Computing and Visualizing Taxi Cab Dynamics as Proxies for Autonomous Mobility on Demand Systems

The Case of the Chicago Taxi Cab System

Dimitris Papanikolaou(✉)

University of North Carolina at Charlotte, Charlotte, NC 28223, USA
dpapanik@uncc.edu

Abstract. Despite the expansion of shared mobility-on-demand (MoD) systems as sustainable modes of urban transport, a growing debate among planners and urban scientists regarding what constitutes cost and how to compute it, divides opinions on the benefits that autonomous MoD systems may bring. We present a comprehensive definition of cost of traveling by MoD systems as the cost of the vehicle hours (VH), the vehicle-hours-traveled (VHT), the vehicle-hours-dispatched (VHD), and the vehicle-hours-parked (VHP) required to serve a pattern of trips. Next, we discuss an approach to estimate empty (dispatch) trips and idle periods from a user trip dataset. Finally, we model, compute, and visualize the relationship between the dynamics of VHP, VHT, and VHD using Chicago's taxi cab system as a case. Our results show that the total fleet of taxis in Chicago can decrease by 51% if all trips, currently served by conventional taxis, were served by autonomous ones.

Keywords: Mobility on Demand systems · Taxi cab systems · Data-driven dynamic modeling · Autonomous Vehicles · System dynamics

1 Introduction

Widely considered as emerging modes of sustainable transport, mobility on demand (MoD) systems utilize shared vehicles, parking spaces, and advanced information technology, allowing citizens to move from point to point on demand while minimizing resource utilization and carbon footprint [1]. With an industry doubling biannually, and pilot autonomous vehicles (AV) already cruising in streets of cities, MoD systems are one of the most rapidly growing sectors of urban transport. Yet, a growing debate among scholars and practitioners on what constitutes cost and how to compute it, has divided opinions on the benefits that autonomous MoD systems may bring. Some studies claim that sharing reduces congestion, improves air quality, and enhances accessibility while reducing dependence on private ownership [2–4]. Others, however, show that vehicle-miles traveled (VMT) and traffic congestion will increase up to 8% and 14.0% respectively due to increased empty trips for charging and pickup [5–7]. Understanding what aspects, and how, affect performance is important, not only for

J.-H. Lee (Ed.): CAAD Futures 2019, CCIS 1028, pp. 183–197, 2019.
https://doi.org/10.1007/978-981-13-8410-3_14

assessing the potential of on-demand mobility, but also for designing, planning, and operating future MoD systems that perform better than today's ones.

For planners and urban scientists, estimating the quantities of vehicles, parking land, and dispatch flows (movements of empty vehicles from one drop-off to the next pickup), required to mobilize a demand for trips, is essential because they determine the cost, the infrastructure size and planning requirements, and the environmental impact of a MoD system. While recent studies have used microsimulation techniques to address such questions for AV MoD systems [4–6, 8], their modeling assumptions of how vehicles move in space and time are often too idealized. It is therefore important to complement results from such theoretical models with similar results from data-driven evidence-based simulation models that describe real systems in high detail.

In this paper, we analyze a dataset of user trips from taxi cabs, asking the question: How many vehicles, parking land, and dispatch work are required to serve the same trips with AVs, and how would these quantities change as technology improves? We consider a taxi system as the closest contemporary analogy to an AV MoD system. Since an AV is a taxi with a restless computerized driver, studying taxi systems can reveal important information about a lower bound in performance of AV MoD systems. At each moment, a taxi cab can be in one of three possible states: *active* (carrying a passenger), *dispatched* (driving empty in search of a new passenger), or *parked* (either temporarily or overnight). In particular, we focus on three questions in estimating the minimum capacity of a MoD system: *how many taxis are required to serve a given demand for trips? How much parking land do these taxis occupy when not in use? How much traffic do these taxis create when in use or being dispatched?*

In previous work [9] we developed a novel method to reconstruct and visualize accumulation dynamics of MoD systems by numerically integrating trip data over time, and we used this method to model the dynamic relationship between the active, dispatched, and parked portions of the fleet in Boston's bike sharing system. In this paper, we expand this method to compute and visualize the dynamics of the actual and minimum quantities of vehicles, parking land, and dispatch work, required to mobilize a pattern of trips in a taxi cab system. We analyze a trip dataset from Chicago's taxi system with over 100 million rides dating back to 2013, and focus on July 13, 2017 as a case.

Our paper makes three contributions in the fields of Smart Cities, Data Analytics, and Urban Modeling and Simulation. (1) We present a holistic definition of cost of traveling by taxi. (2) We discuss a novel method to estimate dispatch trips and idle periods from user trips in taxi systems and we apply this method to estimate the number of dispatch taxi trips and idle periods in Chicago. (3) We model, compute, and visualize the relationship between the dynamics of the vehicle-hours-traveled (VHT), the vehicle-hours-dispatched (VHD) and the vehicle-hours-parked (VHP) for serving users with taxi MoD systems in Chicago.

The rest of the paper is organized as follows. Section 2 discusses related works. Section 3 defines the cost of shared mobility from a dynamics' perspective. Section 4 describes the Chicago taxi trip dataset. Section 5 describes the method to estimate the empty (dispatch) trips and idle periods from the user trip dataset. Section 6 describes the method to reconstruct dynamics of vehicle accumulations and a validation method. Section 7 presents a scenario analysis in Chicago visualizing both the existing and the hypothetical accumulation dynamics if all current trips were served by autonomous taxi cabs that minimize empty trip and idle durations. Finally, Sect. 8 discusses significance and future steps of our work, specifically in relation to MoD systems.

2 Background

Sizing and rebalancing, is the fundamental planning and operations problem of deciding the optimal combination of infrastructure size (number of vehicles, parking spaces, road capacity) and rebalancing work (dispatches of empty vehicles from full to empty locations) for serving a pattern of trips. The two quantities are in a tradeoff relationship. Increasing one, decreases the need for the other. Sizing-rebalancing problems in MoD systems summarize into three directions: sizing optimization [10]; routing optimization [11–13]; and joint sizing-rebalancing optimization which are significantly more complex [4, 14]. State-of-the-art approaches use stochastic microsimulations [4–6, 8, 14], however, developing such models is laborious, data intensive, and computationally expensive, without guaranteeing results [15]. Estimating rebalancing work requires solving routing but routing depends on assumptions about network structure and trip demand, which in MoD systems are often random. This makes micromodeling approaches speculative and, for the scope of our research, unnecessarily complex.

Recent studies on various data sources show that even though individual trip patterns are random, human mobility patterns are macroscopically consistent and predictable [16–21]. In particular, intra-city commuting patterns are found to be periodic [22–24] and universal across cities [25, 26] with a trip duration consistent at 25–35 min [25]. This relationship has been observed in multiple data sources from taxi [16, 23, 27] and bike sharing [28, 29] trips. Finally, other studies on bike sharing [11, 29, 30] and car sharing [8] data found that trip patterns are not only periodic but also cumulatively imbalanced, creating increasingly uneven distributions of vehicles. These studies suggest that accurate dynamic macromodeling is scientifically possible [18].

There are two cultures in building dynamic models from time series data. The first, coming from data science, views time series data as samples from an unknown stochastic process and uses machine learning techniques to invent a mechanism that can emulate them without explicitly knowing the physical system that may have generated them. Despite their usefulness in modeling data patterns when the system that created them is unknown, these techniques work as black boxes without explaining why data have been clustered one way or another [31], or what the resulting model and its parameters mean. As such, they cannot substitute valid theoretical models for planners. The second approach is hypothesis-based and comes from areas of physics, complex systems, and dynamics. It views time series data as measurements of a dynamic physical system which aims to reconstruct by postulating its structure and behavior based on first principles, empirical evidence, or established theories [32]. System dynamic (SD) models, lump accumulations of quantities into compartments representing different states (called stocks), and compute accumulation levels by numerically integrating inflows and outflows to and from each compartment over time.

By describing mathematically flow rates with equations, SD models can simulate a variety of what-if scenarios, with applications in urban studies [33–35].

Our work extends the aforementioned works in three ways. First, we introduce a novel general method to derive empty trips and idle periods from a user trip dataset. Second, we show the relationship between cost and accumulation dynamics, and we introduce a novel method to reconstruct accumulation dynamics and to find the minimum fleet for serving demand, as a time-series dynamic problem. Third, we introduce a novel way to visualize accumulation dynamics and, to our knowledge, for the first time, we visualize the dynamic relationship between accumulation dynamics of all areas in the city for taxicab systems.

3 Cost as Rent and Work

Every shared mobility system requires four fundamental types of resources to mobilize a pattern of trips: vehicles, parking land, work to mobilize user trips, and work to dispatch empty trips (also known as rebalancing). We quantify these quantities as vehicle-hours (VH), vehicle-hours-parked (VHP), vehicle-hours-travelled (VHT), and vehicle-hours-dispatched (VHD), and we define as cost, the cost of these quantities over time, measured as vehicle rent, parking rent, useful work, and rebalancing (empty) work.

From a dynamics' perspective, trip paths are unimportant. If two trips between a pair of locations have the same duration, their effect on accumulations at locations is the same even if their paths are different. What matters instead is where and when departures (outflows) and arrivals (inflows) occur and, consequently, which locations, at what rate, and to what extent, fill with or empty from vehicles. These fluctuations determine minimum fleet size, parking size, and traffic requirements.

If a dataset of user and dispatch trips is available, then VH, VHP, VHT, and VHD can be directly computed by numerically integrating the difference of departures and arrivals over time. The vehicle and land rent of each location during a period, is the cost of the net number of vehicle-hours or parking-hours required to serve a pattern of outgoing and incoming trips during that period. If 10 vehicles depart from a location and arrive after 8 h, the location will require 80 vehicle-hours of vehicle rent for the temporary absence of vehicles. If the 10 vehicles arrive and then depart after 8 h, the location will require 80 parking-hours of land rent to accommodate the temporary presence of vehicles. If each of the 10 vehicles traveled 30 min for each one-way trip, then a total of 10 vehicle-hours-traveled of work was done (5 VHT in each direction) to move them. The vehicle and land rent of a system during a period is the sum of the individual vehicle and land rents of its locations during that period. Since the system is closed and mass is conserved, VHT and VHD can be calculated from the VHP: if a portion of vehicle mass is absent from parking areas then it must be traveling in streets. VHP can be directly computed from locations from their occupancy. We define as occupancy of a location, the aggregate presence or absence of vehicle mass during a period of time. Positive occupancy indicates net presence while negative occupancy indicates net absence of vehicles. In this paper, we show that minimum VH, VHP, VHT, and VHD can be estimated from a user trip dataset, even if no information about dispatch trips is available, using the average dispatch duration as a parameter. The next sections describe a method to derive empty trips from a user trip dataset and a method to compute the dynamics of VHT, VHD, and VHP over time.

4 Chicago Taxi Trip Data

The study is based on a dataset that the City of Chicago started publishing on November 16, 2016. Today, the dataset includes over 100 million taxi ride entries, dating back to 2013. The data can be accessed through the Chicago Socrata API in JSON, XML and CSV formats. Our study focuses on three weekdays of July 12–15, 2017 using the 13th as a reference day. Each entry in the dataset describes a trip and contains 23 attributes:

Trip ID, Taxi ID, Trip Start Timestamp, Trip End Timestamp, Trip Seconds, Trip Miles, Pickup Census Tract, Dropoff Census Tract, Pickup Community Area, Dropoff Community Area, Fare, Tips, Tolls, Extras, Trip Total, Payment Type, Company, Pickup Centroid Latitude, Pickup Centroid Longitude, Pickup Centroid Location, Dropoff Centroid Latitude, Dropoff Centroid Longitude, Dropoff Centroid Location.

All start and end times were rounded to the nearest 15 min. In order to preserve costumer and driver's privacy, location is provided only at Census Tract and Community Area level ("*pickup_community_area*", "*dropoff_community_area*"). Census Tract data is not available for some trips for privacy purposes. The data attributes "*fare*", "*tips*", "*tolls*", "*extras*", "*trip_total*" and "*payement_type*" are all cost related and provide details about the cost of each trip and how costumers paid the cost. The pickup and dropoff centroid latitudes and longitudes, as well as the pickup and dropoff centroid location attributes, provide geographic coordinates for the pick-ups and drop offs. These coordinates are either the center of the census tract in which the trip began and ended or the community area if the census tract has been removed for privacy.

Initial exploration of the data revealed that certain entries seem either implausible or have specific meanings that are not described in the dataset. For example, several trip entries appeared to have either zero-second duration or zero-mile distance. Other entries appeared to have non-zero durations but zero-miles distance or non-zero-miles distance but zero-seconds durations. To avoid unsupported interpretations, we contacted the Chicago Data Portal and inquired about these specific cases. According to Chicago data portal, due to the unavoidable circumstances in data collection having 0 values for trip_seconds and trip_miles may occur and these values are wrong. Therefore, we decided to remove all entries with 0 durations. Another problem with data is that several trips had as start or end locations, areas with no name and 0 latitude and longitude. To address these inconsistencies, we filtered out these data trip entries. In total, from the 105,037 entries of the 3-day focus period, we filtered-out 8,941 such entries.

5 Identification of Empty Trips and Idle Periods

We consider as *empty trip*, the movement that a taxi makes from the dropoff location of a client to the pickup location of the next client. We consider as *idle period*, the duration that a taxi remains parked at a rest area. Therefore, we consider three possible states for each taxi: "full", when the taxi is occupied by a passenger; "empty", when the taxi driver is driving to find the next passenger; and "idle", when the taxi is parked without looking for passengers (sleeping, having lunch, waiting, etc.). Since the dataset contains only user trips, we developed a method to identify empty trips and idle periods from the periods between consecutive rides from the same taxi ID. We labeled each such period as a *combined duration* because it combines both the duration of an empty trip and the duration of a potential idle period. The distribution of full trip durations has one sharp peak that centers at about 900 s and is skewed to the left. This means that while most user trips are short and similar (about 15 min), there are several user trips that are longer and have durations that vary. The histogram in Fig. 1 shows the distribution of combined durations (empty and idle). The distribution has a sharp peak that centers at about 1,000 s and is skewed to the left. A closer look reveals a second peak, shorter and fatter, that centers at about 8,000 s. This means that while many durations between consecutive drop-offs and pickups are similar to the duration of a typical ride, there is a smaller cluster of such durations that are much longer, with a range of 1 to 8 h, and an average of 2.2 h.

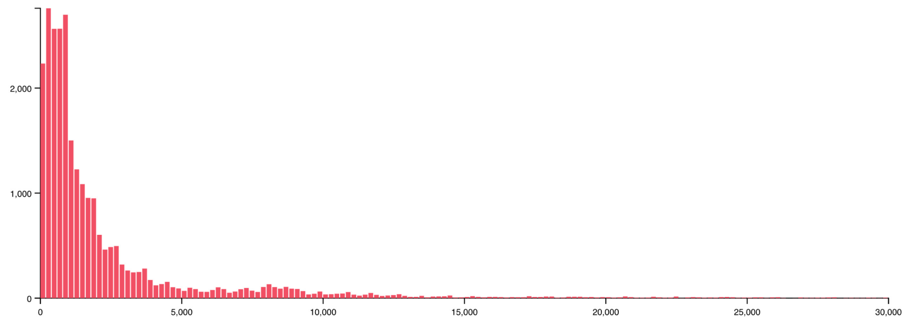

Fig. 1. Histogram of the durations of empty trips.

We interpreted the first peak of short durations as empty trips (trips from a dropoff to a pickup similar in duration to user trips) and we interpreted the second peak of long durations as idle periods. The interpretation is based both on empirical evidence and on common sense: idle periods include overnight rests during sleep hours and can therefore be several hours long. On the other hand, empty trips cannot be overly long: if they were, on average, longer than idle periods, then empty taxis in streets would outnumber occupied ones. This, however, cannot be the case: first, because demand during peak hours is higher than supply, and second, because during off-peak hours, spending more time empty than occupied would be financially unsustainable for taxi operators.

To distinguish between empty trips and idle periods, we analyzed the duration between a user drop-off and the next user pickup for each pair of consecutive trips with the same taxi ID. Some of these durations signify an empty trip while others include idle periods (driver rests temporarily or overnight). We consider three possible cases (Fig. 2).

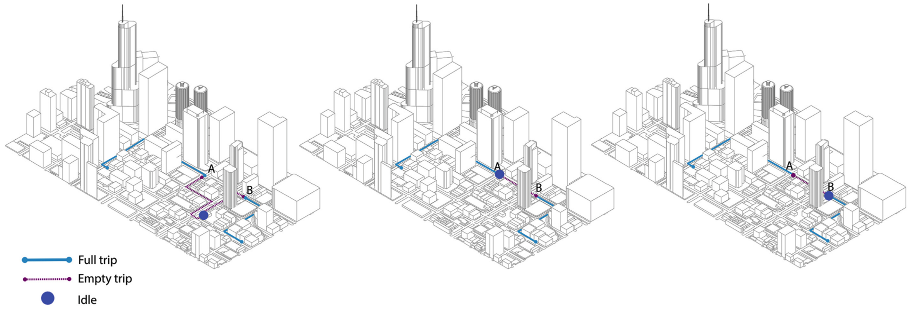

Fig. 2. Three cases for an empty trip. Left: a taxi drives empty from dropoff area A to a rest area (blue dot), remains idle, and then drives empty to pickup area B. Middle: a taxi remains idle after dropping off a client and then drives empty to a pickup area. Right: a taxi drives empty after dropping off a client to a rest area, remains idle, and then picks up the next client. (Color figure online)

In the first case, the taxi is assumed to do an empty trip from the drop-off location of the last client to a parking location (presumably the house of the taxi driver or the central taxi facility) and an empty trip from the parking location to the pickup location of the next client. In the second and third cases, the taxi is assumed to stay idle immediately after dropping-off a client and then perform an empty trip to pick up the next new client (second case) or perform an empty trip to the location of the next client pickup and sojourn at that location until the next pickup (third case). While the first case is more plausible than the second and third cases, we have no data that suggest whether rest areas are different than last dropoff or first pickup areas. On the other hand, while second and third cases are less plausible today, they imply a future scenario in which an AV, if not in use, parks anywhere and stays idle until someone calls it. We therefore decided to choose either the second or third cases. Since these cases make no practical difference for the dynamics of the system, we chose the third case.

To identify empty trips, we developed an algorithm that takes as input a user trip dataset and a duration threshold, and it outputs a dataset of full and empty trips. The duration threshold is used as a classifier. If the duration between a consecutive dropoff and pickup is below the threshold, we assume the duration to signify an empty trip from the dropoff location to the pickup location. If the duration is above the threshold, we assumed that at least one empty trip and an idle duration exist.

To define the threshold duration that is used as a classifier to distinguish empty trips from idle periods, we used a duration of 1,800 s (30 min) which is close to twice the mean trip duration. By filtering the duration values, we labeled trips longer than 5,455.60 s as “idle” and the rest as “empty”. While this method is crude, it provides an empirical separation between empty trips and idle periods and can therefore be used in combination with a sensitivity analysis, in which, a range of datasets are computed and tested for a range of possible thresholds. A future improvement of this method will use the application of a mixture bivariant model and use of Machine Learning (Expectation Maximization) to find the mean times of the two peaks.

6 Reconstruction of Accumulation Dynamics

After calculating the empty and full trips, we used them to reconstruct the accumulation dynamics of the system. A central concept in dynamic analysis and modeling is the reconstruction of the *trajectory* of the system, which is the collection of all the states that the system takes over a time period [19–21]. The state of a taxi system at each moment, is defined by the accumulation levels of stationary, in-use, and in-dispatch taxi vehicles. A dynamic system is defined by its inflow/outflow time series, the difference equations (conservation laws), and the state of the system at a reference time, known as the initial state. Any state can be derived by the initial state by numerical integrating inflows and outflows. In this section, we show that, by combining data (uncontrollable input) with mathematical functions (controllable input), we can build interactive data driven dynamic models that can be used for scenario analysis.

To reconstruct dynamics of VHP, VHT, and VHD, first, we obtain the set of community areas from the origins and destinations of trip data. Next, we create inflow and outflow time series data from user and empty (dispatch) trip data, by computing departures and arrivals per location per unit of time. Next, we compute accumulation level time series data per community area, by integrating numerically inflow and outflow time series over timeline. Finally, we compute the accumulation time series data for the stocks in transit and in dispatch. The total stock of the system is the sum of the stock in locations, in dispatch, and in transit, and remains constant over time. Initial values for accumulation levels may be obtained from any known state of the system with backward numerical integration.

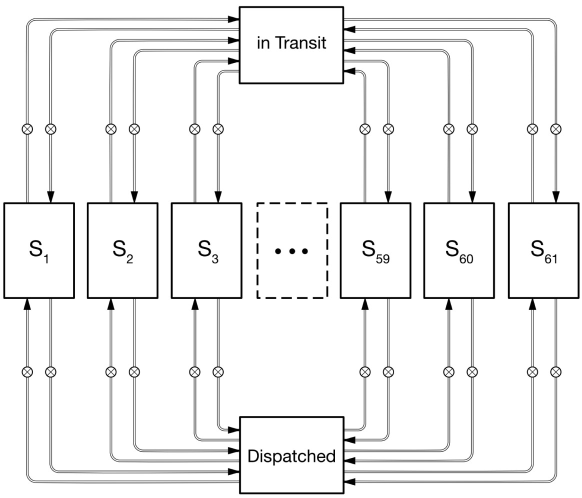

Fig. 3. Stock-flow diagram of a taxi system consisting of all community areas (represented by different stocks) and two stocks representing vehicles in transit (full and empty).

Create Closed System of Locations. To construct a trajectory that conserves mass, it is important that the number of locations derive from trip data. To create locations: (1) Create empty array for locations. (2) For each trip in trips array, for each location in locations array: if location is the same as origin or destination location of trip, go to next trip; else, add origin or destination location of trip in locations array, and go to next trip. (3) End. Figure 3 shows the resulting stock-flow model from this process consisting of stocks for each community area containing parked vehicles (s_1, s_2,…, s_n) and of two additional stocks containing vehicles in transit and dispatched vehicles.

Create User Inflow and Outflow Time Series. To create user inflow and outflow time series per location: (1) Divide timeline into timesteps. (2) For each location in the locations array, create four empty arrays: departures, arrivals, outflows, and inflows. (3) For each trip in trips array, get origin and destination locations of trip; add one departure event in the departures array of the origin location; add one arrival event in the arrivals array of the destination location. (4) For each time-step, for each location in the locations array: add the sum of all departures that occurred within the time-step in the outflows array; add the sum of all arrivals that occurred within time-step in the inflows array. (5) End. The resulting system contains the set S of $|S|$ locations (where $|S|$ is the size of S). Each location s_i contains 4 time-series with $|T|$ data points each (where $|T|$ is number of time-steps within timeline T): user outflows X_i, user inflows Y_i, dispatch outflows $\hat{X}_i$, and dispatch inflows $\hat{Y}_i$:

$$S = \{s_1, s_2, \ldots, s_{|S|}\} : s_i = \left\{ \begin{array}{l} X_i = \{x_1, x_2, \ldots x_{|T|}\} \\ Y_i = \{y_1, y_2, \ldots y_{|T|}\} \\ \hat{X}_i = \{\hat{x}_1, \hat{x}_2, \ldots \hat{x}_{|T|}\} \\ \hat{Y}_i = \{\hat{y}_1, \hat{y}_2, \ldots \hat{y}_{|T|}\} \end{array} \right\} \tag{1}$$

Compute Accumulation Dynamics at Locations. The user accumulation level L_i of location s_i, at time-step j, is computed by adding and subtracting inflows and outflows from $t = 1$ to $t = j$, where $\overline{\overline{L}}_i$ is the initial level $(t = 0)$:

$$L_i(j) = \overline{\overline{L}}_i + \sum\nolimits_{t=1}^{j} x_i(t) - y_i(t) \tag{2}$$

Likewise, the corrected accumulation level $\hat{L}_i$ of location s_i at time-step j is:

$$\hat{L}_i(j) = \overline{\overline{L}}_i + \sum\nolimits_{t=1}^{j} x_i(t) - y_i(t) + \hat{x}_i(t) - \hat{y}_i(t) \tag{3}$$

Compute Accumulation Dynamics of Stocks in Transit. The stock of vehicles in transit (stock-in-transit) contains vehicles in use by users. Since the system is closed, inflows to stock-in-transit equal outflows from locations, and outflows from stock-in-transit equal inflows to locations. In the model diagram, stock-in-transit's compartment connects to each location's compartment through the departure and arrival flow rates (pipelines). The level of stock-in-transit L_T at time-step *j,* is computed by adding and subtracting user outflows and inflows from $t = 1$ to $t = j$ for each of the $|S|$ locations, where $\overline{\overline{L}}_T$ is the initial level:

$$L_T(j) = \overline{\overline{L}}_T + \sum_{i=1}^{|S|} \sum_{t=1}^{j} y_i(t) - x_i(t) \tag{4}$$

Add Dispatch Flows and Reconstruct System Trajectory. To reconstruct the trajectory of the system with dispatching, we add dispatch inflows and outflows and recompute accumulation levels. The dispatch stock contains empty vehicles. In the model diagram, the dispatched compartment connects to each location's compartment through the dispatch departures and dispatch arrivals flow rates (pipelines). The level of the dispatch stock L_D at time-step j, is computed by adding and subtracting dispatch outflows and dispatch inflows from $t = 1$ to $t = j$ for each of the $|S|$ locations where $\overline{\overline{L}}_D$ is the initial level:

$$L_D(j) = \overline{\overline{L}}_D + \sum_{i=1}^{|S|} \sum_{t=1}^{j} \hat{y}_i(t) - \hat{x}_i(t) \tag{5}$$

Set Minimum Initial Values of Accumulation Levels. Initial values for accumulation levels may be obtained from any known state of the system with backward numerical integration. If no such state is known, then the minimum initial level $\overline{\overline{L}}_i$ of any location s_i, is the opposite of the minimum value $L_{i_{min}}$ that L_i had over timeline T:

$$\overline{\overline{L}}_i = -L_{i_{min}}, L_{i_{min}} = min_{t \in T} \sum_{z=1}^{t} x_i(z) - y_i(z) \tag{6}$$

The minimum parking capacity P_i of location s_i is the range (max-min) of its accumulation level values over timeline:

$$P_i = L_{i_{max}} - L_{i_{min}}, L_{i_{max}} = max_{t \in T} \sum_{z=1}^{t} x_i(z) - y_i(z) \tag{7}$$

Set Actual Initial Values of Accumulation Levels. Finding the initial values for the actual accumulation levels was a two-step process. First, we found initial levels of all trips that were considered in the time window. Second, we found initial levels at locations at the beginning of the time window. To find initial levels for each location, for each vehicle ID, we filtered all trips from the same vehicle ID, and we obtained the first trip. Then, we got the start location of the first trip and we incremented unitarily the initial level of that location. This process provided the initial levels of the first trips that either ended or started from the time window. To get the initial levels at time zero, we constructed accumulation dynamics from the start time of the first trip till the end time of the desired range, and we then trimmed the timeseries to the desired length.

6.1 Validation

In reconstructing an empty trip dataset from a given user trip dataset, two possible types of inconsistencies may occur: inconsistencies in time and inconsistencies in space. Inconsistencies in time occur if (1) an estimated empty trip appears to start before the previous full trip of the same vehicle ends, (2) an estimated empty trip appears to end after the next full trip of the same vehicle starts, or (3) an empty trip appears to end before it has started. Inconsistencies in space occur if an estimated empty trip appears to start from, or end to, different locations than its previous and next full trips. Results of such inconsistencies create accumulation dynamics that violate the law of mass preservation. For example, if a vehicle departs from a location before its previous trip arrives at that location, or if a vehicle arrives at a location after its next trip departs from that location, then the quantity of the vehicle at the location appears negative while the quantity of the vehicle in the streets (travelling) appears double. Likewise, if a vehicle arrives at a location but then departs from another location, then there must two vehicles in the system with the same ID. The resulting dataset of both empty and full trips can be validated for topological connectivity as follows. For each chain of trips of the same vehicle, the origin location of each trip is the same as the destination location of the previous trip, and the departure time of each trip is after the arrival time of the previous trip and before the departure time of the next trip. Likewise, the destination location of each trip is the same as the origin location of the next trip, and the arrival time of each trip is after the departure time of the trip and before the departure of the next trip. To validate the reconstructed accumulation dynamics, we analyzed the resulting empty and full trips dataset for mass preservation. If reconstructed empty trips were valid, then the mass of each individual vehicle in the system should remain constant over time. If a vehicle is removed from an area, it must be added to the transitionary stock and if it added in another area, it must be first removed from the transitionary stock. For each vehicle, we traced the chains of empty, idle, and full durations, and we analyzed whether there was any timestep during which the vehicle mass in the system increased or decreased. Moreover, for each location, we found the initial values for each vehicle. In total, the sum of the initial values for all locations for a vehicle were found to be 1, and the sum of all departures and arrivals in the system for that vehicle were found to be zero.

7 Scenario Analysis

An extensive scenario analysis is beyond the scope of this paper. Here, we present two scenario cases: The actual-fleet accumulation dynamics (Fig. 4), and the minimum-fleet accumulation dynamics (Fig. 5). For each case, we used a threshold duration of 1,800 s which classifies every combined duration below 30 min as an empty trip, and every combined duration above 30 min, as a 30-min empty trip and the remainder as idle period. The actual fleet was found to be 2,398 which matches the number of distinct vehicle IDs in the trip dataset. The minimum fleet was found to be 1,220 vehicles or 50.9% smaller than the actual fleet size. In total, for the selected day, taxis in Chicago spent 57,552 VH, 43,550 VHP, 8,071 VHT, and 5,931 VHD.

8 Discussion

We presented a definition of cost for shared mobility and a novel method to compute and visualize dynamics of VHP, VHT and VHD for a taxi cab system. Moreover, we presented a method for exploring how this cost changes in two scenario cases: actual and minimum fleet accumulation dynamics. In this section, we discuss insights from the results, and limitations of the presented method. First of all, the cost of traveling by taxi is substantially dominated by the size of the inactive stationary stock compared to the active transitionary stock. While this hidden stock is a consequence of the current business model for most taxi cab systems around the world, it also constitutes an inefficiency that can be greatly reduced in a future scenario of autonomous vehicles (AVs).

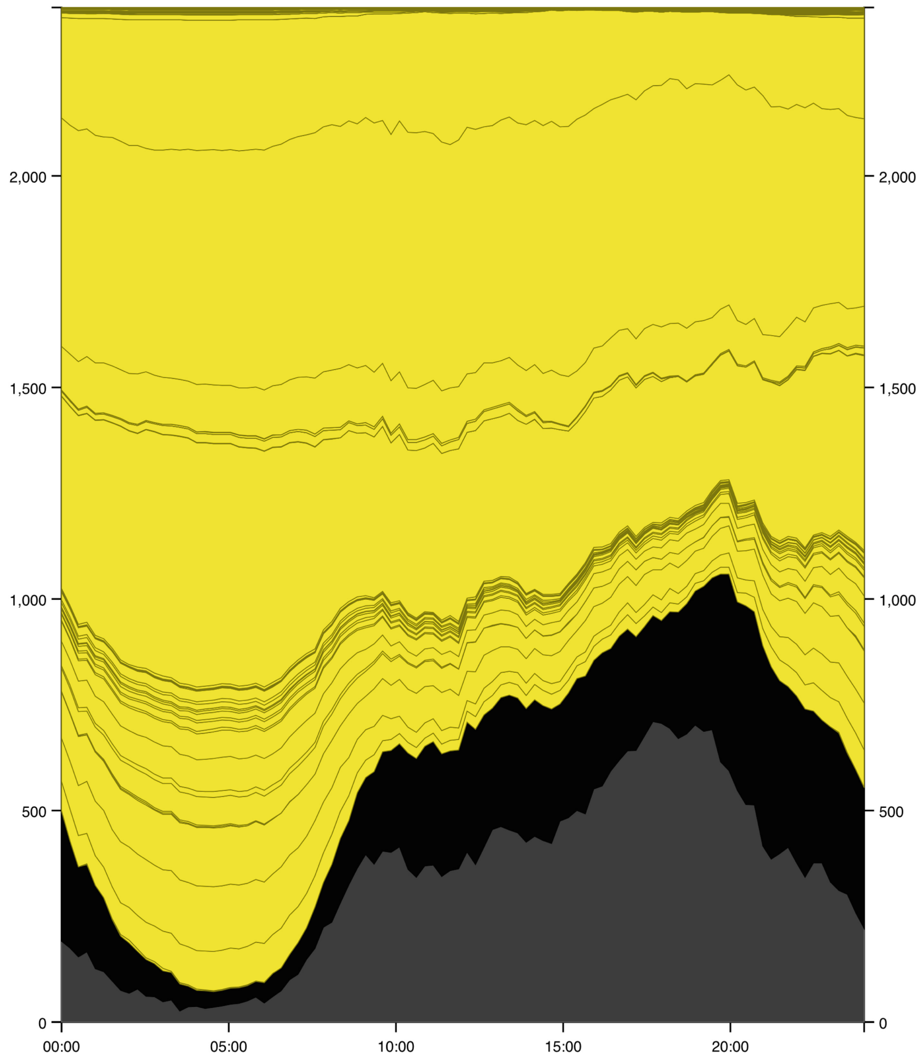

Fig. 4. Reconstruction of the actual accumulation dynamics for the City of Chicago.

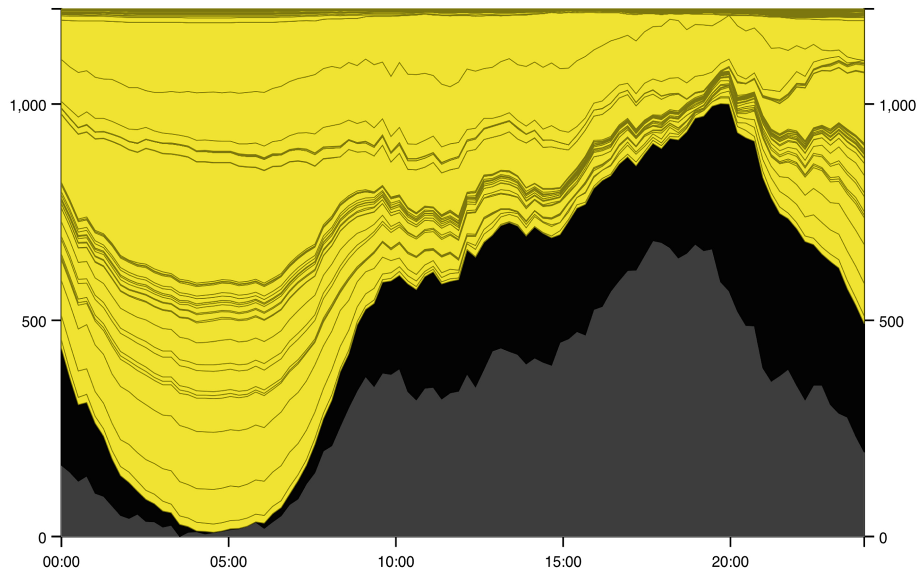

Fig. 5. Reconstruction of the actual accumulation dynamics for the City of Chicago.

In such scenario, an autonomous cab can immediately drive to pick up the next client, reducing the total VHP for the system. In particular, we showed that if each empty vehicle arrival to an idle area coincided with an empty vehicle departure from the same area, the total fleet could reduce by 50% assuming that empty trip duration is not increasing. While our results show that the fleet can reduce by roughly 50%, this reduction reflects only the reduction in vehicle rent. The total cost change i Atefeh Mahdavi Goloujeh <amahdavi@uncc.edu>s different. In particular, the total work that must be done to mobilize full and empty trips (VHT and VHD) either does not change or it increases, assuming that AVs must travel longer distances for empty trips. What decreases instead is the rent that must be paid for locations and parked vehicles (measured as VHP). Second, a strength of the presented method is that, in contrast to existing approaches that rely on solving routing for calculating cost, we showed that by using a very simple assumption, we can show that using MoD, we can significantly reduce the cost of travel in our system by enabling cars to stay idle and wait for other passengers rather than searching for passengers. Constructing empty trips data from full trips allow us to understand approximately how efficient and cost effective the mobility system in Chicago is. First, it shows the true cost of traveling by taxi as the distribution in time of VHP, VHT, and VHD. The main current limitation of the method is that the classification of combined durations into empty trips and idle periods uses an average empty trip duration which is obtained empirically, by observing the shape of the bimodal distribution of combined durations. A future improvement of this approach involves using Machine Learning (ML) techniques to distinguish between empty trips and idle periods.

Acknowledgments. The research presented in this paper has been partially funded by UNCC's Faculty Research Grant (funding cycle 2018-2019). The Chicago taxi trip dataset is publicly available and can be downloaded via the Socrata Open Data API through the following link: https://dev.socrata.com/foundry/data.cityofchicago.org/wrvz-psew. Figure 2 credits: Atefeh Mahdavi Goloujeh.

References

1. Shaheen, S., Cohen, A.: Innovative Mobility Carsharing Outlook. Transportation Sustainability Research Center, University of California, Berkeley (2016)
2. Steininger, K., Vogl, C., Zettl, R.: Car-sharing organizations: the size of the market segment and revealed change in mobility behavior. Transp. Policy **3**, 177–185 (1996)
3. Shaheen, S., Rodier, C.: Travel effects of a suburban commuter carsharing service: CarLink case study. Transp. Res. Rec. J. Transp. Res. Board. **1927**, 182–188 (2005)
4. Spieser, K., Treleaven, K., Zhang, R., Frazzoli, E., Morton, D., Pavone, M.: Toward a Systematic Approach to the Design and Evaluation of Automated Mobility-on-Demand Systems: A Case Study in Singapore. Frazzoli (2014)
5. Fagnant, D.J., Kockelman, K.M.: The travel and environmental implications of shared autonomous vehicles, using agent- based model scenarios. Transp. Res. Part C **40**, 1–13 (2014)
6. Chen, T.D., Kockelman, K.M., Hanna, J.P.: Operations of a shared, autonomous, electric vehicle fleet: implications of vehicle & charging infrastructure decisions. Transp. Res. Part A **94**, 243–254 (2016). https://doi.org/10.1016/j.tra.2016.08.020
7. Papanikolaou, D.: The Potential of On-demand Urban Mobility: Lessons from System Analysis and Data Visualization (2016)
8. Fagnant, D.J., Kockelman, K.M., Bansal, P.: Operations of shared autonomous vehicle fleet for Austin, Texas. Market. Transp. Res. Rec. J. Transp. Res. Board. **2536**, 98–106 (2015). https://doi.org/10.3141/2536-12
9. Papanikolaou, D.: Data-driven state space reconstruction of mobility on demand systems for sizing-rebalancing analysis. In: Proceedings of the 2018 Symposium on Simulation for Architecture and Urban Design (SimAUD 2018). Technical University of Delft, Netherlands (2018)
10. George, D., Xia, C.: Fleet-sizing and service availability for a vehicle rental system via closed queueing networks. Eur. J. Oper. Res. **211**, 198–207 (2011). https://doi.org/10.1016/j.ejor.2010.12.015
11. Dell'Amico, M., Hadjicostantinou, E., Iori, M., Novellani, S.: The bike sharing rebalancing problem: mathematical formulations and benchmark instances. Omega **45**, 7 (2014)
12. Chemla, D., Meunier, F., Wolfler Calvo, R.: Bike sharing systems: solving the static rebalancing problem. Discrete Optim. **10**, 120–146 (2013). https://doi.org/10.1016/j.disopt.2012.11.005
13. Kloimüllner, C., Papazek, P., Hu, B., Raidl, G.R.: Balancing bicycle sharing systems: an approach for the dynamic case. In: Blum, C., Ochoa, G. (eds.) EvoCOP 2014. LNCS, vol. 8600, pp. 73–84. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44320-0_7
14. Pavone, M., Treleaven, K., Frazzoli, E.: Fundamental performance limits and efficient polices for transportation-on-demand systems. In: 49th IEEE Conference on Decision and Control, CDC 2010, pp. 5622–5629 (2010). https://doi.org/10.1109/cdc.2010.5717552
15. McNally, M.G., Rindt, C.: The Activity-Based Approach (2008)

16. Liu, Y., Wang, F., Xiao, Y., Gao, S.: Urban land uses and traffic ‘source-sink areas’: evidence from GPS- enabled taxi data in Shanghai. Landsc. Urban Plan. **106**, 73–87 (2012)
17. González, M.C., Hidalgo, C.A., Barabási, A.-L.: Understanding individual human mobility patterns. Nature **453**, 779 (2008)
18. Song, C., Koren, T., Wang, P., Barabási, A.-L.: Modelling the scaling properties of human mobility. Nat. Phys. **6**, 818 (2010). https://doi.org/10.1038/nphys1760
19. Chowell, G., Hyman, J.M., Eubank, S., Castillo-Chavez, C.: Scaling laws for the movement of people between locations in a large city. Phys. Rev. Stat. Nonlinear Soft Matter Phys. **68**, 066102 (2003)
20. Liang, X., Zheng, X., Lv, W., Zhu, T., Xu, K.: The scaling of human mobility by taxis is exponential. Phys. Stat. Mech. Appl. **391**, 2135–2144 (2012)
21. Phithakkitnukoon, S., Veloso, M., Bento, C., Biderman, A., Ratti, C.: Taxi-aware map: identifying and predicting vacant taxis in the city. In: de Ruyter, B., et al. (eds.) AmI 2010. LNCS, vol. 6439, pp. 86–95. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16917-5_9
22. Kang, C., Ma, X., Tong, D., Liu, Y.: Intra-urban human mobility patterns: an urban morphology perspective. Phys. Stat. Mech. Appl. **391**, 1702–1717 (2012)
23. Scholz, R.W., Lu, Y.: Detection of dynamic activity patterns at a collective level from large-volume trajectory data. Int. J. Geogr. Inf. Sci. **28**, 1–18 (2014)
24. Sun, J.B., Yuan, J., Wang, Y., Si, H.B., Shan, X.M.: Exploring space–time structure of human mobility in urban space. Phys. Stat. Mech. Appl. **390**, 929–942 (2011)
25. Kung, K.S., Greco, K., Sobolevsky, S., Ratti, C.: Exploring universal patterns in human home- work commuting from mobile phone data. PLoS ONE **9**, e96180 (2014)
26. Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., Ratti, C.: Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. Transp. Res. Part C **26**, 301–313 (2013). https://doi.org/10.1016/j.trc.2012.09.009
27. Liu, X., Gong, L., Gong, Y., Liu, Y.: Revealing travel patterns and city structure with taxi trip data. J. Transp. Geogr. **43**, 78–90 (2015). https://doi.org/10.1016/j.jtrangeo.2015.01.016
28. Padgham, M.: Human movement is both diffusive and directed. PLoS ONE **7**, e37754 (2012)
29. Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., Banchs, R.: Urban cycles and mobility patterns: exploring and predicting trends in a bicycle- based public transport system. Pervasive Mob. Comput. **6**, 455–466 (2010). https://doi.org/10.1016/j.pmcj.2010.07.002
30. Vogel, P., Greiser, T., Mattfeld, D.C.: Understanding bike-sharing systems using data mining: exploring activity patterns. Procedia - Soc. Behav. Sci. **20**, 514–523 (2011)
31. Andrienko, N., Andrienko, G.: Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach. Springer, Heidelberg (2006). https://doi.org/10.1007/3-540-31190-4
32. Shalizi, C.R.: Methods and techniques of complex systems science: an overview. In: Thomas, D., Yasha, K. (eds.) Complex Systems Science in Biomedicine, pp. 33–114. Springer, New York (2006). https://doi.org/10.1007/978-0-387-33532-2_2
33. Haghani, A., Lee, S.Y., Byun, J.H.: A system dynamics approach to land use/transportation system performance modeling part I: methodology. J. Adv. Transp. **37**, 1–41 (2003)
34. Batty, M.: The New Science of Cities. MIT Press, Cambridge (2013)
35. Forrester, J.W.: Urban Dynamics. MIT Press, Cambridge (1969)