Transportation Research Part A 190 (2024) 104241

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Analysis of car sharing operation area performance: An idle time prediction approach

Andrea Papu Carrone [a], Jeppe Rich [b,*], David Watling [c]

[a] International Transport Forum, 75016 Paris, France
[b] Technical University of Denmark, Department of Technology, Management and Economics, 2800 Kongens Lyngby, Denmark
[c] Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK

ARTICLE INFO

Keywords:
Free-floating car sharing
Idle time prediction
Hazard duration models
Survival analysis
Operational area performance

ABSTRACT

Free-floating car sharing (FFCS) extends traditional station-based services by providing a more flexible car sharing alternative for users. However, the increased user flexibility introduces challenges from an operator perspective. To make services profitable, the total idle time of vehicles needs to be minimised and available vehicles should be located where demand exists. To increase profitability, it is important to carefully choose the operational area based on the expected idle time that different locations may offer, and only strategically expand into areas where the sustainability of the service can be maintained. In this paper, we present a hazard-based duration model for the idle times of a car sharing vehicle service. It is argued that modelling of idle time as opposed to bookings, which is the common approach, allows to circumvent the problem of latent demand and thereby presents itself as a simpler modelling strategy. In the paper, the model is applied to the city of Copenhagen, where we study the operational performance on the basis of 327,610 electric free-floating car trips in the period 2017-2018. We study the performance over 92 existing zones and predict the expected performance for an additional 28 zones by considering geographical and socio-economic drivers of demand. This enables the prediction of which areas to include as part of an expansion of the operational area, and thus serves the purpose of a strategic planning tool for growing such services. It is found that the additional zones differ substantially in their performance, which is a consequence of zones being more or less aligned with the local FFCS drivers of demand. This leads to a prioritisation of zones for further expansion based on performance, where the idle time of the best performing zones is seen to be as much as one hour less than the worst performing zones.

1. Introduction

Over the past years, commercial car sharing has become increasingly popular in urban areas. The success of car sharing services has been made possible by the growing utilisation of information and communication technology, along with the worldwide prevalence of the sharing economy. While many different service offerings and business models exist today (Roukouni and Correia, 2020), one of the most successful offerings is the free-floating car sharing service (FFCS). This service is made possible through innovations in smartphone and GPS technology and is seen as a more flexible alternative to the traditional station-based services. FFCS enables users to pick up and return vehicles anywhere within a predefined area of operation. It is based on smartphone technology that indicates the real-time location of available vehicles and manages payments on a time-of-use basis.

* Corresponding author.
E-mail address: rich@dtu.dk (J. Rich).

https://doi.org/10.1016/j.tra.2024.104241
Received 6 June 2022; Received in revised form 11 April 2024; Accepted 30 August 2024
Available online 11 September 2024

However, the flexibility that FFCS provides to the users comes at the cost of strategic and operational challenges for the service operators. To ensure the profitability of FFCS services, it is necessary to minimise the idle time of vehicles and always position them in areas with high demand. There are two main challenges for the service operators. First, at a strategic level, when starting to operate in a new city there is a need to define an area of operation. Therefore, it is important to understand to what extent geographic characteristics of the different areas correlate with high demand of FFCS. This question is also relevant for already established services who wish to consider if, and how, to expand the area of operation. Second, at an operational level, since transport patterns are generally asymmetric in cities, service operators need to relocate vehicles in order to coordinate vehicle availability with demand through time. Operators will then need to define relocation strategies by determining which vehicles to move to which areas and when. Despite the widespread focus on the second problem, there has been only limited research conducted on the first strategic issue. This paper focuses on this first problem and uses hazard duration model to; (i) identify successful and unsuccessful geographical areas by analysing geographical drivers of FFCS demand, and (ii) predict the expected time vehicles will remain idle in a given area. Consequently, the term success should here be understood from the operator perspective and their ability to improve their overall profit function by reducing the idle time in the system.

A few previous studies have empirically investigated the demand drivers of FFCS by exploring the socio-demographic characteristics of different zones within the area of operation (Seign et al., 2015; Schmöller et al., 2015; Becker et al., 2017b), the temporal demand patterns (Schmöller et al., 2015) and the effect of weather conditions (Schmöller et al., 2015). They do so by modelling the number of bookings within the different geographical zones. However, such an approach disregards the issue of supply and demand inter-dependency highlighted by Jorge and Correia (2013) for one-way car sharing services. That is to say, the number of bookings is a censored representation of the FFCS car sharing demand which is conditioned by the supply. As a result, the number of bookings only resembles a lower bound for the FFCS demand. Hence, by only considering bookings it is not possible to know if there is latent demand that is not served due to an inefficient location of the vehicles in space and time. A way to overcome this problem is to model the uncensored demand (Gammelli et al., 2020; Evan et al., 2021). Another approach, however, is to model the time between consecutive bookings of FFCS vehicles based on a duration-type model as suggested in this study. While the existence of latent demand for FFCS within a geographical zone yields a lower number of bookings, it does not significantly affect the average time between consecutive bookings of FFCS vehicles. An additional potential vehicle would spend a similar amount of time parked between bookings as other vehicles in the same zone, on average.

The present paper makes three contributions to the existing literature:

(i) From a methodological point of view, it is the first to propose and apply hazard-based methods to a large-scale, Free-Floating Car Sharing (FFCS) system. This is a significant contribution since it mitigates a well-known issue of latent demand when analysing data on bookings, while at the same time allowing an analysis of a Key Performance Indicator that correlates directly with the operating revenue of the system.
(ii) We show how the idle-time prediction model can be applied, through Bayesian inference, to estimate distinct survivor functions for different zones. We show how covariates related to land-use, competing transport options, socio-demographics, availability of FFCS vehicles, weather and temporal attributes may be included to reveal the factors underlying zones that are successful and unsuccessful in terms of the FFCS.
(iii) The developed methodology is applied to a large-scale dataset based on more than half a million FFCS trips, revealing operational and policy insights regarding where such systems may be most successful. This contributes to the growing body of evidence concerning empirical understanding of these systems and the key factors influencing their success or failure. In particular, we demonstrate the application of the methodology for developing a strategic area expansion strategy in Copenhagen, identifying the zones with the greatest potential for expansion.

2. Related work

2.1. Car sharing demand

Car sharing demand and behaviour modelling has received increasing attention from the research community. Overall, evidence presented thus far supports the view that car sharing addresses a diversity of social and environmental challenges by providing a more flexible car-based transport alternative. Some of the benefits and opportunities of car sharing services that have already been presented within the literature are: the reduction of private car ownership (Becker et al., 2018; Le Vine and Polak, 2019), the lower requirement of urban space for parking facilities (Millard-Ball et al., 2006), the reduction of emissions provided shared fleets where electric (Shaheen and Cohen, 2013) and the lower distance travelled by car sharing users than by private car users (Cervero et al., 2007). Several studies have characterised the prototypical car sharing users. According to Firnkorn and Müller (2012), Le Vine et al. (2014), Kopp et al. (2015), Prieto et al. (2017) and Papu Carrone et al. (2020) these are often young educated adults with middle-upper income and no children.

The geographical drivers of demand have mostly been studied with respect to station-based car sharing systems as a means to estimate demand and to locate new stations. Jian et al. (2016) present a spatial hazard formulation, where, instead of modelling the time-to-event they apply a hazard-type model to model the distance to the vehicles. Later, Cheng et al. (2022) presented a spatio-temporal model for predicting demand. As expressed by the authors, underestimation of demand when the observed demand is equal or to supply may occur. This perspective aligns with the findings of Gammelli et al. (2020), who explored latent demand and illustrated its predictability through time-series models, albeit with some methodological distinctions. Notably, Cheng's study is pertinent to our work, as it incorporates land-use data in the prediction process.

In de Lorimier and El-Geneidy (2012) the number of vehicles parked within the station and the density of members in the nearby region of the station were found to be important drivers of demand. Kang et al. (2016) found that demand is likely to be higher in areas with higher proportions of young adults. The relationship between car sharing demand and job density is more difficult to establish. de Lorimier and El-Geneidy (2012) found that employment density was negatively correlated with car sharing demand, while Kang et al. (2016) found a positive correlation between business activity and demand. The effect of accessibility to other modes of transport with respect to the demand for station-based car sharing is not clear either and possibly highly influenced by public transport service quality, as well as cultural and social differences. In a study conducted in the U.S., Stillwater et al. (2009) were unable to not find a clear relation with respect to access to public transport. Kang et al. (2016) found that low accessibility to the subway has a positive effect on demand in Seoul, whereas (Ménoire et al., 2020) based on a study in Montreal found that high accessibility to buses has a positive effect. Another research paper, namely (Lee and Goulias, 2018), examines longitudinal data and identifies a specific group of households characterised by the lowest average household age, the highest number of children, the greatest number of workers, and a preference for residing in densely populated urban areas. This particular group exhibits the highest proportion of participation in car sharing, walking and biking trips.

These findings, however, are difficult to transfer to FFCS due to the structural differences in the way these services operate. Station-based car sharing services generally require the payment of a membership fee and mostly involve round trips with vehicles that, as the name suggests, can only be accessed and returned at fixed locations (stations). Instead, FFCS services are more flexible in that they generally do not involve membership fees and are used mainly for one-way trips within the operation area. Consequently, the attributes that correlate with demand and the usage patterns for station-based and FFCS are generally considered to be different (Schmöller and Bogenberger, 2014; Becker et al., 2017a).

A few studies provide insights on how demand for FFCS is linked to geographical characteristics. There is a general consensus that higher shares of young adults within the population has a positive effect on FFCS demand (Schmöller et al., 2015; Seign et al., 2015; Braun et al., 2016; Becker et al., 2017b; Khan and Machemehl, 2017). In addition, it has been found that areas with high housing rent prices (Schmöller et al., 2015; Seign et al., 2015), high parking fees (for which the FFCS are exempt from paying) (Khan and Machemehl, 2017), a high percentage of households with no cars (Khan and Machemehl, 2017), and a high percentages of single households (Schmöller et al., 2015) tend to have higher demand for FFCS. The influence of density of workplaces and companies is not clear however. While Becker et al. (2017b) and Khan and Machemehl (2017) found a negative correlation, Seign et al. (2015) and Cheng et al. (2021) found a positive correlation with respect to FFCS demand. Cheng et al. (2021) found that the university location as the origin of one-way car sharing trips has a positive correlation with bookings. Also, the effect of accessibility to public transport is ambiguous. Becker et al. (2017b) concludes that FFCS is mainly used for trips for which only inferior public transport alternatives exist, while Khan and Machemehl (2017) found that locations with a greater number of transit stops tend to have higher demand, which is attributed to the increased inter-modal connectivity. Braun et al. (2016) have not found a significant influence of the number of public transit stops but found that short distances to transit stops have a positive effect on car sharing demand. However, a recent paper (Ingvardson et al., 2023) suggests that for the Copenhagen area, FFCS and public transport are mostly substitutes although instances of complementarity exist.

It is not well-understood how zones with different land-use characteristics perform and how statistical models can support the design of the operation area. The present paper takes on these challenges by formulating a geographical hazard-based duration model to understand how the different factors influence local FFCS idle time performance.

2.2. Hazard-based duration models in transport

Hazard-based duration or time-to-event analysis are statistical methods that focus on the modelling of the 'end-of-duration' occurrence, given that the duration has lasted for some specified time. The probability model is conditional in the sense that the likelihood of an occurrence depends on the length of elapsed time since the start of the duration (Bhat and Pinjari, 2007). Hensher and Mannering (1994) and Hensher (1998) studied the concepts of hazard-based duration models and suggested their suitability for transport related problems. The research at the time focused on the modelling of commuting activity participation and vehicle ownership transactions based on duration model. Since then, hazard-based duration models have been applied in a number of studies within the transport research field. The area where duration models have been used most successfully is within the modelling of activity behaviour (Hensher and Mannering, 1994; Bhat, 1996; Anastasopoulos et al., 2012; Dissanayake, 2017). Other applications of hazard-based models concern the modelling of duration between vehicle transactions (Gilbert, 1992; Hensher, 1998; Rashidi and Mohammadian, 2011), duration of time between traffic incidents (Junhua et al., 2013; Shi et al., 2015), duration of time between public transport disruptions (Louie et al., 2017), passenger waiting time tolerance (Rahimi et al., 2019) and crash occurrence risk (Balusu et al., 2020). To the best of our knowledge (Khan and Machemehl, 2017) represent the only attempt to model FFCS rental and parking choice based on a parametric hazard-based duration model. The study identifies the importance of some socio-demographic attributes such as population density, public transport accessibility and household car ownership level on FFCS demand by using a limited dataset for three off-peak hours of a typical weekday.

Previously studies from the transport field have mostly applied parametric hazard functions that allow the incorporation of exogenous variables in the hazard function. However, the hazard function can also be formulated with a non-parametric shape which often provides better model fit, at the cost of losing interpretation of variables. In bio-statistics a popular approach is the semi-parametric Cox proportional hazard model, which does not estimate the hazard distribution itself but only co-variate effects expressed through the Cox partial likelihood function (Cox, 1972). In the literature on shared mobility, applications of non-parametric and semi-parametric models include Guidon et al. (2019) where Cox proportional hazard and survival random forests models are used to predict survival times of shared e-bikes; and Kostic et al. (2021) where Cox proportional hazard and neural network-based survival models are used to predict survival times of shared cars.

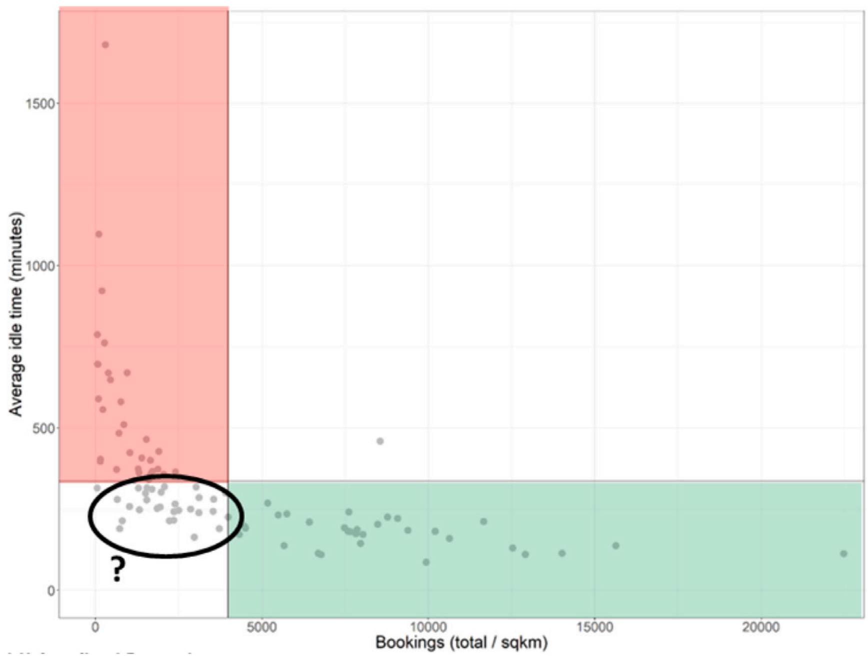


Fig. 1. Zone correlation between idle-time and bookings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Methodology

The main goal of this paper is to investigate the determinants of demand for an electric fleet of FFCS vehicles by observing the amount of time these vehicles remain parked and idle. To this end, we define the end-of-duration occurrence as the time in which the vehicle is booked by a user. The duration model is then based on the number of minutes a FFCS vehicle remains idle (parked in the same location) between two consecutive bookings. The duration of idle time is linked to the survival function, which models the probability of survival, while the related hazard function is linked to the probability of the event (the car being taken) occurring. First, in Section 3.1 we compare the idle time modelling approach with the more commonly used booking modelling approach and argue why the former approach is advantageous from the perspective of car sharing systems. After this first introduction, we outline the theory (from Section 3.2 and on-wards) and propose how, by predicting the survivor and hazard distributions for different zones within the FFCS area, we are able to explore the effect of different attributes for the FFCS demand.

3.1. Idle time or bookings?

As referred in the literature review, the commonly accepted way of modelling car and bike sharing systems is by considering bookings in space and time. However, as also briefly hinted, this comes at the cost of being sensitive to unserved demand as demonstrated and discussed in Gammelli et al. (2020), Jorge and Correia (2013), and Cheng et al. (2022). Without accounting for the censoring of the data, zones for which censoring occur, may be represented as low performance zones (small number of bookings) even if these are high-demand zones with a short idle-time. In these cases, the low number of bookings only reflects that there are supply shortages in the zone.

Fig. 1 is intended to illustrate the censoring observed in our system, even when focusing solely on zone averages. When assessing the performance of different zones in terms of bookings per square km and average idle time, zones can be classified into three categories. Firstly, zones that are appealing to the FFCS operator and perform well with a high number of bookings and short idle times (Green area). Secondly, zones that are unattractive due to a low number of bookings and extended idle times (Red area). Thirdly, zones with a low number of bookings and short idle times. In this third category, situations may arise where no FFCS vehicles are available for customers, leading to a reduction in the total number of bookings. However, the idle time performance of vehicles in these zones consistently remains close to the minimum average idle time, even in situations with a low number of bookings. This observation suggests that while the FFCS bookings are indeed censored, transforming the demand representation into an idle times analysis allows us to overcome this censoring issue.

Additionally, the use of idle time data brings about other advantages, most notably a closer link to the profit function of the FFCS operator. These companies generate revenue based on minutes of bookings rather than the sheer number of bookings.

Table 1
Goodness of fit statistics for identifying a proper distribution for the density function $f(t)$.

Statistics	Gamma	Lognormal	Weibull	Exponential	Normal	Logistic
Kolmogorov–Smirnov (KS)	0.043	0.049	0.026	0.137	0.215	0.228
Anderson–Darling (AD)	1132.8	2491.6	524.1	17 790.1	28 673.8	22 050.8

3.2. Hazard-based duration model

In hazard-based duration or time-to-event analysis the duration time is represented by a non-negative random variable T, for which we define the cumulative distribution $F(t)$ and the corresponding density function $f(t)$. The modelling approach has two main constructs, namely the hazard function $h(t)$ and the survivor function $S(t)$.

The cumulative distribution $F(t) = P(T < t)$ represents the probability of an event T happening before time t. The corresponding density function is $f(t) = dF(t)/dt$. The hazard function $h(t)$ is the conditional probability that an event will occur between time t and $t + dt$ given that the event has not occurred before time t. The function is defined as:

$$h(t) = \frac{f(t)}{1 - F(t)}, \qquad \forall t > 0 \tag{1}$$

The survivor function $S(t)$ is the probability that the duration time will be greater than or equal to t. That is, it is the probability of "surviving" until time t.

$$S(t) = P(T \geq t) = 1 - F(t), \qquad \forall t > 0 \tag{2}$$

The survivor function $S(t)$ is related to the hazard function $h(t)$ through Eq. (3). While the survivor function focuses on the probability of 'surviving', the hazard function focuses on the probability of the event occurring.

$$h(t) = f(t)/S(t) \tag{3}$$

Hazard-based duration models recognise that the likelihood of ending the duration depends on the elapsed time since the start of the duration. This information is provided by the shape of the hazard function, more specifically by the first derivative of the hazard function with respect to time. The shape of the hazard function can be monotonically increasing, monotonically decreasing, monotonically increasing first and then decreasing, or constant over time. Different distributional assumptions for $f(t)$, the probability density function of the random variable T, will yield different shapes for the hazard and survivor function.

3.3. Selection of best fit distribution

The choice of the probability density function $f(t)$ is often based on a combination of theoretical knowledge and empirical assessment. To support the selection of the distribution, several theoretical distributions were evaluated with respect to their ability to fit the empirical idle time distribution for the case study data. Specifically, we considered Exponential, Normal, Lognormal, Weibull, Gamma and Logistic distributions and evaluated these according to Anderson–Darling (AD) and Kolmogorov–Smirnov (KS) test statistics. KS and AD are goodness-of-fit test statistics that are typically considered when fitting continuous distributions, which measure the distance between the fitted parametric distribution and the empirical distribution. The AD statistic is of special interest as a means to evaluate the tail of the distribution. The H_0 hypothesis of the theoretical distribution being equal to the empirical distribution is rejected for larger values of the test statistics. Hence, smaller value indicate a better fit. The result of the fitting exercise is shown in Table 1 and it is indicated that the best fit for our data is the Weibull distribution.

3.4. Including covariates in hazard-based duration models

It is common to include exogenous variables in hazard-based models to be able to use the models for operational planning and for scenario analysis. Hazard duration models are typically classified according to the way in which they incorporate the effect of external variables. This may correspond to proportional hazard models or to the form of accelerated models. In the proportional hazard model, external variables are multiplicative with respect to the underlying hazard function, while in the accelerated form variables represent a direct re-scaling of time. The Accelerated Failure Time (AFT) (also known as accelerated lifetime) model represents the probability that the duration will survive beyond time t and is given by the baseline survivor probability, which is computed with respect to a function of external variables. Its name refers to the fact that it assumes that the effect of the variables on the survival time is to accelerate or decelerate the underlying time scale. Therefore, AFT models are easier to interpret than proportional hazard models because the parameters measure the effect that each variable has on the mean survival time, and for this reason it is adopted in the present study. In order to parameterise the model as an AFT model, the effect of the variable were included as part of the scale parameter of a Weibull distribution.

$$\lambda = exp(\beta^T X) \tag{4}$$

where β is a vector of estimated parameters and X a matrix of variables. This parameterisation ensures that the scale parameter cannot be negative.

3.5. Bayesian Weibull model estimation

The probability density function for idle time is assumed to follow a 2-parameter Weibull distribution given by Eq. (5):

$$f(t \mid \alpha, \lambda) = \frac{\alpha}{\lambda}(\frac{t}{\lambda})^{\alpha-1}e^{-(\frac{t}{\lambda})^{\alpha}}, \qquad t > 0, \lambda > 0, \alpha > 0 \tag{5}$$

where α is the shape and λ is the scale parameter of the distribution. Hence, the mean is equal to:

$$E(t) = \lambda \Gamma(1 + \frac{1}{\alpha}) \tag{6}$$

where Γ represents the gamma function.

The shape parameter α indicates that the Weibull distribution generates a monotonic hazard. When $\alpha > 1$ the hazard rate is monotonically increasing with time; conversely, if $\alpha < 1$ the hazard rate is monotonically decreasing with time; when $\alpha = 1$ the hazard rate is constant, indicating the equivalence to an exponential distribution.

Model inference is performed using a Bayesian methodology. A similar approach is used in Bostanara et al. (2021) in a study of residential location timing Sydney and Chicago residents. The Bayesian method considers the parameters as random variables and represents their uncertainty through the prior and posterior distributions. A Bayesian inference approach based on Markov Chain Monte Carlo (MCMC) was applied to simulate the posterior distribution of the parameters α and λ using the software Stan (Carpenter et al., 2017). Posterior distributions were approximated based on 2000 MCMC samples collected after a burn-in of 500 samples. The mean, standard deviation, and quantiles of each of the estimated parameters were determined from their posterior distribution. Based on Bayes theorem, the posterior distribution of parameters can be estimated from:

$$f(\alpha, \lambda|\mathbf{t}) = \frac{f(t,\alpha,\lambda)}{f(t)} = \frac{f(t|\alpha,\lambda)\pi(\alpha)\pi(\lambda)}{\iint f(t|\alpha,\lambda)\pi(\alpha)\pi(\lambda)d\alpha d\lambda} \propto f(t|\alpha,\lambda)\pi(\alpha)\pi(\lambda) \tag{7}$$

where $f(\alpha, \lambda|t)$ is the posterior distribution of the parameters α and λ conditional on the observed data t, $f(t, \alpha, \lambda)$ is the joint probability distribution of t and parameters α and λ, $f(t|\alpha, \lambda)$ is the likelihood conditional distribution based on parameters α and λ as specified in Eq. (5), and the functions $\pi(\alpha)$ and $\pi(\lambda)$ are the prior distributions of parameters α and λ, respectively. The convergence of the posterior distribution was monitored by visual inspection of the trace plots.

The hazard and survivor functions for the 2-parameter Weibull model are derived in Eqs. (8) and (9). Random samples of the parameters posterior distribution are then used to predict the survivor and hazard functions.

$$h(t, X) = \frac{\alpha}{\lambda}(\frac{t}{\lambda})^{\alpha-1} \tag{8}$$

$$S(t, X) = exp(-(t/\lambda)^{\alpha}) \tag{9}$$

We apply this methodology to model the average idle time of FFCS vehicles parked in a given zone ($i \in I$). The average idle time is modelled as a Weibull distribution as expressed in Eq. (5), which is based on the goodness of fit statistics (KS and AD) in Table 1 and the fact that we require the distribution to be strictly positive. The average idle time distribution is shown in Fig. 2.

The specification of the scale parameter λ of the Weibull distribution includes several covariates related to socio-demographic factors and land-use.

$$\lambda_i = exp(\beta^T X_i + \mu) \qquad \forall i \in I \tag{10}$$

X_i is a matrix of variables associated to zone i, β is the vector of parameters and μ is an intercept. Uninformed prior distributions are considered for the parameters. Normally distributed priors are defined for the socio-demographic and land-use parameters and a gamma distributed prior that ensures positive values and has proved to be suitable for the data is defined for the shape parameter α.

$$\alpha \sim Gamma(3, 1)$$
$$\beta \sim N(0, 1)$$
$$\mu \sim N(0, 10)$$

A baseline model which includes all variables is estimated at first. Subsequently, we use a backward-selection procedure to qualify the final model. In the selection process insignificant variables were excluded from the model by considering the 95% confidence intervals for the parameters and R^2 of the model.

4. Data

The main data source for this study is a large transactions database of electric FFCS vehicles that describes the exact time and position of where trips start and end. This data is further supplemented with socio-economic and demographic data, land use attributes and transport related variables which provide a measure of the level-of-service offered by alternative modes.

4.1. FFCS trips

The transactions database was made available by DriveNow Copenhagen[1] and contains approximately 327.000 FFCS trips performed by a fleet of around 300 electric vehicles during the period August-2017 to May-2018 (43 weeks). While the fleet did not suffer any significant change during the studied period, daily variation of 10%–15% occurred due to maintenance and cleaning. The predefined area of operation in which the trips start and end was unchanged over the period. The data includes a vehicle identifier and the exact time, geographical coordinates and vehicle battery state of charge for the start and end of the trips. The data provides a very detailed and continuous tracking of each vehicle in space and time, which in turn allows a comprehensive spatial and temporal modelling of these patterns. Furthermore, the data includes a user identifier which makes it possible to recognise which trips are carried out by car sharing users and which ones are the result of relocation or maintenance activities.

4.2. Land-use and transport data

Population and employment data for year 2015 was collected from the Danish National Transport Model (Rich and Hansen, 2016). Stops and frequencies of bus and rail services operating within the study area were also derived from the Danish National Transport Model. A categorical parking variable is included to represent the different price levels for hourly parking within Copenhagen city.[2] Hospitals with more than 3000 employees and university campuses in the area were identified. The location of two large parks (Valbyparken and Amagerfaellen) and the Copenhagen airport were also mapped.

4.3. Socio-economic and demographic data

Socio-economic and demographic information was extracted from the Danish National Travel Survey (TU, Transportvaneundersøgelsen) (Christiansen, 2018). TU is a large scale trip diary conducted continuously since 2006 for all Danish residents between the age of 6 and 84 years. This data, in combination with the zone system of the Danish National Transport Model, was used to calculate aggregated variables for each zone. The variables derived for this study were: average household income, average number of cars per household, share of single households, share of households with children, share of inhabitants with high level of education, share of young adults (18 to 37 years' old).

4.4. Zone system and area of operation

To ensure compatibility of the multiple data sets, the start and end coordinates of the trips were geo-coded and matched to geographical zones defined in the Danish National Transport Model. The FFCS operation area was matched to 92 zones. The operation area is composed by a main area that includes the city centre and surrounding neighbourhoods, and several other zones which are isolated from the main area (satellite areas). These satellite areas are either hospitals, university campuses, major shopping locations, important commuting suburban train stations or the airport.

4.5. Pre-processing of data

The transactions dataset which reports all trips performed by the FFCS fleet was used to calculate the idle time dataset. The idle time of each vehicle parked at a given zone was calculated as the amount of minutes between the drop-off and the subsequent pick-up. In order to have a clear definition of the vehicle idle time between bookings, pick-ups performed by the service operator (either for a relocation or a maintenance trip) were removed from the data. Additionally, a small number of observations were excluded from our analysis for various reasons. These exclusions occurred due to low battery state, GPS logging issues, observations outside the designated operation area, and public holidays. In addition, we removed observations with idle times equal to or greater than one day (equivalent to 1440 min). As a result, censored data — representing instances where the event (car pickup) had not occurred by the end of the study period — were omitted from our analysis. However, it is important to note that the number of excluded observations in total amounts only to 2.7% of the raw dataset.

Since the zones used for this study have different sizes, socio-demographic and transport accessibility data was included in the models as averages per zone, percentage of the population per zone and density.

5. Analysis and results

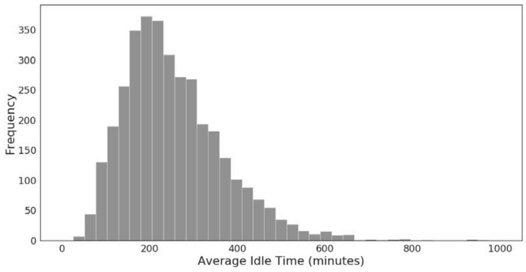
5.1. Descriptive statistics

Fig. 2(a) presents a histogram of the distribution of the weekly average idle time and the geographical distribution of observed idle time per zone. Frequency represents the occurrence of idle time instances over the 92 zones and the 43 weeks of data. The average idle time of zones varies between 35 and 1087 min with a mean of 252 min.

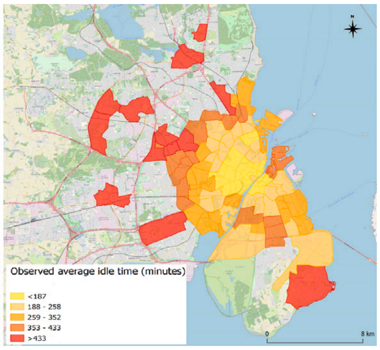
Table 2 presents a summary of the variables used in the modelling process. The full set of variables includes spatial characteristics for each zone classified with respect to socio-demographics and land use.

[1] DriveNow has merged with car2go and re-branded to ShareNow, which follows a similar business model.

[2] Public parking in Copenhagen: https://international.kk.dk/artikel/public-parking-copenhagen.



(a) Observed idle time histogram.



(b) Geographical distribution of observed idle time.

Fig. 2. Visualisation of observed idle time patterns aggregated at the level of zones.

Table 2
Summary of variables considered for modelling.

	Mean	S.D	Min	Max
Socio-demographic				
Jobs density	6394	10 424	93	58 947
Population density	7186	6375	0	28 292
Shopping density	432	746	8	5692
Av. HH car ownership	0.60	0.21	0	1.17
Av. HH income	302	106	0	837
% Young adults	0.41	0.16	0	1
% High level education	0.21	0.10	0	0.64
% Single HH	0.24	0.11	0	0.94
% HH with children	0.32	0.11	0	0.55
Land-use				
Rail departure density	306	596	0	3645
Bus departure density	2504	2953	80	17 371
Parking stress (dummy)	0.79	1.30	0	4
Airport (dummy)	0.01	0.10	0	1
University campus (dummy)	1.08	0.35	1	3
Hospitals (dummy)	0.06	0.24	0	1
Parks (dummy)	0.06	0.24	0	1
Zone location: border (dummy)	0.15	0.36	0	1
Zone location: satellite (dummy)	0.10	0.31	0	1

5.2. Model estimates and posterior summaries

Table 3 presents a summary of the posterior distributions of the model estimated on the basis of the MCMC simulation-based Bayesian approach. These tables show the "Mean" and "S.D". of the posterior distribution of the parameters, as well as the 2.5% and 97.5% percentiles. Only significant parameters are included in the final model. The estimates for shopping density, percentage of single households, percentage of households with children and bus stop density were not found significant at a 95% confidence level and therefore dropped out from the final model.

Given the model parameterisation, a negative sign for any estimated parameter would indicate that an increase in the variable yields a reduction in the FFCS idle time and vice versa for parameters with positive signs. Several variables such as jobs density, population density, number of rail departures included within the scale parameter have a very significant effect on FFCS idle time. The intercept μ of the scale parameter has as well a very significant effect meaning that the average effect of all other variables not included in the model are significant too. In the model, the shape of the hazard α is greater than 1 indicating that the hazard rate is monotonically increasing with time, i.e. the likelihood of a FFCS vehicle of being booked increases as the time since it has been parked increases (see Fig. 5(b)).

The ability of the proposed model to describe the data in the training datasets is evaluated by visual comparison of the observed and predicted survivor curves. Sets of parameters are randomly sampled from their posterior distributions. For each set of parameters, individual survivor curves are simulated based on the variables and are then averaged to derive a population survivor curve. The mean predicted population survivor curve and a 95% confidence interval are calculated based on the samples. Fig. 3 presents the visual evaluation of the model and how well it is predicting the idle time of the training dataset. The empirical survivor curve (red solid line) is plotted along with the simulated mean survivor curve (black dotted line) and the 95% confident intervals (grey areas). To examine out-of-sample accuracy the following procedure is applied. For each element within the test dataset, the expected mean

Table 3
Parameter estimates based on MCMC.

Parameters	Mean	S.D.	2.5%	97.5%
α	3.914	0.057	3.797	4.031
μ	−1.419	0.005	−1.429	−1.409
Socio-demographic				
Jobs density	−0.190	0.006	−0.202	−0.179
Population density	−0.120	0.007	−0.133	−0.107
Av. HH Car ownership	0.051	0.011	0.030	0.072
Av. HH Income	−0.074	0.008	−0.091	−0.058
% Young adults	−0.021	0.009	−0.038	−0.004
% High level education	−0.022	0.008	−0.039	−0.005
Land-use				
Rail departure density	−0.046	0.006	−0.058	−0.034
Parking stress: L2	−0.075	0.006	−0.086	−0.063
Airport	−0.077	0.006	−0.090	−0.064
University campus: DTU	−0.043	0.007	−0.057	−0.028
Hospitals	−0.058	0.006	−0.070	−0.046
Parks	−0.034	0.006	−0.045	−0.022
Zone location: border	0.046	0.006	0.035	0.057
Zone location: satellite	0.142	0.008	0.128	0.157
$R^2 - train$	0.59			
$R^2 - test$	0.56			
$N - train$	2647			
$N - test$	883			

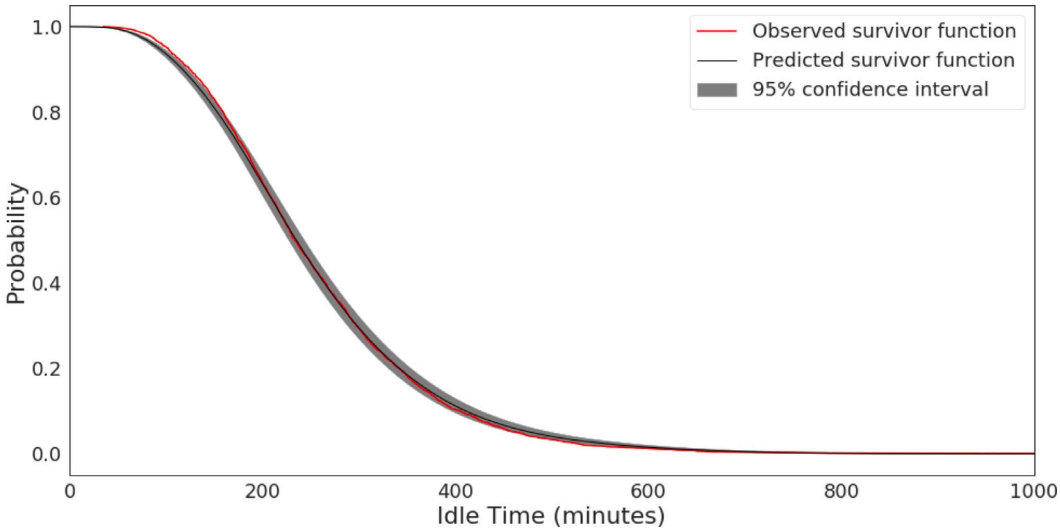


Fig. 3. Fit of the survivor distribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is calculated (as in Eq. (6)) for all sets of random sampled parameters. Then, we consider the average to attain a prediction which is comparable to the observed data based on goodness-of-fit measures such as R^2.

The model can be seen to exhibit a rather acceptable predictive performance when measured according to the R^2. It is clearly shown that the empirical survivor curve falls inside the simulated 95% confidence intervals.

5.3. Zone level predictions

The overall geographical distribution of the average idle time in Fig. 4 exhibits a clear pattern where certain areas appear to be more successful with respect to FFCS demand than others. Zones that are close to the city centre exhibit lower FFCS vehicle idle time (demand hot spots) while vehicles parked in the zones located in the perimeter of the area of operation and in satellite zones have higher idle times.

The estimation approach does not only allow for the prediction of the average idle time per zone but also for the calculation of the complete survivor function and its uncertainty. The mean survivor and hazard functions estimated for the 92 zones within the area of operation are presented in Figs. 5(a) and 5(b), respectively. The survivor curve for a vehicle parked within a given zone

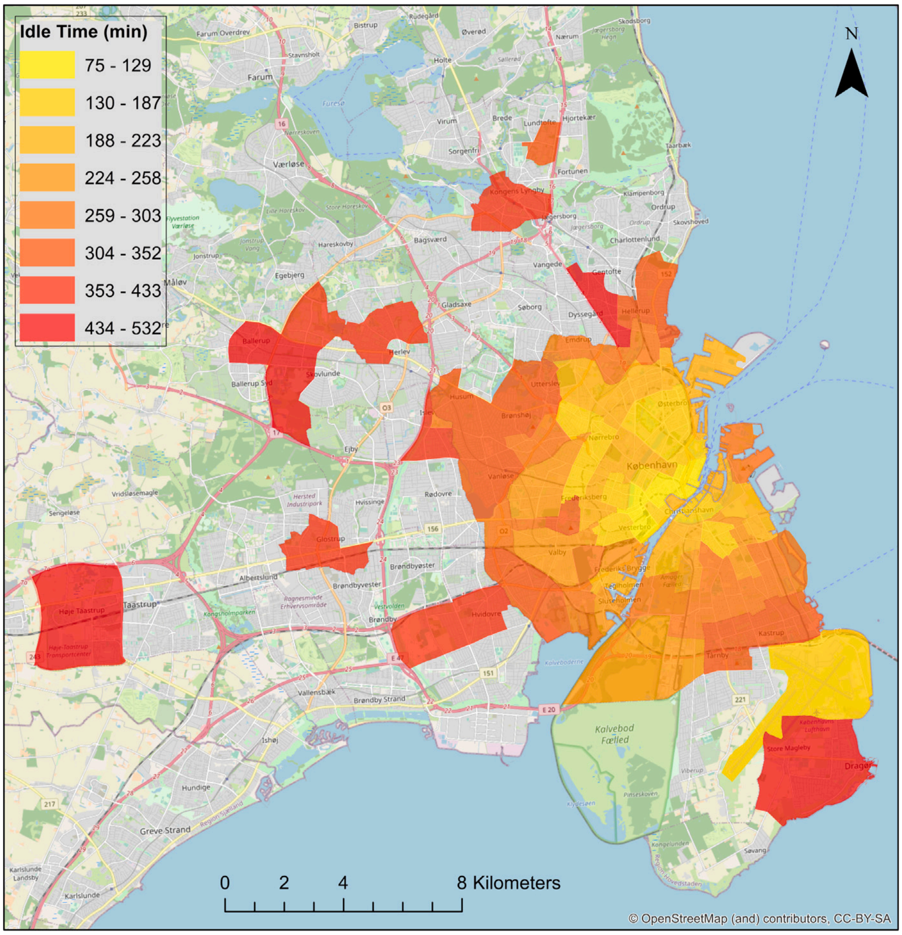


Fig. 4. Geographical distribution of the mean predicted idle time per zone.

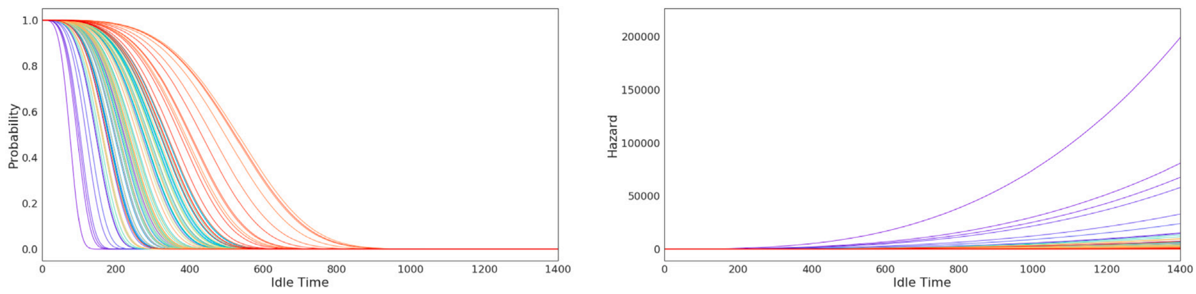
shows a delay or advancement of its survival time compared to the other zones in the area. The steeper the survival curve is, the lower the probability of a vehicle remaining idle for a long time period; and therefore, a demand hot spot for FFCS. The steeper the hazard function is, the higher the probability of an idle vehicle being booked; and therefore this also represents a demand hot spot for FFCS. In order to avoid the overlapping of the curves, and as an example of what is obtained for all zones within the area of operation, four zones were randomly selected and plotted along with their confidence interval in Figs. 6(a) and 6(b). Fig. 6 shows that zones not only present different mean for their survivor and hazard functions, but also they have important differences in the uncertainty associated with the idle time prediction. From Fig. 6(a) it is possible to observe that while survivor predictions for zones with Id 22 and 42 are quite certain, the survivor predictions for zones Id 79 and 82 exhibit a large uncertainty.

6. Case study: Evaluation of operation area expansion

In the following section, in order to illustrate the practical value of the model in a strategic planning context, we consider the expansion of the FFCS operation area of the case-study to new zones within Copenhagen and evaluate their performance according to the predicted vehicle idle time.

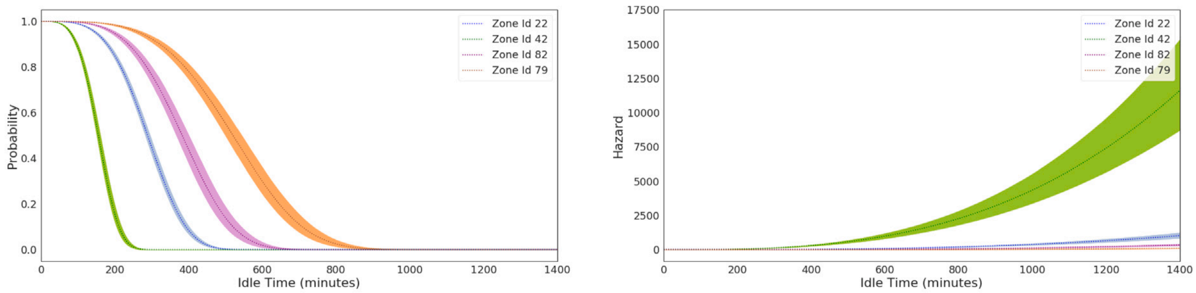
We consider the area of operation to be enlarged by 28 new zones, which will become the new border zones of the operational area. The zone level variables used in the model (as identified in Table 2) are determined for each of the 28 new zones. The expected idle times are simulated on the basis of the estimated parameters presented in Table 3. In total 100 randomly sampled sets of parameters together with the zone level variables were used to simulate the expected idle time for each zone following Eq. (6). The mean idle time of the new zones, as plotted in Fig. 7 is the average of the 100 draws of expected idle time for each zone.

Fig. 7 shows the zones that are already included in the FFCS operation area in a grey colour. Considerable differences in the potential performance of FFCS for these new zones measured by their idle time are observed. It can be seen that there is higher potential to expand the FFCS operation area towards the northern zones rather than to the southwest. This is explained by the fact that the socio-demographic and spatial characteristics of the northern zones are more aligned with the local FFCS drivers of demand.



(a) Survivor function (b) Hazard function

Fig. 5. Zone mean functions for the 92 zones.



(a) Survivor function (b) Hazard function

Fig. 6. Mean and 95% CI for four random selected zones.

Fig. 7 shows that the variation in the performance of the new zones corresponds to approximately 1 h of idle time. Hence, the operational performance for the best and the worst zone translates into a difference of approximately 0.75 trips per day or a potential 20% difference in earnings.

The findings also suggest that from a operator perspective, it would be beneficial to concentrate the operation area rather than allowing satellite zones in the outskirts of the city. When comparing Figs. 4 and 7, the predicted idle time of the most efficient new zone (Fig. 7) is approximately 60% lower than that of the worst performing existing satellite zone (Fig. 4). While there can be other reasons why such a strategy would be chosen, e.g. to attract in the early stage of the service young people that do not own a car by serving universities or to increase accessibility to public facilities by serving hospitals, it comes at a rather substantial cost. Another argument against the current structure is that it tends to be scattered and unconnected, which suggests that an extension to the north of the area would be a more fruitful strategy.

7. Discussion and future research

7.1. Socio-demographic and land-use variables

The model reveals that population and jobs density explain an important share of the variance in idle times between zones. FFCS vehicles parked in zones with higher population density and job density tend to have shorter idle times (ceteris paribus). Although there is no general consensus regarding the effect of job density with respect to FFCS demand, our finding is aligned with previous research by Seign et al. (2015) for FFCS systems in the cities of Berlin and Munich. Instead, findings from Khan and Machemehl (2017) for Austin (Texas, USA) and from Becker et al. (2017b) for Basel (Switzerland) show a negative correlation between FFCS demand and jobs density and discuss that FFCS is less likely to be used for commuting trips. The diversity of findings could possibly relate to differences in cultural backgrounds and the role of private cars in the different cities. As agreed within the FFCS literature, the model shows that areas with a larger fraction of young adults are positively related to a higher demand for shared vehicles. Furthermore, we find that areas with higher than average household income and with a higher share of educated people, are likely to have shorter idle times. Although the negative relation between income and idle time (possibly due to the increased propensity of owning a car) could seem counter intuitive, the finding is consistent with previous FFCS literature that has characterised the prototypical user based on surveys of users and non-users (Kopp et al., 2015; Papu Carrone et al., 2020). Conversely, in areas where the household car ownership levels are high, the idle times of FFCS vehicles tend to be higher, as also evidenced in Khan and Machemehl (2017).

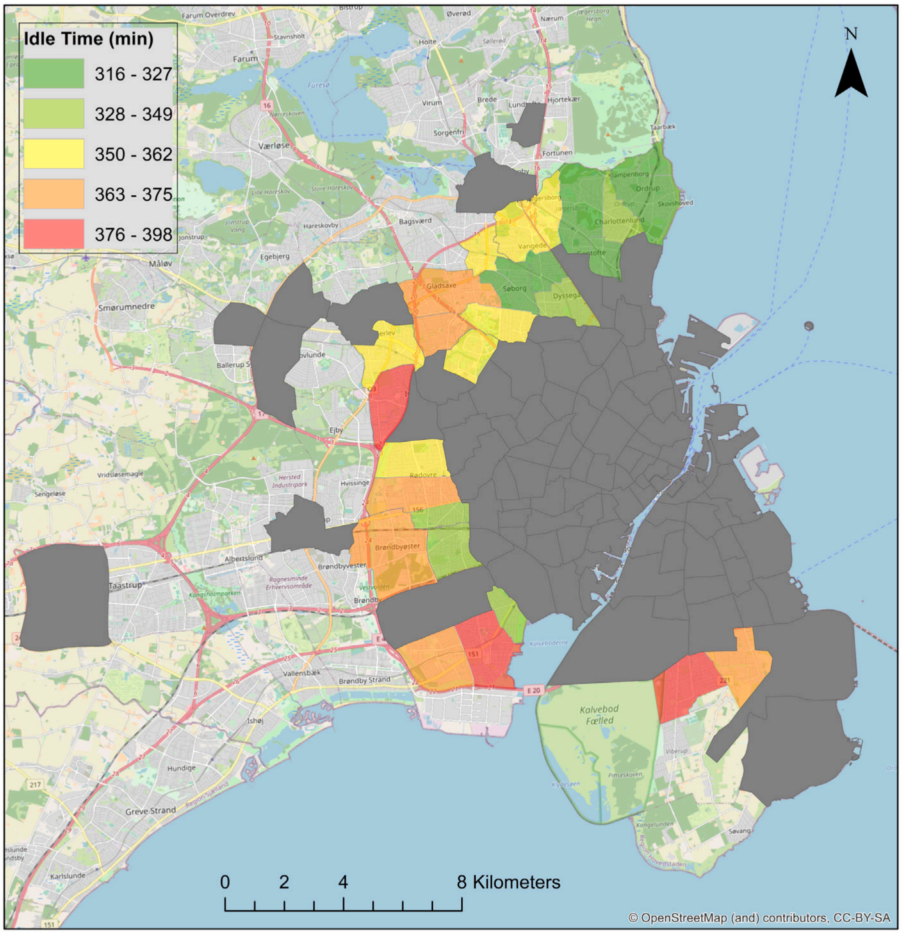


Fig. 7. Predicted mean idle time for zones outside the current area of operation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

With respect to the effect accessibility to other transport modes has on FFCS demand, it is found that higher accessibility to rail transport (metro, urban and suburban trains) generally implies shorter idle times. No significant effect with respect to accessibility to bus stops is confirmed. This result could either be interpreted as FFCS being used as part of inter-modal trip chains involving train services, as discussed by Khan and Machemehl (2017); or it could be understood as a potential substitution to rail transport since both modes compete in the same location (Papu Carrone et al., 2020). However, it most likely reflects that both transport modes are complementary (Ingvardson et al., 2023), with individuals using either rail transport or FFCS for different purposes of travel. Different findings are presented in Becker et al. (2017b) where it is suggested that FFCS is mainly used in areas with inferior public transport alternatives. FFCS vehicles parked in zones with high parking fees are likely to have shorter idle times given that FFCS vehicles are not subjected to the parking cost. However, the highest level of parking fee was not found significant in the model, probably due to the fact that this area is mainly a pedestrian area with very inconvenient access for car transport.

Generally, vehicles parked in zones located close to the border of the operation area have longer idle times compared to those parked in more central locations. Vehicles parked in satellite zones have even longer expected idle times. Yet, the airport, the DTU university campus, the hospitals and parks are all attractions that render shorter idle times than their equivalent without them.

7.2. Implications for real-life systems and limitations

The analysis of the present paper was originally motivated by a discussion with an operator of a FFCS. For those running such systems, the financial viability and sustainability of the business is critically based on a supply of available vehicles being ready at the times and locations where individual drivers demand their use. This could be said to be the weak-point of such a system, since (without load rebalancing) the locations where vehicles are available are determined by where previous users have dropped them. However, the operator has the opportunity to provide incentives, through price reductions, offers or premiums, which could be spatially differentiated, thereby influencing drop-off locations. However, such incentives must also align with locations where there is potential for demand growth, possibly including areas where there is currently low usage of the system. The locations with greatest potential for growth will vary across a city, and so it is important to understand the demand potential across different locations and their determinants/correlates. One of the main limitations of the analysis presented is that it is a 'steady state one', providing tools to model and predict spatial variations in idle time distributions for a fixed fleet size and fixed demand. While this is directly useful for a mature system, it is arguably less applicable to the case of a system that is growing over time. However, in this context we believe the developed average idle time prediction approach could be used to evaluate the expansion of the operational area in an incremental way by investigating marginal expansions. This would respect the limitation of the predictive approach, assuming a constant fleet size and performance when evaluating the addition of new zones to the operational area. Furthermore, the Bayesian approach is highly suited for such a step-wise/incremental application. In practice, the decision process would involve: (i) decide on the best new zone(s) to expand into using model parameters estimated based on the current data; (ii) expand the fleet size correspondingly, with incentives to attract demand to the identified zones from (i); (iii) collect new data under the new conditions of operation (area and fleet), estimating new posterior parameters for these new conditions using estimates from (i) as prior parameters. This process can be repeated as new zones are included within the operational area.

7.3. Future research

Furthermore, there are several interesting avenues for future research. An interesting extension would be to develop a predictive model at the level of the individual vehicles. The methodology presented in this paper should be applicable to disaggregated models that represent the behaviour and characteristics of individual vehicles (such as vehicle fuel type, size or battery state of charge). The operator would be able to predict the expected idle time of each individual vehicle as soon as it is parked and becomes idle. When the predicted idle time for a vehicle exceeds the threshold for the corresponding category, the vehicle should be relocated to a different zone where the demand is expected to be higher. Conversely, when the predicted idle time of a vehicle is lower than its corresponding relocation threshold, the vehicle should remain parked waiting for the next booking. The avenue of research taps into a growing field of research that looks at optimal pricing strategies with the purpose of controlling relocation needs, e.g. such as in Yang et al. (2022), Chang et al. (2022), and Huang et al. (2021). We find it intriguing to enhance relocation strategies by integrating models of idle time performance with methodologies commonly utilised in operational research. Such approaches would also make it possible to integrate weather and time-of-day dependent variables, as well as other attributes that relate to the specific car as its battery state-of-charge. This idle time predictive approach together with a destination choice model would fit nicely into a simulation framework to further investigate different relocation strategies.

Another interesting challenge, which is closely linked to the design of the operational area, is to determine the proper fleet size for a given area. As the design of the operational area and the fleet size are inter-connected this challenge is not easily addressed. Firstly, it would require knowing how additional vehicles will increase demand due to unserved demand, and secondly, knowing the operational cost structure of increasing the fleet. The latter might require the estimation of the economy of scale for this particular system.

The problem of managing and optimising FFCS services can be found in many other parallel applications within transport. This includes the growing number of e-scooters and shared bicycles, for which the presented model can be applied in a straightforward manner. It would be interesting to pursue such studies further.

8. Conclusion

In the paper we present a statistical model for analysing and predicting idle time of a commercial free-floating car sharing service (FFCS). As the idle time performance is a more direct measure of business performance compared to booking-based models, we are able to monitor system performance across the geography in a rather direct manner. Most importantly, however, the focus on idle time as opposed to bookings, allows to circumvent the problem resulting from latent unserved demand. Such latent demand often invalidates the modelling of bookings as a low booking count can result from insufficient supply rather than indicating poor performance. Addressing this challenge from a bookings perspective often necessitates complex methods, whereas the idle time approach appears simpler and more straightforward. The model is formulated as a Weibull hazard-based duration model and is estimated using a Bayesian inference approach based on MCMC simulation.

The model provides insights concerning the geographical distribution of FFCS demand and how this performance is correlated with a number of geographical attributes. The model can be used to identify successful and unsuccessful zones within the area of operation and in turn be used as a tool to understand how the area of operation should be designed or expanded. Furthermore, the Bayesian framework allows the prediction of the survivor function for the average idle time of a given zone and the uncertainty of each prediction.

The model is useful as a means to support strategic decisions concerning the zones that should be included in the operation area when FFCS is deployed for the first time in a city. However, it is also potentially useful as a means to analyse the effect of area expansions for established systems. This is analysed in a case study for Copenhagen, where 28 new zones are considered as potential candidates for extending the area. It is shown that several of these new zones have a better performance compared to many of the existing zones. Considering that many of the new zones are located closer to the core operation area compared to the existing satellite zones, which generally perform poorly, it could suggest that such an expansion strategy would render better business performance.

CRediT authorship contribution statement

Andrea Papu Carrone: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Jeppe Rich:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **David Watling:** Writing – review & editing, Writing – original draft, Supervision, Methodology.

Acknowledgements

We would like to acknowledge Share Now (previously named DriveNow) and, in particular thank Christian Gade Christensen for sharing the data for the purpose of this project. Special thanks to Filipe Rodrigues for fruitful discussions and for introducing us to Stan.

References

Anastasopoulos, P.C., Islam, M.B., Perperidou, D., Karlaftis, M.G., 2012. Hazard-based analysis of travel distance in urban environments: Longitudinal data approach. J. Urban Plann. Dev. 138 (1), 53–61. http://dx.doi.org/10.1061/(ASCE)UP.1943-5444.000010.

Balusu, S.K., Mannering, F., Pinjari, A., 2020. Hazard-based duration analysis of the time between motorcyclists' initial training and their first crash. Anal. Methods Accid. Res. 28, 100143. http://dx.doi.org/10.1016/j.amar.2020.100143.

Becker, H., Ciari, F., Axhausen, K.W., 2017a. Comparing car-sharing schemes in Switzerland: User groups and usage patterns. Transp. Res. A 97, 17–29. http://dx.doi.org/10.1016/j.tra.2017.01.004.

Becker, H., Ciari, F., Axhausen, K.W., 2017b. Modeling free-floating car-sharing use in Switzerland: A spatial regression and conditional logit approach. Transp. Res. C 81, 286–299. http://dx.doi.org/10.1016/j.trc.2017.06.008.

Becker, H., Ciari, F., Axhausen, K.W., 2018. Measuring the car ownership impact of free-floating car-sharing – A case study in Basel, Switzerland. Transp. Res. D 65, 51–62. http://dx.doi.org/10.1016/j.trd.2018.08.003.

Bhat, C.R., 1996. A generalized multiple durations proportional hazard model with an application to activity behavior during the evening work-to-home commute. Transp. Res. B 30 (6), 465–480. http://dx.doi.org/10.1016/0191-2615(96)00007-0.

Bhat, C.R., Pinjari, A.R., 2007. Duration modeling. In: Hensher, D.A. and Button, K. (Ed.), Handbook of Transport Modelling. Emerald Group Publishing Limited, Bingley, pp. 105–131.

Bostanara, M., Taha Hossein, R., Auld, J., Ghasri, M., 2021. A comparison between residential relocation timing of Sydney and Chicago residents: A Bayesian survival analysis. Comput. Environ. Urban Syst. 89, https://api.semanticscholar.org/CorpusID:237418333.

Braun, A., Koch, A., Hochschild, V., 2016. Intraregionale unterschiede in der carsharing-nachfrage. disP - Plan. Rev. 52 (1), 72–85, https://www.tandfonline.com/doi/full/10.1080/02513625.2016.1171051.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: A probabilistic programming language. J. Stat. Softw. 76 (1).

Cervero, R., Golub, A., Nee, B., 2007. City CarShare longer-term travel demand and car ownership impacts. In: Transit Management, Technology, and Planning, one thousand nine hundred ninety-two ed. In: Transportation Research Record, pp. 70–80. http://dx.doi.org/10.3141/1992-09, number 1992.

Chang, X., Wu, J., de Almeida Correia, G.H., Sun, H., Feng, Z., 2022. A cooperative strategy for optimizing vehicle relocations and staff movements in cities where several carsharing companies operate simultaneously. Transp. Res. E 161, http://dx.doi.org/10.1016/j.tre.2022.102711.

Cheng, J., Chen, X., Ye, J., Shan, X., 2021. Flow-based unit is better: exploring factors affecting mid-term OD demand of station-based one-way electric carsharing. Transp. Res. D 98, http://dx.doi.org/10.1016/j.trd.2021.102954.

Cheng, Z., Taha Hossein, R., Sisi, J., Mojtaba, M., Waller, S.T., Dixit, V., 2022. A spatio-temporal autocorrelation model for designing a carshare system using historical heterogeneous data: Policy suggestion. Transp. Res. C 141, http://dx.doi.org/10.1016/j.trc.2022.103758.

Christiansen, H., 2018. Beregning af usikkerhed og varians for Transportvaneundersøgelsen. In: Proceedings from the Annual Transport Conference at Aalborg University. https://www.man.dtu.dk/-/media/institutter/management/scientific-advice/transportvaneundersoegelsen/t-u/tu-udgivelser/faktaark/td18_tuusikkerhedvarians.pdf.

Cox, D.R., 1972. Regression models and life-tables. J. R. Stat. Soc. Ser. B Stat. Methodol. 34 (2), 187–202. http://dx.doi.org/10.1111/j.2517-6161.1972.tb00899.x.

de Lorimier, A., El-Geneidy, A.M., 2012. Understanding the factors affecting vehicle usage and availability in carsharing networks: A case study of communauto carsharing system from Montréal, Canada. Int. J. Sustain. Transp. 7 (1), 35–51. http://dx.doi.org/10.1080/15568318.2012.660104.

Dissanayake, D., 2017. Watching the clock on the way to work? Analysing trends in commuting activities, modes and gender differences in commute times, using hazard-based duration modelling methods. J. Transp. Geogr. 65 (November), 188–199. http://dx.doi.org/10.1016/j.jtrangeo.2017.10.013.

Evan, F., Osorio, C., Zhou, T., 2021. A data-driven method for reconstructing a distribution from a truncated sample with an application to inferring car-sharing demand. Transp. Sci. 55 (3), 616–636. http://dx.doi.org/10.1287/trsc.2020.1028.

Firnkorn, J., Müller, M., 2012. Selling mobility instead of cars: New business strategies of automakers and the impact on private vehicle holding. Bus. Strategy Environ. 21 (4), 264–280. http://dx.doi.org/10.1002/bse.738.

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. Transp. Res. C 120 (September), http://dx.doi.org/10.1016/j.trc.2020.102775.

Gilbert, C.C., 1992. A duration model of automobile ownership. Transp. Res. B 26 (2), 97–114. http://dx.doi.org/10.1016/0191-2615(92)90001-D.

Guidon, S., Becker, H., Axhausen, K., 2019. Avoiding stranded bicycles in free-floating bicycle-sharing systems: using survival analysis to derive operational rules for rebalancing. In: 2019 IEEE Intelligent Transportation Systems Conference. ITSC 2019, IEEE, pp. 1703–1708. http://dx.doi.org/10.1109/ITSC.2019.8916869.

Hensher, D., 1998. The timing of change for automobile transactions : competing risk multispell specification. In: Ortuzar, J.D., Hensher, D., Jara-Diaz, S. (Eds.), Travel Behaviour Research : Updating the State of Play. Elsevier, https://api.semanticscholar.org/CorpusID:106566037.

Hensher, D.A., Mannering, F.L., 1994. Hazard-based duration models and their application to transport analysis. Transp. Rev. 14 (1), 63–82. http://dx.doi.org/10.1080/01441649408716866.

Huang, K., An, K., de Almeida Correia, G.H., Rich, J., Ma, W., 2021. An innovative approach to solve the carsharing demand-supply imbalance problem under demand uncertainty. Transp. Res. C 132, http://dx.doi.org/10.1016/j.trc.2021.103369.

Ingvardson, J.B., Weisshaar, M., Rich, J., 2023. Will good public transport substitute free-floating car sharing? A case study from Copenhagen. Case Stud. Transp. Policy 12, http://dx.doi.org/10.1016/j.cstp.2023.100975.

Jian, S., Taha Hossein, R., Wijayaratna, K.P., Dixit, V.V., 2016. A spatial hazard-based analysis for modelling vehicle selection in station-based carsharing systems. Transp. Res. C 72, http://dx.doi.org/10.1016/j.trc.2016.09.008.

Jorge, D., Correia, G., 2013. Carsharing systems demand estimation and defined operations: A literature review. Eur. J. Transp. Infrastructure Res. 13 (3), 201–220. http://dx.doi.org/10.18757/ejtir.2013.13.3.2999.

Junhua, W., Haozhe, C., Shi, Q., 2013. Estimating freeway incident duration using accelerated failure time modeling. Saf. Sci. 54, 43–50. http://dx.doi.org/10.1016/j.ssci.2012.11.009.

Kang, J., Hwang, K., Park, S., 2016. Finding factors that influence carsharing usage: Case study in seoul. Sustainability (Switzerland) 8 (8), http://dx.doi.org/10.3390/su8080709.

Khan, M., Machemehl, R., 2017. The impact of land-use variables on free-floating carsharing vehicle rental choice and parking duration. In: Springer Geography, pp. 331–347. http://dx.doi.org/10.1007/978-3-319-40902-3_19.

Kopp, J., Gerike, R., Axhausen, K.W., 2015. Do sharing people behave differently? An empirical evaluation of the distinctive mobility patterns of free-floating car-sharing members. Transportation 42 (3), 449–469. http://dx.doi.org/10.1007/s11116-015-9606-1.

Kostic, B., Loft, M.P., Rodrigues, F., Borysov, S.S., 2021. Deep survival modelling for shared mobility. Transp. Res. C 128 (May), http://dx.doi.org/10.1016/j.trc.2021.103213.

Le Vine, S., Lee-Gosselin, M., Sivakumar, A., Polak, J., 2014. A new approach to predict the market and impacts of round-trip and point-to-point carsharing systems: Case study of London. Transp. Res. D 32, 218–229. http://dx.doi.org/10.1016/j.trd.2014.07.005.

Le Vine, S., Polak, J., 2019. The impact of free-floating carsharing on car ownership: Early-stage findings from London. Transp. Policy 75, 119–127. http://dx.doi.org/10.1016/j.tranpol.2017.02.004.

Lee, J.H., Goulias, K.G., 2018. A decade of dynamics of residential location, car ownership, activity, travel and land use in the seattle metropolitan region. Transp. Res. A 114, 272–287. http://dx.doi.org/10.1016/j.tra.2018.01.029, Transportation and Traffic Theory: Behavior and Planning Applications.

Louie, J., Shalaby, A., Habib, K.N., 2017. Modelling the impact of causal and non-causal factors on disruption duration for Toronto's subway system: An exploratory investigation using hazard modelling. Accid. Anal. Prev. 98, 232–240. http://dx.doi.org/10.1016/j.aap.2016.10.008.

Ménoire, M., Wielinski, G., Morency, C., Trépanier, M., 2020. Predicting carsharing station-based trip generation using a growth model. Transp. Res. Procedia 48, 1466–1477. http://dx.doi.org/10.1016/j.trpro.2020.08.192.

Millard-Ball, A., Murray, G., Schure, J.T., 2006. Car sharing as a parking management strategy. Transp. Res. Board 1–15, https://api.semanticscholar.org/CorpusID:166580872.

Papu Carrone, A., Hoening, V.M., Jensen, A.F., Mabit, S.E., Rich, J., 2020. Understanding car sharing preferences and mode substitution patterns: A stated preference experiment. Transp. Policy 98 (March), 139–147. http://dx.doi.org/10.1016/j.tranpol.2020.03.010.

Prieto, M., Baltas, G., Stan, V., 2017. Car sharing adoption intention in urban areas: What are the key sociodemographic drivers? Transp. Res. A 101, 218–227. http://dx.doi.org/10.1016/j.tra.2017.05.012.

Rahimi, E., Shamshiripour, A., Shabanpour, R., Mohammadian, A., Auld, J., 2019. Analysis of transit users' waiting tolerance in response to unplanned service disruptions. Transp. Res. D 77 (October), 639–653. http://dx.doi.org/10.1016/j.trd.2019.10.011.

Rashidi, T., Mohammadian, A., 2011. A competing hazard model of household vehicle transaction behavior with discrete time intervals and unobserved heterogeneity. Transp. Lett. 3 (3), 219–229. http://dx.doi.org/10.3328/TL.2011.03.03.219-229.

Rich, J., Hansen, C.O., 2016. The danish national passenger model. Eur. J. Transp. Infrastructure Res. 4 (December), 573–599. http://dx.doi.org/10.18757/ejtir.2016.16.4.3159.

Roukouni, A., Correia, G.H.d., 2020. Evaluation methods for the impacts of shared mobility: Classification and critical review. Sustainability (Switzerland) 12 (24), 1–22. http://dx.doi.org/10.3390/su122410504.

Schmöller, S., Bogenberger, K., 2014. Analyzing external factors on the spatial and temporal demand of car sharing systems. Procedia - Soc. Behav. Sci. 111, 8–17. http://dx.doi.org/10.1016/j.sbspro.2014.01.033.

Schmöller, S., Weikl, S., Müller, J., Bogenberger, K., 2015. Empirical analysis of free-floating carsharing usage: The munich and berlin case. Transp. Res. C 56, 34–51. http://dx.doi.org/10.1016/j.trc.2015.03.008.

Seign, R., Schüßler, M., Bogenberger, K., 2015. Enabling sustainable transportation: The model-based determination of business/operating areas of free-floating carsharing systems. Res. Transp. Econ. 51, 104–114. http://dx.doi.org/10.1016/j.retrec.2015.10.012.

Shaheen, S.A., Cohen, A.P., 2013. Carsharing and personal vehicle services: Worldwide market developments and emerging trends. Int. J. Sustain. Transp. 7 (1), 5–34. http://dx.doi.org/10.1080/15568318.2012.660103.

Shi, Y., Zhang, L., Liu, P., 2015. Survival analysis of urban traffic incident duration: A case study at shanghai expressways. J. Comput. (Taiwan) 26 (1), 29–39, https://api.semanticscholar.org/CorpusID:40476188.

Stillwater, T., Mokhtarian, P.L., Shaheen, S.A., 2009. Carsharing and the built environment: Geographic information system based study of one U.S. operator. Transp. Res. Rec. (2110), 27–34. http://dx.doi.org/10.3141/2110-04.

Yang, S., Wu, J., Sun, H., Qu, Y., Wang, D.Z., 2022. Integrated optimization of pricing and relocation in the competitive carsharing market: A multi-leader-follower game model. Transp. Res. C 138, http://dx.doi.org/10.1016/j.trc.2022.103613.