

### Investigating the Potential of Ridesharing to Reduce Vehicle Emissions

El-Khatib, Khalil; Jalali, Roozbeh; Li, Heng; Koohi-Fayegh, Seama; Hoorweg, Daniel

Veröffentlichungsversion / Published Version  
Zeitschriftenartikel / journal article

#### Empfohlene Zitierung / Suggested Citation:

El-Khatib, K., Jalali, R., Li, H., Koohi-Fayegh, S., & Hoorweg, D. (2017). Investigating the Potential of Ridesharing to Reduce Vehicle Emissions. *Urban Planning*, 2(2), 26-40. <https://doi.org/10.17645/up.v2i2.937>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:  
<https://creativecommons.org/licenses/by/4.0/deed.de>

#### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:  
<https://creativecommons.org/licenses/by/4.0>

Article

## Investigating the Potential of Ridesharing to Reduce Vehicle Emissions

Roozbeh Jalali <sup>1,\*</sup>, Seama Koochi-Fayegh <sup>2</sup>, Khalil El-Khatib <sup>1</sup>, Daniel Hoornweg <sup>2</sup> and Heng Li <sup>3</sup>

<sup>1</sup> Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, Ontario L1H 7K4, Canada; E-Mails: Roozbeh.Jalali@uoit.ca (R.J.), Khalil.El-khatib@uoit.ca (K.E.-K.)

<sup>2</sup> Faculty of Energy Systems and Nuclear Science, University of Ontario Institute of Technology, Oshawa, Ontario L1H 7K4, Canada; E-Mails: Sima.Kouhi@uoit.ca (S.K-F.), Daniel.Hoornweg@uoit.ca (D.H.)

<sup>3</sup> College of Computer Science and Electronic Engineering, Hunan University, 410082 Changsha, China; E-Mail: lih@hnu.edu.cn

\* Corresponding author

Submitted: 5 March 2017 | Accepted: 12 June 2017 | Published: 29 June 2017

### Abstract

As urban populations grow, cities need new strategies to maintain a good standard of living while enhancing services and infrastructure development. A key area for improving city operations and spatial layout is the transportation of people and goods. While conventional transportation systems (i.e., fossil fuel based) are struggling to serve mobility needs for growing populations, they also represent serious environmental threats. Alternative-fuel vehicles can reduce emissions that contribute to local air pollution and greenhouse gases as mobility needs grow. However, even if alternative-powered vehicles were widely employed, road congestion would still increase. This paper investigates ridesharing as a mobility option to reduce emissions (carbon, particulates and ozone) while accommodating growing transportation needs and reducing overall congestion. The potential of ridesharing to reduce carbon emissions from personal vehicles in Changsha, China, is examined by reviewing mobility patterns of approximately 8,900 privately-owned vehicles over two months. Big data analytics identify ridesharing potential among these drivers by grouping vehicles by their trajectory similarity. The approach includes five steps: data preprocessing, trip recognition, feature vector creation, similarity measurement and clustering. Potential reductions in vehicle emissions through ridesharing among a specific group of drivers are calculated and discussed. While the quantitative results of this analysis are specific to the population of Changsha, they provide useful insights for the potential of ridesharing to reduce vehicle emissions and the congestion expected to grow with mobility needs. Within the study area, ridesharing has the potential to reduce total kilometers driven by about 24% assuming a maximum distance between trips less than 10 kilometers, and schedule time less than 60 minutes. For a more conservative maximum trip distance of 2 kilometers and passenger schedule time of less than 40 minutes, the reductions in traveled kilometers could translate to the equivalent of approximately 4.0 tons CO<sub>2</sub> emission reductions daily.

### Keywords

emission reductions; ridesharing; spatiotemporal data mining; trajectory clustering; trajectory mining

### Issue

This article is part of the issue “Smart Cities—Infrastructure and Information”, edited by Soora Rasouli, Harry Timmermans and Dujuan Yang (Eindhoven University of Technology, The Netherlands).

© 2017 by the authors; licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

### 1. Introduction

Historically, transportation systems improved personal mobility and defined the spatial extent of cities. With advancements in transportation technologies, rising land costs, and growing urban populations, people now make

more and longer trips. However, this increased mobility also increases local air pollution, greenhouse gas (GHG) emissions, and congestion. Globally, the transport sector accounted for 17% of total CO<sub>2</sub> emissions in 2013: a share that is growing (European Commission, 2016). With increased mobility of growing urban populations, global

emissions from the transportation sector are likely to increase in lock step (Organization for Economic Cooperation and Development [OECD], 2011).

In China, the number of on-road vehicles increased dramatically over the past few decades. In response, the national government implemented stringent vehicle emission standards. Several studies have focused on evaluation of the emission factors of various vehicle types in Chinese cities over the past decade (Hu et al., 2012; Huo et al., 2012; Liu, He, Lents, Wang, & Tolvett, 2009; Wang, Westerdahl, Wu, Pan, & Zhang, 2011; Wang et al., 2012; Wu et al., 2012; Zhang, Wu, Liu et al., 2014; Zhang, Wu, Wu et al., 2014; Zheng et al., 2015; Zhou, Wu, Zhang, Fu, & Hao, 2014). They vary in several factors such as the vehicle type under study, urban structure, testing conditions and technologies, and the time of study, which correspond to the emission standards in place. Focusing on vehicle emissions, number of vehicles and emission standards, some studies also provide future trends of on-road vehicle emissions in China (Hao, Liu, Zhao, Li, & Hang, 2015; Wang, Fu, & Bi, 2011; Wu et al., 2017; Zhang, Wu, Wu et al., 2014).

Reduction of GHG emissions from conventionally-fueled vehicles is the main impetus behind alternative fuel vehicles such as hydrogen and electric vehicles, as well as those running on biofuels. Development of more fuel-efficient vehicles and methods to reduce transit delay also reduce air pollution and GHG emissions. However, even if alternative fuel vehicles meet the growing mobility needs and emission reduction targets, the increased number of vehicles could still lead to increased travel time, congestion, particle pollution, and vehicle noise (not for electric vehicles). Such issues could reach a level where only changes in transportation systems could accommodate the growth in mobility demand.

## 2. Recent Findings on Ridesharing Potential

Ridesharing is often defined as the sharing of vehicles by commuters who share common routes and trip schedules to reduce the overall number of trips and travelled distance. In general, the shared vehicles include personal cars, vans, taxis and shuttles, and the shared routes include rides to work and common household trips. In this article, ridesharing encompasses the use of personal vehicles for common routes and schedules. Ridesharing is a potential option to increase mobility while maintaining or reducing vehicle emissions by increasing the effective use (efficiency) of existing transportation resources. There are several options to support ridesharing such as designated carpooling lanes, and web-based applications to connect drivers. Higher vehicle use corresponds to fewer circulating vehicles, increased efficiency of urban traffic, less congestion, reduced local air pollution, and lower overall GHG emissions. Several studies focused on the effectiveness of ridesharing in managing congestion. While most studies identify significant benefits, the potential for rideshar-

ing varies significantly. Alexander and González (2015) suggest a 43% decrease in the number of vehicles in the Boston area with adoption of ridesharing among drivers. They also found that a 14% increase in the number of vehicles would occur if only non-drivers (e.g., transit riders) were to adopt ridesharing. Cici, Markopoulou, Frias-Martinez and Laoutaris (2013, 2014) showed that ridesharing could provide more than 70% reduction in the number of cars in Madrid. Bicocchi and Mamei (2014) showed that the number of trips could decline by over 40% if users within a 1-kilometer distance shared rides in Italy. Goel, Kulik and Ramamohanarao (2016) examined vehicle reductions in Melbourne when passengers are picked-up and dropped-off at predetermined stops. Their model suggested a 23–40% reduction in vehicle kilometers depending on the strategies used in selection of the stops. He, Hwang and Li (2014) found that increasing the number of riders to eight per vehicle, by using a mini van, would increase overall travel savings up to 60%. In an investigation for taxi ridesharing in New York City, Santi et al. (2014) concluded that, with waiting time not exceeding 5 minutes, ridesharing with two or three passengers could reduce total taxi trips by 50% (reaching full potential in trip reduction) and 60%, respectively. This equates to about a 40% reduction in total taxi trip length. Another investigation for taxi ridesharing in New York City (Ota, Vo, Silva, & Freire, 2015) identified 46% and 61% savings, respectively, in taxi trips if rides are shared among two and three passengers with nearby trips within 1.6 kilometers. Using the same algorithm for analysis of taxi ridesharing potential proposed by Santi et al. (2014), Tachet et al. (2017) showed that ridesharing benefits follow the same trends in San Francisco, Vienna and Singapore with total number of taxi trips reduced by 50% for San Francisco and Singapore, and by 42% for Vienna.

Conclusions drawn from past studies on the benefits of ridesharing are affected by the specific land use characteristics of the city under investigation (Kim, Rasouli, & Timmermans, 2017) and pricing preferences (Yang & Timmermans, 2017). Alexander and González (2015) expected ridesharing efficiency to decrease for cities with heterogeneous trip patterns, such as those with multiple major employment centers or with limited residential development. Conversely, simulation results of Tsao and Lin (1999) and Cici et al. (2014) showed that cities with uniform home and work locations provide little potential for ridesharing. Tachet et al. (2017) showed that the potential of ridesharing follows the same trends in New York City and three other major world cities, which differ greatly in traffic characteristics associated with population size and urban extent.

## 3. Factors Affecting Evaluation of Ridesharing Potential

In the analysis of ridesharing potential using driver mobility data, it is crucial to define what data is measured, how it is measured and how the data is analyzed. These parameters can greatly impact findings and are often the

reason behind varied results in various studies. They are reviewed in the next two subsections.

### 3.1. Vehicle Trip Data Set

In order to analyze ridesharing potential to reduce the overall demand for personal vehicles, large-scale data on vehicle mobility patterns in a city is needed. This could include recorded location and time of day for all vehicles for a given time period (e.g., a day). The datasets used in ridesharing models vary depending on the following:

- **Granularity of data** (spatial and temporal). This often depends on the tools used to collect data (e.g., cellphone [Alexander & González, 2015; Cici et al., 2014], GPS systems [He, Hwang, & Li, 2014; Tachet et al., 2017; Trasarti, Pinelli, Nanni, & Gianotti, 2011], surveys [Ghoseiri, Haghani, & Hamed, 2011] and social networking tools [Cici et al., 2014]). In general, cellphone datasets often in the form of Call Detail Records (CDRs) have less granular information in terms of user trajectories since they often record user information when users make calls or send text messages. For the purpose of big data collection on user mobility for ridesharing analysis, cellular data can be collected from network companies. Accuracy of such cellular data, often not specifically designed to indicate accurate location by using cellphone applications, is limited by the density of existing cellular towers in the area of user movements. Cellular telephone towers could in some cases cover a large area (up to several square kilometers) in rural areas resulting in less accurate data. GPS data, on the other hand, rely on satellites and provide more accurate descriptions of user movements. Collection of cellular data from a larger number of users can provide data with acceptable accuracy and comparable to GPS-collected data (Cici et al., 2014). Data from online networks are also unable to reach high granularity, as they can only be collected when users post a geotagged message in a social network.
- **Dataset size.** This corresponds to the number of recorded trips over a period of time that alternatively affects the potential of ridesharing. Santi et al. (2014) studied how the number of shareable trips in a given day varies as a function of the total number of recorded trips. In their study, the average number of daily-recorded trips in New York is around 400,000 and they showed that at approximately 100,000 trips, taxi ridesharing potential reaches its maximum theoretical value.

### 3.2. Data Analysis to Model Ridesharing

Once data on user mobility patterns are collected, extraction of suitable information and analysis to identify potential shared rides is a complex process consisting

of several stages and dependent on several factors. Potential ridesharing opportunities are often presented as the fraction of individual trips that can be shared, sometimes called *shareability* (Tachet et al., 2017). Agatz, Erera, Savelsbergh and Wang (2012) highlighted many of the optimization challenges that arise when developing technology to support ridesharing and reviewed the relevant operations research models in this area.

#### 3.2.1. Spatial and Temporal Constraints

Findings of user trip compatibility analyses are directly affected by the maximum allowed extra distance for each trip as a result of ridesharing and the spatial (i.e., ride potential within a certain distance) and temporal (e.g., pick up and drop off within a time frame) constraints. For example, Cici et al. (2014) found that traffic in the city of Madrid could be reduced by 59% if users are willing to share a ride with people who live and work within 1 kilometer. However, once a pick-up and drop-off delay of up to 10 minutes is placed on the model, this potential benefit drops to 24%. Santi et al. (2014) used a delay time of up to 5 minutes while Ota et al. (2015) used extra distance traveled for recognition of nearby potential rides. He et al. (2014) showed that excessive detouring (i.e., larger than 4 kilometers) reduces ridesharing efficiency to less than 5%.

#### 3.2.2. Number of Users Allowed to Share Rides

Some studies investigate the effect of the maximum number of rides to be shared on the ridesharing potential. He et al. (2014) and Ota et al. (2015) found that as the limit on the number of shared rides increases, *shareability* potential also increases. Results of simulations by Santi et al. (2014), however, showed the number of saved taxi trips is increased from about 50% (maximum theoretical potential) with two shared rides to only about 60% with three (below the 66.7% maximum theoretical potential) suggesting that the benefits of ridesharing do not increase linearly with the number of shared rides. It must be noted that an increase in the number of allowed shared rides is expected to increase extra travel distance and number of extra stops for each trip, two parameters that are often set to limited values in the models. Increasing the number of allowed shared rides would likely be ineffective in increasing *shareability* potential if these parameters are strictly kept at relatively low values. Ota et al. (2015) found that for three shared trips, the total saving in the total distance through ridesharing is 29% on average with the average extra distance of 0.92 kilometers, while for two shared trips the saving is 18.2% with the average extra distance of 0.56 kilometers.

#### 3.2.3. Trip Matching Algorithms: En-Route versus Origin-Destination Ridesharing

Another factor affecting the findings are the trip matching algorithms used in the analysis, and the ability of the

model to capture en-route ridesharing (i.e., ride potential along trips). Studies that analyze user spatial and temporal compatibility based on trip start- and end-points are often not capable of modelling such potential and are expected to report a lower potential for ridesharing. Cici et al. (2014) performed both types of algorithms and found that ridesharing potential increases from 24% to 53% if en-route ridesharing opportunities are modeled as well. Biccocchi and Mamei (2014) also presented a methodology, based on the extraction of suitable information from mobility traces, to identify rides along the same trajectories.

### 3.2.4. Dynamic versus Static Ridesharing

In some models, it is assumed that trips are known in advance, which makes them suitable for carpooling applications but debatable for taxi ridesharing applications where opportunities are computed in real time. Taxi ridesharing requests arrive in real time and the algorithms used in evaluating such potentials need to run large-scale studies that explore a wide range of scenarios through parameter sweeps. This often takes considerable computation time and although many algorithms are capable of evaluating ridesharing potential among users, some are not able to evaluate such potential under the time constraints typically present in applications used for connecting users. Thus, the time constraints affect the calculated potential by the algorithms. In order to model the time-sensitivity of ridesharing potential, a time window is often used in the algorithms, outside of which ridesharing potential is not considered practical in real-time situations (Maciejewski et al., 2016; Shen et al., 2016). Therefore, potential of ridesharing is generally found to be lower in studies that account for this factor. For example, Cici et al. (2013, 2014) showed that a time window of 10–30 minutes results in 10–20% reduction in the number of cars in real-time situations if a delay time of 10 minutes and a detour distance of one kilometer are accepted by users. Without this time restriction, they show a higher ridesharing potential of up to 60%.

### 3.2.5. Techniques Used in Ridesharing Data Analysis

In ridesharing analysis, optimization methods substantially increase the likelihood that ridesharing matches can be found for participants, and lead to ridesharing models that generate larger overall system savings. Agatz, Erera, Savelsbergh and Wang (2011) simulate ridesharing potential (e.g., miles saved) for various optimization objectives such as rider travel time and cost.

Santi et al. (2014) proposed a graph-based approach that is capable of spotting opportunities for en-route ridesharing. The algorithm computes optimal sharing strategies for taxi trips in New York City considering two parameters: the maximum number of trips that can be shared and the minimum time to accommodate all trips.

He et al. (2014) proposed a carpooling system that generates an efficient route for dynamic ridesharing using a GPS-assisted trajectory mining scheme to identify frequent routes taken by participating riders, including private car, taxi, bus, subway and walking. The routing optimization goal is to minimize the driving distance, commute costs, detour distance, social distance and advance time to start the carpool.

Ma, Zheng and Wolfson (2013), Ota et al. (2015) and Ota, Vo, Silva and Freire (2016) proposed a framework that supports the simulation of real-time taxi ridesharing scenarios. Ma et al. (2013) split a region into grid cells such that the distance between any two locations can be computed “heuristically” as the distance between the cells containing them. This allows their system to keep shortest path computations at a minimum. Ota et al. (2015) used a shortest path indexing scheme, where they made use of cache-coherent layout to speed up shortest path queries substantially, and presented a framework that supports the simulation of real-time taxi ridesharing scenarios.

Alexander and González (2015) extracted average daily origin-destination trips from the dataset and matched spatially and temporally similar trips. They evaluated the impacts of congestion network-wide for several adoption scenarios including adoption of ridesharing by non-drivers.

### 3.3. Factors Affecting Adoption of Ridesharing

In the analysis of ridesharing potential, indirect factors affecting adoption of ridesharing, such as passenger safety (i.e., riding with strangers) and privacy (i.e., disclosure of home and work address) are sometimes accounted for. Some studies focused on characterizing crowd mobility and activity patterns using information from social networks (Fujisaka, Lee, & Sumiya, 2010; Noulas, Mascolo, & Frias-Martinez, 2013; Noulas, Scellato, Mascolo, & Pontil, 2011; Wakamiya, Lee, & Sumiya, 2011). Cici et al. (2014) used online social network data to apply social constraints in the analysis of the data for matching drivers (e.g., ridesharing among people who know each other). They found that if users are willing to ride with friends of friends, the potential reduction is up to 31%, but if they are willing to ride only with people they know, the potential of ridesharing becomes negligible. Fixed pick up/drop off locations equipped with video surveillance could improve riders’ safety and protect their privacy (Goel et al., 2016). Such fixed locations can be selected to maximize vehicle occupancy.

## 4. Objective: Estimation of Emission Reductions as a Result of Ridesharing

In most previous studies, ridesharing potential in improving congestion is investigated. Some studies focused on presenting algorithms that are suitable for real-time ridesharing requests, often used for connecting taxi

users. While the studies present the results in the form of ridesharing efficiency or the number of trips saved, they do not provide an approximation of reduced pollutant and/or GHG emissions resulting from the trip savings. Ota et al. (2015) analyzed the savings in CO<sub>2</sub> emissions, but they do not explicitly present their results on CO<sub>2</sub> emissions in their article. In addition, studies that focus on the emissions of vehicles in China often focus on how fuel and engine improvements that comply with the emission standards result in emission reductions. They do not focus on indirect strategies such as increasing vehicle occupancy averages that can reduce overall emissions. In the current study, the potential of ridesharing to reduce pollutant and GHG emissions is investigated. Trip GPS data of approximately 9,000 privately-owned vehicles in Changsha, China, is used. Ridesharing potential is identified based on trip origin and destination. The savings on trip distance as a result of ridesharing is used to provide estimates of pollutant and greenhouse gas emission reductions. The findings suggest the potential of ridesharing to improve local air quality and mitigate GHG emissions.

**5. Methods**

This section introduces a proposed data-driven model that enables the analysis of historical location data to investigate the potential of ridesharing. There are several challenges related to this research, including removal of outliers, noise and false data, investigation of the reliability of data, detection of misrepresented information in terms of location, feature selection, and clustering the data which significantly affect the findings. Figure 1 illustrates the data flow diagram to analyze ridesharing potential that consists of three steps of data processing in-

cluding pre-processing, similarity detection and ridesharing recommendations.

In this study, the vehicles' geographical locations (latitude and longitude) were collected using GPS monitoring systems installed in 8,900 privately-owned vehicles in Changsha, China (population of 7 million). The historical data is processed to determine possible similar rides that could be shared. The potential number of saved kilometers by adopting ridesharing is calculated.

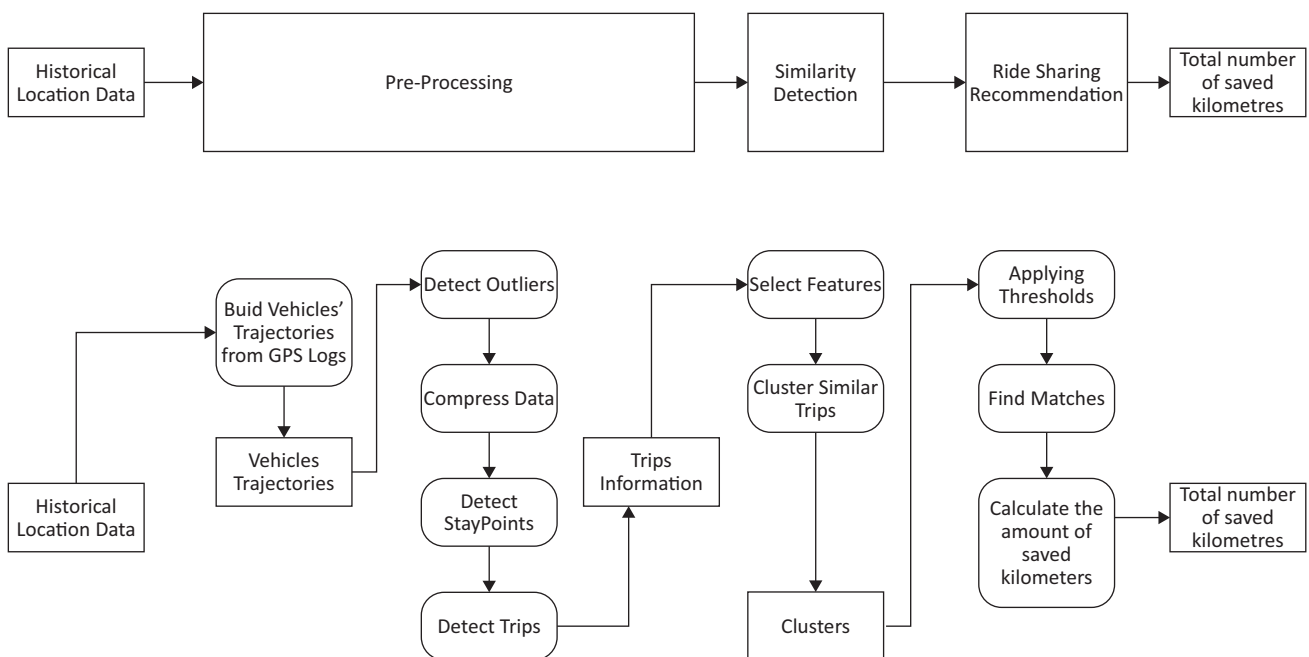
It should be noted that ridesharing in the current analysis is short-distance, static (see Section 3.2.4) and is on a daily basis. It is also assumed that wherever matching trips exist, one car corresponding to the longest trip is selected as the one that provides a ride to others, and is the one setting the origin and destination of the shared trip. Passengers of the cars corresponding to the other trips (i.e., riders) are expected to walk the last part of their trip (also called the last mile) from the driver's destination to theirs.

**5.1. Pre-Processing**

**5.1.1. Trajectory Representation and Location History Modeling**

As depicted in Figure 1, spatial-temporal trajectories are first built from the GPS logs. The data is retrieved from the database for each vehicle and transformed to a series of chronologically ordered points for example,  $P_1 \rightarrow P_2 \rightarrow P_3 \rightarrow \dots \rightarrow P_n$ . Each trajectory point consists of timestamp, geospatial coordinates (latitude, longitude) and the speed of the vehicle.

Data pre-processing is a crucial step as data collection is often loosely controlled, resulting in outliers, noise, and missing information. Thus, to reduce the complex-



**Figure 1.** Data flow diagram for ridesharing.

ity of data analysis and program execution time, the following data pre-processing and representation steps were applied.

### 5.1.2. Noise Filtering and Outlier Detection

The first step in data pre-processing that looks for abnormalities in trajectories is noise filtering and outlier detection. Outliers in trajectories can be a point or series of points that are significantly different from other points. For instance, an outlier can be a point that is far from other points and out of vehicle possible reach within regulated speed and time. An outlier can also be a point observation that does not conform to the expected pattern. In this paper, we used mean filter (Huang, Yang, & Tang, 1979) to detect the noise and outlier. For point  $P_z$ , in a vehicle's trajectories, a true value is the mean of the position of  $P_z$  and the  $n - 1$  predecessor, thus, the mean filter can be a sliding window covering the  $n$  adjacent values of  $P_z$ :

$$\frac{1}{n} \sum_{i=(z-n+1)}^z P_i$$

where  $n$  is the size of sliding window for the mean filter.

### 5.1.3. Compression

While vehicle locations can be constantly sampled and communicated, a high rate of sampling can result in excessive communication overhead, computing and data storage. It is also important to consider that when a vehicle is waiting at a traffic light, or delayed in congestion, its location does not change for a while but sampled continuously. To decrease the amount of data and improve the performance of data processing, the points from trajectories for which there is no updated information are removed.

### 5.1.4. Stay Point Detection

An important part of the analysis is to detect stay points because they can be used in trajectory segmentation and trip detection. Stay points denote the locations where vehicles stay for more than 5 minutes, such as parking lots. There are two different types of stay point: First, single point location where a vehicle remains stationary, and second, when a vehicle location is updated but there is no notable change on a vehicle location. In this study, both types of stay point are detected.

### 5.1.5. Trip Detection

To group similar trips, one needs first to divide a trajectory into different trips. Segmenting trajectories to trips helps to reduce computation cost, delve deeper into vehicle trajectories and find more potential ridesharing options. In this paper, trips are detected based on time in-

terval and stay points. For example, if the time interval between two consecutive points in a vehicle trajectory is larger than a defined threshold, the vehicle trajectory can be divided to two trips. Also, stay points can divide a trajectory into two different segments or trips.

## 5.2. Similarity Detection

The main purpose of our analysis is to detect the similar rides and mark them for potential ridesharing. In this step, clustering detects similar trips and groups them together.

### 5.2.1. Feature Selection

As different trips contain different properties such as length, number of points, and sampling rate, it is difficult to use trip properties for clustering. To solve this issue, one can select useful features from each trip and present them in a uniform way. In this paper, the start time, end time, origin, destination and length of each trip are used to describe the features for each trip and are represented as a vector.

### 5.2.2. Clustering

Clustering in this analysis is the process of grouping similar trips. The trips inside a group are more similar than other trips that are placed in other groups or clusters. The distance between the trips is measured by distance between vectors. Clustering tries to minimize the distance between the trips inside of each cluster and maximize the distance between trips outside of each cluster.

One of the most commonly used algorithms for clustering is the  $k$ -means (Hartigan & Wong, 1979).  $k$ -means is an iterative clustering algorithm that partitions  $n$  observations into a number of clusters ( $k$ ) that is selected before the algorithm starts. In this study,  $K$ -means is used for grouping similar trips.  $K$ -means chooses  $k$  initial cluster centers randomly and calculates the distance of the centroid in each cluster to all the trips and then assigns each trip to the group with closest centroid. After that,  $K$ -means calculates the average distance between trips inside of each cluster and the cluster centroid to find the new centroid.  $K$ -means repeats these steps until the cluster members do not change.

For measuring the similarity between trips and their centroids, we employed multiple similarity functions such as Euclidean, Cosine, City block and Correlation (Deza & Deza, 2009). For each of these functions, we calculated the distance based on the following equations:

$$\text{Euclidean: } d(x, c) = \sqrt{\sum_{i=1}^p (x_i - c_i)^2}$$

$$\text{City block: } d(x, c) = \sum_{i=1}^p |x_i - c_i|$$

$$\text{Cosine: } d(x, c) = 1 - \frac{xc'}{\sqrt{(xx')(cc')}}$$

$$\text{Correlation: } d(x, c) = 1 - \frac{(x - \bar{x})(c - \bar{c})'}{\sqrt{(x - \bar{x})(x - \bar{x})'(c - \bar{c})(c - \bar{c})'}}$$

$$\text{where } \bar{x} = \frac{1}{p} \left( \sum_{j=1}^p x_j \right) \bar{1}_p \text{ and } \bar{c} = \frac{1}{p} \left( \sum_{j=1}^p c_j \right) \bar{1}_p$$

where  $p$  is the dimension,  $x$  is an observation or feature vector for a trip,  $c$  is a centroid and  $\bar{1}_p$  is a row vector of  $p$  ones.

### 5.3. Ridesharing Recommendation

Clustering partitions similar trips into groups but it does not guarantee that all the trips inside each group have the potential for ridesharing. There are still limitations for ridesharing such as the maximum distance between the trip start and end points, the maximum user schedule time, the maximum number of passengers who can share the ride, or the minimum length for which two users prefer to travel together. In this step, such thresholds are considered for each cluster and the potential trips that could be shared are estimated.

## 6. Experimental Analysis

In this section, the performance of our approach is demonstrated using GPS location records of 8,900 privately-owned vehicles in Changsha. In the experiments, the effect of different similarity functions along with different number of clusters on the clustering algorithm are examined to find the best option for ridesharing. We also examined the effect of maximum schedule time and maximum distance between the trip start and end points. The results show that Euclidian similarity function with 11,000 clusters achieves the best performance and there is no notable change on the total saved kilometers if we increase the maximum schedule time to more than one hour and the maximum distance between the trip start points and endpoints to more than 6 kilometers.

### 6.1. Experimental Setup

The historical data of every vehicle was sampled every 10 minutes and stored in a database. Thus, the historical dataset that we studied was also sampled every 10 minutes totaling 65,940,000 records spanning 89 days from February to April 2013. In an ideal situation, each vehicle creates 144 records per day resulting in 114,062,400 for 8,900 vehicles for 89 days but our monitoring system did not collect the data from vehicles that remained stationary for more than 12 hours. Also, there is typically data loss which can be attributed to a variety of reasons. For example, monitoring data was wirelessly communicated to the monitoring platform using cellular GPRS networks which is error-prone due to the nature

of the wireless channel that introduces data loss, delay, and retransmissions.

The experiments ran on a server with Intel 6 cores Xeon E5649 2.53GHz processor, 32 GB RAM and Windows server 2016 operating system running MATLAB R2016b. MATLAB is used as the programming environment for the experiments. We also used MATLAB parallel computing toolbox to get maximum benefit from multiple cores inside the server processor. The toolbox enabled the use of the full power of multicores by executing our program on multiple threads.

### 6.2. Ridesharing in 24 Hours: Case Studies

To demonstrate the performance of the approach, the first day (24 hours) of the dataset which contains 1,080,224 records was selected. This is a weekday including travel typical of all weekdays travels. The total traveled distance on this day is 201,890 kilometers, and the total number of detected trips is 20,018 resulting in an average trip length of 10.53 kilometers. Figure 2 shows the total hourly travel distance driven by the vehicles for 24 hours on the first day of the dataset, and Figure 3 shows the trip start points for 24 hours on an actual map. When rides are shared, we assumed that the maximum capacity of each vehicle, including the driver of the vehicle, is 4 passengers. In addition, we assumed that sharing rides that are shorter than 2 kilometers result in excessive detouring and provide negligible benefits in terms of reductions in overall trip kilometers and therefore, trip data corresponding to such trips were excluded from the experiment.

#### 6.2.1. Effect of the Similarity Function on Ridesharing

##### 6.2.1.1. Case Study 1

In the first case study, the effect of different similarity functions and maximum schedule time on ridesharing potential (indicated in this article as the kilometers and the number of trips that are reduced) were evaluated. Table 1 shows the values assigned for the simulation setup parameters for the first case study. We assumed that the number of clusters is constant and equal to 8,000 clusters. To match trips with ridesharing potential, the maximum time that passengers can wait to get a ride (referred to schedule time in this article) and the maximum allowable distance between trip origins and destinations (also referred to as trip distance in this article) are set. In this case study, the maximum schedule time is varied between 5 and 180 minutes and the maximum distance between trips is set to 2 kilometers. It is found that the Euclidean and City block similarity functions result in the highest values of total saved kilometers (Figure 4a) and total number of saved trips (Figure 4b) if the maximum schedule time is less than an hour. If the maximum schedule time is higher than 60 minutes, the City block similarity function indicates higher values in total



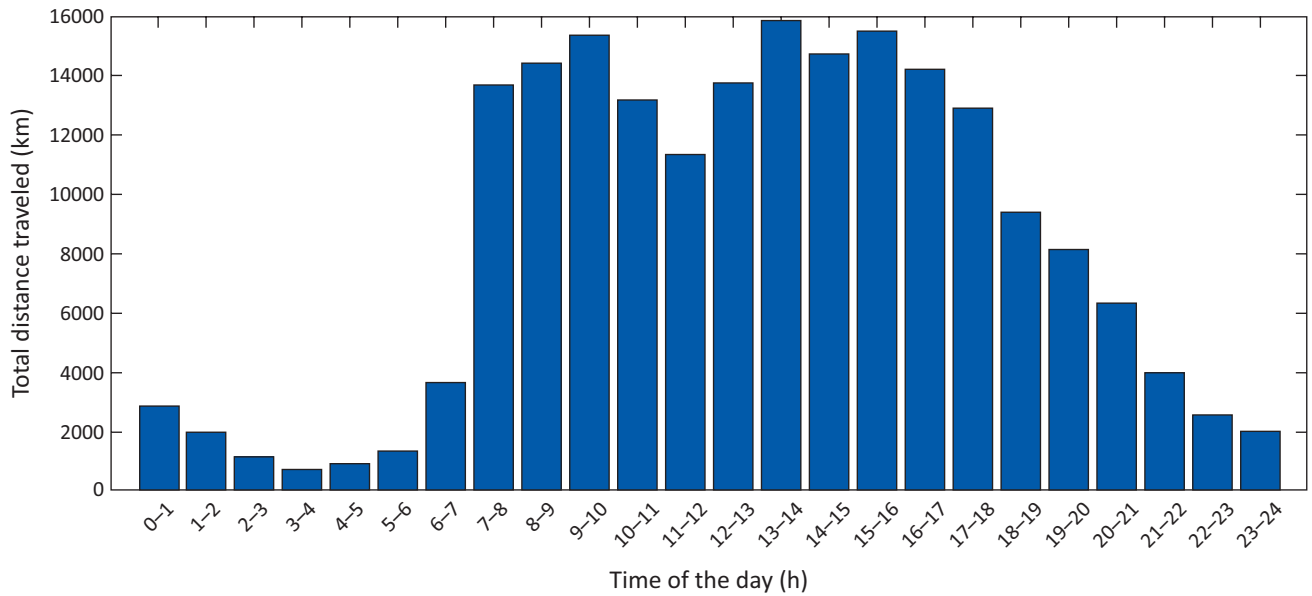


Figure 2. Total hourly distance driven by vehicles for 24 hours.

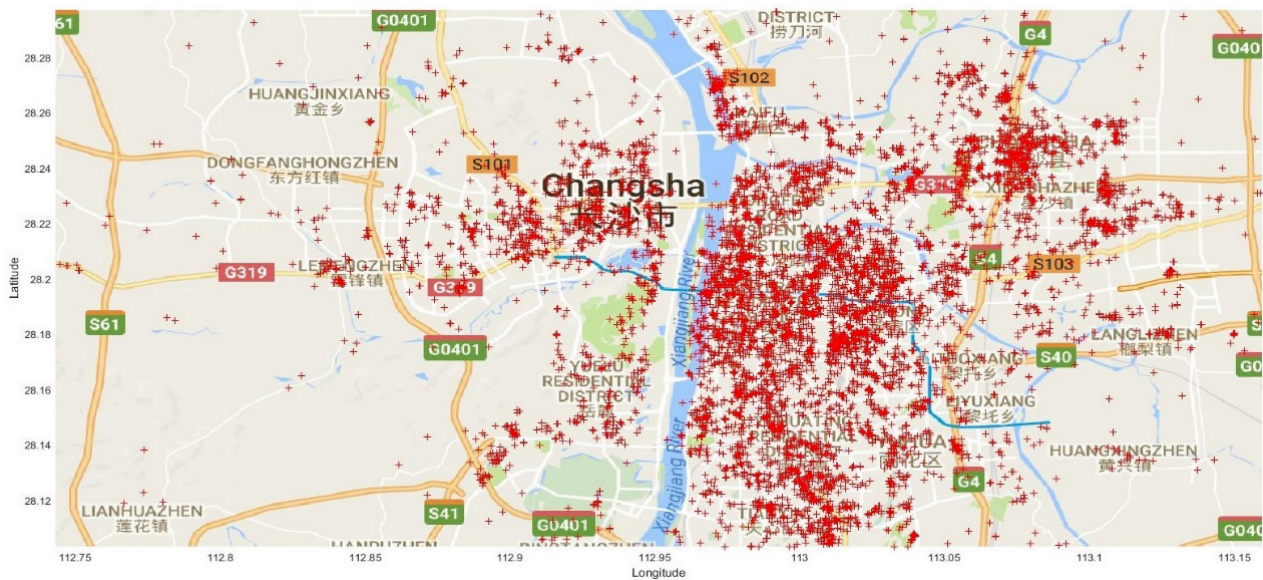


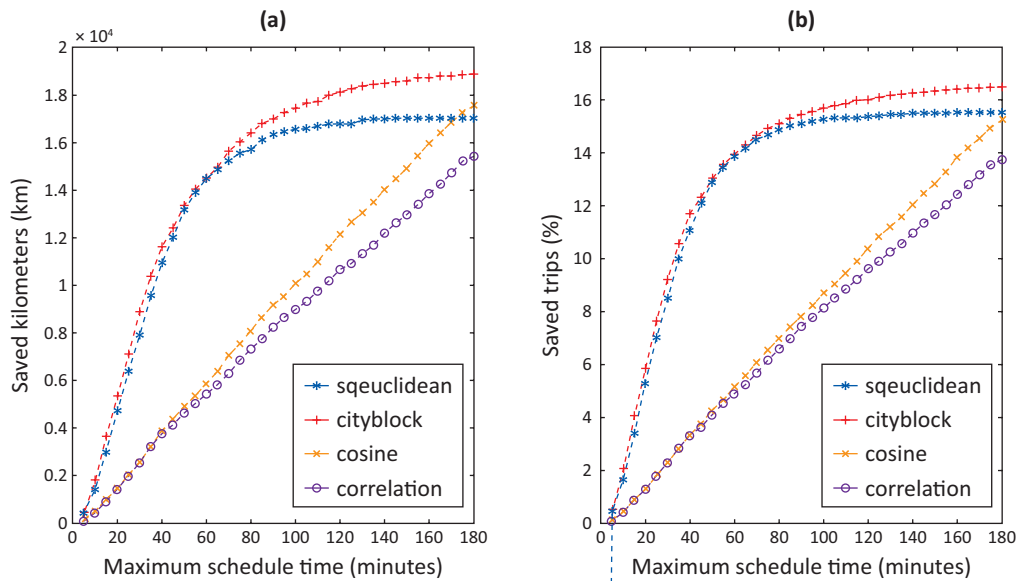
Figure 3. Trips start points (red dots) for 24 hours.

Table 1. Simulation set up parameters (study 1).

Description	Value
Similarity Function	Variable
Maximum distance between trips (Kilometers)	2
Number of clusters	8000
Maximum schedule time (Minutes)	5-180
Total trip length (Kilometers)	210,890
Total number of trips	20,018

saved kilometers and saved number of trips than other functions. Euclidean and City block have better results in terms of saved kilometers because these two similarity functions act better on the data that can be represented as points in a Euclidean space. The cosine similarity mea-

sures the angle between two vectors and while it is a suitable candidate for multi-feature vectors, it did not perform well for the small number of features' vectors. The correlation similarity function is also only suitable for high-dimensional data which is not the case in this study.



**Figure 4.** Effect of similarity function and maximum schedule time on (a) saved kilometers (b) number of saved trips.

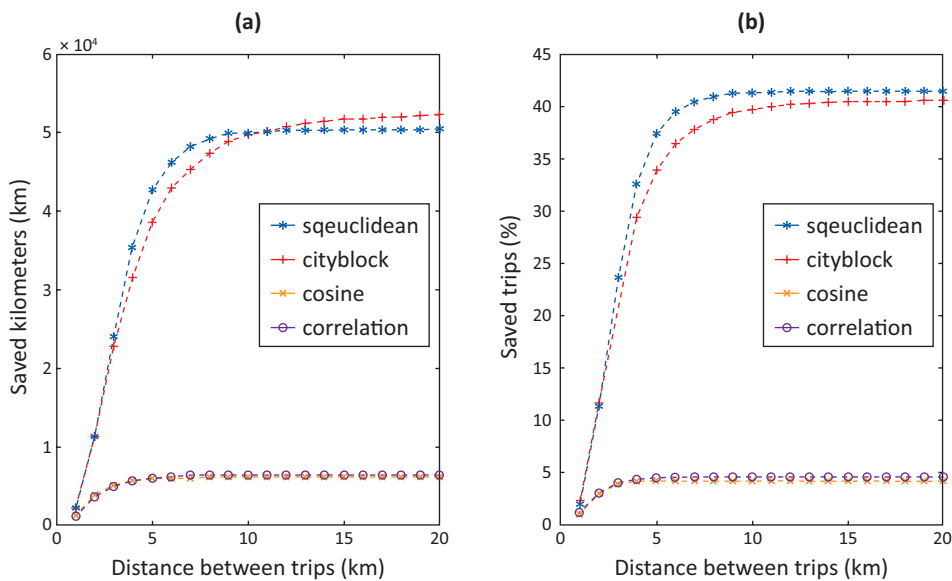
6.2.1.2. Case Study 2

In this case study, the effect of the similarity function and the maximum distance between trips on ridesharing potential are evaluated. Table 2 shows the values assigned for the set-up parameters for the second case study. We assumed that the number of clusters is a constant and

equal to 8,000 clusters. The maximum schedule time is set to 40 minutes, and the maximum distance between trips is a variable between 1 and 20 kilometers. It is found that the Euclidean and city block similarity functions result in higher values of total saved kilometers (Figure 5a) and number of saved trips (Figure 5b) compared to the cosine and correlation functions. As one can see in Fig-

**Table 2.** Simulation set up parameters (case study 2).

Description	Value
Similarity Function	Variable
Maximum distance between trips (Kilometers)	1–20
Number of clusters	8,000
Maximum schedule time (Minutes)	40
Total trip length (Kilometers)	210,890
Total number of trips	20,018



**Figure 5.** Effect of similarity function and distance between trips on (a) saved kilometers (b) number of saved trips.

ure 5, there is no improvement in ridesharing potential if the distance between the trips is more than 6 kilometers. The reason behind this is the decrease in the similarity among trips when the distance among them is increased. Ultimately, when the distance is more than 6 kilometers, there is no similar trip available for matching inside each cluster.

6.2.1.3. Case Study 3

In the third case study, the effect of the similarity function and the number of clusters on ridesharing potential was investigated. Table 3 shows the values assigned for the set-up parameters for the third case study. We assumed that the number of clusters is a variable between 1,000 to 15,000. The maximum schedule time is kept to 40 minutes, and the maximum distance between trips is kept to 3 kilometers. The highest values of saved kilometers and total number of saved trips are achieved with the Euclidean similarity function when the number of clusters is approximately 11,000 (Figure 6). As Figure 6 depicts, increasing the number of clusters to more than 11,000 does not increase the total number of saved kilometers. This can be explained by the decrease in the number of similar trips inside of each cluster as the number of clusters are increased.

6.2.2. Case Study 4: Effect of the Number of Clusters and Schedule Time on Ridesharing

In this case study, we looked at the effect of changing the number of clusters and schedule time on ridesharing potential. Table 4 shows the values assigned for the set-up parameters for the fourth case study. We assumed that the number of clusters is a variable between 1,000 and 15,000, the maximum schedule time is a variable between 5 and 180 minutes, and the maximum distance between trips is a constant, equal to 3 kilometers. The largest reduction in traveled kilometers is achieved with 11,000 clusters if the maximum schedule time is less than an hour. (Figure 7).

6.2.3. Case Study 5: Effect of Maximum Trip Distance and Schedule Time on Ridesharing

In Case study 5, the effect of trip distance and schedule time on ridesharing potential is investigated. Table 5, shows the set-up parameters for this case. As we determined in the previous case studies, the highest values of saved kilometers are achieved using the Euclidean similarity function with 11,000 clusters. In this case study, we kept the number of clusters at 11,000 and used Euclidean distance for the similarity function. The results show that

Table 3. Simulation set up parameters (case study 3)

Description	Value
Similarity Function	Variable
Maximum distance between trips (Kilometers)	3
Number of clusters	1000–15,000
Maximum schedule time (Minutes)	40
Total trip length (Kilometers)	210,890
Total number of trips	20,018

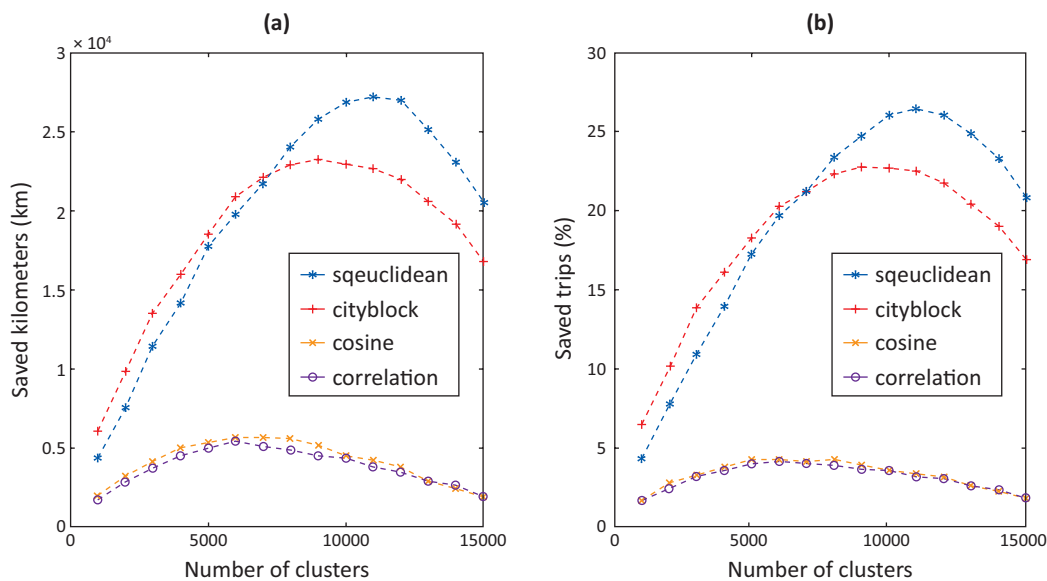
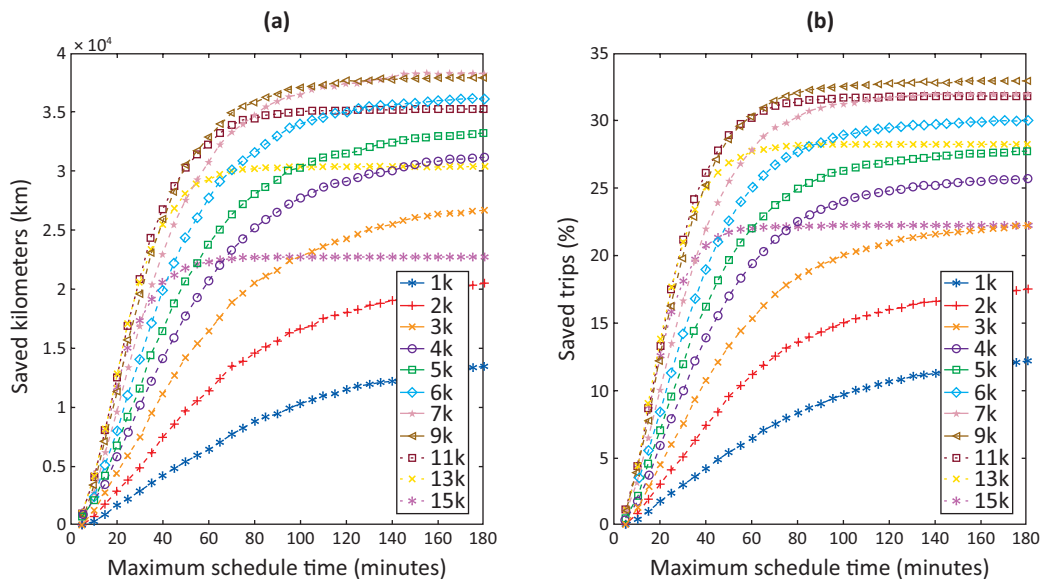


Figure 6. Effect of similarity function and number of clusters on (a) saved kilometers (b) number of saved trips.

**Table 4.** Simulation set up parameters (case study 4).

Description	Value
Similarity Function	Euclidean
Maximum distance between trips (Kilometers)	3
Number of clusters	1,000–15,000
Maximum schedule time (Minutes)	5–180
Total trip length (Kilometers)	210,890
Total number of trips	20,018



**Figure 7.** Effect of number of clusters and maximum schedule time on (a) saved kilometers (b) number of saved trips.

**Table 5.** Simulation set up parameters (case study 5).

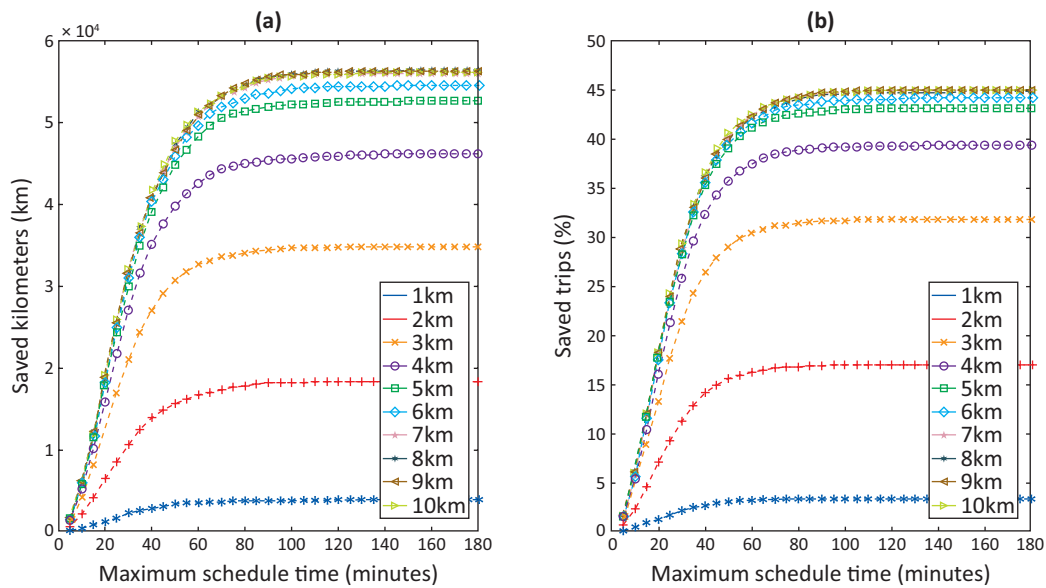
Description	Value
Similarity Function	Euclidean
Maximum distance between trips (Kilometers)	1–10
Number of clusters	11,000
Maximum schedule time (Minutes)	5–180
Total trip length (Kilometers)	210,890
Total number of trips	20,018

we can save more than 15% on total travel distance (Figure 8a) and more than 30% on the number of trips (Figure 8b) if the maximum distance between trips is 3 kilometers and the maximum schedule time is 45 minutes. It is observed that if we increase the maximum schedule time to more than 60 minutes, there is no significant change in the number of saved kilometers and therefore, the maximum time lag between the trips inside any cluster is 60 minutes. Also, by increasing the maximum distance between the trips to more than 6 kilometers there is no change in the number of total saved kilometers.

**7. Estimation of GHG and Pollutant Vehicle Emissions**

In order to estimate the emission reductions resulting from the estimated saved kilometers in Section 6, emission factors are often used. However, the reported emis-

sion factors of vehicles tested under various emission standards in China vary significantly. For example, carbon monoxide (CO), non-methane hydrocarbon (NMHC) and nitrogen oxides (NO<sub>x</sub>) emission factors of pre-Euro 1 gasoline vehicles are 15, 40 and 8 times those of their Euro 3 counterparts, respectively (Huo et al., 2012). Due to these broad variations, there is a high level of uncertainty in emission factors used in this study. Therefore, this study does not attempt to provide accurate figures for emission reductions, but rather an order of magnitude for the emission reductions that could result by adoption of ridesharing. Focusing on vehicle emissions, the number of vehicles and emission standards, Wang et al. (2011) project future trends in HC, CO, NO<sub>x</sub>, particulate matter (PM10) and carbon dioxide (CO<sub>2</sub>) emissions of personal vehicles based on three scenarios. The emission factors projected for light-duty passenger ve-



**Figure 8.** Effect of trip distance and maximum schedule time on (a) saved kilometers (b) number of saved trips.

hicles in 2015 and 2020 under the “recent policy” scenario is used in this study to evaluate the reduction in emissions that could result from ridesharing. In this scenario, it is assumed that PreEuro 1 to Euro 5 emission standards scheduled in stages from 2000 to 2013 are fully implemented.

Using the results presented in Section 6, potential pollutant and GHG emission reductions for a typical day resulting from adoption of ridesharing among the group of vehicles that are investigated are presented in Table 6. The emission reductions are estimated for 2015 based on available emission standards, and projections of the emission standards and their impact on vehicle emissions by 2020 are also used to evaluate the impact of ridesharing under lower emission factors. Assuming the number of vehicle kilometers taken during a typical day in 2020 remains the same, it is observed that ridesharing adoption provides lower pollutant and GHG emission reductions in 2020, but still considerable enough to make it a practical transportation strategy in the future as well. The reduction in the number of kilometers traveled when rides between drivers, within 2 kilometers and 40 minutes of departure location and time, respectively, are shared (see Section 6.2.3.1 and Table 5 for more details), results in approximately 3.1 and 0.0028 tons of CO<sub>2</sub> and NO<sub>x</sub> emission reductions, respectively. This is equivalent to approximately 4.0 tons CO<sub>2</sub> emission reductions [Global warming potential (GWP) of 100 years]. The emis-

sion reductions provided in Table 6 provide estimates for the order of magnitude of emission reductions that can be achieved through adoption of ridesharing. A more accurate estimation of kilometers saved using ridesharing and its corresponding emission reduction is dependent upon several additional factors. However, such rough estimations are useful in providing guidance to regulators and policy development for future planning.

Due to the limited size of the data set, and the dependency of ridesharing potential to the number of drivers, estimated emission reductions are a lower bound to the potential benefit of an overall rideshare system in Changsha, China. Although the results of the current analysis are specific to current mobility patterns in Changsha, they can be used qualitatively to guide the deployment and policy development regarding ridesharing in other cities.

**8. Conclusions**

Adoption of ridesharing among passenger vehicles in Changsha, China, as a potential strategy to reduce vehicle pollutants and GHG emissions is investigated. Historical GPS data of approximately 8,900 privately-owned vehicles in Changsha, China, are collected and is used in an algorithm that is developed to match riders with close temporal and spatial origin and destinations. The developed algorithm is capable of estimating kilometers that

**Table 6.** Average pollutant and GHG emissions projected for the vehicles in the study in 2015 and 2020.

Pollutant and GHG emissions					
Projection year	CO <sub>2</sub> (tons)	HC (kg)	CO (kg)	NO <sub>x</sub> (kg)	PM <sub>10</sub> (kg)
2015	3.14	2.80	24.8	2.80	0.280
2020	3.00	1.68	15.7	1.26	0.252

are reduced among users if ridesharing is adopted. The resulting reductions in vehicle pollutant and GHG emissions are estimated using average projected emissions factors for China.

The results show that the potential of ridesharing to reduce total traveled distance and emissions varies significantly by the users' tolerance towards changes to their original trip route and departure time. For example, the potential of ridesharing in reducing vehicle emissions increases by 94% if riders are willing to walk to drivers within 3 kilometers instead of 2 kilometers to get a ride. Assuming users are able to walk to the drivers, this could translate to an addition of 10–15 minutes to their trip time. In some cases, this delay could be compensated by a reduction in vehicle trip time (e.g., availability of high-occupancy vehicle lanes).

As shown in previous studies, the size of the data set can affect the ridesharing potential among users. Therefore, the results of the current study are dependent on the size of the data set used to identify potential ridesharing opportunities among users. A larger data set (i.e., more participants) would match more riders with ridesharing. As a result, the estimated traveled distance reduction and, associated emission reductions from ridesharing adoption in Changsha, China, are expected to be higher with a larger pool of participants.

While the quantitative results of this analysis are specific to the population under study, they provide useful insights on the potential of ridesharing to improve air quality and reduce emissions associated with climate change. Changsha, China, is one of several cities around the world that uses personal vehicles as a reliable mode of transportation. The methods used in this study to evaluate ridesharing potential in reduction of traveled kilometers in Changsha and reduction in pollutant and GHG emissions can be used in future similar studies on other cities that rely partially or fully on personal vehicle transportation. Analysis of current transportation demand and projection of future trends are key tasks in planning for sustainable transportation modes such as ridesharing that are potentially able to meet future demands.

Within the study area, ridesharing has the potential to reduce total kilometers driven (210,890 kilometers) by about 24% (51,087 kilometers) and vehicle trips (20,018 trips) by approximately 40% (8480). This maximum potential assumes a maximum distance between trips less than 10 kilometers, and schedule-time less than 60 minutes (Figures 8a and 8b). If a more conservative maximum distance of 2 kilometers between trips and schedule time less than 40 minutes is selected, total distance traveled reduces by 7% and total number of trips by 14%. This translates to equivalent of approximately 4.0 tons CO<sub>2</sub> emission reductions daily.

It must be noted that although findings of this study illustrate the potential of ridesharing in reducing pollutants and GHG emissions, its adoption still faces challenges such as passenger safety, privacy and liability for its adoption by users. Furthermore, the success of web-

based applications in connecting potential shared rides are dependent on the number of users. In terms of regulations, they compete with existing regulated taxi companies. Such limitations need to be further analyzed and solutions are needed to overcome these challenges.

### Acknowledgements

The authors acknowledge the support from the Metcalf Foundation, NSERC-CREATE TRANSIT, Mitacs, BlancRide, Plug'n Drive and Veridian Connections.

### Conflict of Interests

The authors declare no conflict of interests.

### References

- Agatz, N., Erera, A., Savelsbergh, M., & Wang, X. (2011). Dynamic ride-sharing: A simulation study in metro Atlanta. *Procedia-Social and Behavioral Sciences*, 17, 532–550.
- Agatz, N., Erera, A., Savelsbergh, M., & Wang, X. (2012). Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 223(2), 295–303.
- Alexander, L. P., & González, M. C. (2015). *Assessing the impact of real-time ridesharing on urban traffic using mobile phone data*. Paper presented at the 4th International Workshop on Urban Computing (UrbComp 2015), Sydney, Australia.
- Bicocchi, N., & Mamei, M. (2014). Investigating ride sharing opportunities through mobility data analysis. *Pervasive and Mobile Computing*, 14, 83–94.
- Cici, B., Markopoulou, A., Frías-Martínez, E., & Laoutaris, N. (2013). Quantifying the potential of ride-sharing using call description records. In *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications* (p. 17). New York: ACM.
- Cici, B., Markopoulou, A., Frías-Martínez, E., & Laoutaris, N. (2014). Assessing the potential of ride-sharing using mobile and social data: A tale of four cities. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 201–211). New York: ACM.
- Deza, M. M., & Deza, E. (2009). *Encyclopedia of distances*. Berlin: Springer.
- European Commission (2016). *Trends in global CO<sub>2</sub> emissions: 2016 Report*. The Hague: Netherlands Environmental Assessment Agency.
- Fujisaka, T., Lee, R., & Sumiya, K. (2010). Exploring urban characteristics using movement history of mass mobile microbloggers. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications* (pp. 13–18). New York: ACM.
- Ghoseiri, K., Haghani, A., & Hamedi, M. (2011). *Real-time rideshare matching problem* (Technical Report UMD-2009-05). College Park, MD: University of Maryland.

- Goel, P., Kulik, L., & Ramamohanarao, K. (2017). Optimal pick up point selection for effective ride sharing. *IEEE Transactions on Big Data*, 3(2), 154–168.
- Hao, H., Liu, Z., Zhao, F., Li, W., & Hang, W. (2015). Scenario analysis of energy consumption and greenhouse gas emissions from China's passenger vehicles. *Energy*, 91, 151–159.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108.
- He, W., Hwang, K., & Li, D. (2014). Intelligent carpool routing for urban ridesharing by mining GPS trajectories. *IEEE Transactions on Intelligent Transportation Systems*, 15(5), 2286–2296.
- Hu, J., Wu, Y., Wang, Z., Li, Z., Zhou, Y., Wang, H., . . . Hao, J. (2012). Real-world fuel efficiency and exhaust emissions of light-duty diesel vehicles and their correlation with road conditions. *Journal of Environmental Sciences*, 24(5), 865–874.
- Huang, T., Yang, G. J. T. G. Y., & Tang, G. (1979). A fast-two-dimensional median filtering algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(1), 13–18.
- Huo, H., Yao, Z., Zhang, Y., Shen, X., Zhang, Q., Ding, Y., & He, K. (2012). On-board measurements of emissions from light-duty gasoline vehicles in three mega-cities of China. *Atmospheric Environment*, 49, 371–377.
- Kim, J., Rasouli, S., & Timmermans, H. (2017). Satisfaction and uncertainty in car-sharing decisions: An integration of hybrid choice and random regret-based models. *Transportation Research Part A: Policy and Practice*, 95, 13–33.
- Liu, H., He, K., Lents, J. M., Wang, Q., & Tolvet, S. (2009). Characteristics of diesel truck emission in China based on portable emissions measurement systems. *Environmental science & technology*, 43(24), 9507–9511.
- Ma, S., Zheng, Y., & Wolfson, O. (2013). T-share: A large-scale dynamic taxi ridesharing service. In *2013 IEEE 29th International Conference on Data Engineering (ICDE)* (pp. 410–421). IEEE.
- Maciejewski, M., Bischoff, J., & Nagel, K. (2016). An assignment-based approach to efficient real-time city-scale taxi dispatching. *IEEE Intelligent Systems*, 31(1), 68–77.
- Noulas, A., Mascolo, C., & Frias-Martinez, E. (2013). Exploiting foursquare and cellular data to infer user activity in urban environments. In *2013 IEEE 14th International Conference on Mobile Data Management (MDM)*, (Vol. 1, pp. 167–176). IEEE.
- Noulas, A., Scellato, S., Mascolo, C., & Pontil, M. (2011). Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. *The Social Mobile Web*, 11(2).
- Organization for Economic Cooperation and Development. (2011). *OECD Environmental Outlook to 2050*. Paris: OECD.
- Ota, M., Vo, H., Silva, C., & Freire, J. (2015). A scalable approach for data-driven taxi ride-sharing simulation. In *2015 IEEE International Conference on Big Data* (pp. 888–897). IEEE.
- Ota, M., Vo, H., Silva, C., & Freire, J. (2016). STARS: Simulating taxi ride sharing at scale. *IEEE Transactions on Big Data*.
- Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S. H., & Ratti, C. (2014). Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences*, 111(37), 13290–13294.
- Shen, B., Huang, Y., & Zhao, Y. (2016). Dynamic ridesharing. *SIGSPATIAL Special*, 7(3), 3–10.
- Tachet, R., Sagarra, O., Santi, P., Resta, G., Szell, M., Strogatz, S. H., & Ratti, C. (2017). Scaling Law of Urban Ride Sharing. *Scientific Reports*, 7, 42868. doi:10.1038/srep42868
- Trasarti, R., Pinelli, F., Nanni, M., & Giannotti, F. (2011). Mining mobility user profiles for car pooling. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1190–1198). New York: ACM.
- Tsao, H. S. J., & Lin, D. J. (1999). *Spatial and temporal factors in estimating the potential of ride-sharing for demand reduction* (California PATH Research Report, UCBITS-PRR-99-2). St. Richmond, CA: California Partners for Advanced Transit and Highways.
- Wakamiya, S., Lee, R., & Sumiya, K. (2011). Urban Area Characterization Based on Semantics of Crowd Activities in Twitter. *GeoSpatial Semantics*, 108–123.
- Wang, H., Fu, L., & Bi, J. (2011). CO<sub>2</sub> and pollutant emissions from passenger cars in China. *Energy Policy*, 39(5), 3005–3011.
- Wang, X., Westerdahl, D., Hu, J., Wu, Y., Yin, H., Pan, X., & Zhang, K. M. (2012). On-road diesel vehicle emission factors for nitrogen oxides and black carbon in two Chinese cities. *Atmospheric Environment*, 46, 45–55.
- Wang, X., Westerdahl, D., Wu, Y., Pan, X., & Zhang, K. M. (2011). On-road emission factor distributions of individual diesel vehicles in and around Beijing, China. *Atmospheric Environment*, 45(2), 503–513.
- Wu, Y., Zhang, S., Hao, J., Liu, H., Wu, X., Hu, J., . . . & Stevanovic, S. (2017). On-road vehicle emissions and their control in China: A review and outlook. *Science of The Total Environment*, 574, 332–349.
- Wu, Y., Yang, Z., Lin, B., Liu, H., Wang, R., Zhou, B., & Hao, J. (2012). Energy consumption and CO<sub>2</sub> emission impacts of vehicle electrification in three developed regions of China. *Energy Policy*, 48, 537–550.
- Yang, D., & Timmermans, H. (2017). Households transport-home energy conservation strategies in response to energy price policies: A stated adaptation experiment based on portfolio choices and cross effects designs. *International Journal of Sustainable Transportation*, 11(2), 133–147.
- Zhang, S., Wu, Y., Liu, H., Huang, R., Un, P., Zhou, Y., . . . & Hao, J. (2014). Real-world fuel consumption and

CO<sub>2</sub> (carbon dioxide) emissions by driving conditions for light-duty passenger vehicles in China. *Energy*, 69, 247–257.

Zhang, S., Wu, Y., Wu, X., Li, M., Ge, Y., Liang, B., . . . & Hao, J. (2014). Historic and future trends of vehicle emissions in Beijing, 1998–2020: A policy assessment for the most stringent vehicle emission control program in China. *Atmospheric Environment*, 89, 216–229.

Zheng, X., Wu, Y., Jiang, J., Zhang, S., Liu, H., Song, S., . . .

& Hao, J. (2015). Characteristics of on-road diesel vehicles: Black carbon emissions in Chinese cities based on portable emissions measurement. *Environmental Science & Technology*, 49(22), 13492–13500.

Zhou, Y., Wu, Y., Zhang, S., Fu, L., & Hao, J. (2014). Evaluating the emission status of light-duty gasoline vehicles and motorcycles in Macao with real-world remote sensing measurement. *Journal of Environmental Sciences*, 26(11), 2240–2248.

### About the Authors



**Roozbeh Jalali** is currently pursuing a PhD degree in Computer Science at the University of Ontario Institute of Technology. He received a master's degree in Computer Engineering from Amirkabir University of Technology, Tehran, Iran in 2011 and bachelor's degree in Software Engineering in 2008. His research interests include Internet of Things, smart cities, and data processing in context of smart cities.



**Seama Koohi-Fayegh** is a Post-Doctoral Fellow at the University of Ontario Institute of Technology (UOIT). Seama received her PhD degree in Mechanical Engineering at UOIT in 2013. Her PhD thesis topic was proposed by the Ontario Ministry of Environment and was focused on thermal sustainability of geothermal energy systems: system interactions and environmental impacts. She completed her Master's in Mechanical Engineering (Energy Conversion) in 2008 at Ferdowsi University of Mashhad, Iran, and worked on entropy generation analysis in heat transfer problems. Dr. Koohi-Fayegh's research interests include thermal analysis, sustainable energy systems and energy technology assessment.



**Khalil El-Khatib** is an Associate Professor at University of Ontario Institute of Technology. Dr. El-Khatib was an Assistant Professor at the University of Western Ontario prior to joining UOIT in July 2006. Between 1992 and 1994, he worked as a Research Assistant in the Computer Science department at AUB. In 1996, he joined the High Capacity Division at Nortel Networks as a software designer. From February 2002, he was a Research Officer in the Network Computing Group (lately renamed the Information Security Group) at the National Research Council of Canada for two years, and continued to be affiliated with the group for another two years. Dr. Khalil's research interests include Parallel Discrete Event Simulation, QoS for multimedia applications, personal and service mobility, IP telephony, feature interaction for VoIP, ubiquitous computing environments, security and privacy issues for the Internet and for the mobile wireless ad hoc networks (MANET).



**Daniel Hoornweg** is an Associate Professor and Richard Marceau Research Chair at the University of Ontario Institute of Technology. Dan received his PhD in Civil Engineering (Cities and Sustainability) from the University of Toronto in 2015. Dan returned to Ontario in 2014 after almost 20 years with the World Bank, where he led the Cities and Climate Change, and Sustainable Cities programs. Dan's research areas include energy and urban systems, waste management, urban resilience and sustainability.



**Heng Li** is currently an Assistant Professor with the College of Computer Science and Electronic Engineering, Hunan University. His research interests include vehicular ad hoc networks, smart cities, and pervasive computing.