

A Space-Time Model for Demand in Free-Floating Car-Sharing Systems

Reto Bürgin^{*a}, Corrado Muratori^a, Mireia Roca-Riu^a, and Christoph Heitz^a

^aZHAW Zurich University of Applied Sciences, Institute of Data Analysis and Process Design, Technikumstrasse 81, 8401 Winterthur, Switzerland

September 8, 2023

Abstract

A novel model approach is proposed to estimate the spatio-temporal distribution of demand for free-floating Carsharing. The proposed model is based on a Poisson regression model for right censored data and estimates possibly time-varying demand rates of small subareas of a service region based on booking data with spatio-temporal information on pickups and dropoffs of cars. The approach allows operators to gain insights into the spatio-temporal distribution of demand for their service, and to estimate the loss of demand due to unavailability of cars. Moreover, it can also be used as an input to improve the design of the service, through relocation techniques or to analyze the service with macro-simulation models. Additionally, the approach is applied to a case study with real data.

1 Introduction

Carsharing is a collaborative mode of transportation that, if used appropriately, can improve urban transport services from a user and environmental perspective. Among the environmental impacts that have attracted the attention of scientists, we can include the reduction in vehicle kilometers traveled (Cervero et al., 2007), emission of pollutants (according to Shaheen & Cohen, 2013, up to 56% reduction), energy consumption (Shaheen & Cohen, 2013) and congestion (Baptista et al., 2014). Among the social impacts, it was highlighted the reduction of the number of privately owned vehicles (according to Shaheen & Cohen, 2013, up to 13 vehicles could be replaced with one shared car).

Carsharing may be classified into station-based and free-floating systems (e.g. Becker et al., 2015). In station-based systems, users start and end their trip at stations distributed within the service region. In free-floating systems, in turn, users pick up a vehicle parked near the origin of their trip using an app for booking, and end their trip by dropping the vehicle at some chosen parking within the service region. In comparison, station-based

^{*}Corresponding Author

Carsharing seems to be easier to operate because vehicles are distributed in a few known locations, while free-floating Carsharing offers the user higher flexibility and, therefore, vehicles are spread throughout the service region.

Due to its flexibility, free-floating Carsharing often suffers from a mismatch between the positioning of supply and the orientation of demand, i.e., the dropoff places do not correspond to where other users want to pick up the car (Weikl, 2016). Therefore, to operate free-floating Carsharing efficiently by positioning vehicles where demand exists, it is crucial to know how demand is distributed across the service region. The contribution of this article is to develop an approach specifically suited to estimate the spatio-temporal distribution of demand in free-floating car-sharing systems.

The challenges that Carsharing research faces may be attributable to the following three categories: (i) The definition of a transport system that is congenial to the needs of users (trying to discover which are these needs), ii) Analysis of the environmental impact of this service, and iii) The economic efficiency of the supply service (aligned with the expected demand). Demand is a crucial input for tackling any of these challenges, and therefore we believe that a good demand model estimation is essential for improving research for free-floating car-sharing systems.

In this paper, we develop an approach to estimate the spatio-temporal distribution of demand in a local free-floating car-sharing system. Thereby, we define the total demand as the number of cars that would have been booked in presence of an infinite number of available cars distributed across the service region, i.e., as the number of car-pickups (observed demand) plus the loss of demand due to the unavailability of cars. However, this problem is complex, as free-floating car-sharing systems deal with significant fluctuations in demand, depending on daytime and the area of a city (e.g. Caggiani et al., 2017), and is stochastic since it varies even between identical circumstances. Additionally, demand can hardly ever be measured directly. In most cases, we only dispose of data on effective bookings, while data on prematurely cancelled bookings are often unstructured or incomplete. Therefore, the total demand needs to be estimated based on incomplete data, using additional assumptions on how people search and decide for booking cars and advanced statistical techniques. Our approach requires the following four inputs, which we believe to be available in most cases: (i) Position and time of available cars, and pickup place and time, (ii) Study area divided into cells and time divided into intervals, with shape and dimension of the cells and the interval length being the model inputs, (iii) Assumptions on how users search and decide for booking a car. Although our approach supports various assumptions, we will assume that users start searching with a preferred pickup place in mind, and book only if a car is available within a circle around that place, with the radius of that circle being the model input, and (iv) Specification of a function for how the total demand varies across time, including unknown parameters to be estimated from the data. This time function can include month, weekdays, and daytime effects.

The remaining part of the paper is structured as follows. In Section 2, the relevant literature review is presented. The postulated statistical model to estimate the spatio-temporal distribution of the total demand is described in Section 3. In Section 4, the approach is tested on real data from a service in a major city in Switzerland, and Section 5 summarizes the work together with the conclusions.

2 Related research

The last 10 years of research have seen great growth in interest in vehicle sharing (Ataç et al., 2021; He et al., 2019), but there are still a lot of questions that need to be addressed. The main reason for these unanswered questions is that Vehicle Sharing is still missing flexible service strategies that can maintain a high level of service and guarantee long-term profitability. Below, we summarize the literature on demand forecast methods for vehicle-sharing Services.

This section is organized according to five approaches. First, research that combines demand forecasting with relocation strategies is discussed. Then, studies using multi-agent simulation tools are described, followed by studies based on the stated preferences of users. After, we focus on studies that use statistical models to estimate demand, and finally, selected research using neural networks and techniques for censored data is presented. The section ends with the identification of the Research Gap that we try to fill with this work.

2.1 Demand estimation combined with relocation strategies

To be successful, Carsharing requires the availability of resources (in terms of available vehicles and available parking spots) in the proximity of desired origin and destination of a trip to keep the service attractive (Stokkink & Geroliminis, 2021). As introduced earlier, our approach results in a loss of demand output, that could be used as an input for a free-floating user-based relocation system, a solution for rebalancing the stochasticity of cars' spatial distribution, orienting them towards the expected demand. The relocation problem is the most studied on the Carsharing's supply side and has its first references in Barth et al. (2004) and Kek et al. (2009). Further information about user-based relocation approaches can be found in Angelopoulos et al. (2018); Ataç et al. (2022); Brendel et al. (2016); Febbraro et al. (2012); Wagner et al. (2015).

Some papers include both a relocation strategy and its relative demand estimation methodology. H. Wang et al. (2010), for example, adapt a model of logistical inventory management, to forecast demand and relocate vehicles in a one-way car-sharing system. This work focuses on the specific forecasting method. Cucu et al. (2009) checks the balance between the stations, investigating them in relation to the time of departure, the day of the week, the weather conditions, and the traffic conditions associated with their addresses. Stokkink and Geroliminis (2021) develop a user-based vehicle relocation approach through the incentivization of customers and a predictive model for the state of the system based on Markov chains, following the concepts of the previous work of Repoux et al. (2019). This approach is specifically designed for one-way station-based car-sharing systems, which are different from our free-floating case. In this approach, the input demand for Markov chains is computed using the approximation method described by Raviv and Kolka (2013). Raviv's method assumes that the arrival processes of renters and returners are non-homogeneous Poisson processes, and estimates the rates using an approximation of a user-defined function. Jian et al. (2019) develop a discrete choice model that includes vehicle availability as a parameter that directly affects the user's mode. In this way, supply and demand are strictly linked together. The model aims to determine the optimal relocation decisions to maximize the Carsharing profit. The deci-

sive variables are: i) Number of vehicles relocated from node i to j at time t , ii) Number of vehicles available at node i at time t , iii) Number of users booking one-way trips from node i to j at time t and iv) Number of users booking round trips from node i to j at time t . This work has an interesting supply-demand focus, but it's not applied to free-floating Carsharing and, as input, it also needs the total travel demand of each origin-destination pair at each time step (sum of demand from Carsharing and other transport services).

2.2 Activity-based multi-agent simulations

A big fraction of current platforms, used for demand estimation of new one-way car-sharing systems, is based on activity-based multi-agent simulations i.e. micro-scale computational models for simulating the actions and interactions of autonomous agents (Vosooghi et al., 2018), that allow modelling the interaction of supply and demand. For station-based operators, Benarbia et al. (2021) propose an agent-based relocation strategy based on real-time inventory control within the framework of generalized stochastic Petri Nets (PN) and a discrete event simulation. The work of Balac et al. (2016), using a multi-agent simulation tool (MATSim), investigates the effects of supply on the demand of the existing round-trip Carsharing (also implemented in the one-way station-based). The use of MATSim with relocation agents is described by Paschke et al. (2016). MATSim was also used by Ciari et al. (2012) to estimate the demand for one-way Carsharing in the urban area of Zurich. This type of solution needs complex inputs, such as the entire transport network (including Public Transit scheduling) and population data, with which the daily plans of each user will be generated. This methodology analyses in detail the convenience of Carsharing for each user, shows the potential demand for this service, and is useful for analyzing the possible impact of new policies and new services. On the other hand, however, it does not allow analyzing the positioning of the vehicles with respect to the users who actually use the service. The analysis tool is therefore more complicated to set up than the one we will propose, and the goal is slightly different. Furthermore, this type of demand estimation has not yet been carried out for free-floating Carsharing.

2.3 Stated preference technique

Many of the initial studies aimed at understanding the potential demand of station-based Carsharing in the urban modal split. For example, Catalano et al. (2008) calibrate a modal split model, stated preference technique (SP). An overview of the literature addressed until 2013 can be found in Jorge and Correia (2013). Until that time, demand estimation had been developed almost exclusively for round-trip station-based Carsharing. Stated preferences technique is still used to catch behavioural patterns, related to specific location. Recent examples can be found in Carrone et al. (2020); Yoon et al. (2017). Lately, Zhou et al. (2020) has also adopted a stated preference methodology to elicit consumers' valuation of vehicle self-driving capability, a factor rarely examined in the literature. Regression models indicate that latent demand for this new technology is associated with respondents' travel patterns, demographics, values, lifestyles, and environmental concern (Long & Axsen, 2022). Stated preference techniques provide real choice data on some individuals and can then be translated to a larger scale of the same environment, on the basis of a series of hypotheses. However, these methods require time to assess the service, they are not adaptable to territories with different characteristics,

and do not provide information on the latent demand related to vehicles' positioning.

2.4 Regression models

Descriptive statistics or regression models may also be used for demand estimation. For example, Wagner et al. (2015) predict future demand for free-floating Carsharing, using neighbourhood data and Point of Interest (POI) data. The technique includes zero-inflated and geographically weighted regression models, from which they derive indicators for the area's attractiveness. Willing et al. (2017) extend that approach, by additionally including daytime and weekday effects in the model. Within this family of methodologies, but without regression models, Gammelli et al. (2020) predict shared mobility demand by incorporating the censored likelihood within a Gaussian Process model, with a censored likelihood function capable of handling time-varying supply. Finally, Negahban (2019) propose a methodology that combines simulation, bootstrapping, and subset selection to estimate the true demand in a bike-sharing Service. Between these approaches, only the last two take into account how supply and demand are interconnected. Compared to the neural network approaches listed below, they may also be less accurate to predict future demand, but easier to be applied in different scenarios.

2.5 Demand forecasting using Neural networks

Neural networks can provide accurate forecasts for future demand, but require a large number of parameters and context validation to be set up. Furthermore, loss of demand due to the unavailability of cars is rarely taken into account by practical implementations. N. Wang et al. (2020) study relevant indicators affecting car-sharing service's demand at the operational level and construct a micro demand forecasting model for one-way electric car-sharing systems, combining long short-term memory networks with the Granger causality test (Granger, 1969). Yu et al. (2020) propose a new approach based on deep learning techniques to assess the operation of a station-based car sharing system. They employed Long Short-Term Memory (LSTM) structure to forecast short-term future vehicle uses. Alencar et al. (2021) evaluate seven state-of-the-art forecasting models on a given free-floating car-sharing service, highlighting the potential of each technique. The assessed models include ARIMA and SARIMA, Prophet, variants of boosting algorithms, and Long Short-Term Memory (LSTM). Guidon et al. (2019) apply Cox proportional-hazards model and random survival forests to a free-floating E-bike-sharing system, using locational characteristics, weather, day of the week, and bikes in the vicinity to predict the time to pick up for each zone. Huttel et al. (2022), instead, have addressed the problem of censored mobility demand and proposed to estimate the entire distribution of latent mobility demand via Multi-Output Censored Quantile Regression Neural Networks. These methods try to model the reality but need parameterization and a context-specific initial study.

2.6 Supply-demand interaction in the analysed literature

Many of the above-mentioned studies ignore situations where there are not enough vehicles available, in which case a part of the demand is lost. Vehicle availability is a key

factor to attract new users for a car-sharing service, and, for this reason, low availability can limit the creation of new demand (Le Vine, 2011). Demand for Carsharing is difficult to model since the availability of vehicles is intrinsically dependent on the number of trips and vice versa (Jorge & Correia, 2013). The supply-demand interaction for shared cars is illustrated by Li et al. (2018), who analyse a free-floating Carsharing in a dynamic user equilibrium model. Among the above-mentioned studies, Stokkink and Geroliminis (2021) and Repoux et al. (2019) focus on the loss of demand, Balac et al. (2016); Ciari et al. (2012); Jian et al. (2019); Raviv and Kolka (2013) focus on how supply influences demand, doing a step over the simple supply-demand balance. Finally, Negahban (2019), Gammelli et al. (2020) and Huttel et al. (2022) (together with our work) model the relationship between supply and demand, taking into consideration the supply influence on the demand: they do it by treating the number of pickups (the observed demand) as a censored measurement of the total demand (the number of pickups plus the loss of demand due to unavailability of cars).

2.7 Related research Conclusions

Data detail, accessibility and reliability, high computational time, calibration and validation still remain major challenges for travel demand estimation for car-sharing systems (Vosooghi et al., 2018). Local characteristics make it complex to standardize many of the listed methods. For this reason, we tried to develop a method that is easy to apply and ductile, keeping high reliability. This ductility allows future integration with an origin-destination commuting matrix (including the estimate of its car-sharing modal split). Between the listed methods, some are more easily transferable than others, but we believe that ours reaches a higher level of reapplication easiness, maintaining convincing results. This transferability does not limit high-detail spatial and temporal analysis. Finally, this is also one of the first car-sharing demand estimation methods suited for a free-floating service.

Carsharing with shared autonomous vehicles can provide the combined benefits of autonomous driving technology and access-based consumption (Merfeld et al., 2019). The advent of self-driving vehicles will address Carsharing's problems related to parking and non-competitive access times. Solving these problems will make Carsharing a service that will be almost equal to the automated-vehicles Taxi service. In these future scenarios, users will not need to walk to pick a car that is parked far from them, because the car will go towards their position. This means that the latent demand connected to vehicles' positioning (related to accessibility) will be highly reduced, but it will still remain important to properly distribute cars following the expected demand, to further reduce the time from the request and the start of the trip.

The problem that we reviewed does not have a resolution methodology that is universally optimal, but a series of parallel methodologies, to be used on the basis of the characteristics and constraints of the analyzed service, the availability of data, and the granularity and logic of the desired outputs. The methodology that we are presenting, compared to the state of the art, brings together different characteristics, and allows to make demand forecasts with High Resolution on Free-Floating Car-Sharing Services, using few data as inputs. When total origin-destination demand data is available, some of the analysed works, such as Jian et al. (2019), could also be used to integrate our method and refine our output. To summarise the gap that we want to fill, our methodology for car-sharing

total-demand estimation covers the following four strengths: 1) Suitability (and application) of the model on a free-floating car-sharing system, 2) High temporal and spatial resolution, 3) Transferability due to a low amount of local-geography-related inputs and 4) Computation of the loss of demand, given a certain supply configuration. As far as we know, previous studies never used methodologies that allowed to focus in parallel on all of the four listed targets. We believe that this combination of strengths makes the proposed methodology a valuable tool for transportation research, especially for contexts that require an agile application and do not allow for time-consuming data preparation.

3 Space-time model

We postulate and implement a statistical model to estimate the total demand for cars of a free-floating car-sharing system at a given time and location within a service region. For this, the service region is divided into a grid of disjoint, bounded and equally sized cells $i = 1, 2, \dots, I$ and time is divided in discrete, consecutive and equally long intervals $t = 1, \dots, T$. The index i refers to a 2-dimensional square-shaped cell defined by center coordinates (e.g. $47^\circ 33' 17''\text{N } 07^\circ 35' 26''\text{E}$) and a common side length (e.g. 250m) and t to the interval $[\tau_t, \tau_t + \Delta_t)$ with τ_t the time stamp (e.g. June 16, 2022, 08:00:00) and Δ_t the interval length (e.g. 1 hour). For illustration, Figure 1 shows a discretization of a major city in Switzerland into 222 adjacent cells with 500m side length.

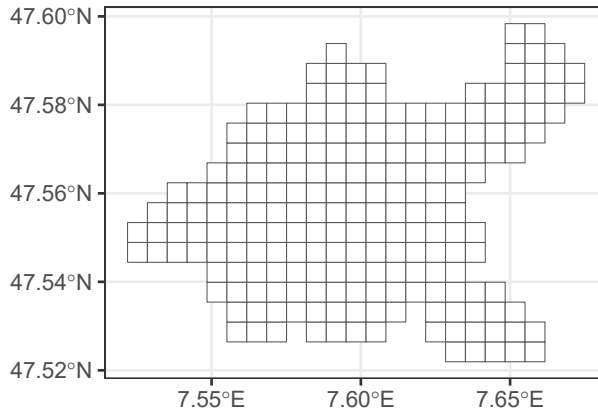


Figure 1: Discretization of a service region in adjacent cells with 500m side length.

Square-shaped cells are chosen for simplicity, and because they can be scaled up easily to any region. The model, however, can use any other shaped cells, like hexagons, or other shaped cells better adapted to the service region. The size of the cells should be chosen just small enough to allow precise conclusions about the spatial distribution of the total demand, but not smaller as smaller cell sizes will increase the number of parameters and therewith the expected computational complexity. For square-shaped cells, we recommend to set the side lengths between 50m and 500m. The duration of the intervals, Δ_t , should be as large as the total demand can be assumed to be constant within the intervals, which will be an implicit model assumption. Setting Δ_t to 1 hour, as in our case study of Section 4, may be a good rule of thumb.

The variable of interest is the total demand for booking a free-floating vehicle. Let D_{it}

be a random variable for the number of users considering booking a car at cell i and time interval t . We assume that D_{it} is Poisson distributed with rate $\lambda_{it} \Delta_t$, so that D_{it} has the density $f(d_{it}) = \frac{(\lambda_{it} \Delta_t)^{d_{it}} e^{-\lambda_{it} \Delta_t}}{d_{it}!}$ and consequently the expectation $E(D_{it}) = \lambda_{it} \Delta_t$. Note that, throughout this document, random variables are denoted with capital letters and associated observations with small letters.

The total demand D_{it} is not directly observable, as when vehicles are not available, the system does not register any information of users looking for a vehicle. We try to estimate the above mentioned rate λ_{it} based on the number of cars available for rent, denoted by C , and the number of pickups observed, denoted by P , in the proximity of cell i at interval t . The model accounts for four situations, and is build sequentially: First, we formulate a model that predicts pickups of available cars by linear combinations of the total demand rates λ_{it} . Second, we extend that model by allowing the rates λ_{it} to vary across time. Third, we take into account situations where the number of available cars was potentially insufficient to satisfy the total demand. Fourth, a smoothing approach is proposed to consider that neighboring cells are expected to have a similar total demand and to simplify the parameter estimation.

To build up the model, we first discuss specific small case examples to outline the logic of how the model spatially links the total demand rates with the observable pickups, and how the model can deal with time-varying total demand. Afterwards, the model is generalized to any grid and extended to situations where not enough cars are available.

Spatially linking the total demand and pickups A picked car at cell i and interval t (that is, $p_{it} = 1$) does not necessarily imply that the demand originates from cell i . It is also possible that the user would have preferred to pickup the car from an adjacent cell, where there was no car available at the time. To model the number of pickups P as a function of the total demand rates λ_{it} , we assume that users have a preferred pickup cell (i.e., the origin of the demand), but choose with equal probability any car standing in a cell not further away than r_{\max} (e.g. 500m) from that origin cell, as measured by the distance between the centers of the cells. If no car is close enough, the demand gets lost. Note that r_{\max} becomes operative only if the center-to-center distances of neighboring cells are smaller than r_{\max} , otherwise a pickup is simply linked to the demand from the same cell.

The r_{\max} assumption is used for its simplicity, while the model below can accept other assumptions better suited for the considered problem. For example, r_{\max} may be set to different values across sections of the service region. The prerequisite for an alternative assumption is that it defines for each cell i an according set of cells that could be the origin of demand for a pickup from cell i . In practice, r_{\max} is generally unknown, and may be determined by using the rule of thumb of Seign and Bogenberger (2021) of about 300-500m, by conducting a survey or by choosing r_{\max} such that a goodness of fit measure (such as the likelihood criterion, see below) is optimized.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	🚗	19	20
21	22	23	24	25

Figure 2: Example situation in a 5×5 grid with cars at cell 18 and the according demand area (yellow).

1	2	3	4	5
6	7	8	🚗	10
11	12	13	14	15
16	17	🚗	19	20
21	22	23	24	25

Figure 3: Example situation in a 5×5 grid with cars at cells 9 and 18 and the according demand areas (orange and yellow).

Figure 2 shows a square grid with 5×5 cells and only one car (or more than one) at cell 18. As an example, we arbitrarily set r_{\max} such that a picked up car can be assigned to a demand from cell 18 or cells around. We call the corresponding set of cells, which is highlighted in yellow in Figure 2, as the *demand area* for cars in cell 18. If we assume temporarily that the rate parameters do not vary over time (i.e., $\lambda_{it} \equiv \lambda_i$) and the number of cars available in cell 18 is higher than necessary to satisfy the total demand for cars (i.e., $C_{18,t} > \sum_{i \in \{12,13,\dots,24\}} D_{it}$), then the expected number of pickups from cell 18 is equal to the sum of the expected total demand of the individual cells,

$$\begin{aligned} E(P_{18,t}) &= E(D_{12,t}) + E(D_{13,t}) + \dots + E(D_{24,t}) & (1) \\ &= (\lambda_{12} + \lambda_{13} + \dots + \lambda_{24}) \Delta_t. & (2) \end{aligned}$$

Since the right hand of the above equation is a sum of Poisson distributed random variables, $P_{18,t}$ is also Poisson distributed with parameter $(\lambda_{12} + \lambda_{13} + \dots + \lambda_{24}) \Delta_t$ (see e.g. Grimmet & Welsh, 2014, Exercise 4.40). Given that the r_{\max} assumption holds, this equation collects all relevant information for estimating the parameters λ_i for the situation shown in Figure 2. The demand of cells further away than r_{\max} from cell 18 cannot be served because there is no car in proximity, and therefore the situation does not provide information on the corresponding parameters λ_{it} . In order to get estimates for each of the parameters λ_{it} , we need to have several data of various moments in time where each cell is part of a demand area of a standing car. Otherwise, if a cell i is never part of a demand area, then the corresponding parameters λ_{it} cannot be identified.

In reality, users might have more than one vehicle in their proximity. For example, Figure 3 presents a situation where two cars are available, one at cell 9 and another at cell 18. Here, the demand areas around cells 9 and 18 overlap at cells 13 and 14. To handle such situations, the total demand rates from cells which have more than one vehicle at reach (i.e. closer than r_{\max} distance) may be split in half, such that each of the two cells with vehicles obtains half of the total demand rates. This results in the following two equations for the expected number of pickups:

$$\begin{aligned} E(P_{9,t}) &= (\lambda_3 + \dots + \lambda_{10} + \lambda_{13}/2 + \lambda_{14}/2 + \lambda_{15}) \Delta_t \quad \text{and} & (3) \\ E(P_{18,t}) &= (\lambda_{12} + \lambda_{13}/2 + \lambda_{14}/2 + \lambda_{17} + \dots + \lambda_{24}) \Delta_t. \end{aligned}$$

Alternative rules for dividing cells of intersections of demand areas could be considered. For example, if we believe that users always pickup the closest car, then we would assign $\lambda_{13,t}$ to the cars of cell 18, and $\lambda_{14,t}$ to the cars of cell 9. Therewith,

$$\begin{aligned} E(P_{9,t}) &= (\lambda_3 + \dots + \lambda_{10} + \lambda_{14} + \lambda_{15}) \Delta_t \quad \text{and} \\ E(P_{18,t}) &= (\lambda_{12} + \lambda_{13} + \lambda_{17} + \dots + \lambda_{24}) \Delta_t. \end{aligned} \quad (4)$$

It is not always possible to fully separate the demand areas with the closest car rule. If, for example, the car of cell 18 in Figure 3 is moved to cell 17, then cell 13 has the same distance to both cars. In these cases we split cell 13 in half as in Eq. (3).

The second step of the model accounts for total demand variation across time. For example, total demand may change between mornings and evenings, weekdays or seasons. To account for these variations, we adapt the equations from above with further parameters. For example, suppose that the rates vary between weekdays (Monday to Friday) and weekends (Saturday and Sunday) so that the total demand rate of cell i is λ_i at weekdays and $\lambda_i + \beta$ at weekends. Let v_t be an indicator with value 1 if time interval t corresponds to a weekend, and 0 otherwise. Eq. (3) for pickups of cars from cell 9 with an additional weekend effect extends to:

$$\begin{aligned} E(P_{9,t}) &= ((\lambda_3 + v_t\beta) + \dots + (\lambda_{10} + v_t\beta) + (\lambda_{13} + v_t\beta)/2 \\ &\quad + (\lambda_{14} + v_t\beta)/2 + (\lambda_{15} + v_t\beta)) \Delta_t \\ &= (\lambda_3 + \dots + \lambda_{10} + \lambda_{13}/2 + \lambda_{14}/2 + \lambda_{15}) \Delta_t + 8v_t\beta\Delta_t \end{aligned} \quad (5)$$

Basic model Eq. (5) refers to a specific situation for the considered 5×5 grid and a simple specification for time effects. For general situations, grids and specifications for time effects, we relate the total demand for available cars in some cell j and time t with a linear combination of the expected total demand of the individual cells at some at a chosen reference time interval and a linear combination of further parameters multiplied with time-related variables Z , as below:

$$E(P_{jt}) = \sum_{i=1}^I u_{jti} E(D_{it}) = \mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta} \quad (6)$$

Elements u_{jti} indicate the share of the total demand rate of cell i that is assigned to cars standing in cell j . They take values between 0 and 1, and the sum over all cells j with vehicles must be 1 (i.e., $\sum_j u_{jti} = 1$). The linear predictor on the right of Eq. (6) has two components. \mathbf{x}_{jt} and $\boldsymbol{\lambda}$ are design and parameter vectors of length I (number of cells) to predict the expected number of pickups of cars from cell j for some reference time interval, and \mathbf{z}_{jt} and $\boldsymbol{\beta}$ are design and parameter vectors of length P that take into account time effects towards the reference time. For example, expressing Eq. (5) with Eq. (6) yields $\mathbf{x}_{jt}^T = (0, 0, \Delta_t, \Delta_t, \Delta_t, \Delta_t, \Delta_t, 0, 0, \Delta_t/2, \Delta_t/2, \Delta_t, 0, \dots, 0)$ and $\mathbf{z}_{jt}^T = 8v_t\Delta_t$.

The proposed model is quite flexible to estimate how total demand is spatially distributed and which time effects are taken into account. The main restriction is that the dependencies between the expected pickups and the possibly time-varying total demand rates

have to be linear regarding the unknown parameters. The linearity restriction is not as limiting as it might appear at first sight. Nonlinear evolution along time may be modeled using dummy variables or polynomials that are still linear in its parameters. Section 4 presents a case study with real data to provide a hands-on specification.

If we assume that the number of cars is always sufficient to satisfy the total demand of the corresponding demand areas, then the parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ can be estimated using tools for Poisson regression models, such as maximum likelihood estimation. For some cell j with cars at interval t , the probability of the observed number of pickups p_{jt} is

$$P(P_{jt} = p_{jt}) = \frac{(\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta})^{p_{jt}} e^{-(\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta})}}{p_{jt}!}, \quad (7)$$

and $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ can be estimated by maximizing $\sum_{t=1}^T \sum_{j \in \{i | c_{it} > 0\}} \log P(P_{jt} = p_{jt})$, with c_{it} the number of cars available for booking in cell i and interval t .

The assumption that the number of cars available is always sufficient to satisfy the total demand from the according demand areas may be realistic if Δ_t is chosen small enough to not expect more than one pickup within the time intervals. In general, however, this assumption does not hold, e.g., when two users want to book the same and only car available practically at the same time. Moreover, bypassing the assumption by decreasing Δ_t blows up the data volume and therewith increases the already considerable computational effort for estimation even more.

Using a Poisson model for right censored data (e.g. Famoye & Wang, 2004) allows to account for situations where the total demand possibly exceeded the number of available cars. The censored Poisson model assigns different probabilities depending on whether the number of pickups is smaller or equal to the number of available cars: In the first case, we assume that the demand was fully satisfied and compute its probability using the Poisson density function. In the second case where the number of picked cars equals the number of available cars, we assume that the number of picked cars is right censored, i.e., could have been larger if there were more cars available. Therefore, we compute its probability as the cumulative Poisson density function from the number of picked cars to infinity. Expressed mathematically,

$$P(P_{jt} = p_{jt}) = \begin{cases} P\left(\left[\sum_{i=1}^I u_{jti} D_{it}\right] = p_{jt}\right) = \frac{(\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta})^{p_{jt}} e^{-(\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta})}}{p_{jt}!} & \text{if } c_{it} > p_{it} \\ P\left(\left[\sum_{i=1}^I u_{jti} D_{it}\right] \geq p_{jt}\right) = 1 - \sum_{b=0}^{p_{jt}-1} \frac{(\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta})^b e^{-(\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta})}}{b!} & \text{otherwise.} \end{cases} \quad (8)$$

Estimations for $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ can be obtained by maximizing the log-likelihood,

$$l(\boldsymbol{\lambda}, \boldsymbol{\beta}) = \sum_{t=1}^T \sum_{j \in \{i | c_{it} > 0\}} \log \left[\left(\frac{(\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta})^{p_{jt}} e^{-(\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta})}}{p_{jt}!} \right)^{1(c_{it} > p_{it})} + \left(1 - \sum_{b=0}^{p_{jt}-1} \frac{(\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta})^b e^{-(\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta})}}{b!} \right)^{1(c_{it} = p_{it})} \right]. \quad (9)$$

Since no closed form solution exists to maximize Eq. (9), we developed a gradient-based implementation in R (R Core Team, 2022) based on the optimizer function `nlmminb()` (Gay, 1990). The implementation allows the parameters $\boldsymbol{\lambda}$ (and possibly $\boldsymbol{\beta}$) to be log-transformed to avoid negative estimates for rate parameters and automatically drops λ_i parameters associated with cells that never were in a demand area and therefore cannot be assessed. The estimating equations and the developed R functions are available on request.

Smoothing The postulated model does not assume any relationship between total demand rates from adjacent cells. In general, it is reasonable to think that adjacent cells might have similar rates, or that the spatial distribution of these parameters should change smoothly across the service region. Only in particular cases, like geographical circumstances (e.g. a river) or other demand singularities (e.g. location of a big demand attractor), the total demand rate parameters might experience an abrupt spatial change. We propose to use a kernel smoothing approach (e.g. Hastie et al., 2009, Chap. 6) to construct dependencies between the total demand rate parameters. The idea is to estimate “pseudo” total demand rates $\tilde{\lambda}_k$ for $K < I$ chosen supporting points, and calculate the total demand rates λ_i of the I cells as weighted sums of these λ_k , that is,

$$\lambda_i = \sum_{k=1}^K w_{ik} \tilde{\lambda}_k = \mathbf{w}_i \tilde{\boldsymbol{\lambda}}, \quad \text{with} \quad \sum_{k=1}^K w_{ik} = 1. \quad (10)$$

Figure 4 shows as an example the use of $K = 9$ nine supporting points for a grid with 5×5 cells. The supporting points are located at the edges of the cells, which is not a requirement.

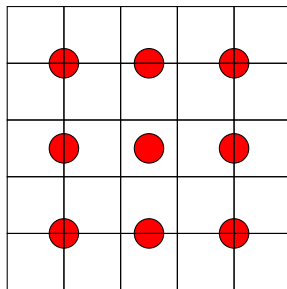


Figure 4: Example of $K = 9$ supporting points (red) for a grid with 5×5 cells.

Any kernel function can be used to compute the weights w_{ik} , such as the Epanechnikov (Epanechnikov, 1969) or the Gaussian kernel. We propose an implementation that uses higher weights if a supporting point is closer to cell i : Let r_{ik} be the euclidean distance between the center coordinates of cell i and the support point k . Using the standard Gaussian kernel, the w_{ik} are computed as:

$$w_{ik} = \frac{\phi(r_{ik})}{\sum_{k=1}^K \phi(r_{ik})}, \quad \phi(r_{ik}) = \frac{e^{-r_{ik}^2/2}}{\sqrt{2\pi}}. \quad (11)$$

Now we can rewrite the linear predictor of our model to:

$$\mathbf{x}_{jt}^T \boldsymbol{\lambda} + \mathbf{z}_{jt}^T \boldsymbol{\beta} = (\mathbf{x}_{jt}^T \mathbf{W}) \tilde{\boldsymbol{\lambda}} + \mathbf{z}_{jt}^T \boldsymbol{\beta}, \quad \text{with } \mathbf{W} = \begin{bmatrix} \mathbf{w}_1^T \\ \vdots \\ \mathbf{w}_I^T \end{bmatrix} \quad (12)$$

which is again linear in its parameters $\tilde{\boldsymbol{\lambda}}$ and $\boldsymbol{\beta}$ and can be estimated with the previously described tools.

The simulation studies in the Appendix show that estimates for $\boldsymbol{\lambda}$ from the smoothing approach can have lower variance than those from the original model, due to the smaller number of unknown parameters involved. The downside of this are potential biases for the λ_i s, which can also be seen in the Appendix.

The smoothing approach involves specifying the location and the number of supporting points. An equally-spaced grid is used most frequently, for simplicity reasons. More supporting points allow capturing demand distributions with finer structures (cf. Wand, 1994), but decrease the wiggleness and increase the computational effort. A practical implementation is provided in Section 4.

4 Case study

We consider data provided by the Swiss commercial company Mobility¹ from their so-called Mobility-Go free-floating car-sharing service in a major city of Switzerland during 2021, where the service was operated with about 128 cars. The raw data consisted of 28'682 records on individual rentals without service trips and include information on the vehicle number, coordinates and time stamps of the pickup and dropoff. To discretize space and time, each record was assigned to one of 747 cells of 250m side length, based on the pickup coordinates, and to one of 8'759 hourly intervals (e.g. June 16, 8 to 9 o'clock) based on the pickup time. The number of cars available for an interval was computed as the number of cars at the beginning of that interval plus the number of cars dropped during the interval.

¹<https://www.mobility.ch>

Descriptive analysis

To provide an overview of the used data, we divided Basel into nine equally sized districts, divided according to the cardinal direction from the center of the city. Figure 5 shows the average number of pickups per hour along daytime, weekdays and months for each area, together with the average of the nine districts.

Daytime presents a classical temporal demand pattern with few demand during night period (below 0.2 PU/h per district), a first strong increase in the morning between 7:00 and 9:00, and a moderated continuous increase until the daily peak at 19:30 in the afternoon. The values are higher in the afternoon because the graph also includes non-working days. If we only consider working days, this daytime profile is much more balanced between morning and afternoon.

Regarding the days of the week, we can see that, as expected, Saturday is the busiest day with about 0.45 pickups per hour and area, followed by Friday and Sunday. For workdays we see a slight growth from Monday to Thursday, while we could have expected a flat profile for these days. Along the months we see a tendency of higher values in cold months and lower values during summer. The relative maximum in May it is not self-explanatory, but it is important to remember that from this graph we can extrapolate little information, considering that that we have a one-time pattern and not a pattern that was repeated many times, such as the daily profile. The pattern along the months could have been influenced by factors such as pricing, policies, information campaigns or the COVID-19 pandemic that was still relevant in 2021.

The nine districts into which we divided the city differ in some cases from the average profile, both in terms of frequency and in terms of profile shape. Readers can examine Figure 5 to understand differences between the districts.

As mentioned above (Sec. 2), some relocation models (classified as non-predictive) rely on few indicators to characterize the demand. Reiss and Bogenberger (2017) used three indicators for each district to detect the attractiveness of a district: demand factor, origin destination factor, and idle times. Here, we check the balance between supply and demand in the nine districts by using their demand factor, which is defined as the ratio between rentals and vehicles in a district.

	South-West	West	North-West	South	Center	North	South-East	East	North-East	All-district Average
Pickups	7.24	11.24	9.17	13.62	10.79	7.71	5.87	7.28	3.82	8.53
Cars	13.69	16.27	12.85	19.67	12.77	9.95	11.99	10.79	6.93	12.77
Pickups/ Cars	0.53	0.69	0.71	0.69	0.84	0.77	0.49	0.67	0.55	0.66

Table 1: Daily average of the number of pickups (P), number of cars available (C) for rent and demand factor (P/C) by district and average over the districts.

In Table 1 we can consider the demand factor as a parameter to evaluate where relocation should be carried out (from the districts with many cars and low demand factor to the districts with a high demand factor). We note that the South-West and North-West districts have higher availabilities of cars than the center, despite having a

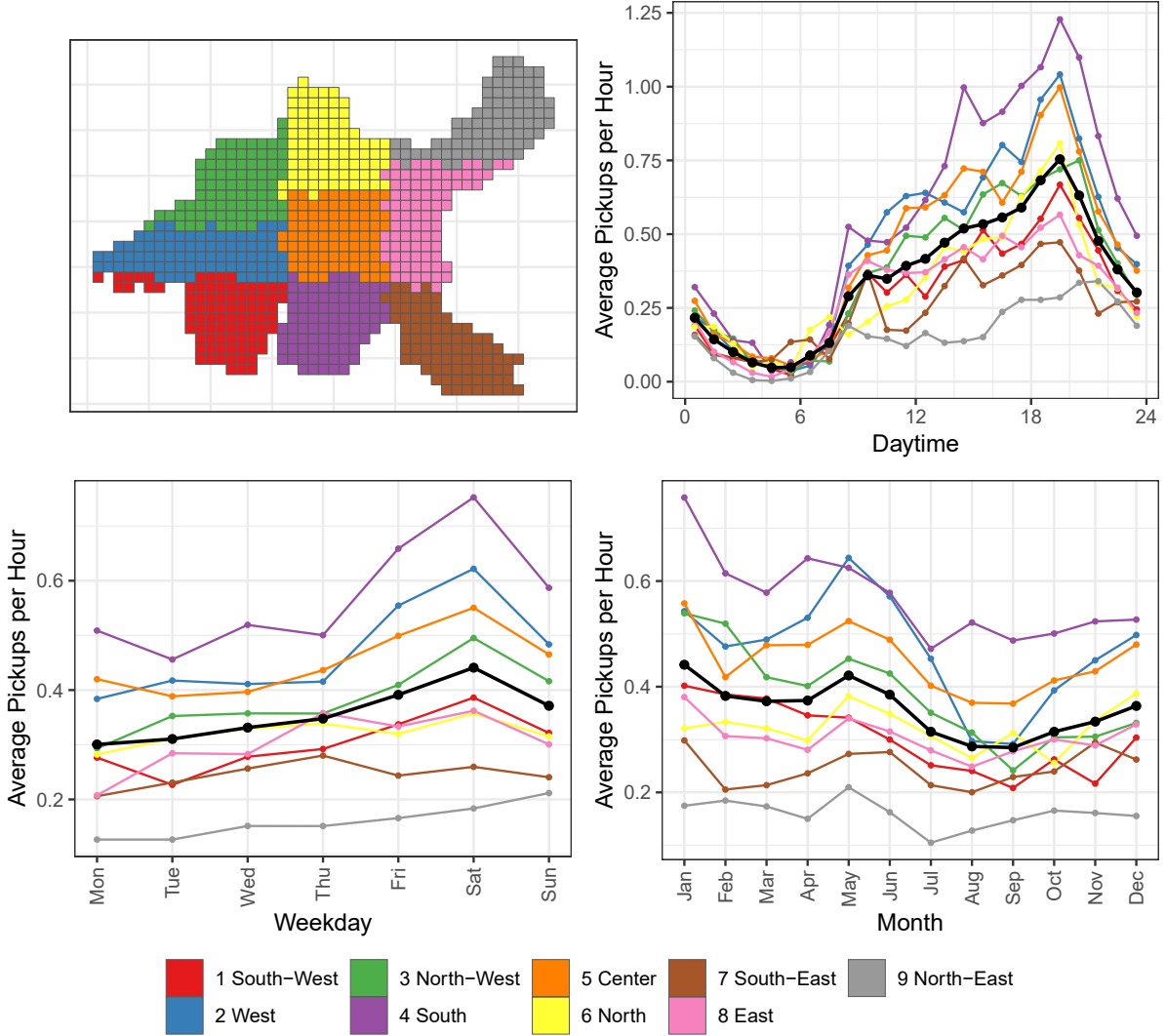


Figure 5: Average number of car pickups per hour for the Mobility-Go data 2021, along daytime, weekday and month and separately by 9 self-defined districts of the city (south-west, ...). The black line represents the average of the 9 colored lines.

considerably lower number of pickups. Candidates for receiving cars through relocation are the Center and the North districts, exhibiting with 84% and 77% the highest ratios between pickups made and car availability. On the contrary, the South-East district has, in average, twice as much cars available than car pickups, and is therefore the best candidate to take cars away for relocation. Here, the demand factor is applied on a large scale but could give further information if applied to smaller districts. Similar conclusions on relocation can be obtained from the proposed modelling approach applied in the subsequent analyses, which uses a much more detailed spatial resolution by design.

Spatio-temporal model specifications

In this section, a spatio-temporal model is estimated based on the previously described data. Based on Seign and Bogenberger (2021) suggesting that cars should be available within 300-500m walking-distance, we set $r_{\max} = 500\text{m}$ for all models. This implies that

users book a car only if they find an available vehicle not further away than 500m from the preferred cell, as measured by the distance between the centers of the cells. Furthermore, we assumed the total demand rates to vary across months (Jan, Feb, . . . , Dec), weekdays (Mon, Tue, . . . , Sun) and day intervals (night: [0-6), morning: [6-12), afternoon: [12-18) and evening: [18-24)). Such time effects were implemented with dummy variables, where the reference categories (September, Monday and night) refer to the interval with the lowest average number of pickups. Using the dummy variable specification allows to deal with nonlinear evolution across time, which can be identified from Figure 5.

We implemented four specifications of the proposed model (M_1 to M_4), which differ by how overlaps of demand areas treated (cf. Sec. 3, Spatially linking the total demand and pickups), and whether or not smoothing (cf. Sec. 3, Smoothing) was applied. The models M_1 and M_2 use overlapping demand areas (total demand rates from cells which have more than one vehicle at reach are distributed evenly across the intersecting demand areas), whereas M_3 and M_4 use the closest car demand areas (total demand rates are assigned, whenever possible, to the demand area of the closest car). Smoothing was applied only on the models M_2 and M_4 . A grid of 81 uniformly spread supporting points was used for smoothing, reducing the total number of unknown parameters from 767 to 101.

To estimate the models, corresponding design vectors \mathbf{x}_{jt} and \mathbf{z}_{jt} had to be prepared based on the stated assumptions for the demand areas around cars and time effects. Smoothing additionally required the preparation of the weights \mathbf{w} , see Eq. (12).

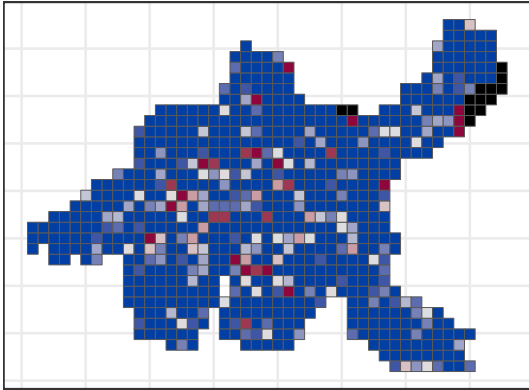
The number of unknown parameters is 767 for basic models (747 one for each zone and 20 temporal parameters) and 101 for smoothing models. For estimation, all parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ were estimated on the log scale. This ensures the estimated total demand rates to be always larger than zero.

Spatial distribution of the total demand rates Figure 6 shows the estimated $\boldsymbol{\lambda}$ coefficients, which refer to the estimated total demand rates (per hour) at a Monday in September 2021 from 0 to 6 o'clock AM. Most obvious is that basic models result a patchwork of estimated rates, while smoothing models do not. Comparing the two basic models it can be seen that overlapping demand areas result more peaks than the closest car demand areas. This may be related to our findings from the simulation studies in the Appendix, pursuant to which the total demand rate estimates of overlapping demand areas have higher variance, see Figure A.13.

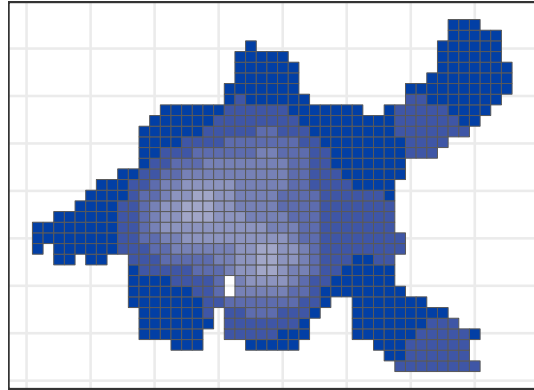
Figure 6 exhibits that the basic models may not be able to estimate the total demand rates for all cells, see the black squares in the north-east. This is because there was never a car available for rent within circle r_{\max} around these cells. Smoothing results estimates for these cells, however, these estimates should be interpreted carefully since kernel smoothing approaches are known for boundary bias (e.g. Hastie et al., 2009, Sec. 6.1).

In terms of model fit, we found that the log-likelihood, the Akaike information criterion (AIC, Akaike, 1974), the root mean squared error (RMSE) and the mean absolute error (MAE) of the basic models are slightly superior to smoothing, see Table 2. While the superiority of the basic model regarding the log-likelihood, RMSE and MAE was expected because the smoothing models are merely restricted submodels with fewer parameters, the superiority regarding the AIC indicates that the smoothing is too strong and should be improved, e.g., by adding supporting points or placing them more more

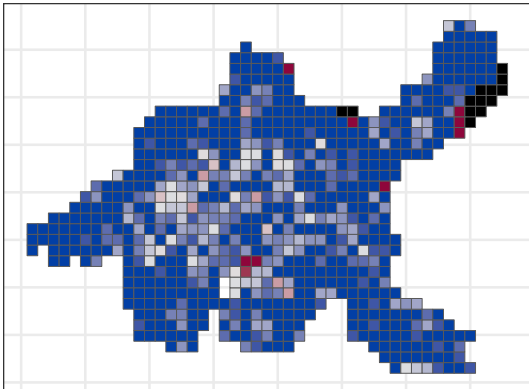
M₁ : Overlapping demand areas, basic



M₂ : Overlapping demand areas, smoothing



M₃ : Closest car demand areas, basic



M₄ : Closest car demand areas, smoothing

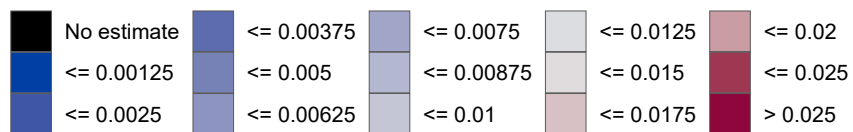
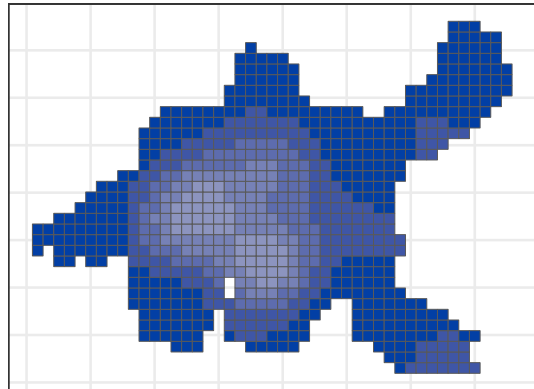


Figure 6: Estimated spatial distribution of total demand rates per hour for renting cars at the baseline setting (Mondays in September 2021, 0 to 6 o'clock AM) from the different models M₁ to M₄.

efficiently. Furthermore, the models M_1 and M_2 with overlapping demand areas perform insignificantly better than the according models M_3 and M_4 with closest car demand areas regarding the log-likelihood, AIC and RMSE, but insignificantly worse regarding the MAE.

Model	# Params.	Log-Likelihood	AIC	RMSE	MAE
M_1 : Overlapping, basic	754	-117605	236719	0.032080	0.06298
M_2 : Overlapping, smooth	101	-117612	236732	0.032214	0.06316
M_3 : Closest car, basic	754	-118920	238043	0.032084	0.06294
M_4 : Closest car, smoothing	101	-119030	238261	0.032222	0.06313

Table 2: Goodness of fit measures of the models M_1 to M_4 . The RMSE and the MAE were computed using the differences between the observed and the expected number of pickups for each cell and interval with cars, $p_{jt} - E_{\hat{M}}(P_{jt}|c_{jt})$.

Figure 7 shows the estimated total demand rates of the model M_2 with overlapping demand areas and smoothing, which is the best among the smoothing models according to the log-likelihood. Figure 7 is identical to the top right plot of Figure 6, but with a finer color scale to facilitate closer examinations. The plot highlights two regions with higher total demand within the center of the service region, which can be attributed to regions close of the train station and the old town, and two local peaks at north-east and south-east.

M_2 : Overlapping demand areas, smoothing

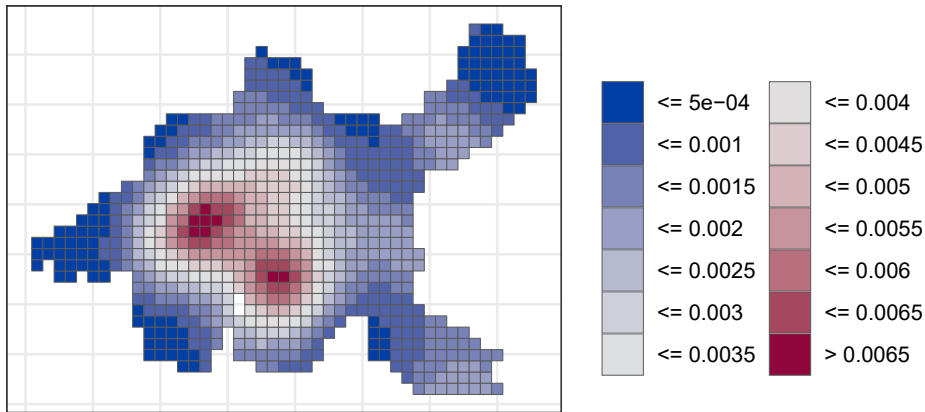


Figure 7: Estimated spatial distribution of total demand rates per hour for renting cars at the baseline setting (Mondays in September 2021, 0 to 6 o'clock AM). Same as top right panel in Figure 6, but with adjusted color scale.

The Figures 6 and 7 present the estimated distribution of the total demand rates for the reference time Monday in September 2021 from 0 to 6 o'clock AM. The time effects discussed below allow total demand rates to vary over time, e.g. because the demand might vary across weekdays. However, because the considered models assume that time effects are constant across the whole service region, the estimated spatial distribution will not change, only the rates will be increased in every cell by the same factor.

Temporal distribution of total demand rates Figure 8 shows the estimates of the four models regarding the three considered types of time effects. The shape of the coefficients along time is very similar between the four models. The plot on month effects on the top left reveals that the total demand was highest in January, and there was a temporary peak in May to June. We expect this pattern to be related to the Covid-19 pandemic and to not be repeated in 2022. Estimates for weekday and daytime effects can better accommodate the expected: We find a clear peak for total demand on Saturdays, and higher total demand at afternoon and evenings than at night and mornings.

Some coefficients almost reach value zero, which is the lowest possible value due to estimating the coefficients on the log scale. This is especially notable for month effects where the estimates indicate that the total demand of July, August, October and November was practically the same as in the reference month September. To find out if those estimates with values close to 0 relate to convergence difficulties, we used different optimizer routines and applied a number of small model modifications, such as changing the reference categories and the side lengths of the cells. However, the optimizer routines reported to converge, and the model modifications did neither clear out the close-to-zero coefficients nor change the findings for time effects fundamentally. Moreover, the order of the estimated time effects is consistent with the results from the descriptive analyses, cf. Figure 5. For these reasons, we assume that the estimated coefficients are reliable.

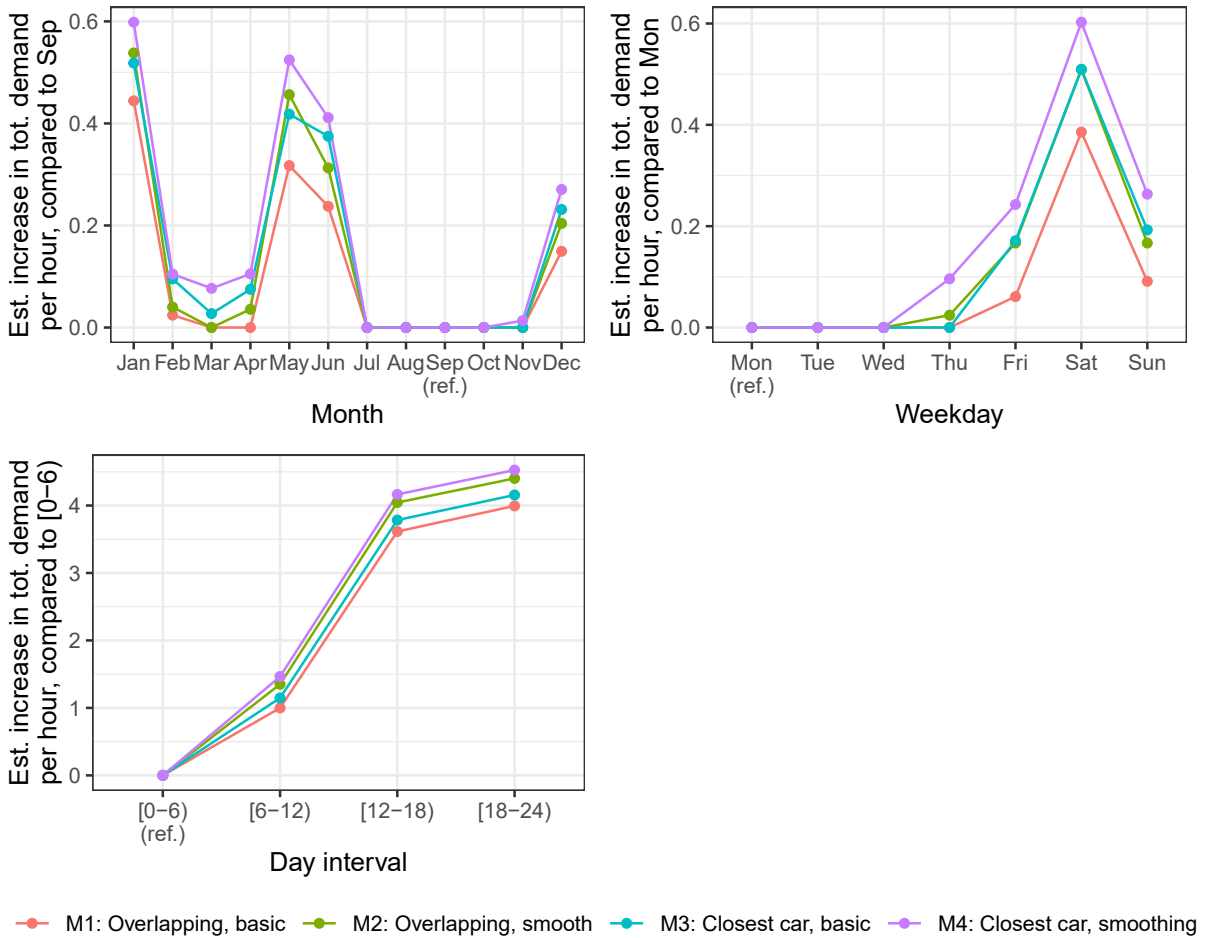


Figure 8: Estimated total increase in total demand of months, weekdays and day intervals as compared a Monday at September between 0 and 6 o'clock (reference).

Loss of demand The estimated models may be used to estimate the loss of demand. In line with our model, we distinguish between two types of loss of demand:

1. **No cars in proximity** In situations where there is no offer, the entire demand gets lost. For some cell i and time interval t that is further away than r_{\max} from the nearest cell with cars, we estimate this type of loss of demand as $\hat{l}_{it}^{(1)} = \hat{\lambda}_i + \mathbf{z}_{it}^T \hat{\beta}$, i.e., as the estimated total demand at the baseline setting (in our case: Mondays in September 2021, 0 to 6 o'clock AM) plus the estimated time effect for interval t .
2. **Not enough cars in proximity** In situations where all available cars are picked up, loss of demand occurs because more cars could have been rented with a larger offer. Consider some cell j with $c_{jt} > 0$ cars and $p_{jt} = c_{jt}$ pickups at time t . The conditional expectation for the total demand D_{jt} is in this situation

$$\mathbb{E}(D_{jt} | D_{jt} \geq p_{jt}) = \frac{\sum_{k=p_{jt}}^{\infty} k \cdot \mathbb{P}(D_{jt} = k)}{\mathbb{P}(D_{jt} \geq p_{jt})}, \quad (13)$$

where the model estimate for $\mathbb{P}(D_{jt} = k)$ is $\hat{\mathbb{P}}(D_{jt} = k) = \frac{(\mathbf{x}_{jt}^T \hat{\lambda} + \mathbf{z}_{jt}^T \hat{\beta})^k e^{-(\mathbf{x}_{jt}^T \hat{\lambda} + \mathbf{z}_{jt}^T \hat{\beta})}}{k!}$. Therefore, we estimate this type of loss of demand as $\hat{l}_{jt}^{(2)} = \hat{\mathbb{E}}(D_{jt} | D_{jt} \geq p_{jt}) - p_{jt}$, i.e., the estimated conditional expectation minus the number of pickups. The peculiarity of this type of loss of demand is that our model implies that it cannot solely be attributed to cell j , but to all cells not further away than r_{\max} from cell j . However, we did not find an analytical formula of how to divide $\hat{l}_{jt}^{(2)}$ to the neighboring cells, and therefore they are attributed to the cell j car in the following results. It should be noted that in the presented case study this second type of loss of demand is practically negligible compared to the first type.

The two proposed estimates for loss of demand above refer to the number of cars not rented by the car-sharing system compared to the same system with the same demand but an infinite number of cars available. The following results on the estimated loss of demand are based on model M_2 (overlapping demand areas, smoothing) and the 2021 data used for estimating the model.

For the entire service region, we estimated an average loss of demand of 14.2 cars per day, thereof 12.9 because there were no cars in proximity, and 1.31 because there were not enough cars in proximity. Compared to the average number of pickups per day of 77.0, this means that the number of rentals could be increased by about 18.5% by providing an unlimited number of cars, assuming that increased offer would not increase the demand.

To detect the loss of demand locally, Figure 9 shows the spatial distribution of the average loss of demand per day, based on the model M_2 . According to Figure 9, loss of demand is especially pronounced in the center of the city, and not exactly at the total demand peaks situated on the left and bottom of the center, cf. Figure 7.

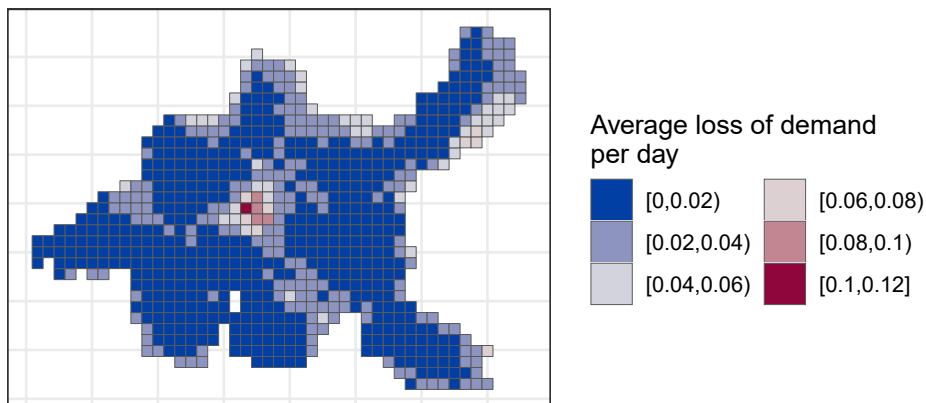


Figure 9: Spatial distribution of the estimated average loss of demand per day in 2021, based on the model M_2 .

5 Summary and conclusions

This article proposes a novel model approach to estimate the spatial and temporal distribution of total demand rates for free-floating Carsharing. The proposed model is based on a Poisson regression model for right censored data and estimates possibly time-varying demand rates of discrete cells of the service region based on booking data with spatio-temporal information on pickups and dropoffs of cars. The model is quite flexible as it can accommodate various shapes of cells of selectable size and different temporal effects. The model was successfully applied for a case study in a major city of Switzerland with data from year 2021.

The proposed model is useful for the following purposes: First, the model provides insights to operators on how total demand was spatially distributed and evolved over time. This insight can hardly be gained using simple descriptive statistics, because total demand is often not directly observable, and therefore must be estimated using auxiliary variables such as the number of pickups, and an advanced modelling technique such as regression. And second, the model may be used to estimate the loss of demand due to unavailability of cars. These insights may prove useful to designate convenient dropoff places in incentive schemes for user-based relocations or to extract input parameters for macro-simulation models.

Limitations The total demand rates estimated with our approach refer to the free-floating car-sharing service that provided the data. Therefore, for competitive situations with multiple services, they cannot be interpreted as the global demand rates of the considered service region. If global demand rates are of interest, the model must be estimated using data that combine the competing services. Moreover, the estimated total demand rates do not take into account for other transport services such as public

transport. Therefore, they refer to a given split of available transport services, and may be sensitive towards launches or discontinuations of other transport services.

Future work Further investigations could focus on practical aspects of the model. Implementations for larger and more frequented service regions would help to define the scope of our approach and to improve guidelines for model specification. Furthermore, operators may be interested into forecasting future total demand. Forecasting involves extrapolation and has yet not been elaborated with our model approach, partly because it seemed difficult to be implemented for data from the Covid-19 era. A forecasting approach should additionally take into account for auxiliary predictor variables such as weather, and should be able to deal with temporal correlation (e.g., by using a model with autoregressive errors) to provide reliable prediction intervals.

Data availability

Access to the booking data used to support the findings of this study requires the approval of the data owner Mobility and can be requested from the corresponding author.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding statement

This paper represents independent research funded by the Swiss Innovation Agency Innosuisse (project title: *User-based Redistribution für Free-floating Carsharing*). The authors are grateful to the Innosuisse for its financial support.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Alencar, V. A., Pessamilio, L. R., Rooke, F., Bernardino, H. S., & Vieira, A. B. (2021, 12). Forecasting the carsharing service demand using uni and multivariable models. *Journal of Internet Services and Applications*, 12, 1-20. doi: 10.1186/S13174-021-00137-8/TABLES/7
- Angelopoulos, A., Gavalas, D., Konstantopoulos, C., Kypriadis, D., & Pantziou, G. (2018, 12). Incentivized vehicle relocation in vehicle sharing systems. *Transportation Research Part C: Emerging Technologies*, 97, 175-193. doi: 10.1016/J.TRC.2018.10.016
- Ataç, S., Obrenović, N., & Bierlaire, M. (2022). Evaluating different strategies to solve rebalancing operations in car sharing systems. In *Strc conference paper 2022*.

- Ataç, S., Obrenović, N., & Bierlaire, M. (2021, 1). Vehicle sharing systems: A review and a holistic management framework. *EURO Journal on Transportation and Logistics*, *10*, 100033. doi: 10.1016/J.EJTL.2021.100033
- Balac, M., Ciari, F., & Axhausen, K. W. (2016, 4). Carsharing demand estimation: Zurich, switzerland, area case study. *Transportation Research Record*, *2536*, 10-18. doi: 10.3141/2536-02
- Baptista, P., Melo, S., & Rolim, C. (2014, 2). Energy, environmental and mobility impacts of car-sharing systems. empirical results from lisbon, portugal. *Procedia - Social and Behavioral Sciences*, *111*, 28-37. doi: 10.1016/J.SBSPRO.2014.01.035
- Barth, M., Todd, M., & Xue, L. (2004). User-based vehicle relocation techniques for multiple-station shared-use vehicle systems..
- Becker, H., Ciari, F., Brignoni, M., & Axhausen, K. W. (2015). Impacts of a new free-floating car-sharing system in basel. In *Eth library impacts of a new free-floating car-sharing system in basel working paper*. ETH Zürich. doi: 10.3929/ethz-b-000250448
- Benarbia, T., Axhausen, K. W., & Farooq, B. (2021, 5). Modeling, relocation, and real-time inventory control of one-way electric cars sharing systems in a stochastic petri nets framework. *IEEE Transactions on Intelligent Transportation Systems*, *22*, 2846-2861. doi: 10.1109/TITS.2020.2976568
- Brendel, A. B., Brauer, B., & Hildebrandt, B. (2016). Toward user-based relocation information systems in station-based one-way car sharing. In *Amcis*.
- Caggiani, L., Camporeale, R., & Ottomanelli, M. (2017, 1). A dynamic clustering method for relocation process in free-floating vehicle sharing systems. *Transportation Research Procedia*, *27*, 278-285. doi: 10.1016/J.TRPRO.2017.12.146
- Carrone, A. P., Hoening, V. M., Jensen, A. F., Mabit, S. E., & Rich, J. (2020, 11). Understanding car sharing preferences and mode substitution patterns: A stated preference experiment. *Transport Policy*, *98*, 139-147. doi: 10.1016/j.tranpol.2020.03.010
- Catalano, M., Casto, B. L., & Migliore, M. (2008). Car sharing demand estimation and urban transport demand modelling using stated preference techniques. *European Transport Trasporti Europei n*, *40*, 33-50.
- Cervero, R., Golub, A., & Nee, B. (2007). City carshare longer-term travel demand and car ownership impacts. *Transportation Research Record*, 70-80. doi: 10.3141/1992-09
- Ciari, F., Schuessler, N., & Axhausen, K. W. (2012, 6). Estimation of carsharing demand using an activity-based microsimulation approach: Model discussion and some results. *International Journal of Sustainable Transportation*, *7*, 70-84. doi: 10.1080/15568318.2012.660113
- Cucu, T., Ion, L., Ducq, Y., & Boussier, J. M. (2009). Management of a public transportation service: Car-sharing service..
- Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, *14*(1). doi: 10.1137/1114019
- Famoye, F., & Wang, W. (2004). Censored generalized poisson regression model. *Computational Statistics & Data Analysis*, *46*, 47 – 560. doi: 10.1016/j.csda.2003.08.007
- Febbraro, A., Sacco, N., & Saeednia, M. (2012). One-way carsharing: Solving the relocation problem. *Transportation Research Record*, *2319*(1), 113-120. doi: 10.3141/2319-13
- Gammelli, D., Peled, I., Rodrigues, F., Pacino, D., Kurtaran, H. A., & Pereira, F. C. (2020). Estimating latent demand of shared mobility through censored gaussian

- processes. *Transportation Research Part C: Emerging Technologies*, 120, 102775. doi: 10.1016/j.trc.2020.102775
- Gay, D. M. (1990). *Usage summary for selected optimization routines* (Tech. Rep.). Murray Hill: AT&T Bell Laboratories.
- Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438.
- Grimmet, G., & Welsh, D. (2014). *Probability: An introduction* (2nd ed.). Oxford Univeristy Press.
- Guidon, S., Becker, H., & Axhausen, K. (2019, 10). Avoiding stranded bicycles in free-floating bicycle-sharing systems: using survival analysis to derive operational rules for rebalancing. *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019*, 1703-1708. doi: 10.1109/ITSC.2019.8916869
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). New York, NY: Springer. doi: 10.1007/978-0-387-84858-7
- He, L., Mak, H.-Y., & Rong, Y. (2019). Operations management of vehicle sharing systems. In M. Hu (Ed.), *Sharing economy: Making supply meet demand* (pp. 461–484). Cham: Springer International Publishing. doi: 10.1007/978-3-030-01863-4_19
- Huttel, F. B., Peled, I., Rodrigues, F., & Pereira, F. C. (2022). IEEE transactions on intelligent transportation systems. *Modeling Censored Mobility Demand Through Censored Quantile Regression Neural Networks*, 23(11), 21753-21765. doi: 10.1109/TITS.2022.3190194
- Jian, S., Rey, D., & Dixit, V. (2019, 6). An integrated supply-demand approach to solving optimal relocations in station-based carsharing systems. *Networks and Spatial Economics*, 19, 611-632. doi: 10.1007/s11067-018-9401-6
- Jorge, D., & Correia, G. (2013, 6). Carsharing systems demand estimation and defined operations: a literature review. *European Journal of Transport and Infrastructure Research*, 13, 201-220. doi: 10.18757/EJTIR.2013.13.3.2999
- Kek, A. G., Cheu, R. L., Meng, Q., & Fung, C. H. (2009, 1). A decision support system for vehicle relocation operations in carsharing systems. *Transportation Research Part E: Logistics and Transportation Review*, 45, 149-158. doi: 10.1016/J.TRE.2008.02.008
- Le Vine, S. (2011). Strategies for personal mobility: A study of consumer acceptance of subscription drive-it-yourself car services..
- Li, Q., Liao, F., Timmermans, H. J., Huang, H., & Zhou, J. (2018, 1). Incorporating free-floating car-sharing into an activity-based dynamic user equilibrium model: A demand-side model. *Transportation Research Part B: Methodological*, 107, 102-123. doi: 10.1016/J.TRB.2017.11.011
- Long, Z., & Axsen, J. (2022, 6). Who will use new mobility technologies? exploring demand for shared, electric, and automated vehicles in three canadian metropolitan regions. *Energy Research & Social Science*, 88, 102506. doi: 10.1016/J.ERSS.2022.102506
- Merfeld, K., Wilhelms, M.-P., Henkel, S., & Kreutzer, K. (2019, 7). Carsharing with shared autonomous vehicles: Uncovering drivers, barriers and future developments – a four-stage delphi study. *Technological Forecasting and Social Change*, 144, 66-81. doi: 10.1016/j.techfore.2019.03.012
- Negahban, A. (2019). Simulation-based estimation of the real demand in bike-sharing systems in the presence of censoring. *European Journal of Operational Research*,

- 277(1), 317-332. doi: 10.1016/j.ejor.2019.02.013
- Paschke, S., Balać, M., & Ciari, F. (2016, 8). Implementation of vehicle relocation for carsharing services in the multi-agent transport simulation matsim. *Arbeitsberichte Verkehrs- und Raumplanung*, 1188. doi: 10.3929/ETHZ-B-000119010
- R Core Team. (2022). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org>
- Raviv, T., & Kolka, O. (2013, 10). Optimal inventory management of a bike-sharing station. *Operations Engineering & Analytics*, 45, 1077-1093. doi: 10.1080/0740817X.2013.770186
- Reiss, S., & Bogenberger, K. (2017, 1). A relocation strategy for munich's bike sharing system: Combining an operator-based and a user-based scheme. *Transportation Research Procedia*, 22, 105-114. doi: 10.1016/J.TRPRO.2017.03.016
- Repoux, M., Kaspi, M., Boyacı, B., & Geroliminis, N. (2019, 12). Dynamic prediction-based relocation policies in one-way station-based carsharing systems with complete journey reservations. *Transportation Research Part B: Methodological*, 130, 82-104. doi: 10.1016/J.TRB.2019.10.004
- Seign, R., & Bogenberger, K. (2021, March). Prescriptions for the successful diffusion of carsharing with electric vehicles. In *Conference on future automotive technology*. München.
- Shaheen, S. A., & Cohen, A. P. (2013, 6). Carsharing and personal vehicle services: Worldwide market developments and emerging trends. *International Journal of Sustainable Transportation*, 7, 5-34. doi: 10.1080/15568318.2012.660103
- Stokkink, P., & Geroliminis, N. (2021, 7). Predictive user-based relocation through incentives in one-way car-sharing systems. *Transportation Research Part B: Methodological*, 149, 230-249. doi: 10.1016/J.TRB.2021.05.008
- Vosooghi, R., Puchinger, J., Jankovic, M., & Sirin, G. (2018, 3). A critical analysis of travel demand estimation for new one-way carsharing systems. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2018-March*, 199-205. doi: 10.1109/ITSC.2017.8317917
- Wagner, S., Brandt, T., & Neumann, D. (2015, 1). Data analytics in free-floating carsharing: Evidence from the city of berlin. In (p. 897-907). IEEE. doi: 10.1109/HICSS.2015.112
- Wand, M. (1994). Fast computation of multivariate kernel estimators. *Journal of Computational and Graphical Statistics*, 3(4), 433-445.
- Wang, H., Cheu, R., & Lee, D. H. (2010). Logistical inventory approach in forecasting and relocating share-use vehicles. *Proceedings - 2nd IEEE International Conference on Advanced Computer Control, ICACC 2010*, 5, 314-318. doi: 10.1109/ICACC.2010.5487071
- Wang, N., Guo, J., Liu, X., & Fang, T. (2020, 1). A service demand forecasting model for one-way electric car-sharing systems combining long short-term memory networks with granger causality test. *Journal of Cleaner Production*, 244, 118812. doi: 10.1016/J.JCLEPRO.2019.118812
- Weikl, S. (2016). *A mesoscopic relocation model for free-floating carsharing systems* (Unpublished doctoral dissertation). Munich University of the Federal Armed Forces.
- Willing, C., Klemmer, K., Brandt, T., & Neumann, D. (2017, 7). Moving in time and space – location intelligence for carsharing decision support. *Decision Support Systems*, 99, 75-85. doi: 10.1016/J.DSS.2017.05.005

- Yoon, T., Cherry, C. R., & Jones, L. R. (2017, 6). One-way and round-trip carsharing: A stated preference experiment in beijing. *Transportation Research Part D: Transport and Environment*, 53, 102-114. doi: 10.1016/j.trd.2017.04.009
- Yu, D., Li, Z., Zhong, Q., Ai, Y., & Chen, W. (2020). Demand management of station-based car sharing system based on deep learning forecasting. *Journal of Advanced Transportation*, 2020. doi: 10.1155/2020/8935857
- Zhou, F., Zheng, Z., Whitehead, J., Washington, S., Perrons, R. K., & Page, L. (2020, 2). Preference heterogeneity in mode choice for car-sharing and shared automated vehicles. *Transportation Research Part A: Policy and Practice*, 132, 633-650. doi: 10.1016/J.TRA.2019.12.004

Appendix

To validate our method, we performed a simulation study using a 5×5 grid of square-shaped cells with side lengths 0.2. The total demand rates of the cells were proportional to a multivariate normal distribution centered at the center cell 13 and varying between 0.05 and 0.3 cars per hour. To include time effects, the individual total demand rates were increased at evenings (18-24 o'clock) by 0.1 on weekend days by 0.2. Figure A.10 illustrates the specified total demand rates by a map and a scatterplot.

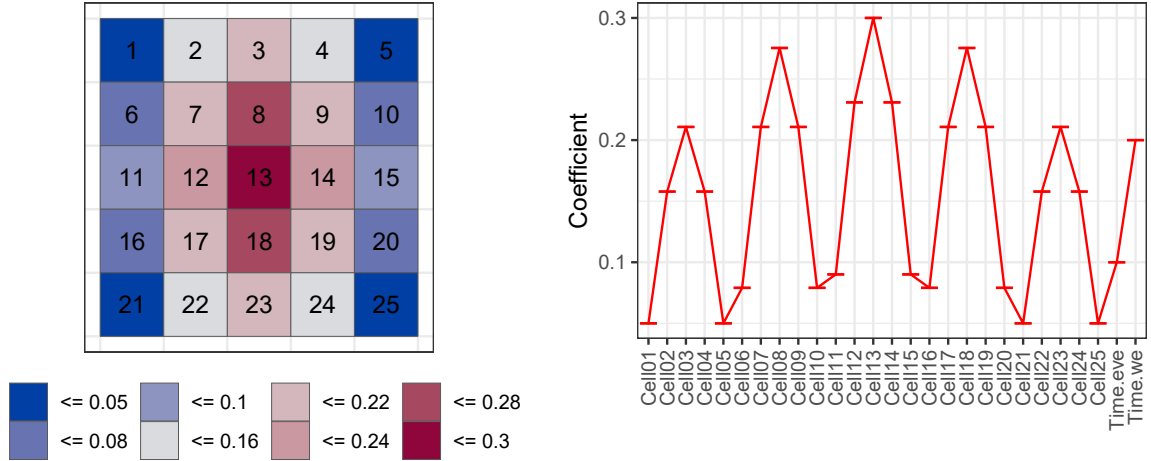


Figure A.10: Used total demand rates for the simulation study. Left: Spatial distribution of the demand rates on weekdays between 0 and 18 o'clock. Right: The same demand rates along the cell numbers and effects for evenings (*Time.eve*) and weekend days (*Time.we*).

We generated pickups for independent hourly intervals. For this, we first generated for each interval and cell the number of cars available using a Poisson distribution with a common rate for all cells, and then generated the corresponding pickups based on our postulated model (Eq. 6) and the specified total demand rates. Simulations were performed for six scenarios regarding data generation and model specification. Each scenario was replicated 512 times, resulting 512 estimated models per scenario.

For the baseline scenario, we used a car rate per cell of 0.16, which corresponds to the average of the total demand rates on weekdays between 0 and 18 o'clock used in this simulation study. Demand areas, which need to be found to construct the vectors \mathbf{x}_{jt} of the postulated model, included adjacent and diagonally adjacent cells of the cells with cars ($r_{\max} = \sqrt{2} \cdot 0.2^2 \approx 0.28$). Data for 4321 hourly intervals were generated, which corresponds to the number of hours of the first half year of 2021 (including summer time changeover). Fitted models from the baseline scenario are correctly specified and therefore should identify the data generating total demand rates.

For the alternative scenarios, we halved and doubled the car rate, misspecified the r_{\max} parameter for estimation ($r_{\max} = 0.2$ and $r_{\max} = 0.4$ instead of $r_{\max} = 0.28$), halved and doubled the number of time intervals, used the smoothing approach with 3×3 supporting points and considered demand areas that include all adjacent or diagonally adjacent cells including cells have a closer car somewhere else.

Results Figure A.11 shows the distribution of the estimated parameters for the baseline scenario. The estimates vary around the predefined total demand rates, suggesting that the estimation procedure is able to identify the data generating total demand rates if the model is correctly specified.

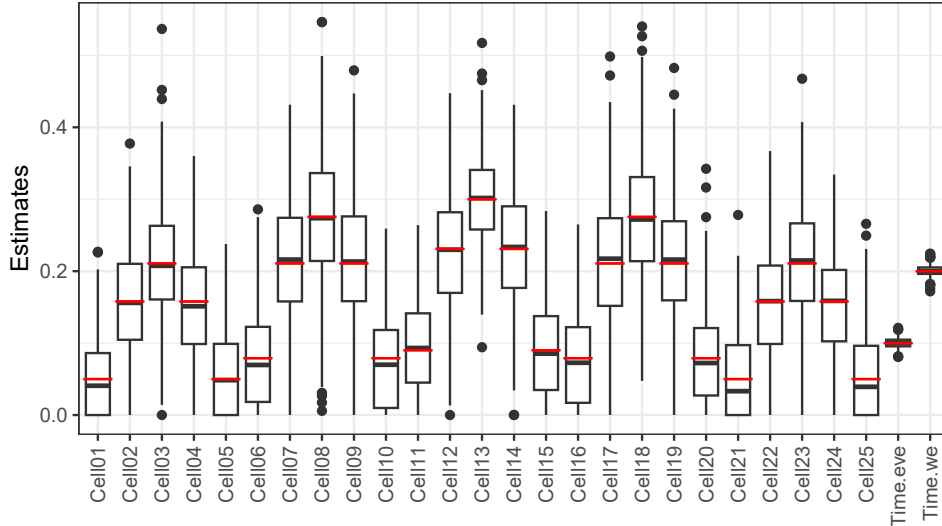


Figure A.11: Simulation results for the baseline scenario: Estimated parameters (box-plots, black) and according true values (red).

Figure A.12 compares the estimated parameters for cells 1 (lowest total demand rate) and 13 (highest total demand rate) between the baseline and three alternative scenarios. The top left panel shows the effect of increasing the number of cars available. It can be seen that increasing the number of cars increases the accuracy of the estimates. Interestingly, the accuracy of the estimates for fewer cars are about the same or slightly better than for the baseline scenario.

The middle panel of Figure A.12 shows the effect of misspecifying the parameter r_{\max} , i.e. the maximum deviation users would accept from the preferred pick up cell. While misspecifying r_{\max} seems not to affect the estimation of the total demand rate of cell 1, it does for cell 13. Specifically, choosing r_{\max} too small results a upward bias, and choosing r_{\max} to large a downward bias. This seems plausible because increasing r_{\max} implies that the total demand is spread over more cells.

The right panel of Figure A.12 shows the effect of decreasing or increasing the number of observations. As expected, the accuracy improves with a increasing data size.

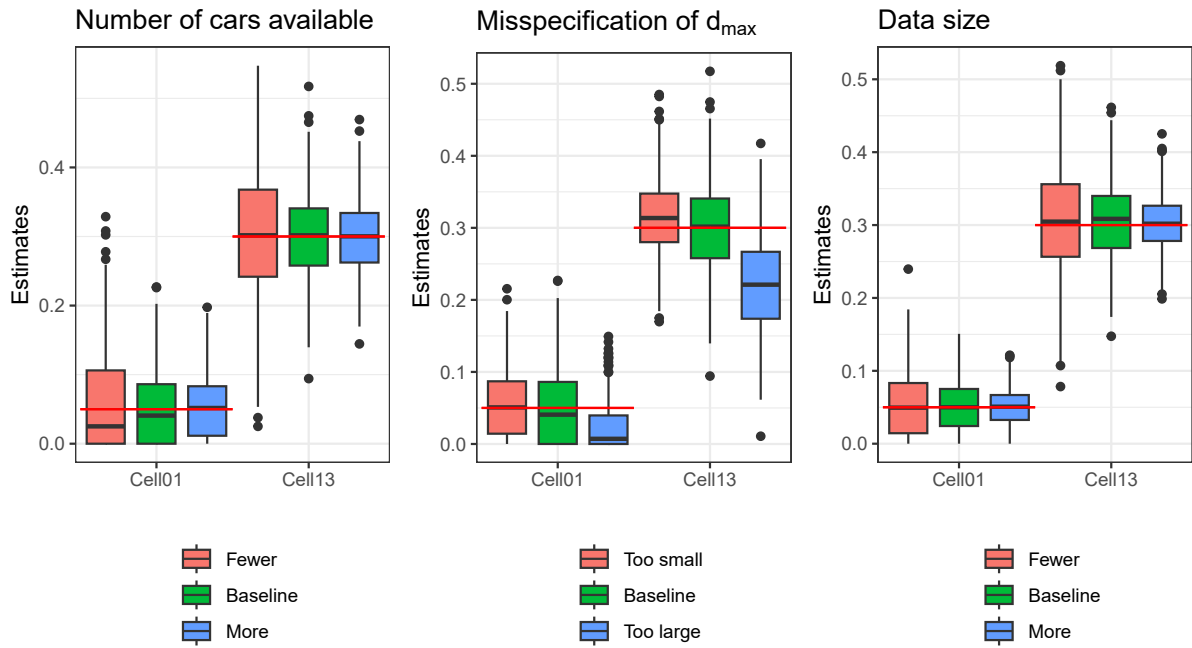


Figure A.12: Simulation results on comparing the baseline with alternative scenarios (car rate, misspecification of d_{\max} and data size): Estimated parameters (boxplots) and according true values (red). For reasons of clarity, only the parameters for the cells 1 (lowest total demand rate) and 13 (highest total demand rate) are shown.

The left hand of Figure A.13 compares parameter estimates between the baseline scenario and a smoothed estimation with 9 supporting points, which were evenly distributed within the surface of the 5×5 grid. As could have been expected, the smoothing approach decreases the variance of the estimates, however, in case cell 1, it introduces a bias by overestimating the total demand rate.

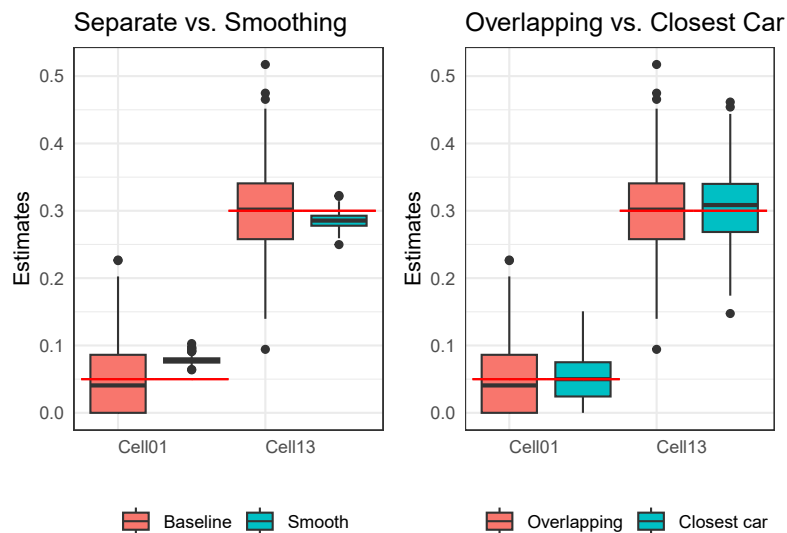


Figure A.13: Simulation results on comparing the baseline with alternative scenarios (smoothing and specification of demand areas): Estimated parameters (boxplots) and according true values (red).

The right hand of Figure A.13 compares the parameter estimates between the overlapping and the closest car demand areas. In both cases the estimates vary around the data generating total demand rates. The estimates for the overlapping areas around cars have slightly higher variance.