2019 Proceedings of SIG GREEN Workshop

2019

# An Online Learning and Optimization Approach for Competitor-Aware Management of Shared Mobility Systems

Karsten Schroer

Wolfgang Ketter

Thomas Y. Lee

Alok Gupta

Micha Kahlen

# An online learning and optimization approach for competitor-aware management of shared mobility systems

**Karsten Schroer**
University of Cologne
karsten.schroer@uni-koeln.de

**Wolfgang Ketter**
University of Cologne
ketter@wiso.uni-koeln.de

**Thomas Y. Lee**
University of California, Berkeley
thomasyl@berkeley.edu

**Alok Gupta**
University of Minnesota
gupta037@umn.edu

**Micha Kahlen**
Erasmus University Rotterdam
micha.kahlen@gmail.com

## Abstract

An important trend in mobility is the consumption of mobility as-a-service heralding in the age of free-floating vehicle sharing (FFVS) systems. In many markets such fleets compete. We investigate how real-time competitor information can create value for operators in this context. We focus on the vehicle supply decision which is a large operational concern. We show empirically that local market shares directly depend on the share of available vehicles in a location, which underlines the value potential of competitor awareness. We leverage this insight by proposing a novel decision support system for optimal management of FFVS systems under competition. We proceed in two phases, (1) a predictive phase and (2) a prescriptive phase. In phase (1), we compile a spatio-temporal dataset based on Car2Go and DriveNow transactions in Berlin, which we supplement with temporal, geographical and weather data. We partition the city into hexagonal tiles and observe vehicle supply per tile at the start of each period. We train machine learning models to predict vehicle inflows and vehicle outflows during the next period to derive total supply and demand. We find that inflows and outflows can be predicted with high accuracy using similar models. We test different temporal and spatial resolutions and find that spatial resolution incurs larger performance penalties. In phase (2), we formulate a myopic mixed integer non-linear programming model with a margin-maximizing objective function. The model trades off additional market share gains against the cost of re-locating vehicles, which enables operators to assign vehicles optimally across the service network. Our numerical studies on the case of Car2Go and DriveNow demonstrate that this competitor-aware model is capable of profitably improving market share by up to 1.4% or 3.4% for human-based and autonomous re-location respectively in a prefect foresight scenario and by up to 0.8% and 1.8% respectively when using predicted values.

### Keywords

shared mobility, Car2Go, machine learning, online optimization, Green IS

## Introduction and Background

The mobility landscape as we know it is undergoing a deep transformation. Tomorrow's mobility system will likely feature a multi-modal portfolio of connected, autonomous, shared and electric (CASE) mobility options (Sperling 2018). A parallel trend is the consumption of mobility as-a-service (MaaS) and on-demand (MoD) (Laporte et al. 2018) heralding in the age of shared, fleet-based transportation companies operating digital, smart market-type platforms (Bichler et al. 2010). Car2Go and DriveNow, the world's leading free-floating carsharing operators are excellent manifestations of this. Similar platforms are emerging for bikes (e.g. Nextbike) or e-scooters (e.g. Lime and Bird). In many markets multiple mobility platform operators compete. Literature on multi-homing suggests that users actively switch between platforms. Real-time competitor insights such as vehicle positions may therefore proof to be of strategic value. In the new digitalized mobility system, this previously confidential competitive information is now available for the first time. In many cases transportation companies provide open APIs that offer real-time access to fleet status data. Alternatively, such competitor fleet information can be obtained by scraping mobile app interfaces. We explore how fleet operators can leverage real-time competitor information along with other large-scale urban data sources in their operational decision making. We focus on the problem of fleet repositioning, which aims to optimally allocate fleet vehicles. It seems intuitive that failing to take competitive information into account may result in sub-optimal repositioning decisions. For example, the fleet operator may choose to re-locate to a location where competitor supply is already abundant or to reposition away from a location where insufficient competitor assets are located and an overall supply shortage exists. Two steps are required to enable competitor-aware data-driven repositioning. First (1), a spatio-temporal machine learning model to predict total combined demand and supply of shared mobility services is needed. This enables fleet operators to identify pockets of relative over- and under-supply. We refer to this phase as predictive analytics phase. Second (2), a competitor-aware fleet repositioning routine must be developed on top of this prediction (prescriptive analytics phase). We report results on these two phases and formulate our research objective as follows:

> *Develop a spatio-temporal machine learning model to predict combined demand and supply of shared mobility services in a given market (predictive analytics) and demonstrate how a mathematical model can leverage these predictions to facilitate competitor-aware repositioning decision making for shared mobility operators (prescriptive analytics).*

The contributions of our research are the following: We are the first to consider the value potential of real-time competitor information in operational decision making of shared mobility companies. We also develop a novel, highly accurate and flexible prediction model for total shared mobility demand and vehicle supply in a given market and provide comprehensive test metrics as well as recommendations regarding spatial and temporal resolutions for fleet operators. We simultaneously predict supply and demand of vehicles per period and location – A key prerequisite for a prescriptive repositioning algorithm. Finally, as far as we are aware, our work is among the first to bridge the gap between predictive analytics research of vehicle repositioning, which is mainly driven by IS researchers as well as mathematical programming-focused research on optimal shared vehicle repositioning. We take a first step in merging these approaches. It follows from these contributions that our work ties in with the IS Design Science research philosophy (Hevner et al. 2004) in that we present and thoroughly evaluate an IS artefact for the relevant fleet management challenge which generalizes to all types of shared vehicle systems. Shared mobility will play an increasingly large part in the mobility landscape of the future.

The remainder of this paper is structured as follows: In the next section we present a short review of related work. We then describe our data and present selected descriptive statistics. We proceed by presenting and evaluating the predictive and prescriptive modeling approaches, the two core contributions of this research. We end with a discussion of our findings and an outlook on future work.

## Related Work

### *Free floating vehicle sharing (FFVS) systems*

Vehicle sharing has attracted increased interest by the research community and the public due to its perceived potential to contribute to a sustainable society (Ketter, Peters, and Collins 2016; Ketter, Peters, Collins, et al. 2016; Sperling 2018). The claimed benefits include less car ownership per household, reduced emissions, and a more efficient transport system in general. However, car sharing itself is not a novel concept. In fact it dates back to the 80s and became popular in Europe in the 90s (Shaheen and Cohen 2007). The popular providers at that time used station-based system, where cars have to be collected and dropped off at a certain number of fixed stations. While initial costs for a station-based system are low, consumer adoption is far exceeded by another type of vehicle sharing concept: free floating vehicle sharing. In such flexible one-way systems, customers are free to pick up and drop off a vehicle at any point within a predefined operating area, which provides the benefit of direct point-to-point travel. As a result, FFVS schemes have become increasingly popular and are emerging across different types of transportation modes including cars (FFCS), bikes and (e-) scooters and will likely constitute a major building block in the connected autonomous, shared and electric (CASE) future mobility system (Sperling 2018). Research on FFVS systems is therefore characterized by high relevance especially from a Green IS and analytics for sustainability perspective (vom Brocke et al. 2013; Ketter et al. 2018; Melville 2010; Watson et al. 2010).

## *Predictive analytics for FFVS*

The emergence of big urban data (Arribas-Bel and Tranos 2018) and spatial big behavioral data (Shmueli 2017) provide novel insights into human behavior in time and space. This has resulted in new research opportunities for IS both in explanatory analytics (e.g. Pelechrinis et al. 2016; Zhang et al. 2016) as well as predictive analytics (e.g. Kahlen et al. 2017; Willing et al. 2017). We focus here on a review of the predictive analytics stream. Müller and Bogenberger (2015) are among the first to attempt a spatial demand prediction of shared vehicle demand using a simple time series approach and ZIP codes to discretize the geographic areas. Relevant other studies in this area include Wagner et al. (2016), who use Google points-of-interest (POI) data to predict popular rental areas for free-floating car sharing systems. However, the authors neglect any time-specific patterns, which would be a key requirement in a repositioning application. Willing, Klemmer, et al. (2017) build on this research by including a time dimension. However, they work with fixed intervals, which is suitable for their application of service area definition but not granular and flexible enough for a repositioning application. They also do not consider any competitive effects and only predict rental end points. Brendel et al. (2018) propose a Generic Vehicle Relocation Information System which also includes a demand prediction module. The prediction is purely based on historic rental demand. The model also does not forecast absolute demand but identifies demand scores instead. The approach, which is most closely related to our work is presented by Kahlen et al. (2018). This research develops a mobility demand prediction based on a statistical model that predicts rental outflows per location of a single fleet operator using not only POI data but also temporal and meteorological features. However, the authors do not consider competition, a highly relevant aspect in competitive markets. Additionally, they only predict rental outflows per tile. In an application where we are interested in economic regimes (Ketter et al. 2009, 2012) per time and location based on which a downstream prescriptive model could formulate pre-emptive repositioning decisions, simply predicting outflows per period is insufficient. To sum up, the existing body of research on shared mobility demand prediction takes an operator-centric perspective without taking competition into account. All existing models are also limited in the sense that they either predict inflows or outflows. We also see gaps regarding the flexibility of the model in terms of spatial and temporal resolution which might be required for different modes. In this research we address these gaps.

## *Operations management of FFVS*

The problem of vehicle (asset) allocation and repositioning is a well-established area of research (He, Mak, et al. 2019). It is especially significant in one-way vehicle sharing systems. Due to the unknown target destination of one-way trips, demand and supply imbalances may occur locally. This is of particular concern in free-floating systems where trips can be terminated virtually anywhere within a given operating area. Two distinct approaches to facilitate vehicle repositioning exists: (1) operator-based repositioning and (2) user-based repositioning. Operator-based re-location is executed via the operator while in user-based repositioning users take over the physical task of moving vehicles to more favorable locations – Usually in exchange for some sort of reward (Laporte et al. 2018). In both cases, an effective repositioning decision support system requires at least two components. First (1), a granular supply and demand prediction model to estimate localized supply surpluses or shortages. Second (2), based on predicted demand and usage patterns, a repositioning algorithm is required. Prescriptive re-location analytics and optimization for fleet operators is a well-known problem. The state of the art in this stream of research has moved towards optimization under uncertainty which uses (adaptive) robust optimization to deal with uncertainty in demand (He, Mak, et al. 2019; Laporte et al. 2018). A recent example is provided by He, Hu, et al. (2019) who produce a myopic and multi-period robust optimization based on a simple time-series prediction of fleet dynamics. In another study Lu et al. (2018) develop two-stage stochastic integer programming models for optimizing strategic parking planning and vehicle allocation for carshare systems under uncertain demand. These studies usually employ (if any) very simple timeseries prediction models. They also do not consider competition. There is a clear opportunity for contributions by merging state-of-the-art competitor-aware predictive analytics and mathematical modeling techniques to mitigate uncertainty and enable competitor-aware decision making, which this research focuses on.

# Data

## *Raw Data*

We utilize a circa 5-months window of carsharing demand data covering the period from December 1st, 2016 to April 26th, 2017. The data was obtained from Car2Go and DriveNow and contains information on availability of vehicles and their position at a granularity of five minutes. Neither of the two service providers currently performs active re-locations. This allows us to infer trip start and end times along with departure and destination locations by tracking individual vehicles disappearing and reappearing in our records. We obtain historic weather data for the same period from the German Meteorological Service (DWD). The dataset contains temperature, wind and precipitation information in hourly resolution from five different weather stations in Berlin for which we compute mean values. We use this data as proxy data for weather forecasts in our model, which is common practice (Kahlen et al. 2017) and is well warranted given the very high quality of state-of-the-art weather forecasts for the short forecast horizons (<24h) that we use in our work. POI data was obtained via Google Maps. The Google Maps API returns up to 60 POI for an input of coordinates and a radius. To collect this data, we partition the city into overlapping circles with 64m radii. At this setting only 0.4% of API calls reach the limit of 60 POIs. We also test coarser (96m, 1.2% calls returned with 60 POI limit) and more granular radii (48m, 0.2% calls returned with 60 POI limit) but deem a rate of 0.4% to be satisfactory. We obtain geographic vector data in the form of polygon shapefiles of the city of Berlin from Berlin's open data platform.

## *Data Compilation*

To combine the various datasets into a single database we draw on techniques from geographic information systems (GIS) and urban analytics (Arribas-Bel and Tranos 2018). We create spatial and temporal joins between the individual datasets. To do so, we partition the city into hexagonal tiles. Hexagonal tiles have the key advantage of uniform adjacency. Each hexagon has exactly six neighbors with whom it shares an edge and whose centers are exactly the same distance apart. These properties have resulted in a strong uptake of hexagons in discrete spatial simulations (Sahr et al. 2004), to which clear methodological parallels exist in our work. We construct a hexagonal geodesic Discrete Global Grid Systems (DGGS) as proposed by Sahr et al. (2004) to partition the city. This approach results in less distortion compared to planar grids as employed in previous research (Kahlen et al. 2017; Wagner et al. 2016). It is also more flexible and allows for relatively easy adjustment of resolution due to the hierarchical nature of tile sizes, a property which we take advantage of to test different model setups. To construct the hexagon grid we employ a Python binding of the open-source H3 library. To create spatial merges between the datasets and the constructed hexagon grids we draw on open source GIS libraries in Python, particularly Geopandas. We match the geo-referenced mobility demand and POI data to the individual hexagons. We then construct spatio-temporal keys combining both geographical and temporal information to compile all datasets into one. We repeat this process for three different resolutions of tile grids. We opt for a coarse granularity (circa 3.0km edge length), a medium-coarse granularity (circa 1.5km edge length), and a fine granularity (circa 0.5km edge length). We also compile three different sets of temporal resolutions: 0.5h, 1h and 6h. We choose these to explore trade-offs between temporal resolution and predictive performance. From a practical perspective, different applications may require different resolutions.

## *Descriptive Statistics*

Selected descriptive statistics of the combined dataset are shown in Figures 1 and 2. We explore the similarities in demand patterns between the two mobility platforms, which can be thought of as a proxy for homogeneity between the provided mobility services. We characterize a mobility service based on three attributes: (1) rental start location (see Figure 1), (2) rental start time (see Figure 2) and (3) total trip length. Our assumption is that, if large similarities exist across these attributes, both platforms can be thought of as providing substitute services and therefore compete for the same customer base. Figure 1 illustrates the considerable similarities in spatial rental patterns. Figure 2 provides an overview of the temporal patterns in the rental demand for the two competing fleet operators. Daily patterns are highly correlated between the two platforms (Pearson correlation coefficient = 0.9626). A weekly trend can be observed in the distributions of daily demand with peaks on Fridays and Saturdays followed by lows on Sundays. Intra-daily rental demand patterns exhibit similarly strong correlation (Pearson correlation coefficient = 0. 9761). In sum, both fleets seem to offer substitute services to customers with largely similar preference patterns. Finally, we test the core assumption of our work: we assume that the captured amounts of rides per location is directly proportional to the local density share (i.e. the share of vehicles on the ground). We find proof for this assumption in the data. Regressing local share in rides captured versus the local density share (at h=3.0km and p=6h) yields a clearly linear relationship with slope of 0.9248 and intercept of 0.0563 (both highly significant) at an $R^2$ of 0.723 – an almost directly proportional relationship.

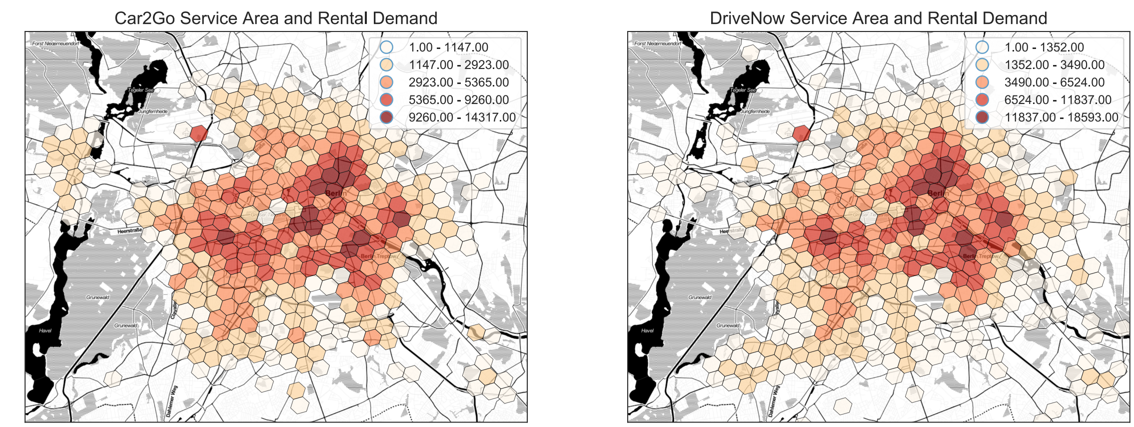


**Figure 1. Geographical rental patterns for Car2Go (left) and DriveNow (right)**

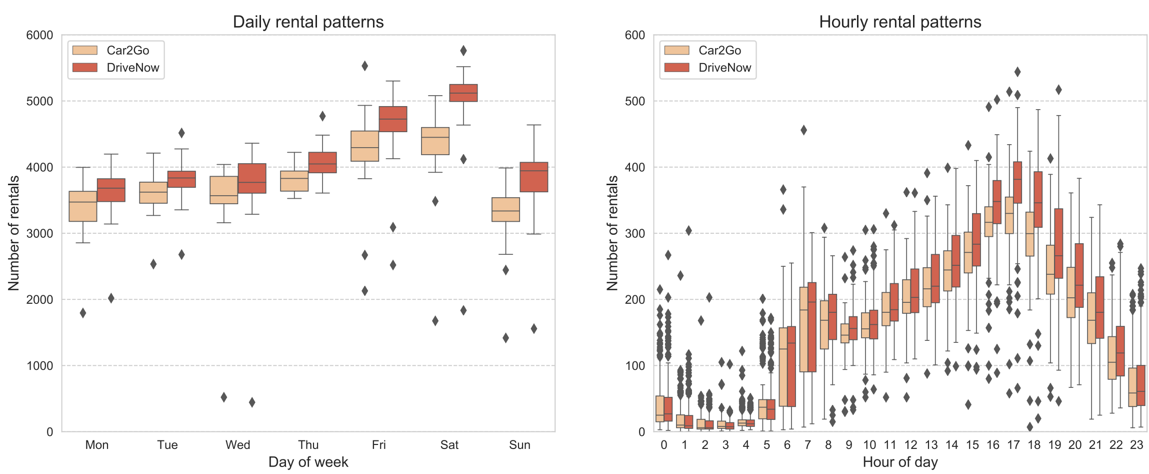


**Figure 2. Temporal rental patterns for Car2Go and DriveNow in daily (left) and hourly (right) resolution**

## Predictive model

### *Feature Engineering and Selection*

We compile a feature set which embodies both spatial and temporal aspects of mobility demand. We utilize 170 features which can be grouped into five sets.

- **Temporal feature set ($T$)**: We employ a total of 35 temporal features on the hour of the day, the day of the week or whether the day is a weekend or holiday.
- **Meteorological feature set ($W$)**: We include six meteorological features, such as temperature, whether it will rain or not and how much, cloud cover and wind speed.
- **Geographical feature set ($G$)**: We use 105 features related to POIs. These include counts for each of the 99 POI-types contained in our dataset (e.g. number of restaurants, cafes, doctors, etc.). We also capture the relative popularity of the POIs by counting the number of POIs that have a rating and a price.
- **Competitor fleet feature set ($C$)**: We include two different types of competitor features. First, we include vehicle densities per tile for which we distinguish between electric and conventional vehicles. Second, we include lagged target features. We use lags of one, two and three periods prior as well as 1 to 7 days prior to the period in scope, which were select based on autocorrelation. We also include the mean and variance in the target variable over of the past 7 days. Finally we include differencing variables, to captured any trends.
- **Own fleet feature set ($O$)**: We use exactly the same features as for the competitor fleet also for our own fleet. Since we are predicting total demand/inflows, lagged target features are combined into the sum of both fleets and therefore only appear once.

### *Machine Learning Model Formulation*

We opt for a linear machine learning model. We choose this model for several reasons. First, its relatively low complexity results in a high level of interpretability, which may be preferred by practitioners. Second, previous studies have successfully employed linear models for mobility demand prediction (Kahlen et al. 2017; Willing et al. 2017). Third, our focus is on identifying relevant predictive feature sets, demonstrating trade-offs between different spatial and temporal resolutions and model performance and to show the value of competitor information in shared mobility operations. For this purpose the choice of model is secondary. Fourth, we already achieve high performance with the chosen model for most setups. We do suspect, that more complex models such as random forests or deep neural networks as well as ensembles of different models may yield even better results. This could be subject of future work, but is not the focus of the current work. To avoid overfitting and to increase the computational performance of our model we employ LASSO ($\ell$1-) regularization to our linear model, which adds a penalty for absolute coefficient size to the objective function. The advantage of this approach is the embedded feature selection capability which result from the fact that coefficients of those features that are not adding to model performance are set to zero, a commonly used property for feature selection (Miao and Niu 2016). We opt for a mild level of regularization ($\lambda = 0.01$). At this level we do not incur any performance penalties across any resolution setting. On overage, this approach allows us to reduce the feature space by circa 55% (or to about 77 features).

We use our model to predict two different target values:

- **Vehicle outflows $\hat{D}^{M}_{i,p}$:** Vehicle outflows per tile $i \in H$ and period $p$, where $H$ is the set of tiles in the network. These are a proxy for the constrained market demand (denoted by superscript $M$) for vehicles. Our target value for this predictive task is the number of rental start points per tile and location.
- **Vehicle inflows $\hat{I}^{M}_{i,p}$**: Rental inflows into a tile $i \in H$ during period $p$ provide information on the added vehicle supply in that tile. Our target value for this predictive task is the total market-level number of rental end points per tile and location.

We choose the same general model architecture to predict vehicle outflows and inflows. The only difference between the model setups is the lagged target features that are used as independent variables. Our model predicts $\hat{y}_{i,p}$, the total market vehicle outflows or inflows in tile $i \in H$ and period $p$. The formulation of the model is summarized in the below equation. We use subscripts to denote whether a variable is time- ($p$) or location- dependent ($i \in H$). Superscripts denote the association of a variable with one of our five feature sets.

$$\hat{y}_{i,p} = \sum_{t \in T} \beta^t \alpha_p^t + \sum_{w \in W} \beta^w \alpha_p^w + \sum_{g \in G} \beta^g \alpha_i^g + \sum_{c \in C} \beta^c \alpha_{i,p}^c + \sum_{o \in O} \beta^o \alpha_{i,p}^o + \mathrm{b_{i,p}}$$

Where $T$ = set of temporal features, $W$ = set of meteorological features, $G$ = set of geographical features, $C$ = set of competitor fleet features, $O$ = set of own fleet features. As mentioned we add an $\ell 1$-regularization term to the objective function which results in only the most relevant features per feature set being used in the prediction.

## *Predictive Model Results and Evaluation*

To evaluate our model, we use a form of moving window validation. We train the model on a fixed amount of observed data and evaluate it on the next period which most accurately resembles the intended real-world application. We keep our training window fixed at 60 days of data, which was selected based on a sensitivity analysis. We do not achieve significant performance improvements with longer training windows. All prediction results are post-processed by rounding to the nearest integer. We evaluate our model on a one-week period. We randomly select this week from our data and use the same week for the evaluation of all spatial and temporal resolutions to ensure comparability. The spatio-temporal flexibility adopted in our modeling serves three purposes. First, it allows us to demonstrates the trade-off between temporal and spatial resolution and predictive accuracy. Second, different fleet applications may require different model setups, which our approach can help choose. Third, we wish to use the different magnitudes in prediction error in later research to compute the marginal penalty of prediction error on re-allocation decisions. In terms of error statistics, we focus on the mean absolute error (MAE), which is the appropriate choice of error metric for our application because it returns the absolute number of vehicles the model is off on average. One downside of using the MAE is that comparisons between different resolutions is made difficult, since the error is not scaled in accordance with the baseline value. A very popular relative metric in the prediction literature is the mean absolute percentage error (MAPE), which is scale-independent, but suffers from several problems, its inability to handle cases where the true value is zero being the most serious one for this work (Chen et al. 2017). Therefore the symmetric mean absolute percentage error (SMAPE) is used. Note that one drawback of SMAPE is its inflation caused by zero-valued observations as is the case in our application (Chen et al. 2017). This may cause it to behave counter-intuitively at times (e.g. higher SMAPE at higher $R^2$). We also report the $R^2$-scores, another highly popular relative error metric. All results are reported in Table 1.

| | | **Rental Outflows** | | | **Rental Inflows** | | |
|---|---|---|---|---|---|---|---|
| | | p = 0.5h | p = 1h | p = 6h | p = 0.5h | p = 1h | p = 6h |
| h = 0.5 km | MAE | 0.65 | 0.97 | 2.98 | 0.66 | 1.03 | 3.37 |
| | **SMAPE** | **63.37%** | **63.10%** | **46.10%** | **63.97%** | **65.39%** | **46.10%** |
| | **$R^2$** | **0.58** | **0.73** | **0.90** | **0.56** | **0.70** | **0.88** |
| h = 1.5 km | MAE | 1.39 | 2.13 | 7.12 | 1.42 | 2.19 | 7.89 |
| | **SMAPE** | **52.09%** | **47.65%** | **31.18%** | **53.05%** | **48.48%** | **31.62%** |
| | **$R^2$** | **0.89** | **0.93** | **0.97** | **0.88** | **0.92** | **0.96** |
| h = 3.0 km | MAE | 2.86 | 4.53 | 17.48 | 2.96 | 4.70 | 18.90 |
| | **SMAPE** | **47.23%** | **41.93%** | **33.04%** | **47.97%** | **42.02%** | **27.19%** |
| | **$R^2$** | **0.97** | **0.98** | **0.99** | **0.97** | **0.98** | **0.99** |

**Table 1: Overview of model predictive performance per resolution**

We find that vehicle inflows and outflows are highly predictable with our model – particularly so at lower spatial and temporal resolutions. We obtain very high predictive accuracy achieving $R^2$-scores of up to 0.99 and SMAPE-scores of down to 27% using the coarsest resolution setting (p=6h, h=3km). This is in line with our expectations as high aggregation levels tend to smooth out much of the variance in the data, thus leading to better predictability. Overall, we achieve very similar predictive performance for outflows and inflows, which seems intuitive given their strong time-shifted correlation per location. Popular inflow destinations are likely to be popular outflow destinations at a later point as mobility users move towards and away from a destination. Our model captures many observed and unobserved patterns and can be flexibly trained to capture the temporal lag between inflows and outflows, thus delivering comparable and highly accurate prediction results. Figure 3 provides insights into the trade-off effects of temporal and spatial resolutions on predictive performance. We observe that prediction errors seem to be particularly high when moving from medium to high spatial resolutions. At a high spatial resolution, changes in temporal resolution also have a large predictive performance penalty. At spatial resolutions of 1.5km and higher, predictive performance plateaus reaching $R^2$ scores of nearly 0.9 (p=0.5h, h=1.5) or higher at all temporal resolutions. Resolution – particularly spatial resolution – comes at a cost, which is quantified here and needs to be managed carefully. Fleet operators must carefully weigh up the acceptable loss in predictive accuracy versus a higher predictive resolution.

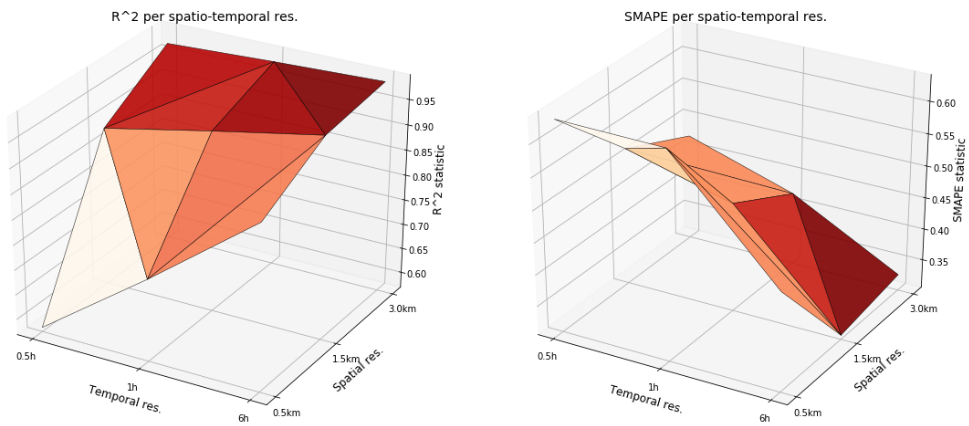


**Figure 3. Trade-off in model performance (outflows)**

Our results provide indications of how to manage this trade-off. For example, we find that penalties are higher for higher spatial resolutions compared to temporal resolutions. Depending on the use case it may be acceptable to opt for longer prediction windows but instead increase spatial resolution. For a fleet of autonomous vehicles, spatial resolution may be coarser as the car could actively approach a customer. In this case, repositioning would need to be less accurate but could happen much more frequently during the day.

## Prescriptive model

### *Mathematical Formulation*

Similar to He, Hu, et al. (2019) we take a myopic view of the relocation problem by looking one period ahead. Our optimization objective can be formulated as follows: "*For the upcoming period, re-locate vehicles across the network in a competitor-aware fashion such that fleet contribution margin is maximized*". We choose the contribution margin as our objective, as it neglects fixed cost associated with the fleet size, which we assume to be fixed. In such a scenario the fleet operator has an incentive to maximize contribution margin, i.e. maximize the proportion of sales revenue that is not consumed by variable costs and can thus contribute to cover the fixed cost base. What results is a special profit-maximizing form of the assignment problem where $i \in H$ origin tiles are matched by the same number of destination tiles $j \in H$, with $H$ being the set of tiles (Kuhn 1955). Let us define the decision variable $x_{ij,p}$ as the number of vehicles that flow from tile $i$ to $j$ during the period $p$, where $i, j \in H$. These decision variables have both revenue and cost implications. Our general mathematical model is formulated as follows:

$$\max \quad \Pi_p = \sum_{i \in H} \left( D_{i,p}^{M} * s_{i,p}^{O} * \left(r^{rent} - c^{rent}\right) \right) - \sum_{i \in H} \sum_{j \in H} \left( x_{ij,p} c_{ij}^{reloc} \right)$$

$$s.t. \qquad x_{ij,p} \geq 0, int \qquad\qquad \forall\, i, j \in H$$

$$x_{ij,p} \leq A^{O}_{i,p-1} + I^{O}_{i,p} \qquad \forall\, i,j \in H$$

$$D^{M}_{i,p} \geq A^{M}_{i,p-1} + I^{M}_{i,p} + \sum_{j \in H} (x_{ji} - x_{ij}) \qquad \forall\, i \in H$$

Where $D^{M}_{i,p}$ is the market demand in tile $i \in H$, $s^{O}_{i,p}$ is the own local density share in tile $i \in H$, $r^{rent}$ is the expected revenue per rental, $c^{rent}$ is the expected variable cost per rental (fuel and degradation), $c^{reloc}_{ij}$ is the cost of relocating from tile $i$ to $j$. $A^{O}_{i,p-1}$ and $A^{M}_{i,p-1}$ are the vehicle availabilities in tile $i$ in the previous period for the own fleet ($O$) and for all fleets combined ($M$) respectively. $I^{O}_{i,p}$ and $I^{M}_{i,p}$ are the inflows of own vehicles ($O$) and vehicles of all fleets combined ($M$) into $i$ during period $p$ respectively.

The first term of the objective describes the contribution margin of each captured trip, where the number of captured trips is determined by the own local density share $s^{O}_{i,p}$. $s^{O}_{i,p}$ is directly controlled by the fleet operator's fleet re-location decision and is defined as follows:

$$s^{O}_{i,p} = \frac{A^{O}_{i,p-1} + I^{O}_{i,p} + \sum_{j \in H}(x_{ji,p} - x_{ij,p})}{A^{M}_{i,p-1} + I^{M}_{i,p} + \sum_{j \in H}(x_{ji,p} - x_{ij,p})}$$

Note that this definition of $s^{O}_{i,p}$ results in a non-linear concave objective function that is maximized. In combination with the linear mixed integer constraints this results in a convex, mixed integer non-linear program (MINLP). Such problems, although considered complex, can be solved with state-of-the-art modeling frameworks (Kronqvist et al. 2019).

The second term of the objective is the re-location cost incurred by the individual relocation decision $x_{ij}$, which is summed over all individual relocations. The constraints (in consecutive order) ensure (1) that the decision variables (i.e. the relocations) are positive integer values, (2) that the re-locations from a tile do not exceed the available own vehicles within that tile and (3) that the total demand within each tile is always met. Further practical constraints, such as e.g. a maximum on the number of re-locations during a period may be added as appropriate.

## *Prescriptive Model Results and Evaluation*

For the purpose of this research we demonstrate the use of our model by evaluating its proficiency over a one-week period (the same we have also evaluated our predictive model on). For illustrative purposes we use a resolution of h=3km and p=6h. At this resolution, tile centers are a distance of $d_{ij} = 6.1\ km$ apart (measured in terms of great circle distance). We parameterize two types of relocation cost $c^{reloc}_{ij}$: one cost parameter for human operator-based relocation ($c^{reloc,hum}_{ij} = d_{ij} * (c^{var} + c^{hum})$) and one cost for an autonomous vehicle case ($c^{reloc,aut}_{ij} = d_{ij} * (c^{var} + 0)$). We let the specific variable cost per km $c^{var}$ equal to USD 0.2 per km (includes fuel and degradation cost) and the human re-location cost $c^{hum}$ equal to USD 0.5 per km, which may be realized via an equal share of free ride offers (USD 0 per km) and relocation workers (USD 1 per km). Finally, we define $r^{rent}$ = USD 20 (avg. rental length of 50 min and revenue of USD 0.40 per min) and $c^{rent} = d^{rent} * c^{var}$ = USD 2.

We explore two cases: (1) a "perfect foresight case" and (2) a "predicted foresight case". While in the first case we observe the true values for all parameters, we use predicted values for demand $\hat{D}^{M}_{i,p}$ and for inflows $\hat{I}^{M}_{i,p}$ and $\hat{I}^{O}_{i,p}$ in the second case. Since we are predicting inflows only at a market level ($\hat{I}^{M}_{i,p}$), we use the following equation to derive own inflows: $\hat{I}^{O}_{i,p} = \hat{I}^{M}_{i,p} * s^{O}_{i,p-1}$. This proves to be a good heuristic. Note that constraint (2) remains unchanged in a predicted foresight case as relocations would be limited by the true number of available vehicles in a tile $i \in H$.

We implement the optimization with the above parameterization in Julia using the JuMP library and the Juniper branch and bound MINLP solver with IPOPT as the non-linear programming solver and CBC as the mixed-integer programming solver. Table 2 summarizes the results.

| | | Human Relocation | | | | Autonomous Relocation | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Weekdays | | Weekends | | Weekdays | | Weekends | |
| | | Perf. | Pred. | Perf. | Pred. | Perf. | Pred. | Perf. | Pred. |
| Night (00:00-06:00) | Avg. relocations | 0 | 1 | 1 | 17 | 58 | 69 | 139 | 188 |
| | Avg. rides gained | 0 | 0 | 1 | 3 | 7 | 2 | 24 | 9 |
| | Mkt. share gain (%p) | 0.0% | 0.0% | 0.1% | 0.4% | 1.5% | 0.4% | 3.4% | 1.3% |
| | Margin gain (%) | 0.0% | -0.1% | 0.0% | -0.2% | 0.5% | -0.9% | 1.5% | 0.2% |
| Morning (06:00-12:00) | Avg. relocations | 41 | 49 | 59 | 68 | 131 | 168 | 131 | 150 |
| | Avg. rides gained | 15 | 14 | 21 | 11 | 29 | 26 | 31 | 26 |
| | Mkt. share gain (%p) | 0.8% | 0.7% | 1.4% | 0.8% | 1.5% | 1.4% | 2.1% | 1.8% |
| | Margin gain (%) | 0.2% | 0.1% | 0.3% | 0.1% | 1.0% | 0.6% | 1.0% | 0.7% |
| Afternoon (12:00-18:00) | Avg. relocations | 50 | 46 | 56 | 58 | 138 | 151 | 131 | 117 |
| | Avg. rides gained | 17 | 13 | 23 | 13 | 32 | 18 | 37 | 32 |
| | Mkt. share gain (%p) | 0.7% | 0.5% | 0.8% | 0.5% | 1.4% | 0.8% | 1.3% | 1.1% |
| | Margin gain (%) | 0.2% | 0.0% | 0.3% | 0.3% | 0.9% | 0.3% | 0.9% | 0.8% |
| Evening (18:00-24:00) | Avg. relocations | 30 | 74 | 47 | 90 | 160 | 206 | 130 | 162 |
| | Avg. rides gained | 9 | 15 | 17 | 23 | 30 | 23 | 30 | 16 |
| | Mkt. share gain (%p) | 0.5% | 0.8% | 0.8% | 1.1% | 1.6% | 1.2% | 1.4% | 0.8% |
| | Margin gain (%) | 0.1% | -0.1% | 0.2% | -0.1% | 0.9% | 0.4% | 0.9% | 0.2% |

**Table 2: Optimization results overview**

Our model enables fleet operators to capture a “larger slice of the cake” by strategically controlling local market shares via relocations in order to capture an optimal amount of additional rides in a profit maximizing manner. It does so quite effectively for the periods we test. In a perfect foresight case, our model is capable of achieving profitable market share gains of up to 1.4% in a human-based relocation case and up to 3.4% in an autonomous relocation case. These market shares are achieved in a profitable manner (margin gains of up to 0.3% and 1.5% respectively). In a human-relocation case, up to 59 re-locations are conducted over a period, while in an autonomous case, where re-locations are considerably cheaper, up to 160 relocations are carried out. Yet, driven by increased relocation cost, the human-based re-location is considerably more efficient, requiring only 2.8 relocations on average per additional ride gained versus 4.6 in an autonomous scenario (both perfect foresight). We note that the effectiveness of our model depends strongly on the overall level of demand (e.g. more rentals captured and more relocations conducted during weekday afternoon rush hours versus night hours on weekdays). Other factors such as initial distribution of the vehicles would also significantly impact model effectiveness. We suspect that higher spatial resolutions could better leverage these potentials and improve model performance.

When compared against the perfect foresight scenario, relocation decisions based on predicted inputs are not optimal given the true realization of demand and supply. Re-locations are less efficient (4.4 (human-based) and 8.0 (autonomous) relocations per additional) and fail to capture the same market share, or alternatively incur profitability penalties. This is especially pronounced where the prefect foresight case already achieves low gains (i.e. limited slack) and/or where demand is low and predictive model performance less accurate. Yet, seen over all periods, re-location decisions based on our predictive model outputs outperform the benchmark “do nothing” case (against which all performance indicators are computed) both in terms of market share gains (profitable gains of up to 0.8% (human-based) and 1.8% (autonomous)) and profitability (overall positive margin gain vs. “do nothing” case in both scenarios). This shows the effectiveness of the combined predictive/prescriptive approach.

## Discussion & Future Research

This research illustrates the value of real-time competitor data for operational decision-making in shared mobility systems. We focus on the challenge of pre-emptive fleet repositioning. We select this focus because it is of large operational concern for FFVS system operators (He, Mak, et al. 2019). We proceed in two phases: (1) a predictive phase and (2) a prescriptive phase. In the first phase we develop a model to predict total market supply and total market demand for shared mobility in time and space, two core inputs for any repositioning algorithm. To do this, we leverage a unique set of big urban data such as vehicle positions of own and competitor fleet as well as relevant temporal, meteorological and geographical (POI) features. We achieve high predictive accuracy for both outflows and inflows. We also explore the trade-off between resolution and predictive performance and find that spatial resolution incurs larger penalties compared to temporal resolution. In the second phase we then formulate a prescriptive MINLP optimization framework that leverages outputs from our predictive model to derive optimal vehicle supply decisions by trading off additional density share gains (which may translate into additional rides captured) with the cost of re-locating vehicles. We achieve profitable market share gains of 1.4% (human-based) and 3.4% (autonomous) in a perfect information case. Using the output of our predictive model, profitable market share gains are still substantial at up to 0.8% and 1.8% respectively. Using higher spatial and temporal resolutions is expected to further enhance the model performance. From an academic standpoint our work has various contributions. We are the first to introduce the notion of competitor awareness in the management of FFVS systems. We combine predictive methods with mathematical programming approaches. In terms of predictive analytics we are the first to simultaneously predict supply and demand of vehicles. We also demonstrate and quantify the performance impacts of spatial and temporal resolution on predictive performance which yields novel insights. To the best of our knowledge, our work is also the first to bridge the gap between predictive analytics research of vehicle repositioning and operations research. While leveraging existing work from the OR community, we do not take a cost minimizing approach in vehicle repositioning to satisfy some stochastic or deterministic demand but recognize that re-locations take place in a competitive market where local market shares influence demand for own vehicles, a novel approach. Finally we demonstrate the effectiveness of our approach by means of a rich and unique dataset from two free-floating car sharing operators. Yet, our work generalizes beyond the case of cars to other FFVS systems such as bike sharing or e-scooter sharing systems. Practitioners can leverage our results to implement competitor aware re-location strategies. We show that such strategies are profitable today with human-based re-location. Having chosen a spatial and temporal resolution which suits their application (which our evaluation of performance trade-offs can help choose), our model will offer optimal recommendations regarding how many vehicles should be re-located when and to which locations. Benefits for early adopter are expected to be particularly high. In future work we plan to evaluate our model in different geographies and for different fleets (bike fleets). We also plan to further develop our prescriptive model by adding a dynamic component (multi-period model) and by simulating a range of counterfactuals in a large-scale simulation.

## References


Arribas-Bel, D., and Tranos, E. 2018. "Big Urban Data: Challenges and Opportunities for Geographical Analysis," *Geographical Analysis* (50:2), pp. 123–124.

Bichler, M., Gupta, A., and Ketter, W. 2010. "Designing Smart Markets," *Information Systems Research* (21:4), pp. 688–699.

Brendel, A. B., Brennecke, J. T., and Nastjuk, I. 2018. "Applying Econophysics in the Context of Carsharing - Development of a Vehicle Relocation Algorithm and Decision Support System," *Thirty Ninth International Conference on Information Systems*, pp. 1–17.

vom Brocke, J., Watson, R. T., Dwyer, C., Elliot, S., and Melville, N. 2013. "Green Information Systems: Directives for the IS Discipline," *Communications of the Association for Information Systems* (33), pp. 509–520.

Chen, C., Twycross, J., and Garibaldi, J. M. 2017. "A New Accuracy Measure Based on Bounded Relative Error for Time Series Forecasting," *PLoS ONE* (12:3), pp. 1–23.

He, L., Hu, Z., and Zhang, M. 2019. "Robust Repositioning for Vehicle Sharing," *Manufacturing & Service Operations Management* (May), Msom.2018.0734.

He, L., Mak, H., and Rong, Y. 2019. "Operations Management of Vehicle Sharing Systems," in *Sharing Economy*, M. Hu (ed.), Berlin: Springer Nature, pp. 461–484.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75–105.

Kahlen, M., Lee, T. Y., Ketter, W., and Gupta, A. 2017. "Optimal Prepositioning and Fleet Sizing to Maximize Profits for One-Way Transportation Companies," in *Thirty Eighth International Conference on Information Systems, South Korea 2017*, pp. 0–17.

Ketter, W., Collins, J., Gini, M., Gupta, A., and Schrater, P. 2009. "Detecting and Forecasting Economic Regimes in Multi-Agent Automated Exchanges," *Decision Support Systems* (47:4), Elsevier B.V., pp. 307–318.

Ketter, W., Collins, J., Gini, M., Gupta, A., and Schrater, P. 2012. "Real-Time Tactical and Strategic Sales Management for Intelligent Agents Guided by Economic Regimes," *Information Systems Research*

(23:4), pp. 1263–1283.
Ketter, W., Collins, J., Saar-Tsechansky, M., and Marom, O. 2018. "Information Systems for a Smart Electricity Grid," *ACM Transactions on Management Information Systems* (9:3), pp. 1–22.
Ketter, W., Peters, M., and Collins, J. 2016. "A Multiagent Competitive Gaming Platform to Address Societal Challenges," *MIS Quarterly* (40:2), pp. 447–460.
Ketter, W., Peters, M., Collins, J., and Gupta, A. 2016. "Competitive Benchmarking: An IS Research Approach to Address Wicked Problems with Big Data and Analytics," *MIS Quarterly* (40:4), pp. 1–53.
Kronqvist, J., Bernal, D. E., Lundell, A., and Grossmann, I. E. 2019. "A Review and Comparison of Solvers for Convex MINLP," *Optimization and Engineering* (Vol. 20), Springer US.
Kuhn, H. W. 1955. "The Hungarian Method for the Assignment Problem," *Naval Research Logistics Quarterly* (2:1–2), pp. 83–97.
Laporte, G., Meunier, F., and Wolfler Calvo, R. 2018. "Shared Mobility Systems: An Updated Survey," *Annals of Operations Research* (271:1), Springer US, pp. 105–126.
Lu, M., Chen, Z., and Shen, S. 2018. "Optimizing the Profitability and Quality of Service in Carshare Systems Under Demand Uncertainty," *Manufacturing & Service Operations Management* (20:2), pp. 162–180.
Melville, N. 2010. "Information Systems Innovation for Environmental Sustainability," *MIS Quarterly* (34:1), pp. 1–21.
Miao, J., and Niu, L. 2016. "A Survey on Feature Selection," *Procedia Computer Science* (91:Itqm), Elsevier Masson SAS, pp. 919–926.
Müller, J., and Bogenberger, K. 2015. "Time Series Analysis of Booking Data of a Free-Floating Carsharing System in Berlin," *Transportation Research Procedia* (10:July), pp. 345–354.
Pelechrinis, K., Li, B., and Qian, S. 2016. "Bike Sharing and Car Trips in the City: The Case of Healthy Ride Pittsburgh," *SSRN Electronic Journal*.
Sahr, K., White, D., and Kimerling, A. J. 2004. "Geodesic Discrete Global Grid Systems," *Cartography and Geographic Information Science* (30:2), pp. 121–134.
Shaheen, S., and Cohen, A. 2007. "Growth in Worldwide Carsharing: An International Comparison," *Transportation Research Record: Journal of the Transportation Research Board* (1992), pp. 81–89.
Shmueli, G. 2017. "Research Dilemmas with Behavioral Big Data," *Big Data* (5:2), pp. 98–119.
Sperling, D. 2018. *Three Revolutions: Steering Automated, Shared, and Electric Vehicles to a Better Future*, Washington, D.C., USA: Island Press.
Wagner, S., Brandt, T., and Neumann, D. 2016. "In Free Float: Developing Business Analytics Support for Carsharing Providers," *Omega (United Kingdom)* (59), Elsevier, pp. 4–14.
Watson, R. T., Boudreau, M.-C., and Chen, A. J. 2010. "Information Systems and Environmentally Sustainable Development: Energy Informatics and New Directions for the IS Community," *MIS Quarterly* (24:4), pp. 665–694.
Willing, C., Klemmer, K., Brandt, T., and Neumann, D. 2017. "Moving in Time and Space – Location Intelligence for Carsharing Decision Support," *Decision Support Systems* (99), Elsevier B.V., pp. 75–85.
Zhang, Y., Li, B., and Hong, J. 2016. "Understanding User Economic Behavior in the City Using Large-Scale Geotagged and Crowdsourced Data," *Proceedings of the 25th International Conference on World Wide Web - WWW '16*, pp. 205–214.