*, 111(37):13290–13294, 2014.

[41] Xinwu Qian, Wenbo Zhang, Satish V Ukkusuri, and Chao Yang. Optimal assignment and incentive design in the taxi group ride problem. *Transportation Research Part B: Methodological*, 103:208–226, 2017.

[42] Uber says that 20% of its rides globally are now on UberPool, accessed June, 2019. Available online at `https://techcrunch.com/2016/05/10/uber-says-that-20-of-its-rides-globally-are-now-on-uber-pool/?ncid=rss`.

[43] Judge approves congestion pricing for New York City taxi, Uber and Lyft rides, accessed June, 2019. Available online at `https://abcnews.go.com/Business/judge-approves-congestion-pricing-york-city-taxi-uber/story?id=60778450`.