Dual-Horizon Forecasts and Repositioning Strategies for Operating Shared Autonomous Mobility Fleets

Florian Dandl
Bundeswehr University Munich
Institute for Intelligent Transportation Systems
Werner-Heisenberg-Weg 39
85779 Neubiberg, Germany
Email: florian.dandl@unibw.de
ORCiD: 0000-0003-3706-9725

Michael Hyland
University of California-Irvine
Civil and Environmental Engineering
Institute of Transportation Studies, 4000 Anteater Instruction and Research Bldg. (AIRB)
Irvine, CA 92697-3600
Phone: (949) 824-5084
Email: hylandm@uci.edu
ORCiD: 0000-0001-8394-8064

Klaus Bogenberger
Bundeswehr University Munich
Institute for Intelligent Transportation Systems
Werner-Heisenberg-Weg 39
85779 Neubiberg, Germany
Email: klaus.bogenberger@unibw.de

Hani S. Mahmassani (Corresponding Author)
William A. Patterson Distinguished Chair in Transportation, Northwestern University
215 Chambers Hall,
600 Foster Street,
Evanston, IL 60208
Phone: (847)-467-2276; Fax: (847)-491-3090
Email: masmah@northwestern.edu

Word count: 6,959 words text + 2 tables x 250 words (each) = 7,459 words
6 figures

Submitted for *presentation* at the 99th Annual Meeting of the Transportation Research Board and *publication* in the Transportation Research Record

Submission Date: August 1, 2019

ABSTRACT

The goal of this study is to improve the operational performance of shared-use automated vehicle (AV) mobility services (SAMSs) via combining less frequent mid-term demand forecasts and vehicle repositioning (i.e. 1-3 hour look-ahead every 15 minutes) with short-term demand forecasts and vehicle repositioning (i.e. 15 minute look-ahead every 30 seconds). The short-term forecasts help the fleet controller reposition vehicles to handle small spatial discrepancies in vehicle supply and travel demand, in the near term over a limited spatial scale. Incorporating mid-term forecasts allows the fleet controller to reposition vehicles to overcome larger systemic spatio-temporal demand imbalances that exist over large spatial distances in the service region (e.g. the regional airport and the central business district). This study proposes a new math programming formulation for the mid-term repositioning policy, in addition to an existing joint user-vehicle assignment and short-term repositioning policy. To test the ability of the mid-term forecast and repositioning strategy to improve the SAMS's operational performance, this study employs an agent-based simulation model and runs a variety of scenarios using taxi data from Manhattan, NY and transportation network company data from Chicago, IL. The results indicate that incorporating mid-term repositioning strategies substantially improves the SAMS fleet performance. Without the mid-term repositioning strategy, many vehicles become stuck in low-demand areas over the course of the 7-day simulations resulting in the fleet serving less than 40% of requests; whereas, the dual-horizon approach serves 90% of requests with the same fleet size, at the expense of more empty vehicle miles.

Keywords: Autonomous Vehicles, Mobility Service, Autonomous Mobility On-Demand, Demand Forecasting, Fleet Operations

INTRODUCTION

The emergence and growth of shared-use vehicle-based mobility service providers (MSPs) like Uber, Lyft, and Didi as well as the expected advent of fully-automated vehicles (AVs) has motivated significant research related to shared-use AV-enabled mobility services (SAMSs), including at the fleet operations level. Several studies aim to maximize the operational efficiency of SAMS fleets via proposing and comparing operational strategies (*1*–*9*). These studies focus on the online matching or dynamic assignment of vehicles to user requests and some also incorporate empty vehicle repositioning strategies. However, as far as the authors are aware, no SAMS fleet operational strategy exists in the literature that includes multiple time horizons for forecasting demand and repositioning empty vehicles. This study aims to fill this gap and improve upon the state-of-the-art SAMS fleet operational strategies. Although Mitrovic-Minic et al. (*10*) propose a double horizon approach to dynamically operate a vehicle fleet, their approach applies to the dynamic courier delivery problem which is quite different from operating a SAMS fleet serving on-demand passenger requests without shared rides.

Like most firms, managerial decisions for MSPs can be separated into different time scales. At the longest term scale (annually to quarterly), the MSP must decide what type(s) of services to offer and how to price the services. At the second longest term scale (monthly to weekly), the MSP must determine the appropriate fleet size, assuming the MSP owns and operates its own vehicles (*13*–*15*). At the third longest term scale (weekly to daily), the MSP must decide when and where to perform maintenance and vehicle recharging/refueling. At the mid-term scale (hourly), the MSP must consider repositioning vehicles, long distances, to decrease systemic imbalances arising from uneven spatio-temporal demand distributions. Finally, at the short-term scale (minutes to seconds), the MSP must assign AVs to user requests and reposition AVs to adjust for small imbalances between supply and demand. The decisions made at each of these time scales are highly inter-related.

In theory, making current decisions in order to optimize an objective function considering both current and future rewards within a time horizon can be addressed by (approximate) dynamic programming (*11, 12*). However, the scale of problems created by existing mobility services and complicated demand-supply interactions require approximations in order to allow real-time performant algorithms.

Maciejewski et al. (*2*) and Hyland and Mahmassani (*1*) present several strategies to assign vehicles to user requests. Other research focuses on pooled services where multiple users share a ride at the same time (*8, 16*). Research also shows that the SAMS service design can constrain the vehicle-user assignment optimization potential (*5*). Even the time that the operator and users need to make decisions affects the fleet performance (*6*).

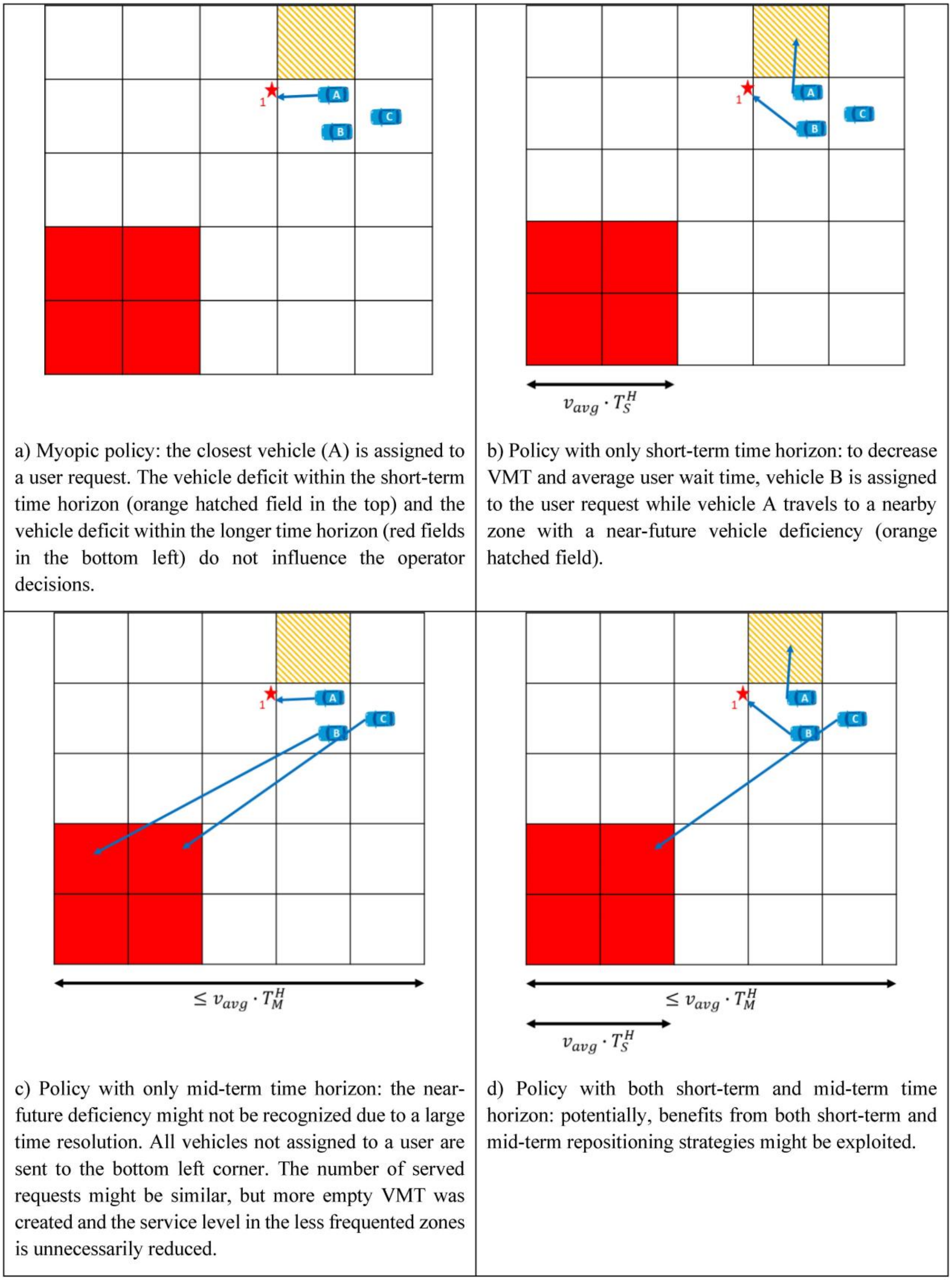
These vehicle-user assignment strategies involve solving static optimization problems periodically. Since new requests are revealed over time, the available information changes over time and suboptimal decisions of a previous optimization period can lead to better system states over the course of time. This is the dynamic nature of the SAMS operation problem (*17, 18*). In order to make decisions that are beneficial for the near future, waiting strategies can be applied to save empty vehicle miles traveled (VMT) in case users are not very time sensitive (*19, 20*). Another approach to address the dynamic nature of the SAMS operation problem and improve myopic vehicle-user assignments is to utilize short-term demand forecasts and jointly assign vehicles to users and reposition tasks (*9, 21*).

Dandl et al. (*21*) argue that choosing spatial forecast resolutions should consider fleet performance rather than just forecast errors and that forecast accuracy significantly impacts performance at high spatial resolutions. Wen et al. (*22*) discuss more generally about the value of demand information and conclude from simulations that aggregate demand information are valuable to SAMS operators. Sayarshad and Chow (*23*) survey common short term demand forecast methods. Moreira-Matias et al. (*24*) use online streaming data to improve the forecast quality of a historic average model.

Finally, there are studies that separate the repositioning problem from the user-assignment problem altogether (*7*, *25*, *26*). These approaches typically solve an optimization problem to determine the number of vehicles that should reposition from one area to another. However, these mid-term reposition strategies do not affect current assignments and do not consider the uncertainty of real demand forecasts.

The main contribution of this research involves using a dual-horizon (short-term and mid-term) demand forecasting and AV repositioning strategy to improve the operational efficiency of SAMS fleets via anticipating and responding to (i) local and short-term fluctuations in the vehicle supply and traveler demand and (ii) systemic spatio-temporal demand imbalances across the service region. The motivation for the dual-horizon approach is summarized in FIGURE 1.

The following section defines the research problem and hypothesizes why incorporating mid-term forecasts and repositioning should improve the efficiency of SAMS fleets. The next section presents the research methodology. The following section presents the experiments designed to test the operational policy. The penultimate section presents computational results and the final section concludes the paper with a summary and list of potential extensions.



a) Myopic policy: the closest vehicle (A) is assigned to a user request. The vehicle deficit within the short-term time horizon (orange hatched field in the top) and the vehicle deficit within the longer time horizon (red fields in the bottom left) do not influence the operator decisions.

b) Policy with only short-term time horizon: to decrease VMT and average user wait time, vehicle B is assigned to the user request while vehicle A travels to a nearby zone with a near-future vehicle deficiency (orange hatched field).

c) Policy with only mid-term time horizon: the near-future deficiency might not be recognized due to a large time resolution. All vehicles not assigned to a user are sent to the bottom left corner. The number of served requests might be similar, but more empty VMT was created and the service level in the less frequented zones is unnecessarily reduced.

d) Policy with both short-term and mid-term time horizon: potentially, benefits from both short-term and mid-term repositioning strategies might be exploited.

FIGURE 1 Motivation for dual time-horizon approach for SAMS fleet operation.

SERVICE DESCRIPTION AND PROBLEM DEFINITION

SAMSs can have many different service specifications regarding user-operator interaction (who accepts/rejects, communication response times, and communicated information) and service considerations (pricing, ridepooling and vehicle capacity, waiting and detour time limitations, and ability to re-assign). The concept of operations and the underlying operational problem for the SAMS in this study is described as follows:

The fleet controller continuously receives new on-demand requests from users. An incoming request $r = (o_r, d_r, \tau_r)$ at time τ_r contains the origin o_r and the destination d_r of a user. These users are not willing to share their ride and only accept wait times smaller than $T^{w,max}$; moreover, the operator must accept or decline a user request within one minute. The price of the SAMS in this study is exogenous and solely distance based; a user must pay f^D per kilometer between her origin and destination d_r^{od}. In the envisioned SAMS, the operator can re-assign a user to another vehicle up to three minutes before this user's currently communicated pick-up time. After that, the vehicle-user assignment is locked and cannot be changed anymore. Additionally, this study assumes an environment in which users, whose request is rejected, are unforgiving and unwilling to use the service in the future. This expected future loss is modeled by a large assignment reward ξ.

The objective of the studied SAMS is to maximize the sum of profit and additional reward for satisfying the wishes of served users $r \in R^S$, in Eqn. 1.

$$Obj = \left(\sum_{r \in R^S} \xi + f^D \cdot d_r^{od} \right) - \sum_{v \in V} c^D \cdot d_v \qquad (1)$$

To maximize the objective function, the SAMS fleet controller can dynamically assign vehicles at each decision period: to new requests, to reposition to nearby zones, or reposition to far away zones, or remain in their current location.

RESEARCH METHODOLOGY

Vehicle-User Assignment

An operator only needs to consider the set of active requests R^A consisting of new requests and the ones with unlocked vehicle-user assignments. The operator re-optimizes vehicle assignments every $T^{D,s} = 30$ seconds, the short-term decision time step. Requests that were received within the last optimization period are either accepted or rejected in order to keep the response time within the limit.

The operator checks the availability of its vehicles. Vehicle availability refers to the position p_v^{av} and time t_v^{av} at which vehicle v can be assigned to new tasks. When a vehicle has locked tasks, these are the position at the end of the vehicle's final task and the expected arrival time there. Idle vehicles in the set V^I, and vehicles without locked assignment constitute the set V^A of vehicles available right away at their current position.

At time t, the operator can compute the estimated waiting time for a hypothetical vehicle-user assignment of vehicle v with user request r by

$$t_{vr} = t[p_v^{av} \rightarrow o_r] + t_v^{av} - t \qquad (2)$$

where $t[p_v^{av} \rightarrow o_r]$ denotes the travel time from p_v^{av} to o_r. We define d_{vr} to be the distance associated with that path and create a decision variable x_{vr} if $t_{vr} \leq T^{w,max}$, the defining property of the set $V(r)$ of vehicles capable of serving request r. New user assignments are rewarded with the sum of the revenue of a user trip and the assignment reward ξ. Rejecting a previously assigned user is practically unacceptable and is penalized by a very large value of Ξ. This study defines χ_r to be ξ for new and Ξ for previously assigned user requests r. The vehicle-user profit optimization then reads:

$$\max_{x_{vr}} \sum_{r \in R^A} \sum_{v \in V(r)} \left(\chi_r + (f^D - c^D) d_r^{od} - c^D d_{vr}\right) x_{vr} \tag{3a}$$

$$\sum_{v \in V(r)} x_{vr} \leq 1 \qquad \forall r \in R^A \tag{3b}$$

$$\sum_{r \in R^A} x_{vr} \leq 1 \qquad \forall v \in V \tag{3c}$$

$$x_{vr} \in \{0,1\} \qquad \forall v \in V, \forall r \in R^A \tag{3d}$$

The constraints limit the number of assignments for each request to one (3b) and the number of unlocked assignments per vehicle to one (3c).

Combined Vehicle-User and Short-Term Reposition Assignment

An operator can use short-term demand forecasts to make decisions considering current and expected future demand within a time horizon $T^{H,s}$. In the following, a superscript s/m will denote short-term/mid-term variables, respectively. The additional demand information can affect vehicle-user assignments as illustrated in FIGURE 1b. In order to make predictions about demand, the operator typically divides the operating area into a set A of disjoint areas a. When the operator needs to make decisions at time t, it predicts the number of trip requests originating from an area a between t and $t + T^{H,s}$, which we denote by λ_a^s. For $T^{H,s}$ in the range of the average trip duration, the vehicle supply can be predicted well by the vehicle availability constrained by current assignments. For each zone, we count the number of vehicles that will be available before the time horizon ends ($t_v^{av} \leq t_i + T^{H,s}$). We define the short-term imbalance of zone a as the difference between supply and demand:

$$I_a^s = \left(\sum_{\substack{v: p_v^{av} \in a \\ t_v^{av} \leq T^{H,s}}} 1 \right) - \lambda_a^s \qquad \forall a \in A \tag{4}$$

We divide the operating area A into three disjoint subsets: A_+^s, A_0^s, and A_-^s, where the subscript indicates vehicle surplus (+), deficit (-) or balanced supply and demand (0). Due to uncertainty in forecasts, it is reasonable to define the balanced interval to be bigger than just zero, which can be done by introducing a vehicle-imbalance buffer B^s (for short-term strategy):

$$A_+^s = \{a \in A: I_a^s > B^s\} \tag{5a}$$

$$A_-^s = \{a \in A: I_a^s < -B^s\} \tag{5b}$$

While it makes sense to assign a new user request to a vehicle currently en-route to drop off a user in order to make an accept/reject decision, it is unnecessary to consider these vehicles for repositioning at the current decision time step.

Since the demand forecast does not extend beyond the time horizon, it is possible that a zone showing vehicle deficit within $T^{H,s}$ no longer is a deficit zone thereafter. Hence, it only makes sense to assign a vehicle reposition when vehicle v would arrive in zone a before the time horizon ends, i.e. set the decision variable z_{va} to zero if the vehicle would arrive after the time horizon:

$$z_{va} = 0 \qquad \forall v \in V^A, \forall a \in A: t_{va} = t[p_v^{av} \to a] > T^{H,s} \tag{6}$$

Furthermore, the set of decision variables z_{va} can be reduced by acknowledging that vehicle repositioning should be from surplus areas $o \in A_+^s$ to deficit areas $a \in A_-^s$. Therefore, we denote the set of currently available vehicles within all surplus areas by V_+^A and a specific surplus area a_+ by $V_{a_+}^A$. Not moving more vehicles to/from an area than there is vehicle deficit/surplus (including the buffer) generates additional constraints, namely equations (7d/e). Finally, the reward for an assignment is set to the sum of ξ and the expected profit from the additional trip, which we approximate with the average profit generated by a user trip P^{avg}. Finally, we introduce a discount factor $\gamma^s \in [0,1]$ to weigh future rewards against current rewards of actual request assignments and costs of driving.

Equations (7a-f) display the joint AV-user assignment and short-term repositioning problem:

$$\max_{x_{vr}, z_{va}} \sum_{r \in R^A} \sum_{v \in V(r)} u_{vr} \cdot x_{vr} + \sum_{v \in V_+^A} \sum_{a \in A_-^s} (\gamma^s(\xi + P^{avg}) - c^D d_{va}) z_{va} \tag{7a}$$

$$\sum_{v \in V(r)} x_{vr} \leq 1 \qquad \forall r \in R^A \tag{7b}$$

$$\sum_{r \in R^A} x_{vr} + \sum_{a \in A_-^s} z_{va} \leq 1 \qquad \forall v \in V \tag{7c}$$

$$\sum_{v \in V_+^A} z_{va} \leq -(I_a^s - B^s) \qquad \forall a \in A_-^s \tag{7d}$$

$$\sum_{v \in V_{a+}^A} \sum_{a \in A_-^s} z_{va} \leq I_a^s - B^s \qquad \forall a_+ \in A_+^s \tag{7e}$$

$$x_{vr} \in \{0,1\}, z_{va} \in \{0,1\} \qquad \forall v \in V, r \in R^A, a \in A_-^s \tag{7f}$$

where u_{vr} denotes the reward/cost associated with x_{vr} from the user-assignment problem equation (3a) and d_{va} the distance between vehicle v and area a. Equation (7b) limits the number of assignments per user to one, while equation (7c) constrains the number of unlocked assignments, which can be either a user-assignment or a reposition-assignment to another zone, to one.

In this study, we randomly assign a node within area a as the destination for a reposition assignment to a and use the average distance of the fastest paths from the vehicle position to each node in a to approximate d_{va} (the same method is used to determine t_{va}).

Mid-Term Reposition Assignments

Short-term repositioning uses a time horizon $T^{H,s}$ in the range of average trip duration, which allows a very good prediction of the supply side since the current assignments determine vehicle availability very well. However, the short-time horizon limits the capability of balancing the fleet over larger operating areas. Vehicles will not reach distant areas in time and equation (6) will prohibit repositions.

Considering a longer time horizon, called the mid-term time horizon, $T^{H,m}$ will allow vehicles to reach distant areas; however, future predictions are less reliable. While the quality of the prediction of future requests per area depends on the underlying demand patterns and could potentially even become better (which usually it does not), the troublesome part is the prediction of vehicle supply throughout that time horizon. Vehicles will be able to serve multiple trip requests during the longer time horizon and their current availability at t is not very valuable as a long-term supply forecast (i.e. at $t + 30$ minutes). The availability of vehicles at a future point in time is also dependent on all decisions between the present and then, which becomes infeasible to predict. An approximation to determine the imbalance inside an area $a \in A$ is to predict the number of requests λ_a^m and the number of vehicle arrivals (at destinations of user-trips) μ_a^m within the time horizon: $I_a^m = \mu_a^m - \lambda_a^m$.

Since serving current demand should be prioritized over possible future rewards, we separate the mid-term repositioning problem from the short-term problem. As the forecasts of supply and demand over this longer time horizon are not prone to change every 30 seconds, it is sufficient to solve mid-term reposition problems less frequently. We denote the mid-term interval as $T^{D,m}$. We introduce a discount factor $\gamma^m \in [0,1]$, which generally should be smaller than γ^s because possible rewards are even further in the future.

Moreover, we define sets for surplus, balanced, and deficit areas A_+^m, A_0^m, A_-^m, respectively. Even though the formal definition of the areas is the same (replacing $s \to m$ in equations (5a) and (5b)), the buffer B^m to account for forecast uncertainty should scale with the size of the forecasted values. The scale of the buffer should also reflect the current utilization of the fleet: if there are hardly any idle vehicles, the buffer should prohibit extensive repositioning. We choose a simple linear function to account for both effects:

$$B^m = \alpha \cdot \left(\frac{1}{2|A|} \sum_{a \in A} \mu_a^m + \lambda_a^m \right) \cdot \frac{|V|}{|V^I|} \tag{8}$$

The second factor scales the buffer with the average forecast value of both requests and arrivals and the third factor is the inverse share of idle vehicles. The hyper-parameter α determines the “aggressiveness” of the repositioning strategy: a large value expects higher uncertainty in the forecasts and therefore generates less reposition assignments.

In order to reduce computational complexity, we propose an area-based optimization problem instead of a vehicle-based optimization problem. The number of repositions from an area o are constrained either by the number of idle vehicles within this area or by the expected vehicle surplus. We define the constraining condition for zone $o \in A_+^m$ by

$$C_o^+ = \min(|V_{o+}^A|, I_o^m - B^m) \qquad \forall o \in A_+^m \tag{9}$$

The number of vehicles n_{od} that will be repositioned from area $o \in A_+^m$ to area $d \in A_-^m$ should not be rewarded equally for each missing or surplus vehicle. Instead, the reward should depend on the magnitude of the vehicle imbalances in o and d. We propose to include the following surplus (F_o^+) and deficit (F_d^-) share factors in the reward formulation:

$$F_o^+ = 1 + C_o^+ \Big/ \sum_{a \in A_+^m} C_a^+ \qquad \forall o \in A_+^m \tag{10a}$$

$$F_d^- = 1 + (B^m - I_d^m) \Big/ \left(\sum_{a \in A_-^m} B^m - I_a^m \right) \qquad \forall d \in A_-^m \qquad (10b)$$

Linear Model

With these definitions, we formulate the linear mid-term repositioning model:

$$\max_{x_{vr}} \sum_{o \in A_+^m} \sum_{d \in A_-^m} (\gamma^m(\xi + P^{avg}) \cdot F_o^+ \cdot F_d^- - c^D d_{od}) n_{od} \qquad (11a)$$

$$\sum_{d \in A_-^m} n_{od} \leq C_o^+ \qquad \forall o \in A_+^m \qquad (11b)$$

$$\sum_{o \in A_+^m} n_{od} \leq B^m - I_d^m \qquad \forall d \in A_-^m \qquad (11c)$$

$$n_{od} = 0 \qquad \forall o \in A_+^m, d \in A_-^m : t_{od} > T^{H,m} \qquad (11d)$$

$$n_{od} \in \mathbb{N}_0^+ \qquad \forall o \in A_+^m, \forall d \in A_-^m \qquad (11e)$$

Constraint (11b) limits the number of repositions originating from an area, equation (11c) constrains the number of repositioning vehicles ending in an area, constraint (11d) only allows repositioning a vehicle if it arrives at the destination area within the time horizon. The travel distance d_{od} and time t_{od} are approximated by the average of the values of the respective fastest path from nodes in area o to nodes in area d.

Adding a repositioning trip between two areas trades off the reward and the distance-related costs in the objective function (11a). Due to the surplus and deficit share factors, the areas with the largest imbalances and vehicle stock are prioritized. The form of the factors in equation (11a) and (11b) is chosen such that the factor remains in the scale ~ 1 in order to keep the balance between total rewards and distance costs rather constant. FIGURE 2a) illustrates the effect of the buffer and the deficit share factors on the marginal reward of vehicle repositioning.

As a final step, the operator needs to assign the repositioning trips to vehicles. We use a greedy algorithm that sequentially chooses a random single trip from $o \in A_+^m$ to $d \in A_-^m$ from the solution to (11a-e) and assigns it to the idle vehicle in o, which has the shortest distance to region d. These mid-term assignments are locked in order to keep them from being changed during the next short-term decision process.

Non-linear model

Even with the surplus and deficit share factors, the linear model cannot consider how many trips have already been assigned to a deficit area as long as there is still a deficit and all trips between an origin and a destination are rewarded the same. It makes sense to value the first assignments to a deficit area stronger than the following ones, i.e. make the marginal reward of an additional hypothetical assignment dependent on other assignments to this area, as shown in FIGURE 2b). For this reason, we replace the deficit share factor with the term $\beta(B^m - I_d^m - n_{od})$, thereby generating a quadratic term in the objective function.

$$\max_{x_{vr}} \sum_{o \in A_+^m} \sum_{d \in A_-^m} (\gamma^m (\xi + P^{avg}) \cdot F_o^+ \cdot \beta (B^m - I_d^m - n_{od}) - c^D d_{od}) n_{od} \tag{12a}$$

$$\sum_{d \in A_-^m} n_{od} \leq C_o^+ \qquad \forall o \in A_+^m \tag{12b}$$

$$\sum_{o \in A_+^m} n_{od} \leq B^m - I_d^m \qquad \forall d \in A_-^m \tag{12c}$$

$$n_{od} = 0 \qquad \forall o \in A_+^m, d \in A_-^m : t_{od} > T^{H,m} \tag{12d}$$

$$n_{od} \in \mathbb{N}_0^+ \qquad \forall o \in A_+^m, \forall d \in A_-^m \tag{12e}$$

The meaning of the hyper-parameters, which helps to set an initial value and ultimately to control the trade-offs between balancing the fleet and creating additional driving costs, becomes less interpretable with the introduction of this nonlinearity. From a mathematical point of view, β is not required if we allow choosing γ^m different from the linear model.

The constraints of the optimization problem remain the same as in the linear case.

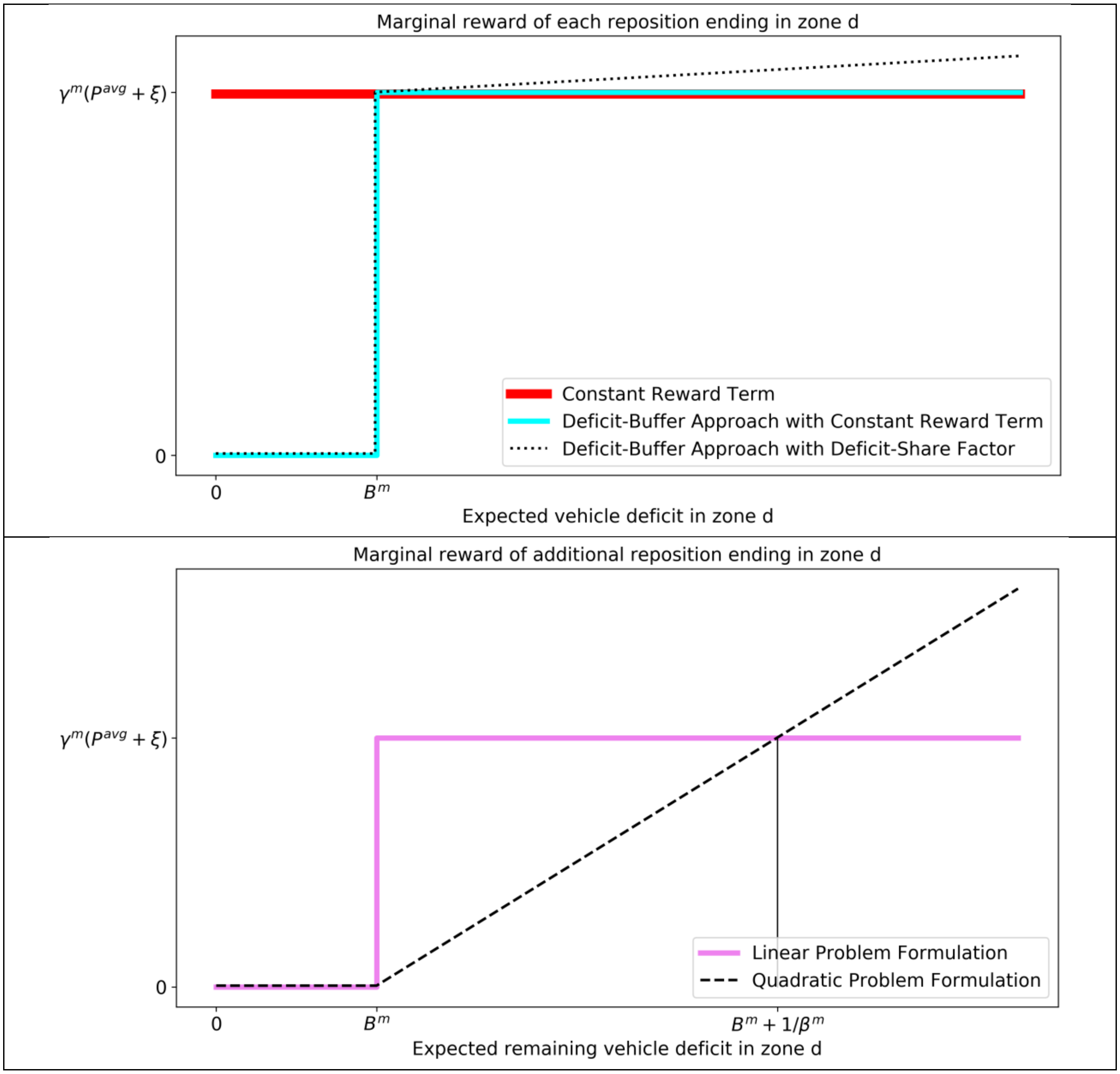


FIGURE 2 Marginal benefit of mid-term repositioning depending on the initial expected vehicle deficit of the zone (top) and the remaining vehicle deficit depending on other repositioning assignments to that zone (bottom).

Forecast Methodology

The previously described SAMS operator strategies require frequent demand forecasts for the time horizon $T^{H,s}$, and less frequent forecasts for number of requests and vehicle arrivals for time horizon $T^{H,m}$.

In this study, we implement an online model correcting historic day-of-week period-of-day forecasts and test the operating strategies with perfect forecasts to test the robustness of the fleet operator algorithms. The perfect forecasts are generated by aggregating the trip request data. Averaging the recorded trip numbers y of a period p for a weekday d of the past 6 weeks determines the historic data forecast $\bar{y}_{w,d,p}$:

$$\bar{y}_{w,d,p} = \frac{1}{6}\sum_{j=1}^{6} y_{w-j,d,p} \tag{13}$$

The online forecast aims to improve that value by comparing the day-of-week period-of-day forecast ($\bar{y}$) and recorded (y) trip data of the past few forecast periods. We assume an exponential decrease in importance of previous periods and use the values of the last four forecast periods:

$$\eta_j = \begin{cases} \frac{y_{wd,p-j}}{\bar{y}_{w,d,p-j}} & \bar{y}_{w,d,p-j} > 0 \\ 1 & \bar{y}_{w,d,p-j} = 0 \end{cases} \tag{14a}$$

$$\hat{y}_{w,d,p}(t_i) = y_{w,d,p} \cdot \left(\sum_{j=1}^{4} \eta_j \, e^{-j} \right) \tag{14b}$$

We expect that forecasts combining more advanced machine-learning methods and the inclusion other data sources for event and weather data would improve the quality of forecasts tremendously.

Agent-Based Simulation Framework

Our model contains three classes of agents: users, the operator, and the vehicles that the operator controls. A simulation starts at one second before 00:00 and runs for 24 hours in one-second time-steps. The initial vehicle positions and open tasks from the simulation of the previous day are used such that the end-of-day vehicle distribution matters.

FIGURE 3 displays the procedure of the simulation. In each time step, the system state is updated based on the current assignments. Vehicles drive for one of four reasons: to drive an on-board user from her origin to her destination, to pick-up a passenger, to perform a short-term repositioning trip, or to perform a long-term repositioning trip. When a vehicle arrives at a pick-up or drop-off location, users can start boarding or disembarking, respectively. The state update phase terminates with new users revealing their trip requests to the operator.

These trip requests are batched until the next short-term decision time step ($t\% T^{D,s} = 0$). Depending on the studied SAMS strategy, the operator assigns requests and possibly short-term repositions to vehicles, then the operator accepts a new user request if it was assigned or rejects it.

This framework assumes $T^{D,m} \,\%\, T^{D,s} = 0$ and possibly runs a mid-term decision time step after short-term assignments in order to not block user-assignments by locked mid-term repositions.

After the final time step, the vehicle positions and unfinished vehicle assignments are recorded for the simulation of the next day.

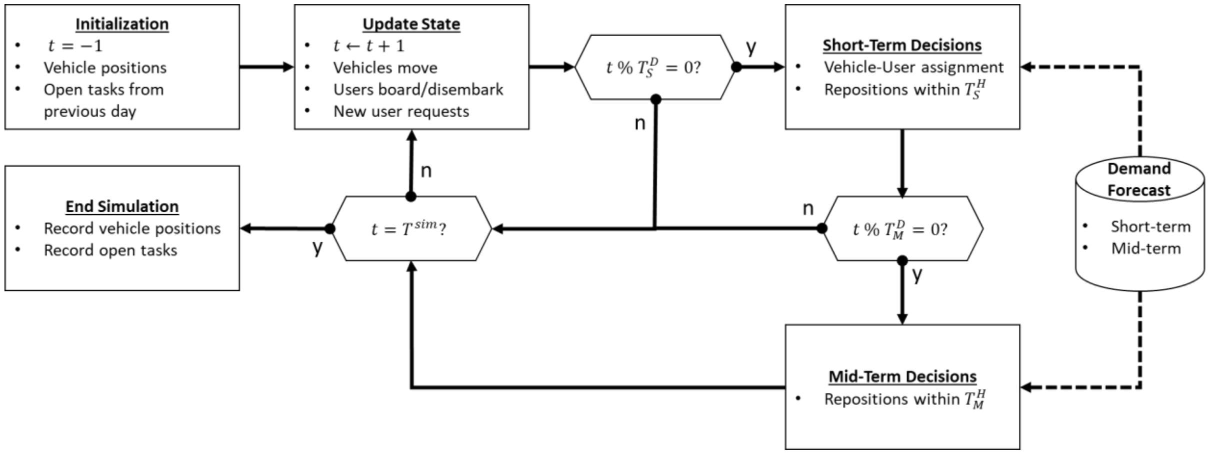


FIGURE 3 Flowchart describing the simulation framework.

EXPERIMENTAL DESIGN

We expect the benefit generated by the dual-horizon approach to be depending on the operating area and the demand patterns therein. Hence, we conduct simulations in Chicago and Manhattan as representatives of large cities with quite different spatial characteristics.

Chicago TNC and New York City Yellow Taxi Data

Demand forecasts and requests for the simulations are based on Chicago TNC data from 2019-02-01 to 2019-03-31 and New York City Taxi data from 2018-10-01 to 2018-11-30. The first six weeks of each data set are solely used to create the historic forecasts $\bar{y}_{w,d,p}$. This study sets the period for forecasts to the temporal accuracy of the Chicago TNC data, i.e. 15 minutes. The city of Chicago states privacy reasons for the 15-minute aggregation intervals, while the New York City taxi data provide time stamps down to seconds. We disaggregate the Chicago TNC data by random processes in order to get request times on the second level as well. The average recorded trip duration is approximately 13 minutes in Manhattan and 16 minutes in Chicago.

The set A of areas covering the respective operating areas were chosen based on the trip record data as well. Both cities publish user pick-up and drop-off locations not with GPS data, but on an area level, once again for privacy reasons. In Chicago, most trips contain origin and destination information at the census tract level. However, about 18% only have valid information on community area level and 12% do not have any spatial information. Hence, we decided to use the community areas as set A for Chicago. Furthermore, Chicago covers more than 600 km^2 and the division into 800 TAZ-level areas would likely lead to computational problems during the optimization processes. We filtered the New York City data set for trips originating and ending in Manhattan, thereby removing 16% of trips of the original data set. User pick-up and drop-off points are recorded within Taxi zones, which are therefore used as the set A for Manhattan. We applied two more filters to exclude possible trips with intermediary stops or even round trips and probably false records by removing trips with speeds below 1 m/s and above 30 m/s. After filtering, the trip data contain approximately 900k trips (2019-03-25 to 2019-03-31 in Chicago) and 1500k (2018-11-12 to 2018-11-18 in Manhattan) in the respective 1-week periods.

Both data sets show an imbalance between trip origins and destinations, as illustrated in FIGURE 4. Trip origins and destinations were counted for the respective 1-week period. Coincidently, the scales of imbalances after one, three, five and 7 days are similar in both cities. The highest imbalances in Chicago, namely downtown in the east and the O'Hare airport in the northwest are more than 25 km apart; as a comparison, the scales in Manhattan are much shorter with the width of Manhattan around Central Park only being approximately 3 km.

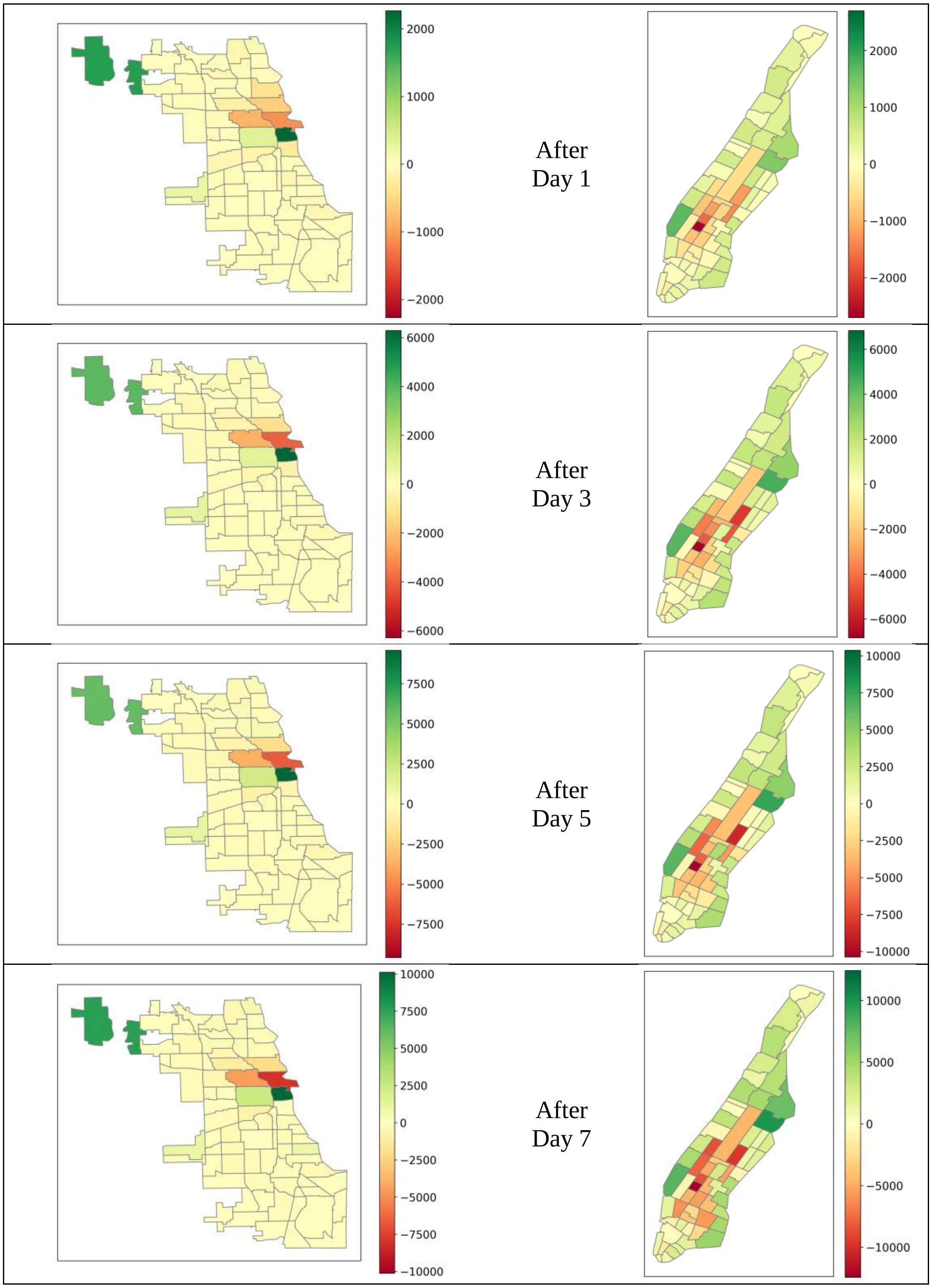


FIGURE 4 Cumulative recorded trip arrivals minus departures in Chicago (left) and in Manhattan (right).

Chicago and Manhattan Street Network Data

This study employs OpenStreetMap network data for both study areas and removes links that cannot be used by cars. Additionally, in Chicago, streets of type "living_street" and "residential" were removed to keep the number of links manageable. After the link filters, Chicago and Manhattan networks contain 3963 nodes with 10059 unidirectional links and 4498 nodes with 9757 unidirectional links, respectively. Since SAMS vehicles likely will not be allowed to pick-up or drop-off customers everywhere (e.g. on motorways), the simulation model only allows boarding at nodes that are only connected to primary, secondary, and tertiary streets in Chicago and additionally residential and living streets in Manhattan. This process also excludes streets that are unclassified or have multiple entries. In order to derive free flow travel times in the network, the speed limit is set at 30 mph (Chicago) and 25 mph (Manhattan) on links without speed limit information. The nodes were projected onto the areas A of the respective study area and the trip origins and destinations were randomly assigned to nodes within the recorded areas. In order to approximate realistic travel times, the velocities in the network were adapted to match the average velocities of trip records in 15-minute intervals.

After setting the origin and destination nodes, we evaluated the trip distance distribution. The average trip distance is 6.7 km in Chicago, with quartiles of trip distance at 2.5 km, 4.6 km and 8.4 km; in Manhattan, the average is 3.1 km and the quartiles are at 1.7 km, 2.6 km and 3.9 km respectively.

General Settings

Many parameters need to be set in the agent-based simulation model and additionally in the SAMS operating strategy models. Requests are generated based on the spatially projected trip data, where the request time is set to the start time of the original trip in Manhattan and randomly drawn from the start time interval in Chicago. The fare for users is \$0.5 per km on the fastest route between the user's trip origin and destination. The study assumes the actual costs of driving are \$0.25 per km and the fixed vehicle costs per day are \$25. For simplicity, the average user-profit parameter P^{avg} in the operator strategies is set to the average trip distance multiplied by \$0.25, thereby ignoring the costs of pick-up trips. Its value is \$1.68 in Chicago and \$0.77 in Manhattan. This study assumes a very high user dissatisfaction cost ξ of \$10, thereby modelling a system where users are likely not to use the system again if they cannot be picked-up within 6 minutes.

Scenarios

This study conducts 7-day scenarios/simulations by sequentially simulating single days and using the final vehicle positions and utilization as initial conditions for the simulations of the next day. Evaluating the number of concurrent travelers allows estimating the range of reasonable fleet sizes. This study compares the following operator strategies: "vehicle-user assignment only" (AO, equations 2a-d), "short-term combined vehicle-user and reposition strategy" (ST, equations 6a-f), "short-term combined vehicle-user strategy with mid-term linear optimization reposition" (ST-MTL, equations 6a-f and 10a-e) and "short-term combined vehicle-user strategy with mid-term quadratic optimization reposition" (ST-MTQ, equations 6a-f and 11a-e). Furthermore, the study compares different mid-term time-horizons and the impact of forecast accuracy on the mid-term repositioning policies. In total, this study conducts $2 \cdot 7 \cdot 3 \cdot (2 + 2 \cdot 3 \cdot 2) = 588$ single-day simulations. All general and scenario-specific parameters are summarized in TABLE 1.

TABLE 1 Parameters for Simulation Scenarios

Parameter	Math Notation	Values	Unit
General Parameter Settings			
Simulation Period		00:00 – 24:00	
Max. User Wait Time	$T^{w,max}$	6	min
User Re-Assignment Limitation Before Pick-Up		3	min
Distance-Based User Fare	f^D	0.50	\$/km
User Dissatisfaction Cost / Assignment Reward	ξ	10	\$
Daily Fixed Vehicle Cost Per Vehicle		25	\$
Vehicle Operating Cost	c^D	0.25	\$/km
Short-Term Decision Time Step	$T^{D,s}$	0.5	min
Short-Term Time Horizon	$T^{H,s}$	15	min
Short-Term Discount Factor	γ^s	0.5	
Short-Term Vehicle-Imbalance Buffer	B^s	1	veh
Mid-Term Decision Time Step	$T^{D,m}$	15	min
Mid-Term Discount Factor	γ^m	0.25	
Mid-Term Vehicle-Imbalance Buffer Coefficient	α	0.01	
Scenario Definition Parameters			
Study Area		Chicago (C), Manhattan (M)	
Simulated Days (Request Data Sets)		2019-03-[25 to 31] (C) 2018-11-[12 to 18] (M)	
Fleet Size	$\|V\|$	2750, 3250, 3750 (C) 4000, 4500, 5000 (M)	veh
SAMS Operating Strategy	**AO**: user assignment only **ST**: AO + short-term reposition **ST-MTL**: ST + mid-term linear reposition **ST-MTQ**: ST + mid-term nonlinear reposition		
Mid-Term Time Horizon	$T^{H,m}$	1,2,3	hour
Forecast Methodology		online, perfect	

RESULTS

This section briefly presents demand forecast results and then a detailed analysis and comparison of SAMS fleet operational strategies.

Forecasting Results

The demand forecasting results indicate superiority of the online model predictions over simply using historic data for both demand requests $\lambda_a^{s/m}$ and vehicle arrivals $\mu_a^{s/m}$. The analysis employs the root mean squared error per minute of the prediction time horizon (RSME/TH) rather than a relative error measure as the latter measure over-emphasizes zone-time pairs with just a few predicted requests/vehicle arrivals.

In Chicago, the RSME/TH over all time horizons is 0.42 for both the predictions of requests and vehicle arrivals. In Manhattan, the RSME/TH over all time horizons is 0.62 for vehicle arrival predictions and 0.71 for request predictions. As expected, the advantage of the online model forecast over a historic data forecast decreases with the length of the time horizon. The online correction improves the average RMSE/TH of requests and arrivals from 0.90 to 0.60 in Manhattan and from 0.55 to 0.42 in Chicago for 15-minute time horizon. The RMSE/TH value only decreases from 0.73 to 0.58 in Manhattan and from 0.40 to 0.37 in Chicago for a 3-hour time horizon. The larger forecast errors and the bigger gap between the historic data and the online forecast in Manhattan can partly be explained by a strong storm that caused part of public transit system to stop service.

SAMS Fleet Operational Performance Results

Since the total values of the objective function are hard to interpret, this study uses the definition of value of information (VOI) in Eqn. 15, to measure the relative gain in the objective function of strategy x using short/mid-term forecasts compared to the AO strategy (*10, 22*):

$$VOI[x] = \frac{Obj[x] - Obj[AO]}{Obj[AO]} \tag{15}$$

TABLE 2 displays the fleet performance results where fleet size is exogenous and only the best mid-term horizon length results are shown. In both cities, the AO strategies struggle to serve demand; many vehicles get stuck at the airport of Chicago and in the northeast part of Manhattan, which are outside of the maximum waiting time range of demand centers. The addition of short-term repositioning minimally impacts the results in Chicago, but more than doubles the rate of served requests in Manhattan while only producing 2-3% more empty VMT. In Manhattan, the ST strategy has a VOI greater than 1.5. The smaller benefit in Chicago relates to the airport being more than T^H away from the other demand centers.

TABLE 2 also indicates that the inclusion of a mid-term strategy significantly improves the objective function in both cities by orders of magnitude. In Manhattan, the choice of time horizon has a minimal impact on the VOI. However, in Chicago the linear model performs significantly better for a longer mid-term time horizon, while the nonlinear model shows similar performance for all three tested time horizons. Unfortunately, the large difference in the objective function comes at the cost of significant increases in empty VMT. The MTL scenarios with a 60-minute time horizon that produce considerably lower VMT are also the ones that serve many fewer requests. With the chosen hyper-parameters, the ST-MTQ generally produces more empty VMT, but the solution quality with perfect forecasts is always at least as good as the linear model. The operator should check whether a less sensitive MQT strategy, by modification of the hyper-parameters, could produce pareto-improving results.

The analysis also compares the performance of the strategies under forecast uncertainty using online forecasts and perfect forecasts. All key performance indicators of the ST-MTQ model show robust results indicating the usefulness of the vehicle imbalance buffer B^m. The empty VMT increases by TABLE 2approximately 1%, while serving a few more requests. A rather surprising result manifests for the ST-MTL model, as the online forecasts scenarios with errors perform better than perfect forecast scenarios. The inaccurate forecasts cause the operator to reposition more vehicles, which happens to improve the overall fleet performance. The reason for this unexpected result is that the exact destination area determined by the forecasts is not as important, as long as vehicles drive in the direction of the demand centers.

TABLE 2 Results of 7-Day Simulations for Selected Scenarios (Best Mid-Term Time Horizon of Strategy)

Study Area	Fleet Size	SAMS Operating Strategy	Mid-Term Horizon [hours]	Forecast Method	Value of Info	Profit $	Served Requests [%]	Empty VMT [%]	Average Wait Time [min]
Chicago	2750	AO	-	-	0	237k	33.3	6.4	5.1
		ST	-	perfect	0.03	225k	34.4	8.5	5.1
		ST-MTL	3	perfect	1.36	480k	78.7	17.0	4.7
			3	online	1.48	533k	83.1	19.6	4.7
		ST-MTQ	1	perfect	1.51	444k	85.2	23.5	4.4
			1	online	1.50	416k	85.3	24.7	4.5
	3250	AO	-	-	0	167k	34.8	6.5	5.1
		ST	-	perfect	0.05	165k	36.7	15.1	5.0
		ST-MTL	3	perfect	1.54	500k	87.6	18.7	4.4
			3	online	1.60	450k	90.4	21.5	4.3
		ST-MTQ	1	perfect	1.60	323k	91.5	25.9	4.1
			1	online	1.60	314k	91.7	26.6	4.1
	3750	AO	-	-	0	107k	37.0	6.6	5.1
		ST	-	perfect	0.07	101k	39.8	8.8	5.1
		ST-MTL	3	perfect	1.55	408k	92.7	20.3	4.1
			3	online	1.58	339k	94.5	23.5	4.0
		ST-MTQ	1	perfect	1.55	244k	94.4	26.5	3.8
			1	online	1.53	195k	94.5	28.4	3.8
Manhattan	3500	AO	-	-	0	-370k	17.3	9.7	6.2
		ST	-	perfect	1.81	-70k	42.3	12.6	6.4
		ST-MTL	3	perfect	4.82	242k	84.7	15.9	5.8
			3	online	5.13	255k	89.3	16.7	5.6
		ST-MTQ	3	perfect	5.13	200k	89.6	20.1	5.7
			3	online	5.17	189k	90.3	21.0	5.6
	4000	AO	-	-	0	-421k	20.6	9.9	6.3
		ST	-	perfect	1.64	-104k	47.7	12.6	6.3
		ST-MTL	3	perfect	4.34	190k	93.9	16.8	5.3
			3	online	4.51	187k	96.9	17.9	5.0
		ST-MTQ	3	perfect	4.47	114k	96.6	21.9	5.2
			3	online	4.49	98k	97.1	22.9	5.2
	4500	AO	-	-	0	-478k	23.6	10.0	6.4
		ST	-	perfect	1.58	-133k	53.7	12.5	6.2
		ST-MTL	3	perfect	3.84	104k	98.4	18.0	4.7
			2	online	3.90	91k	99.6	19.1	4.4
		ST-MTQ	1	perfect	3.87	22k	99.5	22.8	4.5
			1	online	3.87	8k	99.6	23.5	4.5

FIGURE 5 illustrates the spatial distribution of the origins of unserved requests. The scale of the colorbar reflects the absolute number of unserved requests and is much higher for AO and ST than ST-MTL and ST-MTQ. The spatial distribution of unserved requests mirrors the demand imbalance in FIGURE 4; however, the total number is much higher for the AO and ST strategies because the vehicles are mostly stuck at the airport in Chicago and the northeast of Manhattan.

The inclusion of short-term forecasts and repositioning in the ST strategy mitigates part of the AO strategy problem. In Manhattan, $T^{H,s}$ is large enough to allow repositions from the northeast corner to the some areas in the center, thereby providing a possibility to reduce the number of stuck vehicles.

FIGURE 6 displays fleet utilization over seven days for all four strategies in addition to the time-dependent travel demand. MSP fleet operators can analyze the fleet utilization over time to determine when to perform maintenance and recharging/refueling. Typically, the vehicle utilization rate will follow the number of concurrent travelers in the system, which is the convolution of number of travel requests and the trip duration of these requests. This quantity describes the number of AVs with on-board users very well when service rates are high.

FIGURE 6b) depicts the very strong degradation of the AO strategy in Manhattan in the first two days and the much slower degradation with ST repositioning. Both MT strategies generate very similar utilization rates in Manhattan that start increasing before the demand peak due to the rolling time-horizon approach. In comparison with the demand curve, the utilization plateaus because of vehicle repositions. The demand in Chicago generates the classical two peaks per day in FIGURE 6a). Vehicle repositions tend to take place in the time between these two peaks. The worse performance of the MTL model becomes evident in the inability to serve demand in the afternoon/evening peak, as the vehicle utilization does not even reach the level of the demand curve. The MTQ scenarios show very good results in this case with the majority of unserved requests originating from the time on March 29, when the theoretic demand even exceeded the fleet size.

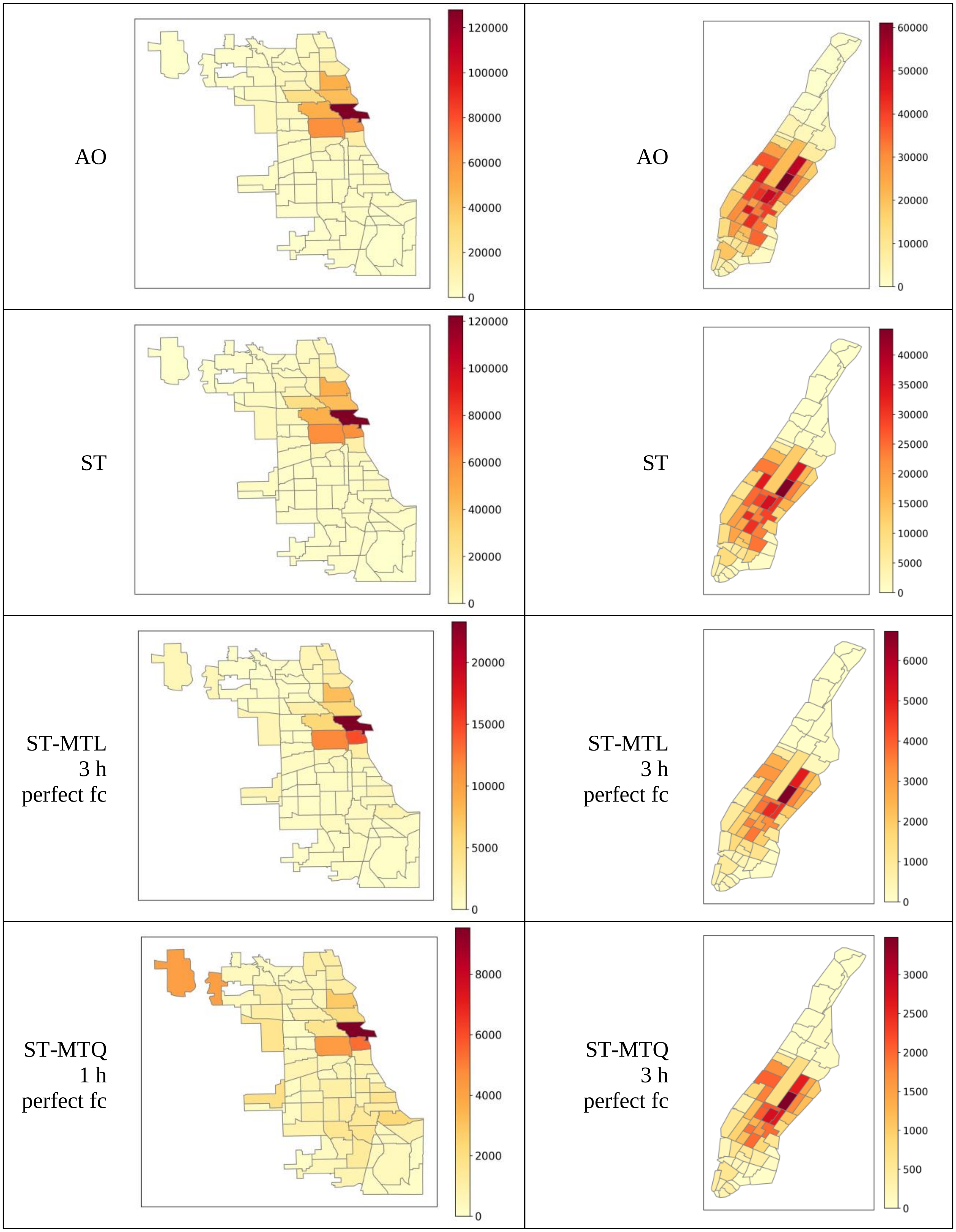
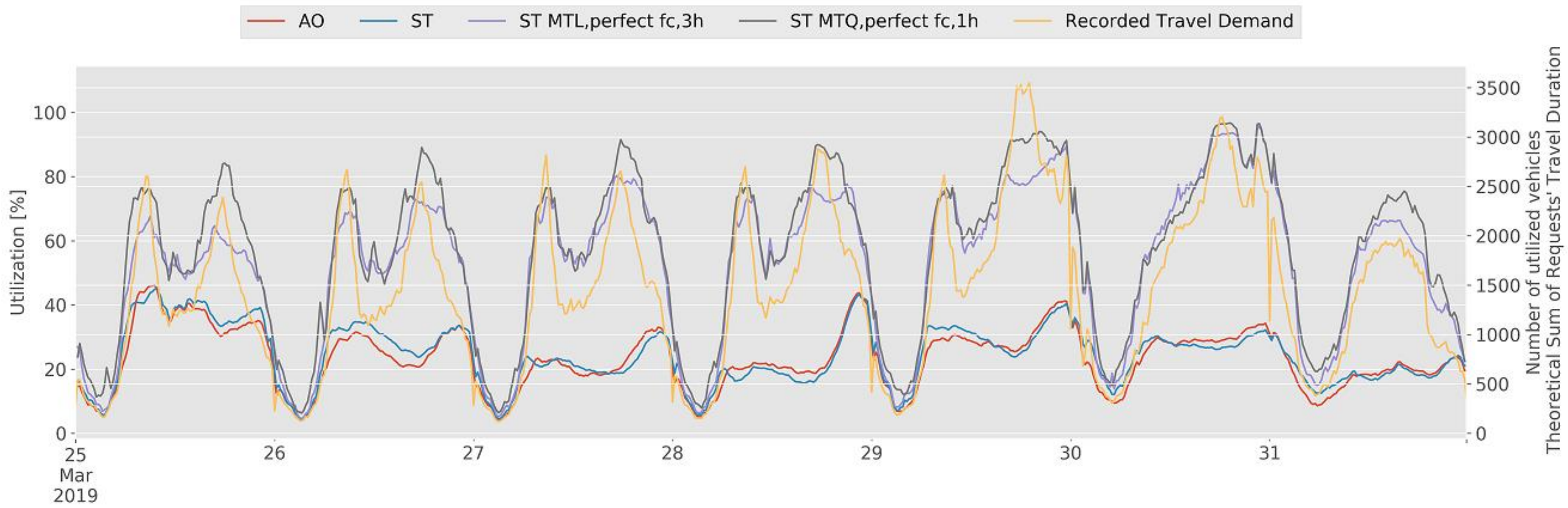
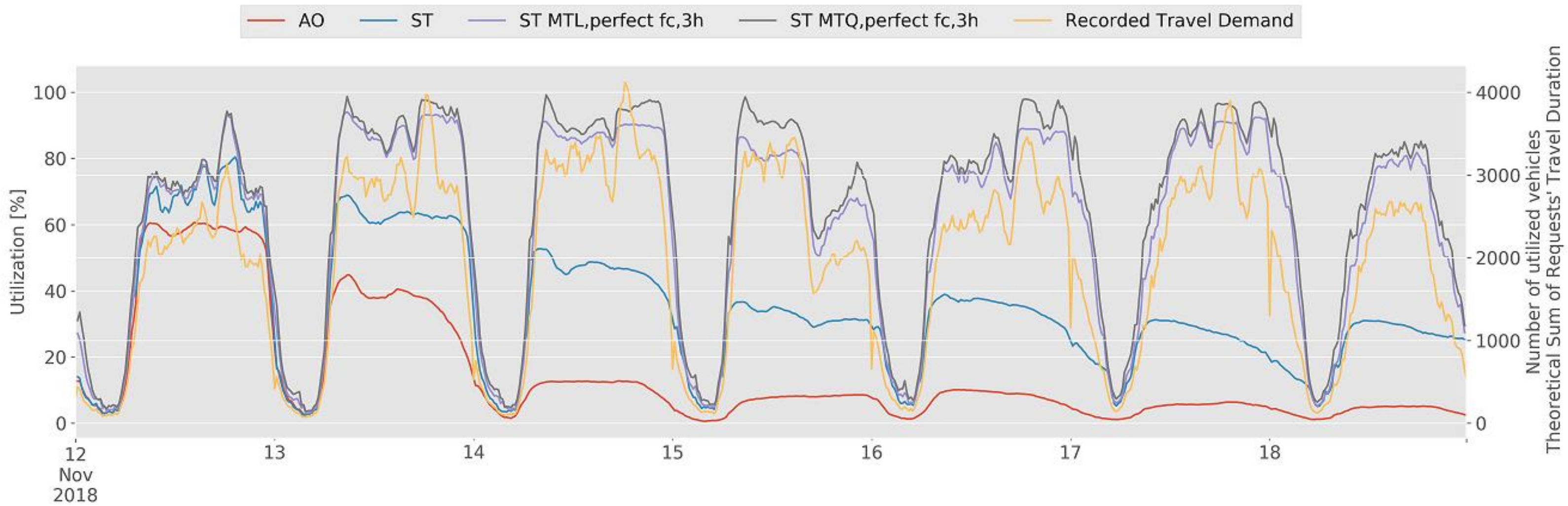


FIGURE 5 Not served requests in 7-day simulations with fleet size 3250 in Chicago and 4000 in Manhattan.



a) Fleet size of 3250 vehicles in Chicago.



b) Fleet size of 4000 vehicles in Manhattan.

FIGURE 6 Fleet utilization plots for different operator strategies. The "Recorded Travel Demand" curve represents concurrent travelers from the original trip data.

CONCLUSISON

Summary

This study presents the idea of a dual-horizon forecast and reposition approach to improve the efficiency of SAMS fleet operations. In order to trade off the costs generated by vehicles driving empty against the rewards of moving vehicles from expected vehicle surplus to expected vehicle deficit areas, two new mid-term repositioning problem formulations are introduced. The rewards in both the linear and the nonlinear optimization problem take the strength of the deficit and the surplus into account, whereas the nonlinear problem formulation even allows adapting the marginal reward of each additional vehicle sent to a deficit area. The introduction of a buffer parameter prohibits the algorithm from being too sensitive to uncertainties in demand forecasts.

7-day simulations based on taxi data in Manhattan and TNC data in Chicago show that by applying this strategy an SAMS operator can balance the vehicle distribution in the operating area and thereby serve demand much better. This study assumes a situation where the operator puts a lot of emphasis on serving user requests, expressed in a large reward for an assignment or alternative penalty for a request that cannot be served within 6 minutes. With this assumption, the total objective can be improved by 150% in Chicago and more than 350% in Manhattan with the dual-horizon repositioning approach.

Future Work

This study assumed a certain SAMS design with a high assignment reward value, which could represent a situation where an operator needs to strengthen its market position and user satisfaction is more important than immediate profitability. However, the methodology also must be tested for an operator having more focus on the immediate profitability of the system. This change will affect the choice of hyper-parameters of the reposition models, which control the balance of the tradeoff between driving costs and expected future reward. This study suggests reasonable values, but a real operator would optimize them.

Finally, new reposition formulations that consider the spatio-temporal relations between areas rather than just trying to balance demand and supply in each area, might be another interesting approach to tackle the reposition problem.

ACKNOWLEDGEMENTS

The German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety provides partial funding for the first author through the project “City2Share”.

The authors remain responsible for all findings and opinions presented in the paper.

AUTHOR CONTRIBUTION

All authors contributed to all aspects of the study from conception and design, to data collection, to analysis and interpretation of results, and manuscript preparation. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Hyland, M. F., and H. S. Mahmassani. Dynamic Autonomous Vehicle Fleet Operations: Optimization-Based Strategies to Assign AVs to Immediate Traveler Demand Requests. *TRB*, 2018.
2. Maciejewski, M., J. Bischoff, and K. Nagel. An Assignment-Based Approach to Efficient Real-Time City-Scale Taxi Dispatching. *IEEE Intelligent Systems*, Vol. 31, No. 1, 2016, pp. 68–77.
3. Hörl, S., C. Ruch, F. Becker, E. Frazzoli, and K. W. Axhausen. Fleet Control Algorithms for Automated Mobility: A Simulation Assessment for Zurich. Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017
4. Fagnant, D. J., and K. Kockelman. Dynamic Ride-Sharing and Optimal Fleet Sizing for a System of Shared Autonomous Vehicles. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015
5. Dandl, F., and K. Bogenberger. Booking Processes in Autonomous Carsharing and Taxi Systems. *Proceedings of 7th Transport Research Arena, Vienna*, 2018.
6. Dandl, F., K. Bogenberger, and H. S. Mahmassani. Autonomous Mobility-on-Demand Real-Time Gaming Framework. *IEEE International Conference on Models and Technologies for Intelligent Transportation Systems*, 2019.
7. Azevedo, C. L., K. Marczuk, S. Raveau, H. Soh, M. Adnan, K. Basak, H. Loganathan, N. Deshmunkh, D.-H. Lee, E. Frazzoli, and M. Ben-Akiva. Microsimulation of Demand and Supply of Autonomous Mobility on Demand. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1*, 2016, pp. 21–30.
8. Alonso-Mora, J., S. Samaranayake, A. Wallar, E. Frazzoli, and D. Rus. On-Demand High-Capacity Ride-Sharing via Dynamic Trip-Vehicle Assignment. *Proceedings of the National Academy of Sciences*, 2017, pp. 462–467.
9. Alonso-Mora, J., A. Wallar, and D. Rus. Predictive Routing for Autonomous Mobility-on-Demand Systems with Ride-Sharing. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Vancouver, BC, pp. 3583–3590.
10. Mitrović-Minić, S., R. Krishnamurti, and G. Laporte. Double-Horizon Based Heuristics for the Dynamic Pickup and Delivery Problem with Time Windows. *Transportation Research Part B: Methodological*, Vol. 38, No. 8, 2004, pp. 669–685.
11. Bertsekas, D. P. *Dynamic Programming and Optimal Control.* Athena Scientific, Belmont, Mass., 2005.
12. Powell, W. B. *Approximate Dynamic Programming*. Solving the Curses of Dimensionality. Wiley, Hoboken, New Jersey, 2011.
13. Spieser, K., S. Samaranayake, W. Gruel, and E. Frazzoli. Shared-Vehicle Mobility-on-Demand Systems: Fleet Operator's Guide to Rebalancing Empty Vehicles. Presented at 95th Annual Meeting of the Transportation Research Board, Washington, D.C., 2016

14. Dandl, F., and K. Bogenberger. Comparing Future Autonomous Electric Taxis with an Existing Free-Floating Carsharing System. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 6, 2019, pp. 2037–2047.

15. Vosooghi, R., J. Kamel, J. Puchinger, V. Leblond, and M. Jankovic. Robo-Taxi Service Fleet Sizing: Assessing the Impact of User Trust and Willingness-to-Use. *Transportation*, Vol. 67, 2019, p. 1.

16. Engelhardt, R., F. Dandl, A. Bilali, and K. Bogenberger. Quantifying the Benefits of Autonomous on-Demand Ride-Pooling. A Simulation Study for Munich, Germany. *accepted to IEEE ITSC*, 2019.

17. Psaraftis, H. N., M. Wen, and C. A. Kontovas. Dynamic Vehicle Routing Problems. Three Decades and Counting. *Networks*, Vol. 67, No. 1, 2016, pp. 3–31.

18. Hyland, M. F., and H. S. Mahmassani. Taxonomy of Shared Autonomous Vehicle Fleet Management Problems to Inform Future Transportation Mobility. In *Transportation Research Record: Journal of the Transportation Research Board*, 2017, pp. 26–34.

19. Pureza, V., and G. Laporte. Waiting and Buffering Strategies for the Dynamic Pickup and Delivery Problem with Time Windows. *INFOR: Information Systems and Operational Research*, Vol. 46, No. 3, 2008, pp. 165–175.

20. Mitrović-Minić, S., and G. Laporte. Waiting Strategies for the Dynamic Pickup and Delivery Problem with Time Windows. *Transportation Research Part B: Methodological*, Vol. 38, No. 7, 2004, pp. 635–655.

21. Dandl, F., M. Hyland, K. Bogenberger, and H. S. Mahmassani. Evaluating the Impact of Spatio-Temporal Demand Forecast Aggregation on the Operational Performance of Shared Autonomous Mobility Fleets. *Transportation*, Vol. 114, 2019, pp. 462–484.

22. Wen, J., N. Nassir, and J. Zhao. Value of Demand Information in Autonomous Mobility-on-Demand Systems. *Transportation Research Part A: Policy and Practice*, Vol. 121, 2019, pp. 346–359.

23. Sayarshad, H. R., and J. Y. J. Chow. Survey and Empirical Evaluation of Nonhomogeneous Arrival Process Models with Taxi Data. *Journal of Advanced Transportation*, Vol. 50, No. 7, 2016, pp. 1275–1294.

24. Moreira-Matias, L., J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas. Predicting Taxi–Passenger Demand Using Streaming Data. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 3, 2013, pp. 1393–1402.

25. Fagnant, D. J., K. Kockelman, and P. Bansal. Operations of Shared Autonomous Vehicle Fleet for Austin, Texas Market. In *Transportation Research Record: Journal of the Transportation Research Board*, 2015, pp. 98–106.

26. Pavone, M., S. L. Smith, E. Frazzoli, and D. Rus. Robotic Load Balancing for Mobility-on-Demand Systems. *The International Journal of Robotics Research*, Vol. 31, No. 7, 2012, pp. 839–854.