



A simple and explainable model for park-and-ride car park occupancy prediction

Andreas Kaltenbrunner^{1,2*}, Josep Ferrer², David Moreno² and Vicenç Gómez²

Handling Editor: Luca Aiello

*Correspondence:

kaltenbrunner@gmail.com

¹Internet Interdisciplinary Institute,
Universitat Oberta de Catalunya,
Rambla del Poblenou 154-156,
08018 Barcelona, Spain

²Department of Engineering,
Universitat Pompeu Fabra, Carrer de
Tànger 122-140, 08018 Barcelona,
Spain

Abstract

In a scenario of growing usage of park-and-ride facilities, understanding and predicting car park occupancy is becoming increasingly important. This study presents a model that effectively captures the occupancy patterns of park-and-ride car parks for commuters using truncated normal distributions for vehicle arrival and departure times. The objective is to develop a predictive model with minimal parameters corresponding to commuter behaviour, enabling the estimation of parking saturation and unfulfilled demand. The proposed model successfully identifies the regular, periodic nature of commuter parking behaviour, where vehicles arrive in the morning and depart in the afternoon. It operates using aggregate data, eliminating the need for individual tracking of arrivals and departures. The model's predictive and nowcasting capabilities are demonstrated through real-world data from car parks in the Barcelona Metropolitan Area. A simple model extension furthermore enables the prediction of when a car park will reach its occupancy limit and estimates the additional spaces required to accommodate such excess demand. Thus, beyond forecasting, the model serves as a valuable tool for evaluating interventions, such as expanding parking capacity, to optimise park-and-ride facilities.

Keywords: Car Park; Parking Lot; Occupancy Prediction; Demand estimation; Park-and-ride; Commuter Behaviour

1 Introduction

The rapid increase in energy prices and the growing awareness of the need to reduce greenhouse gas emissions have led an increasing number of commuters to shift from private to public transport. Additionally, major urban centres such as London, Barcelona, and Vienna are progressively restricting access to city centres for private vehicles or limiting free-of-charge parking by expanding regulated parking zones. In response to these developments, park-and-ride systems have become a crucial solution, allowing commuters to leave their private vehicles at designated facilities outside city limits and complete their journeys using public transport.

Effective park-and-ride systems require large parking facilities strategically located near major transport hubs, such as train stations. Accurately estimating the required capacity of these parking facilities and expanding them when demand exceeds available space is a critical aspect of urban planning. A thorough understanding of commuter behaviour and

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

the ability to predict parking demand are essential for making informed decisions in this context.

In this work, we use data collected from a set of eight park-and-ride facilities located in the vicinity of Barcelona, Catalonia. These car parks¹ primarily serve commuters on working days, who typically park their vehicles in the morning before transferring to public transport for work or study in the city. In the afternoon, they return to retrieve their cars and drive home.

The observed periodic patterns in car park occupancy are highly regular, making them well-suited for analysis and prediction using simple models. Our goal is to develop models that improve the planning and management of park-and-ride infrastructure. The specific objectives of our modelling approach are as follows:

- Use as few parameters as possible.
- Ensure that the model parameters are interpretable and correspond to behavioural metrics.
- Enable the model to predict when a parking lot will be full.
- Allow the model to assess the unmet demand for free parking spaces in a car park.

To achieve these goals, we will model car arrivals and departures separately, each using a different truncated normal distribution. This approach allows us to use aggregate data without the need to monitor individual arrivals and departures. Before presenting the proposed model in Sect. 4, we first review relevant literature in Sect. 2. In Sect. 3, we describe the dataset used in our study. Section 5 then analyses the model's ability to fit the observed data and evaluates its performance in both prediction and nowcasting tasks. Finally, we conclude with policy recommendations in Sect. 6 and a discussion of our findings in Sect. 7.

2 Related work

The literature on smart parking solutions is extensive, spanning multiple disciplines and problem types, as highlighted in the survey by Lin et al. [1]. Our work focuses on a specific aspect within this area: modelling car park occupancy. Traditionally, this has been approached primarily as a prediction task (see [2] for a recent overview). Prediction problems can be classified into two main categories: on-street parking and off-street parking (e.g., garages and dedicated parking lots). Our research specifically addresses the latter.

This section is organised around three main limitations identified in existing studies: limited model interpretability, challenges in handling aggregate data instead of individual vehicle arrivals and departures, and insufficient capacity to account for unmet demand, i.e., reasoning about counterfactual scenarios.

Notable studies that focus specifically on car park occupancy prediction include [3–5], and [6]. In [3], the authors compared regression trees, support vector regression, and single-layer neural networks for predicting occupancy in sensor-equipped on-street parking in San Francisco (US) and Melbourne (Australia). They concluded that regression trees, using historical occupancy rates and weekdays as input features, provided the best performance with the least computational cost.

Building on this, recurrent neural networks with multiple hidden layers were employed by [4] to predict occupancy rates in 29 off-street car parks located in Birmingham, UK.

¹Note: throughout this paper, we use the term “car park” following British English usage, which corresponds to “parking lot” in American English.

In another comparative analysis, Zhao et al. [5] reviewed various prediction methods, including linear regression, support vector machines (SVM), neural networks, and autoregressive integrated moving average (ARIMA). Their evaluation, based on data from four car parks across three Chinese cities, also explored the impact of differentiating between weekdays and weekends. They found that SVM consistently provided stable and accurate predictions across nearly all car park types and sizes, while distinguishing weekdays from weekends improved results mainly for medium-sized and office-related parking lots.

Similarly, Awan et al. [6] analysed sensor-collected occupancy data from Santander, a smart city in Spain, over a period of nine months. Their study concluded that simpler algorithms, such as random forests, systematically outperformed more complex deep learning models, such as neural networks.

All these approaches share the three limitations we aim to address in our work.

Several other works relax the assumptions regarding data requirements. However, these approaches still face important limitations in terms of interpretability and their ability to model unmet demand.

Chen [7] consider different levels of data aggregation to predict parking space availability without relying on an explicit behavioural model. They analyse complex usage patterns, including multiple daily peaks, which are more intricate than those addressed in our work.

Ji et al. [8] propose a wavelet neural network (WNN) model for short-term parking prediction, demonstrating higher accuracy and computational efficiency compared to traditional methods using real-world data from Newcastle, UK.

Chawathe [9] introduce a data-driven predictive approach that leverages historical occupancy rates to estimate current and future availability. A key advantage of this method is that it requires minimal or no additional instrumentation of parking facilities, while still achieving sufficient accuracy for practical guidance.

Fokker et al. [10] conduct long-term (6-months ahead) off-street occupancy forecasting for 57 parking facilities in Amsterdam. Their models incorporate external factors such as weather, local events, parking tariffs, and public transport changes. While this work incorporates rich contextual data, the models used offer limited interpretability.

Vakrinou et al. [11] address both short- and long-term occupancy forecasting across multiple off-street parking facilities in central Athens, using a multisource dataset that includes traffic, public transport operations, and weather conditions. Their approach employs LSTMs, Transformers, and XGBoost. Although XGBoost provides a degree of interpretability, the models are not grounded in an explicit behavioural framework, and cannot be used to reason about unmet demand.

Some recent works introduce interpretable models, but they still rely on fine-grained data and are unable to reason about unmet demand.

Tavafoghi et al. [12] propose a queuing framework with non-homogeneous arrival rates for predicting occupancy in truck parking lots. Their method leverages individual arrival and departure times to estimate the service time distribution, assuming four non-overlapping heterogeneous populations of arrivals, each with a time-independent distribution. They also assume that no arriving vehicle finds the parking lot full, i.e., there are infinitely many parking spots.

Schneble and Kauermann [14] model on-street individual parking lot occupancy in the city centre of Melbourne as a two-state stochastic process using a semi-Markov frame-

Table 1 Overview of limitations in existing work on car park occupancy modelling. The table summarises whether each group of studies provides interpretable models, operates without requiring high-granularity data (e.g., individual arrival and departure times), and whether they can account for unmet demand. Our proposed framework addresses all three aspects

Related work	Provides interpretable models	Does not require high-granularity data	Able to handle unmet demand
Zheng et al. [3], Camero et al. [4], Zhao et al. [5], Awan et al. [6]	X	X	X
Chen [7], Ji et al. [8], Chawathe [9], Zhao et al. [5], Fokker et al. [10], Vakrinou et al. [11]	X	✓	X
Tavafoghi et al. [12], Daniotti et al. [13], Schneble and Kauermann [14]	✓	X	X

work. While the model is interpretable and statistically grounded, it requires detailed sensor-level data at the level of individual parking spots.

Similarly, Daniotti et al. [13] analyses patterns in the number of parked car-sharing vehicles across specific areas in the Milan metropolitan region. They employ a Maximum Entropy modelling approach that enables the identification of rare or extreme events, such as weather anomalies or transit strikes. Despite its interpretability, the method also depends on high-resolution, disaggregated data.

Table 1 summarises the main limitations of the reviewed approaches along the three dimensions that motivate our work: model interpretability, ability to operate with aggregate data, and capacity to reason about unmet demand.

Our proposed framework is inspired by prediction techniques based on average activity profiles, such as those used in [15, 16]. Although these methods were originally developed for non-parking mobility data, they illustrate how temporal regularities in aggregate behaviour can be leveraged for forecasting. Similar ideas have been applied in the parking domain, where historical data has proven effective in predicting future occupancy levels even with limited data availability [9].

We end this section by reviewing conceptually related work which considers behavioural aspects of parking demand using questionnaire data. For example, Xue et al. [17] models commuter departure time choices under parking capacity constraints using bounded rationality principles. While their goal is not occupancy prediction per se, their approach highlights the role of limited parking availability in shaping temporal demand patterns and motivates the use of bell-shaped temporal distributions, a key component of our modelling framework.

Similarly, Zhao et al. [18] analyses parking behaviour in the context of Park-and-Ride facilities, using survey data to study how personal attributes, travel characteristics, and user intentions influence the decision to use such infrastructure. Their findings reveal that shortage of parking space is among the top three reasons users avoid Park-and-Ride, emphasising the importance of modelling not only observed demand but also unmet demand.

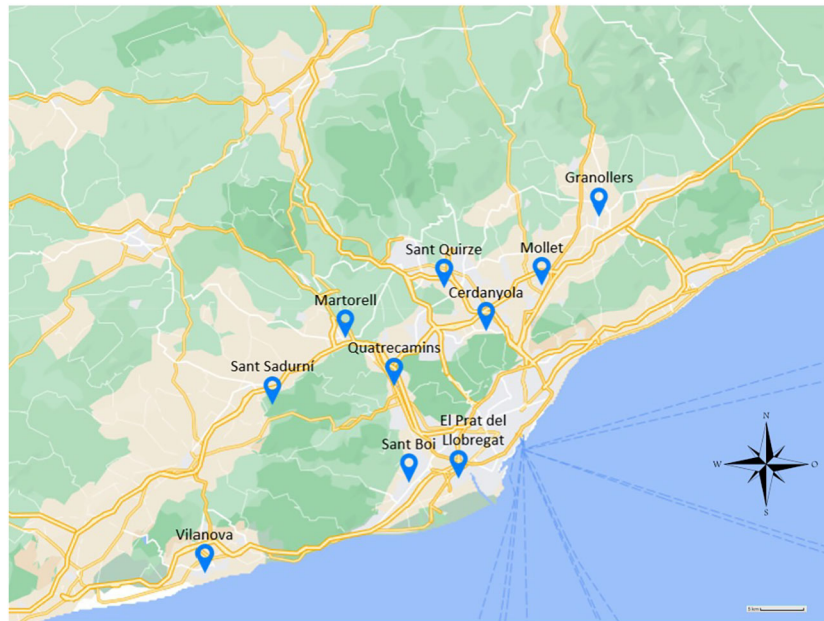


Figure 1 Locations of the park-and-ride car parks analysed, all of them in the vicinity of Barcelona (Catalonia). Two of the locations, Martorell and San Quirze, have not been used due to errors in the data collection process. Scale indicates a 5 km distance

3 Dataset description

We use data provided by the metropolitan transport authority² of Barcelona (Catalonia) which has installed sensors in different park-and-ride parking locations with the aim to better understand their usage patterns and to develop a long-term strategy that allows to improve the interurban mobility from and to Barcelona. Each of the car parks is located in the vicinity of a regional train station, which allows people to park their cars and change directly there to the public transportation system, working as a bridge between both public and private transportation means [19].

Data from a total of 10 parking lots have been provided. Their geographical location is depicted in Fig. 1. Data is continuously gathered from physical sensors placed in the entrances and exits of the car parks and stored in occupancy counters. The entrance of a vehicle increases the occupancy of the particular car park, and the departure of a vehicle decreases that counter. No data about the precise physical location where a vehicle is parked is recorded. This occupancy data is sampled in 30-minute intervals from every parking during a 3-month period from January 1st 2020, 0:00 to April 1st 2020, 0:00. Note that Catalonia entered a COVID-19-induced nationwide lockdown on March 15, 2020, and the car park usage decreased significantly. We will omit the data posterior to this date in the subsequent analysis. We have furthermore removed data from two car parks (Martorell and San Quirze) due to defective sensors and storage errors, which led to irregular readings.

Figure 2 gives an example of the raw data from one of the car parks (Vilanova). The observed occupancy daily pattern is the aggregate result of arrivals and departures to the

²Autoritat del Transport Metropolità (ATM): <https://www.atm.cat/en/atm>.

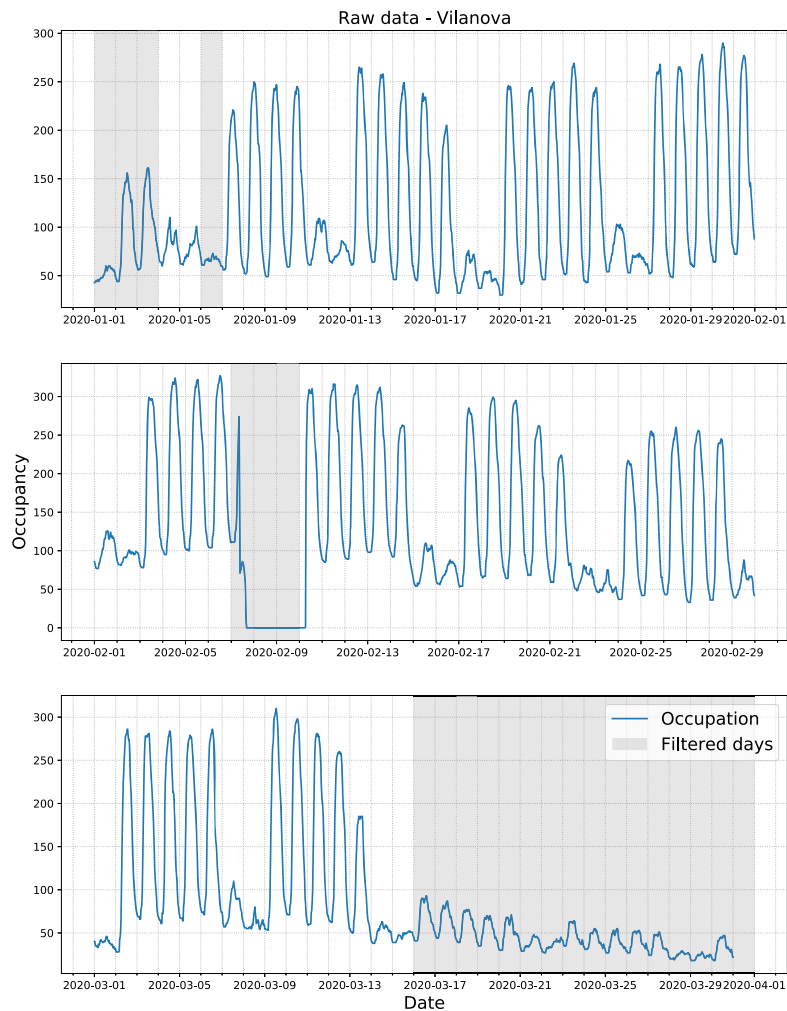


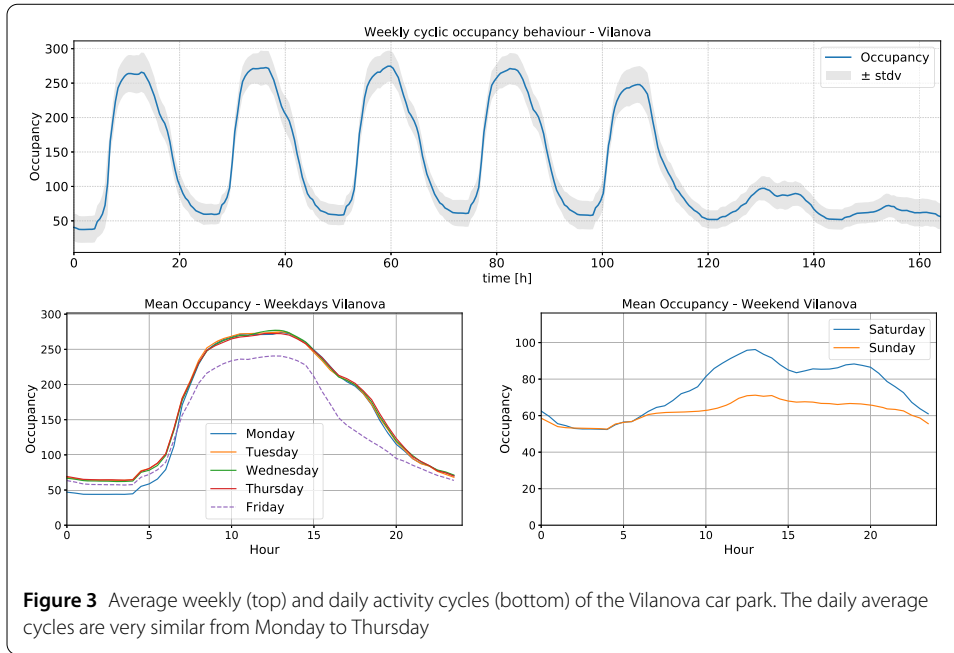
Figure 2 Example of input data from the Vilanova car park, days with grey background have been removed from the analysis due to sensor failure (7th to 9th of February), COVID-19 induced lockdown (days after March 15th) or reduced activity due to Christmas holidays (1st to 3rd and 6th of January)

car park location. In the subsequent analysis, we have manually removed days where the occupancy profiles show errors due to sensor or storage failures. An example of such days are the days from February 7 to February 9 in Fig. 2 or days that are influenced by holidays or vacation periods (e.g. days 1 to 3 and 6 of January in Fig. 2, both periods together with the also removed COVID-19 lockdown period are shown with a grey background).

This raw data already hints at the regularity in the parking behaviour. If we aggregate this data into weekly or daily activity cycles as is done in Fig. 3, we can observe the regularity of the corresponding circadian patterns. The average behaviour from Monday to Thursday is nearly identical, and activity on Fridays shows a small decay.

In what follows, we aggregate the data in three groups: weekdays from Monday to Thursday, Fridays, and weekends (Saturdays and Sundays)

Finally, in Fig. 4 we look at how close the different car parks get to their capacity limit (indicated by the red vertical line). For every day (excluding the days we have filtered out as mentioned above), we record the maximum occupancy of the car park and depict the



corresponding distributions. Four of the car parks (Sant Sadurni, Sant Boi, Quatre Camins and Mollet) reach their maximal occupancy limit regularly during weekdays (orange bars) and also Fridays (blue bars). In the case of the example we used above, the Vilanova car park, this is not the case. We will use this distinction and develop a model extension for the stations where the capacity limit is reached.

4 Modelling car park occupancy

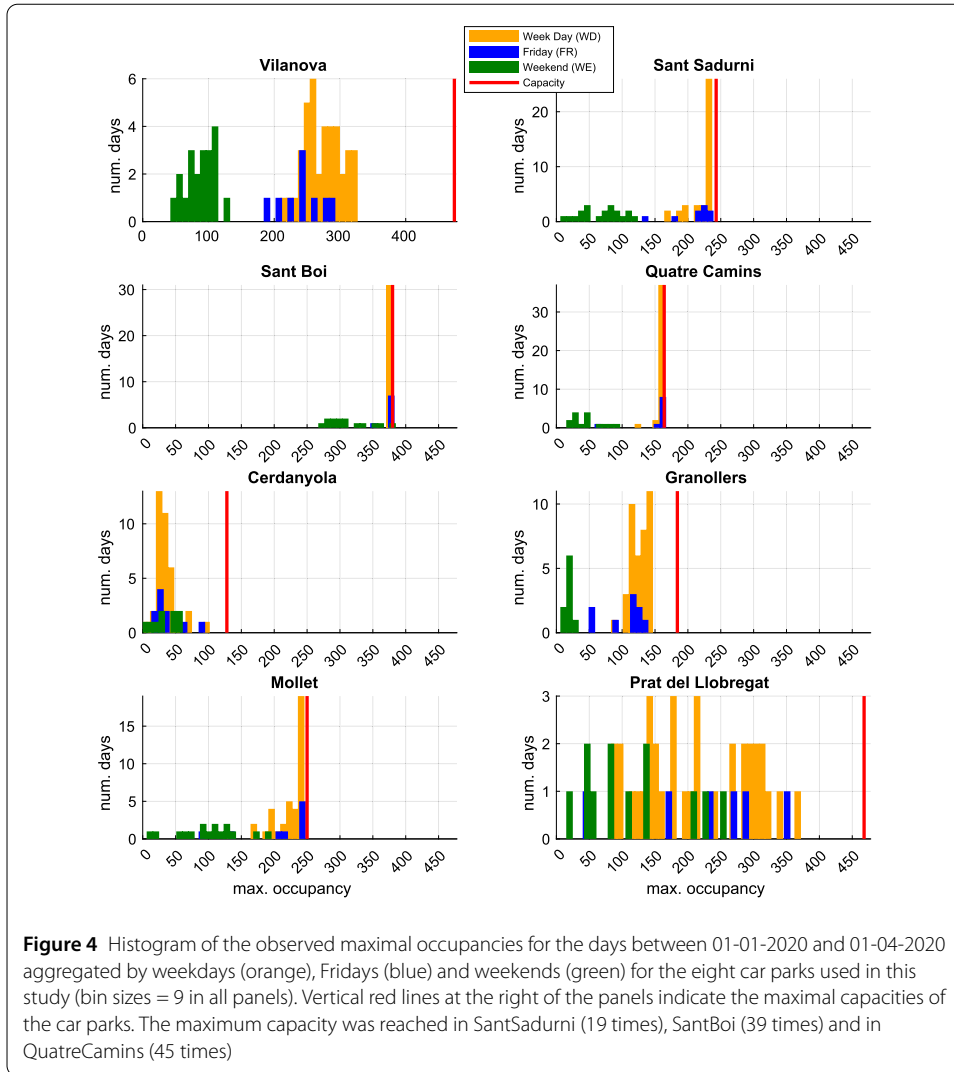
Given the regularity of the activity cycles we have seen in Fig. 3 we now present statistical models able to describe the daily occupancy patterns of an individual parking lot. The model considers the parking lot occupancy as a combination of two independent processes: arrivals and departures. At any time, the occupancy is simply the aggregated number of arrivals minus the aggregated number of departures.

We introduce two types of models: a *basic model* which assumes an unlimited amount of free parking spaces, and an *extended model with occupancy limit*. The former is suitable for parking lots which do not fill up completely, whereas the latter takes into account that a parking lot may reach its capacity limit, and cars arriving after this moment will have to find parking space elsewhere.

4.1 Basic model (unlimited parking spaces)

The proposed model accounts for two event types: car arrivals and departures. For simplicity, these events are assumed to be independent. Each process is modelled as a random variable following a (truncated) normal distribution. The truncation is necessary to constrain the arrival and departure times to occur within the period corresponding to one day (which we normalise to the interval $[0, 1]$). The probability density functions for arrivals ϕ_a and departures ϕ_d are thus:

$$\phi_a(t; \mu_a, \sigma_a) = \frac{\phi\left(\frac{t - \mu_a}{\sigma_a}\right)}{\sigma_a \cdot \left(\Phi\left(\frac{1 - \mu_a}{\sigma_a}\right) - \Phi\left(\frac{-\mu_a}{\sigma_a}\right)\right)}, \quad (1)$$

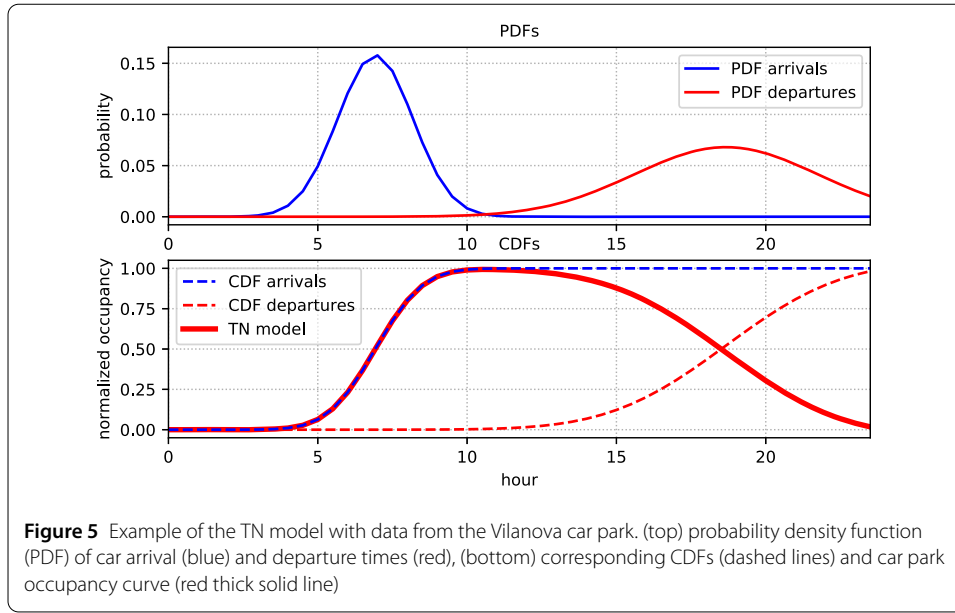


$$\phi_d(t; \mu_d, \sigma_d) = \frac{\phi\left(\frac{t - \mu_d}{\sigma_d}\right)}{\sigma_d \cdot \left(\Phi\left(\frac{1 - \mu_d}{\sigma_d}\right) - \Phi\left(\frac{-\mu_d}{\sigma_d}\right)\right)}, \quad (2)$$

where $\phi(\cdot)$ is the probability density function of the standard normal distribution and $\Phi(\cdot)$ its cumulative distribution function (CDF). We denote the truncated CDFs as $\Phi_a(t; \mu_a, \sigma_a)$ for the arrival and $\Phi_d(t; \mu_d, \sigma_d)$ for the departure process, which are determined by parameters (μ_a, σ_a) and (μ_d, σ_d) respectively.

At any time, the occupancy at the parking lot is given by the number of arrivals minus the number of departures. We can generate a daily occupancy realization by drawing M sample times from both processes $\{A_n\}_{n=1}^M \sim \phi_a$ and $\{D_n\}_{n=1}^M \sim \phi_d$ and subtracting the corresponding counting processes:

$$o(t) = \sum_{n \geq 1} \mathbb{1}_{\{t \geq A_n\}} - \sum_{n \geq 1} \mathbb{1}_{\{t \geq D_n\}}, \quad t = [0, 1], \quad (3)$$



where $\mathbb{1}_{\{\cdot\}}$ is the indicator function that takes the value 1 when the condition is true and 0 otherwise. To know the occupancy of a car park at a given time t , we calculate the difference between all arrivals and departures that happened before.

A (normalized) continuous approximation to Eq. (3) takes the difference between the CDFs of the arrivals $\Phi_a(\cdot)$ and departures $\Phi_d(\cdot)$:

$$f_{\text{TN}}(t; \theta) = \Phi_a(t; \mu_a, \sigma_a) - \Phi_d(t; \mu_d, \sigma_d). \quad (4)$$

We name such a basic model the *Truncated Normal (TN) model*. The TN model has parameters $\theta = (\mu_a, \mu_d, \sigma_a, \sigma_d)$, where μ_a and σ_a denote the location time and the scale of the arrivals, respectively, and μ_d and σ_d , which denote the average time and the scale of the departures, respectively.

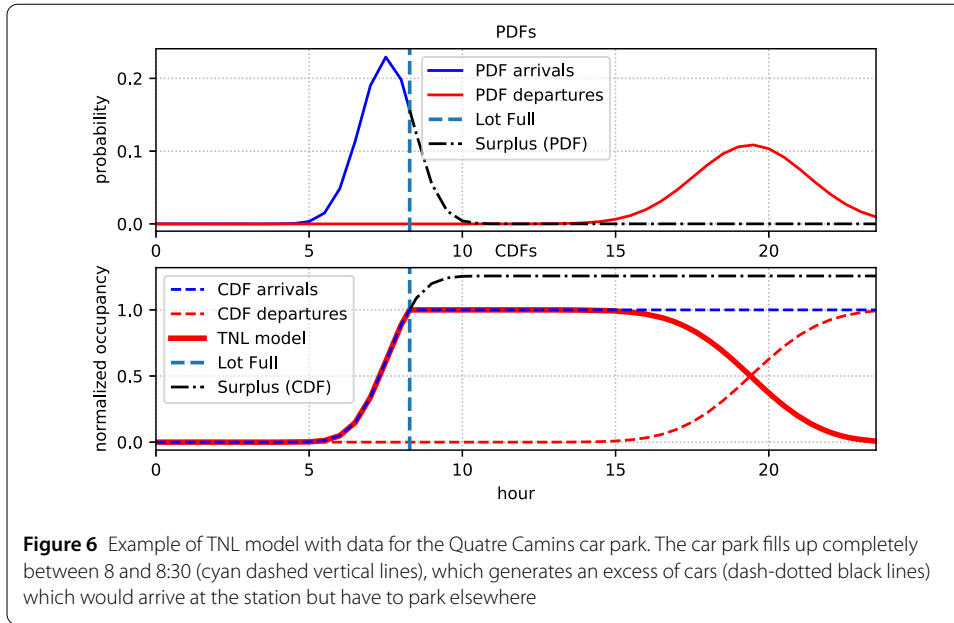
The top panel of Fig. 5 shows an example with a density function for arrivals (in blue) and departures (in red). The bottom panel shows the corresponding CDFs (dashed lines) together with their difference (red solid line) as the aggregate result of cars parked in the parking lot.

The TN model provides a continuous temporal profile that can be easily adjusted if the available data consists of aggregated occupancies. Additionally, it provides a clear interpretability of the underlying mechanisms shaping the occupancy profile. Despite its simplicity, we will show that the TN model achieves predictive performance comparable to or superior to more complex approaches.

4.2 Model with occupancy limit

We can extend the basic TN model to account for a capacity limit in the parking lot by incorporating a new parameter into the model. Consider the total number of cars arriving in one day, and let τ be the fraction of those cars that fit in the parking lot. We define the *Truncated Normal with Limit (TNL) model* as

$$f_{\text{TNL}}(t; \theta) = \bar{\Phi}_a(t; \mu_a, \sigma_a, \tau) - \Phi_d(t; \mu_d, \sigma_d),$$



$$\text{with } \bar{\Phi}_a(t; \mu_a, \sigma_a, \tau) = \min \left(\frac{\Phi_a(t; \mu_a, \sigma_a)}{\tau}, 1 \right) \quad (5)$$

where $\bar{\Phi}_a(\cdot)$ is a truncated and re-scaled version of $\Phi_a(\cdot)$ accounting for the fraction τ only. In other words $\bar{\Phi}_a(\cdot)$ uses only the part of $\Phi_a(\cdot)$ which is smaller than τ .

The TNL model allows not only to characterise the moment t_L when the capacity limit is reached (i.e. $\Phi_a(t_L; \mu_a, \sigma_a) = \tau$), but also to have an estimate of the number of cars which did not find a free parking space. This is very relevant in scenarios where informed decision-making is required to determine policies to resize crowded parking lots.

Figure 6 illustrates the TNL model for a day when the parking lot capacity limit was reached between 8 and 8:30 (vertical dashed line in cyan). The top panel shows an example of a density function for arrivals (in blue) and departures (in red). The arrival distribution is truncated when the parking fills up. The dashed-dotted black line of the arrival distribution corresponds to the estimated density of cars arriving afterwards, which no longer fit in the parking lot. The bottom panel shows the corresponding modified CDFs $\bar{\Phi}_a$ (dashed line in blue) and Φ_d (dashed line in red) together with the difference $\bar{\Phi}_a - \Phi_d$ (red solid line), as the aggregate result of cars parked in the lot. Again, the proportion of surplus cars is indicated by the dashed-dotted line.

We will allow a different value for τ for every day d of training data while using the same values for other parameters of the two distributions. This way the parameters of the arrival distribution are independent of the actual moment a parking lot reaches its capacity limit and can be used, not only to predict the moment when the capacity limit will be reached, but also to have an estimate of the number of cars which did not find a free parking space, as we will show later in Sect. 5.3.

4.3 Optimising model parameters from observed data

In this section, we describe the procedure for estimating the model parameters using observed car park occupancy data.

4.3.1 Fitting the TN model

For the basic TN model, the objective is to estimate the parameters $\theta = (\mu_a, \mu_d, \sigma_a, \sigma_d)$ from a dataset consisting of N daily occupancy profiles, $\mathcal{D} = \{o_{1:T}\}_{n=1}^N$, recorded at T discrete timestamps.³ To ensure consistency, all occupancy profiles are normalised such that $\sum_{i=1}^T o_i = 1$.

If occupancy profiles exhibit stability within a group, meaning there is minimal variation across weekdays or weekends, we can reasonably assume that occupancy at time t is independently and identically distributed. For simplicity, we model occupancies as Gaussian-distributed with mean $f_\theta(t)$ and variance β^2 . The likelihood of the parameters θ is given by

$$p(o_t^{(n)} | t, \theta, \beta) = \frac{1}{\beta\sqrt{2\pi}} \exp\left(-\frac{1}{2\beta^2} \left(o_t^{(n)} - f_\theta(t)\right)^2\right) \quad (6)$$

$$p(\mathcal{D} | \theta, \beta) = \prod_{n=1}^N \prod_{t=1}^T p(o_t^{(n)} | t, \theta, \beta). \quad (7)$$

The optimisation can be carried out by minimising the sum of squares loss

$$\mathcal{L}^* = \min \sum_{n=1}^N \sum_{t=1}^T \left(o_t^{(n)} - f_\theta(t)\right)^2, \quad (8)$$

and the variance is given by $\beta^2 = \mathcal{L}^* / NT$.

4.3.2 Fitting the TNL model

For the TNL model with occupancy limit, the set of fitted parameters θ also includes a set of threshold parameters $\{\tau_i\}_{i=1}^N$. Thus, we allow τ to take a different value for each training day i while keeping the other parameters fixed. This approach ensures that the parameters of the arrival distribution remain independent of the exact moment a parking lot reaches capacity. As a result, the model can not only predict when the capacity limit will be reached but also estimate the number of cars that were unable to find a free parking space. Note, however, that this approach requires a different normalisation of the data. In this case, all occupancy profiles are scaled such that $\max(o_{1:T}) = 1$. This normalisation is essential for identifying the moments when the parking lots reach full capacity.

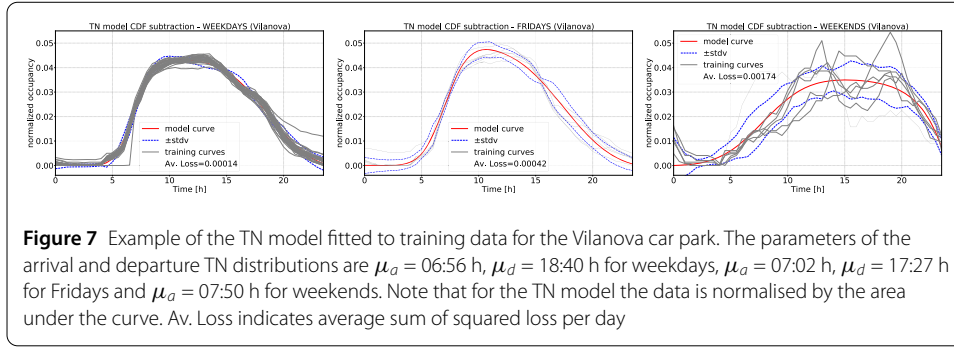
4.3.3 Test-training split

To train the model, we split the available data into training and test sets after the data-cleaning process. The test set is fixed at three weeks, while the remaining data is used for training. For most stations, this results in an approximate 30/70 test-training split.

4.4 Alternative approaches

In this section, we briefly describe two alternative methods against which we compare our proposed modelling approaches.

³In our case, $T = 48$ since the dataset has a 30-minute time resolution.



4.4.1 Average activity profile

A straightforward method for predicting patterns influenced by circadian rhythms is the use of average activity profiles, as employed in [16]. We adapt this approach by computing the average over normalised, aggregated data, distinguishing between three groups: weekdays (Monday to Thursday), Fridays, and weekends (Saturday and Sunday).

4.4.2 Linear regression

As another baseline, we use a linear regression model to predict car park occupancy based on prior activity. Formally, given a dataset of N normalized daily occupancy profiles, $\mathcal{D} = \{o_{1:T}\}_{n=1}^N$, we aim to predict o_y using past observations $o_{1:x}$, where $x < y$. Since occupancy profiles represent cumulative arrivals and departures, we apply the `diff` function to obtain changes in occupancy at each 30-minute interval. Thus, we define the transformed input as $\hat{o}_{1:x} = \{0, \text{diff}(o_{1:x})\}$.

The regression model optimises the following objective:

$$\mathcal{L}_{\text{lr}}^* = \min \sum_{n=1}^N \left(\sum_{t=1}^x \beta_t \hat{o}_t^{(n)} + \beta_0 - o_y^{(n)} \right)^2. \quad (9)$$

5 Results and discussion

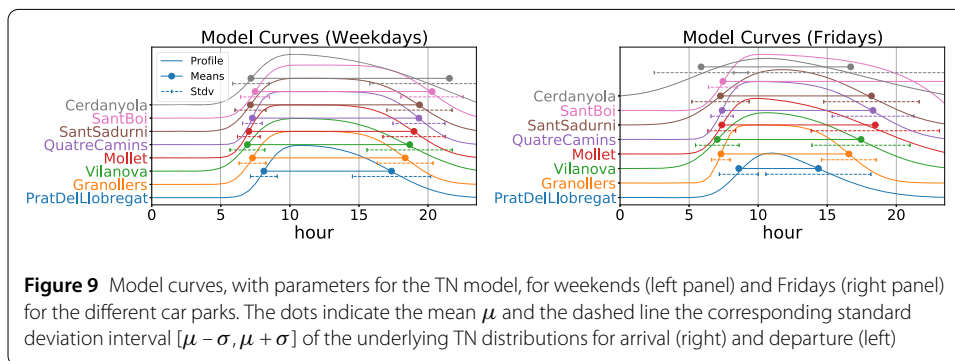
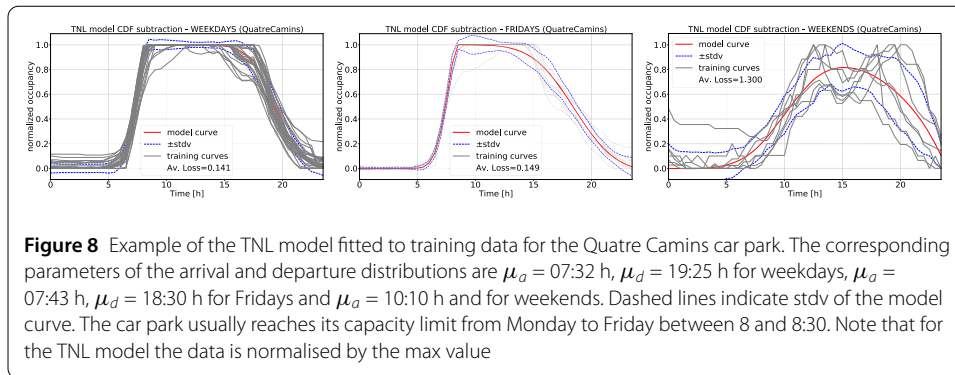
In this section, we first evaluate how well the model fits the data, followed by an analysis of its effectiveness in predicting or nowcasting parking lot occupancy at a given moment.

5.1 Model fit quality

We begin by assessing how well the model represents both the training and test data. For this, we use the mean relative proportional error, computed relative to the maximum occupancy of a parking lot. For clarity, we group the test data into three categories based on observed activity patterns (as illustrated in Fig. 3): weekdays (Monday to Thursday), Fridays, and weekends (Saturday and Sunday). The goodness of fit of the models in terms of the average loss per training day is given in Tables 5 and 6 in Appendix B.

5.1.1 Example fits on training data

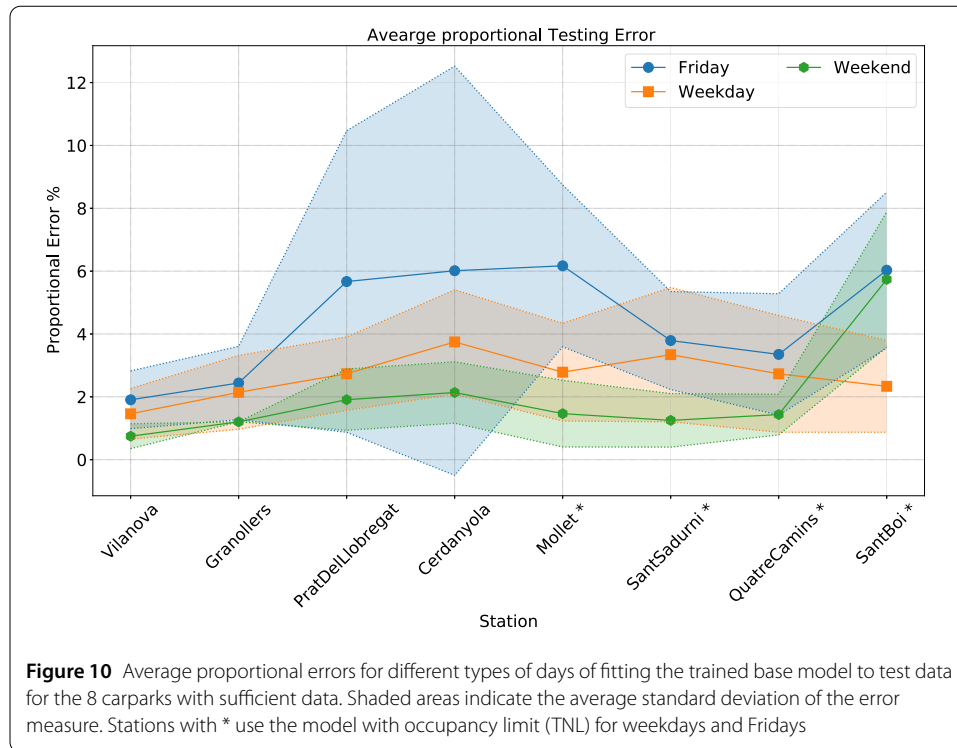
We first illustrate model fits on the training data using two example car parks. Figure 7 shows the extracted TN model curves for the Vilanova car park, while Fig. 8 shows the TNL model curves, incorporating an occupancy limit, for the Quatre Camins car park. The latter reaches maximum capacity on most weekdays and Fridays (since for the weekend profile, this does not happen, the basic TN model is used there).



The dark grey lines represent actual occupancy profiles from the training data. The best-fitting model curves are shown as solid red lines, while blue dashed lines indicate the standard deviation of the model fit. All curves are normalised: in the TN model, by the area under the curve, and in the TNL model, by the maximum value. The latter normalisation is necessary to estimate excess demand in a given car park. For Fig. 8, the model curve is plotted using the average excess across all training curves. We observe the best fits for weekday profiles, while Fridays, especially weekends, exhibit greater variability.

The fitted TN model curves and their corresponding parameters are shown in Fig. 9. The figure presents model curves for weekdays (left panel) and Fridays (right panel) across all eight car parks, ordered by the difference between the location parameters μ of the arrival and departure distributions (represented as circle markers). The interval corresponding to $\pm$ one standard deviation σ around μ is indicated by dashed lines and horizontal markers for both the arrival and departure distributions. The exact values of the parameters are given in Appendix B in Table 3 for the TN model and in Table 4 for the TNL model.

A smaller time difference between arrivals and departures is observed on Fridays, primarily due to earlier departure times. This pattern aligns with the common practice of shorter work schedules on Fridays, often without lunch breaks. The most pronounced example is the Prat del Llobregat car park, where the average departure time shifts from approximately 17:20 to 14:20—a three-hour advancement. We exclude the corresponding curves for weekends, which are characterised by more extreme parameter values and greater variability.



5.1.2 Evaluating model performance on test data

We now analyse the model's performance by evaluating errors in the test data. Examples of two individual stations are given in Appendix A.

Taking the averages of these errors on the testing days and aggregating them into the three weekday groups, we obtain the errors depicted in Fig. 10. The four rightmost stations reach their occupancy limit during weekdays (including Fridays), and for these, we depict the error of the TNL model for the weekday and Friday profiles. The shaded areas represent the average daily standard deviation of the prediction errors.

We find that, apart from the four stations that show significant variability on Fridays (Prat Del Llobregat, Cerdanyola, Mollet, and Sant Boi, the latter also on weekends), the average error is always below 4%. The high error on weekends for the Sant Boi car park is caused by party-goers visiting nearby nightlife attractions. This results in a high-occupancy phase during the early morning hours, which is not well captured by the current models. To address this issue, future work could extend the model and use cycles that last longer than 24 hours on weekends and end when the car parks reach their minimum occupancy.

In general, the models perform better on weekdays from Monday to Thursday, which exhibit more regular activity compared to Fridays. The error is nearly always lower for the weekend profile, but this is due to the lower activity on these days relative to the stations' maximum occupancy.

5.2 Prediction

Next, we analyse the predictive quality of our model. More precisely, at a given moment in time, we aim to predict the evolution of occupancy in the subsequent hours of the day. Again, we measure prediction quality using the relative absolute error with respect to the

maximum occupancy of a car park and aggregate weekdays into three groups. We also compare the results with average profiles, which represent the typical activity in a given car park for the corresponding types of days in the training data.

We make predictions by adjusting the model and the average profile to match the observed occupancy at the initial time. To achieve this, we simply shift and rescale the corresponding curves to best fit the known data. Formally, given an occupancy profile $o_{1:T}$ at a certain hour of the day $h < T$, we determine the parameters β_0 and β_1 that minimize

$$\mathcal{L}^* = \min \sum_{t=1}^h (o_t - \beta_0 - \beta_1 f(t))^2, \quad (10)$$

where $f(t)$ represents either a model or the average activity profile, and β_0 corresponds to the number of cars parked at the station at the beginning of the day.

In the case of the TNL model with an occupancy limit, only the arrival distribution Φ_a is used for fitting. This allows us to calculate the excess number of cars, resulting in the following procedure:

$$\mathcal{L}^* = \min \sum_{t=1}^{\min(h, t_m)} (o_t - \beta_0 - \beta_1 \Phi_a(t; \mu_a, \sigma_a))^2, \quad (11)$$

where t_m is the time slot at which o_t first reaches its peak value:

$$t_m = \min\{t \mid o_t = \max(o_{1:h})\}$$

In other words, we only use the phase up until the car park reaches its maximum capacity for fitting.

To determine the fraction τ_i of the arrival distribution that fits within the car park, we solve:

$$\tau_i = \max\{\Phi_a(t; \mu_a, \sigma_a) \mid \beta_0 + \beta_1 \Phi_a(t; \mu_a, \sigma_a, \tau_i) \leq \max(o_{1:T})\}. \quad (12)$$

This can be simplified to:

$$\Phi_a(t; \mu_a, \sigma_a) \leq \frac{\max(o_{1:T}) - \beta_0}{\beta_1} \quad (13)$$

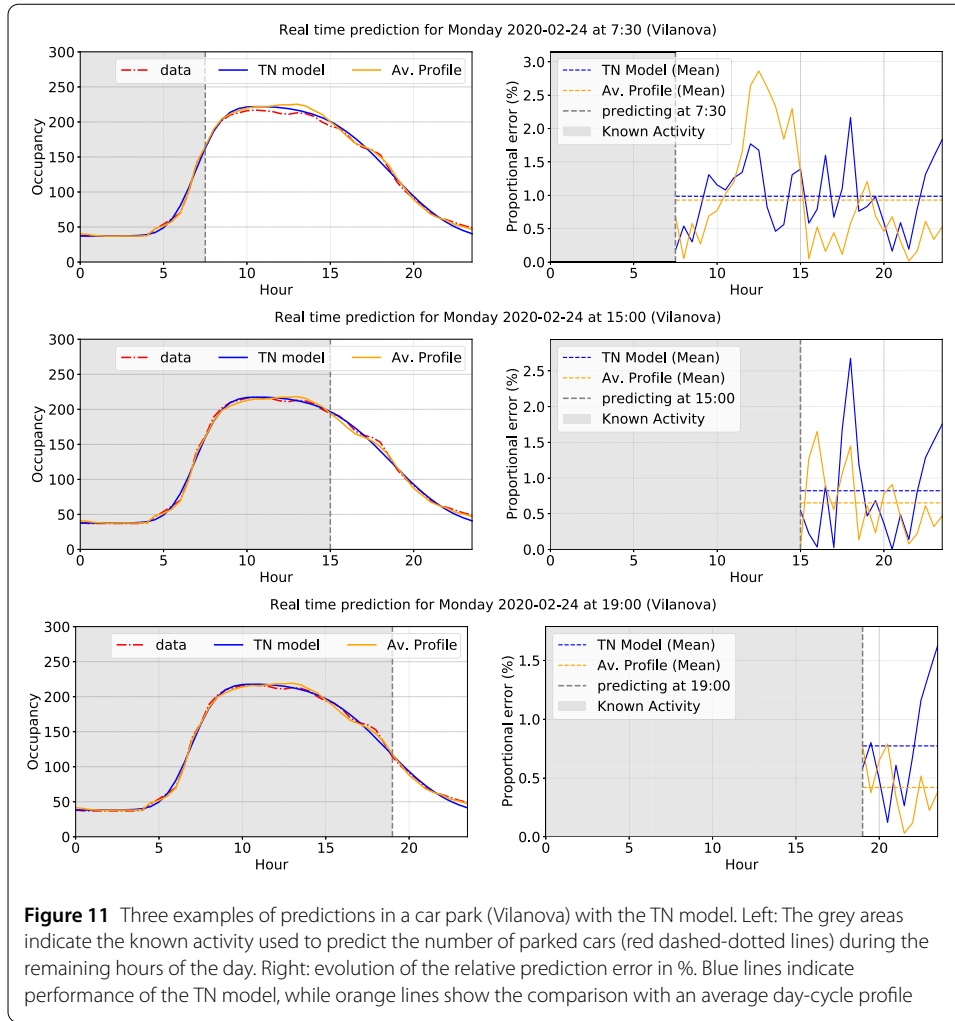
which leads to:

$$\tau_i = \frac{\max(o_{1:T}) - \beta_0}{\beta_1}. \quad (14)$$

The departure distribution $\Phi_d(t; \mu_d, \sigma_d)$ is then rescaled as follows:

$$\bar{\Phi}_d(t; \mu_d, \sigma_d) = \Phi_d(t; \mu_d, \sigma_d) \cdot (\max(o_{1:T}) - \beta_0). \quad (15)$$

Essentially, this ensures that the departure distribution returns to a baseline of β_0 cars at the end of the day.



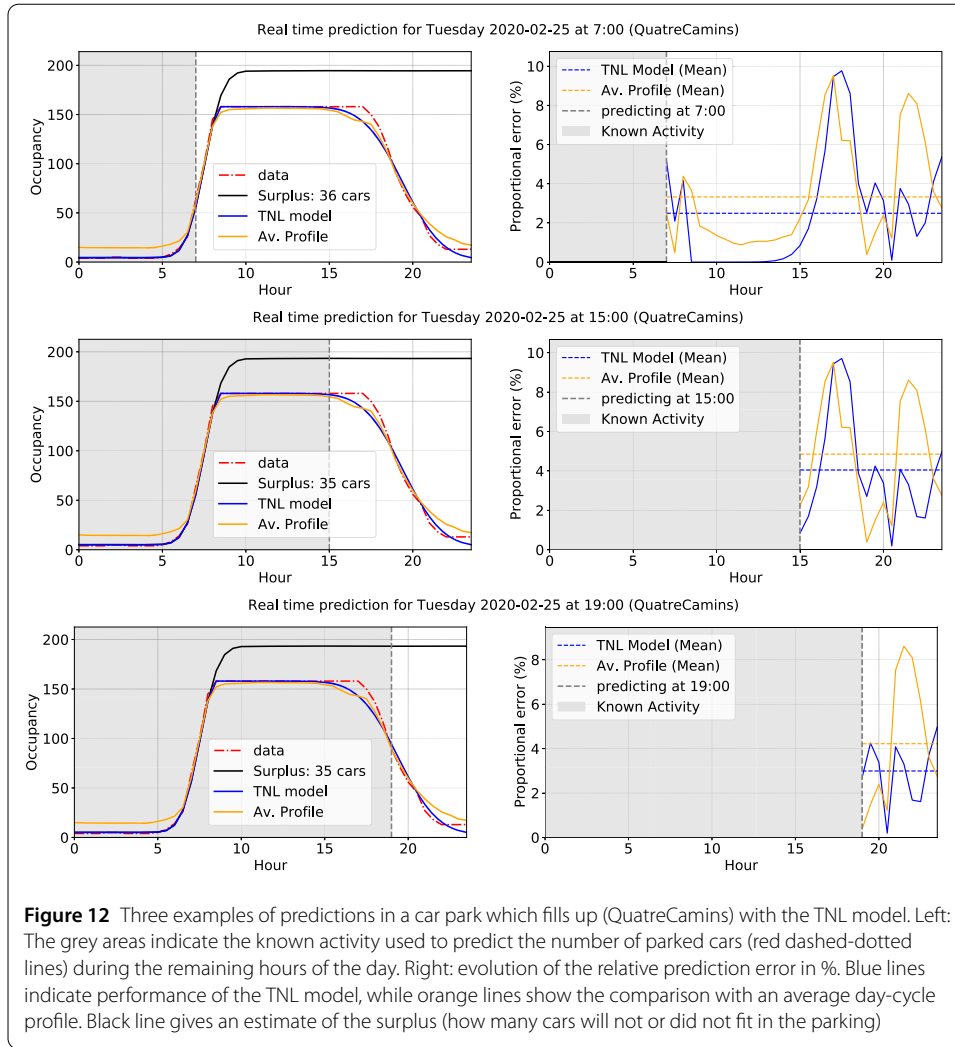
The excess, i.e., the number of cars that the arrival distribution predicts will arrive at the parking lot after it reaches full capacity, can be calculated as

$$\text{Excess}(T) = \beta_1 \cdot \max(\Phi_a(t; \mu_a, \sigma_a)) + \beta_0 - \max(o_{1:T}) = \beta_1 + \beta_0 - \max(o_{1:T}). \quad (16)$$

Figures 11 and 12 provide examples of predictions performed at 7:00, 15:00, and 19:00 for the Vilanova (TN model) and Quatre Camins (TNL model) car parks.

The left panels display real data (dash-dotted lines) along with the predicted occupancy, while the right panels show the corresponding prediction errors for both the model curve (in blue) and the average profile (in orange). The average errors are depicted as dashed lines. The dark grey area represents the data used to generate the prediction.

Figure 12 also includes an estimate of the surplus of cars that could not fit in the car park, indicated by the black continuous line. We observe that the performance of the TNL model is slightly better than that of the average profile for the Quatre Camins car park, and the surplus prediction remains stable across the three depicted time points.



5.3 Nowcasting

To further evaluate the model's performance, we apply it to a nowcasting task, i.e., we assess how well it predicts the number of cars in the car parks one hour ahead.

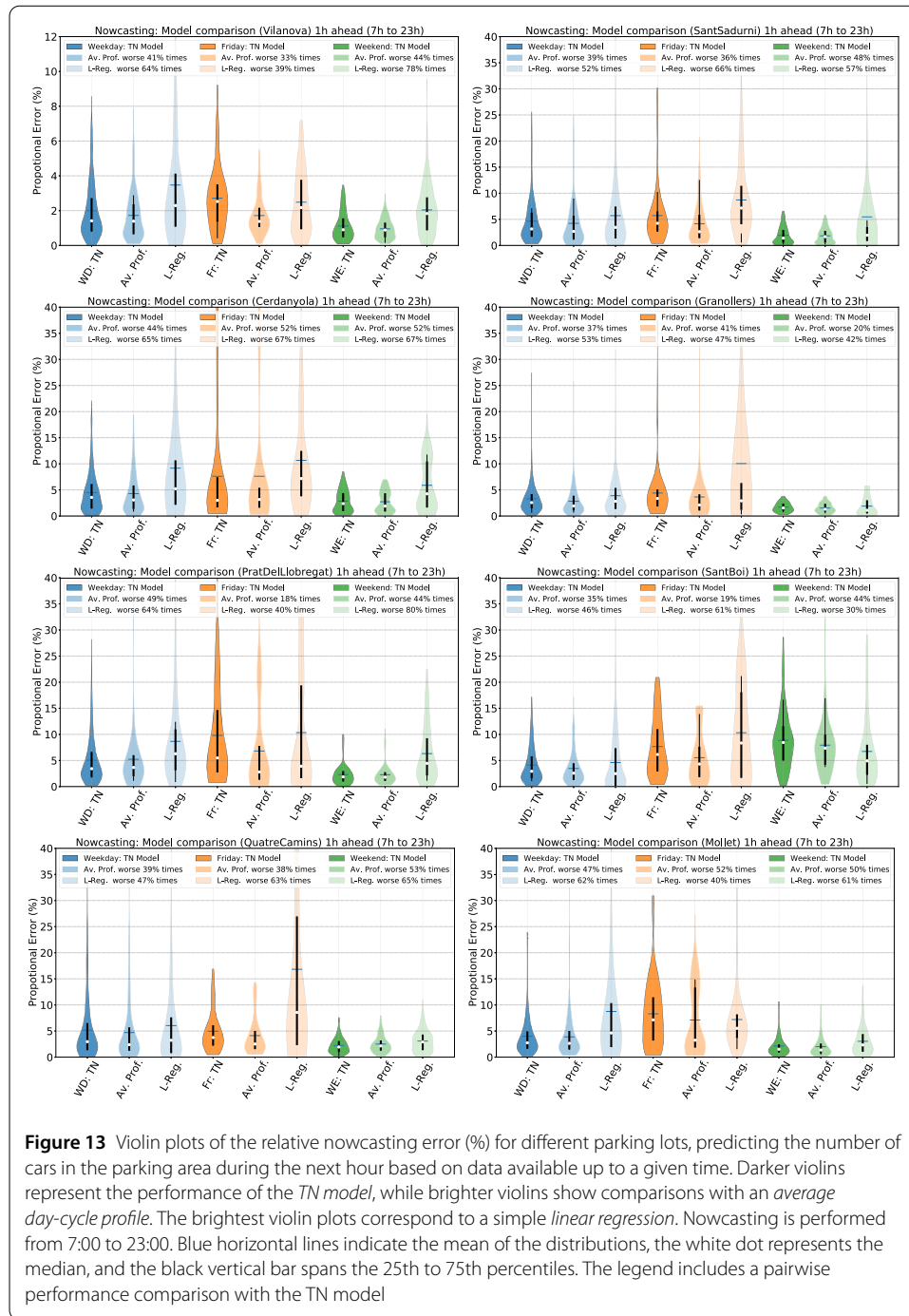
We quantify this by measuring the following prediction error, using the parameters obtained in Eq. (10):

$$\mathcal{E} = \sum_{t=h}^{h+w} \frac{|o_t - \beta_0 - \beta_1 f(t)|}{w \cdot \max(o_{1:T})}, \quad (17)$$

where $w = 2$ (corresponding to a prediction window of 1 hour with a 30-minute data resolution).

Figure 13 shows violin plots of the distribution of nowcasting errors, aggregated by car park and activity profile. The figure compares prediction errors between the basic model (darker colours) and the average activity profile (brighter colours).

Both approaches exhibit similar performance, with the average profile performing slightly better on average. As before, performance is worst on Fridays (orange tones), improves on weekdays from Monday to Thursday (blue tones), and is best on weekends



(green tones), although the latter is likely caused by the lower number of parked cars on weekends in general.

Nowcasting is performed from 7:00 to 23:00 in 30-minute intervals, with predictions made one hour ahead.

Figure 14 compares the performance of the model with activity thresholds for the four stations that reach their capacity limit. In this case, we display nowcasting errors only for the period between 7:00 and 15:00 (with 30-minute intervals). After 15:00, car parks begin to empty, reducing the advantage of using the threshold-based model.

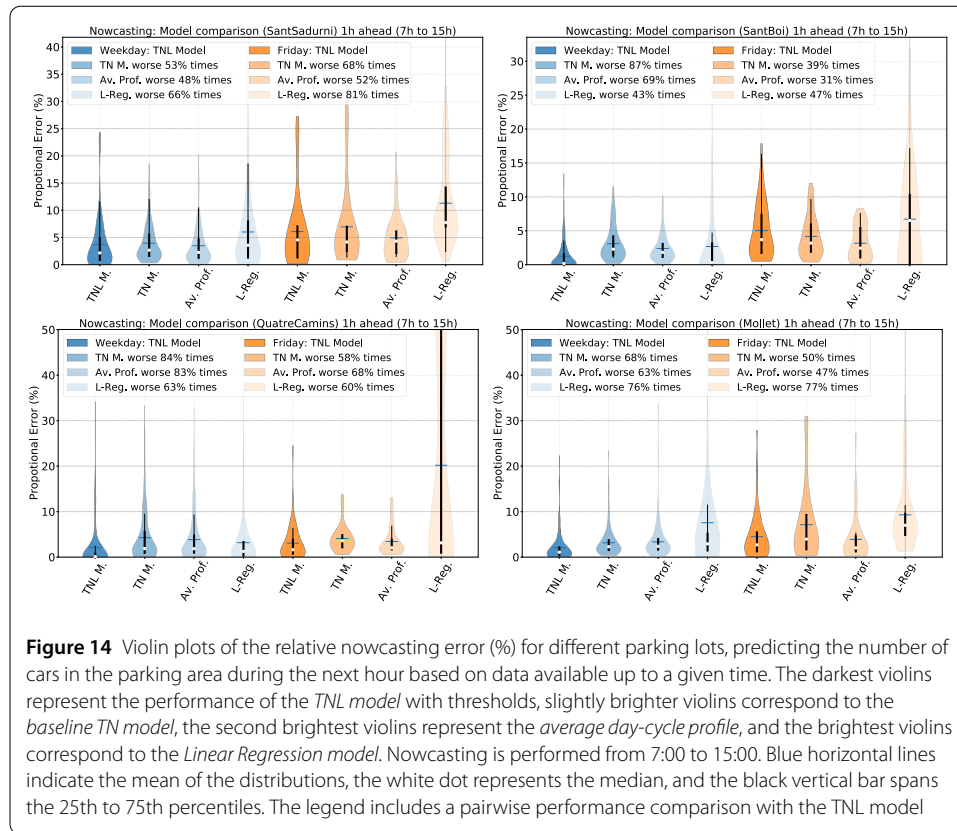


Table 2 Median nowcasting error in % for the four stations depicted in Fig. 14 for weekdays and Fridays comparing the Truncated Normal with Limit (TNL), basic (TN), and Average Profile (Av. Prof.) model with a Linear Regression (L-Reg.). Nowcasting is performed between 7:00 and 15:00 in 30-minute intervals

Car park	Weekdays				Fridays			
	TNL	TN	Av. Prof.	L-Reg.	TNL	TN	Av. Prof.	L-Reg.
Sant Sadurni	2.16	2.67	2.26	3.59	4.55	4.17	4.33	7.74
Sant Boi	0.18	2.29	1.82	0.31	3.69	3.19	2.44	6.60
Quatre Camins	0.08	1.88	1.84	1.26	1.62	3.60	2.04	3.14
Mollet	1.09	2.29	2.38	2.92	2.70	3.96	2.03	7.01

Violin plots with intermediate colour tones represent the models with thresholds, which consistently outperform the other models on weekdays across all three stations. However, the improvement on Fridays is less pronounced, with only a slight advantage observed for the Quatre Camins car park.

The figures also include a pairwise comparison of prediction improvement (shown in the legend), where we count how often the TNL model outperforms the other two models in each nowcasting instance. For example, on weekdays at the Quatre Camins station, the TNL model performs better than the baseline model in 81% of cases and surpasses the average profile model in 79% of the tested instances.

For better comparison, Table 2 presents the numerical values of the corresponding median prediction errors.

We observe that the Linear Regression Model generally exhibits the highest error values, particularly on Fridays, suggesting that it may not effectively capture key behavioural

patterns. In contrast, the TNL Model demonstrates superior performance in predicting parking occupancy, especially on weekdays, where it consistently yields lower error values than the other models.

Notably, it performs exceptionally well for the Sant Boi and Mollet car parks, achieving median prediction errors as low as 0.18% and 0.08%, respectively, possibly due to more regular commuter patterns.

The Basic Model, in comparison, generally performs worse than the TNL Model, often showing significantly higher errors. For example, at Sant Boi, its weekday error (2.29) is more than ten times higher than that of the TNL Model (0.18). This highlights the importance of considering capacity limits in the modelling approach for improved accuracy.

Errors tend to increase across all models on Fridays, suggesting greater variability in parking behaviour. Additionally, since the difference between the TNL and Basic Model becomes less pronounced, this may indicate that, due to lower demand, the capacity limit of car parks is less frequently reached on Fridays in most stations.

6 Policy recommendations

Based on the findings of this study, we propose the following policy recommendations for transportation planners and municipal authorities involved in managing park-and-ride infrastructure:

- *Use simple, explainable models for decision support:* The proposed Truncated Normal (TN) and Truncated Normal with Limit (TNL) models provide accurate occupancy predictions using a minimal set of interpretable parameters. This simplicity facilitates communication with non-technical stakeholders and supports transparent, evidence-based decision-making.
- *Monitor and address unmet demand:* The TNL model explicitly estimates the number of vehicles unable to find parking during peak times. Authorities can use this to identify high-demand stations and prioritise infrastructure investments (e.g., expanding parking capacity or improving alternative access) where overflow is consistently observed.
- *Incorporate behavioural insights in planning:* Because the model parameters correspond to actual commuter behaviour (e.g., average arrival and departure times), planners can use these insights to design time-specific interventions (such as staggered work hours, variable pricing, or promotional campaigns encouraging off-peak usage) to better balance demand.
- *Enable low-cost scalability using aggregate data:* Our approach requires only aggregate occupancy data, which is often easier and cheaper to collect than individual vehicle tracking. This enables scalable and privacy-preserving deployment across a broad set of car parks without major investments in infrastructure.
- *Prioritise regularly saturated Facilities:* As shown in our results (e.g., Fig. 4), some car parks consistently reach their capacity limits. We recommend incorporating our model outputs into multi-criteria prioritisation frameworks to guide expansions, especially in facilities critical for multimodal commuting corridors.

By aligning parking infrastructure planning with behavioural patterns and unmet demand forecasting, authorities could improve efficiency, user satisfaction, and the overall effectiveness of park-and-ride systems.

7 Conclusion

We have introduced a simple yet effective model for predicting car park occupancy in park-and-ride facilities. The model leverages truncated normal distributions to represent car arrival and departure times. This approach enables efficient and accurate parking demand prediction and estimation using only a small set of parameters.

A key advantage of the model is that it relies solely on aggregate data, eliminating the need to monitor individual arrivals and departures. Validation using data from Barcelona's metropolitan area demonstrates the model's robustness, even when applied to heterogeneous datasets. Its effectiveness is evident in both prediction and nowcasting tasks.

Furthermore, an extended version of the model, Truncated Normal with Limit (TNL), can estimate the additional parking spaces required to accommodate demand that remains unmet when car parks reach capacity. The extended model explicitly accounts for thresholding behaviour in station overflow, allowing it to determine the moment at which the capacity limit is reached and estimate the number of vehicles unable to find a free parking space.

This capability is particularly relevant for making informed decisions regarding the expansion of crowded parking lots. This estimation of unmet demand is a particular novelty of our study. To our knowledge, previous research has not directly addressed this issue, despite the fact that parking shortages rank among the top three reasons for not using Park-and-Ride facilities [18].

Additionally, our work differentiates itself from related studies through its specific focus on the modelling approach. Unlike some studies that prioritise prediction using techniques without an underlying behavioural model [3–5], our approach is more ambitious, aiming to understand the fundamental behavioural patterns of commuters rather than merely forecasting occupancy trends.

Moreover, the use of truncated normal distributions provides two key advantages over more complex, overparametrized models, such as neural networks [4]. First, it relies on a minimal number of parameters, ensuring computational efficiency. Second, these parameters correspond to behavioural metrics that are inherently interpretable, offering valuable insights into commuter patterns.

In contrast with the prevailing trend of requiring large-scale, fine-grained, high-precision datasets [2], our approach demonstrates that accurate predictions can be achieved using aggregated and limited data.

However, we also have to acknowledge two limitations of our study. First, our model assumes that arrival and departure times are independent random processes. This simplification enables tractability and parameter identifiability using aggregate data. While it may not fully capture dependencies present in individual commuter behaviour (e.g., fixed work shifts), the strong regularity observed in occupancy patterns suggests that, at the aggregate level, this assumption remains a reasonable approximation. Future work could explore coupled distributions or joint modelling frameworks if individual-level data becomes available. Secondly, the model is tailored to regular commuting behaviour and may be less suitable in non-routine contexts such as shopping centres or event-driven demand, where more flexible or hybrid models may be required.

Nonetheless, and ultimately, we believe this work will contribute to improved urban planning by offering a practical tool for understanding and predicting commuter behaviour in park-and-ride systems. By balancing simplicity, interpretability, and predictive

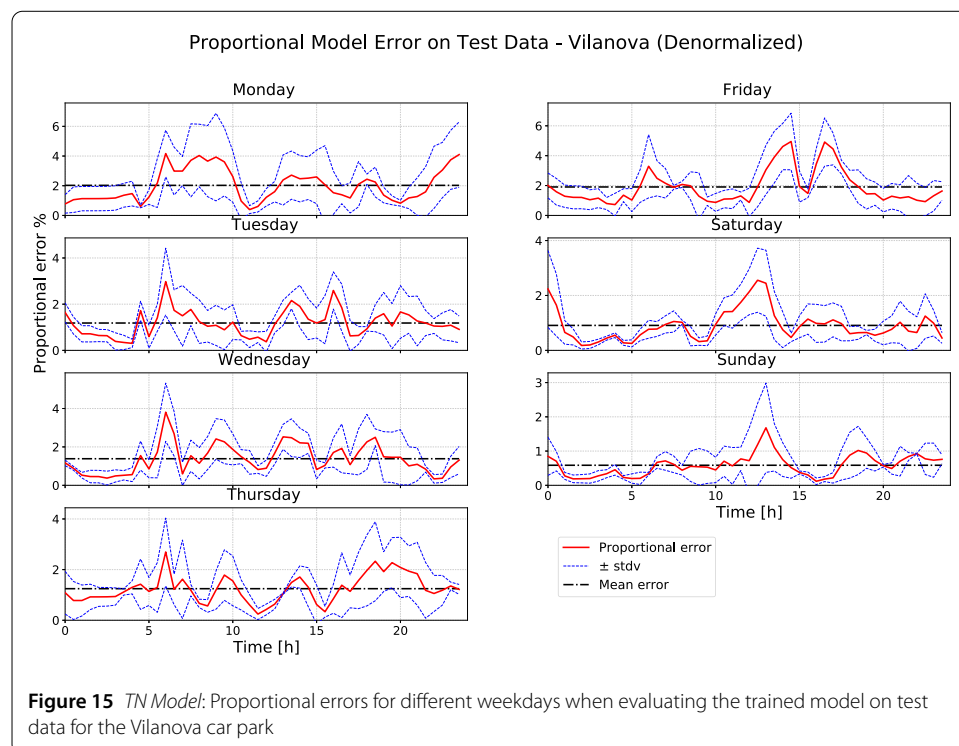
accuracy, our approach can support data-driven decision-making for optimising parking infrastructure and enhancing sustainable mobility solutions.

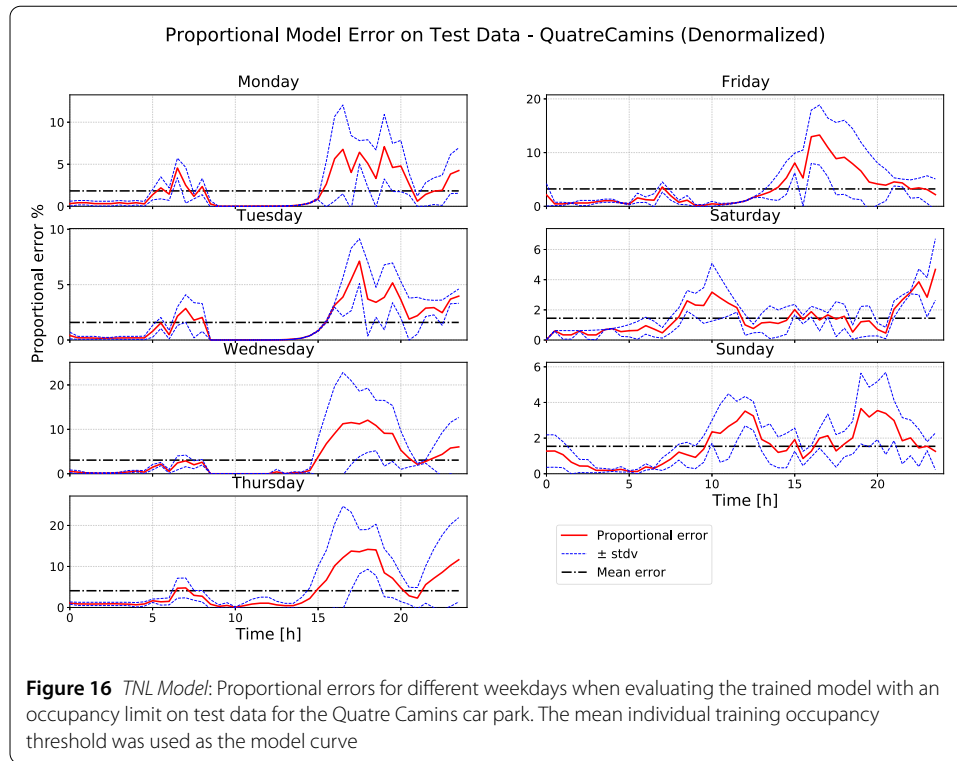
Appendix A: Error analysis for two representative car parks

In this section, we present the relative errors of the two models for two example car parks.

For the TN model, we examine the relative errors depicted in Fig. 15 for the Vilanova car park, aggregated by days of the week and hours of the day. The figure displays errors separately for each day of the week, with the average error per hour represented by red continuous lines and the corresponding standard deviations indicated by dashed blue lines. The black dashed-dotted line represents the average error across all hours of the day.

Similarly, Fig. 16 provides an evaluation of the TNL model's performance for the Quatre Camins car park. We observe how this model results in large periods with very small errors but exhibits larger errors in the afternoon during the departure phase of the cars, which is not influenced by the addition of a threshold to the model.





Appendix B: Model parameters and loss

In this section, we display the values of the parameters obtained when fitting the training data and also the sum of squares loss of the corresponding optimisation processes.

Table 3 lists the parameters for the TN model for all car parks, and Table 4 lists the parameters of the TNL model for the four car parks which reach their capacity limits

Table 3 Parameters $\mu_a \pm \sigma_a$, $\mu_d \pm \sigma_d$ (in hh:mm units) of the TN Model for the different stations. * indicates cases where higher early morning occupancy during weekends (as explained in Sect. 5.1.2) causes non-interpretable parameter values

Car park	Weekdays		Fridays	
	Arrival	Departure	Arrival	Departure
PratDelLlobregat	08:07 ±00h58	17:21 ±02h50	08:36 ±01h25	14:22 ±03h48
Granollers	07:18 ±00h58	18:21 ±02h01	07:18 ±00h40	16:34 ±01h59
Vilanova	06:56 ±01h16	18:40 ±03h05	07:02 ±01h35	17:27 ±03h33
Mollet	07:02 ±00h49	18:59 ±02h15	07:22 ±01h00	18:29 ±04h39
QuatreCamins	07:17 ±00h43	19:20 ±01h53	07:23 ±00h48	18:19 ±02h57
SantSadurni	07:09 ±01h07	19:23 ±02h22	07:16 ±02h04	18:12 ±03h27
SantBoi	07:29 ±01h03	20:17 ±02h15	07:27 ±01h02	*6809:53 ±164h43
Cerdanyola	07:11 ±01h19	21:33 ±05h46	05:52 ±03h23	16:42 ±08h27
Weekends				
PratDelLlobregat	08:06 ±03h09	22:55 ±07h54		
Granollers	07:26 ±02h34	*10,786:37 ±249h47		
Vilanova	07:50 ±03h00	*248,917:52 ±730h15		
Mollet	08:34 ±02h46	22:39 ±02h13		
QuatreCamins	09:29 ±02h08	22:53 ±02h31		
SantSadurni	09:49 ±01h58	19:32 ±04h47		
SantBoi	*00:00 ±07h47	*33:11 ±02h09		
Cerdanyola	08:29 ±04h08	18:26 ±07h24		

Table 4 Parameters $\mu_a \pm \sigma_a$, $\mu_d \pm \sigma_d$ (in hh:mm units) and τ (in %) of the TNL Model for the different stations which reach the occupancy limits. * indicates a case where higher early morning occupancy during weekends (as explained in Sect. 5.1.2) causes non-interpretable parameter values

Car park	Weekdays			Fridays		
	Arrival	Departure	Avg. τ	Arrival	Departure	Avg. τ
Mollet	07:06 $\pm$ 00h52	19:00 $\pm$ 02h16	80.66%	07:20 $\pm$ 00h56	19:28 $\pm$ 04h07	72.48%
QuatreCamins	07:32 $\pm$ 00h52	19:25 $\pm$ 01h51	79.58%	07:43 $\pm$ 00h55	18:30 $\pm$ 02h50	73.08%
SantSadurni	07:20 $\pm$ 01h17	19:20 $\pm$ 02h30	76.90%	07:08 $\pm$ 01h47	18:17 $\pm$ 02h57	66.69%
SantBoi	08:00 $\pm$ 01h22	20:19 $\pm$ 02h07	75.75%	11:05 $\pm$ 02h13	*32:07 $\pm$ 07h27	*12.29%

Table 5 Average Loss per day of the TN Model for the different stations

Car park	Weekdays	Fridays	Weekends
PratDelLlobregat	0.00179	0.00638	0.00541
Granollers	0.00029	0.00246	0.00346
Vilanova	0.00014	0.00042	0.00174
Mollet	0.00058	0.00269	0.00215
QuatreCamins	0.00026	0.00031	0.00342
SantSadurni	0.00021	0.00089	0.00589
SantBoi	0.00025	0.00057	0.01242
Cerdanyola	0.00209	0.00474	0.01082

Table 6 Average Loss per day of the TNL Model for the different stations

Car park	Weekdays	Fridays
Mollet	0.331	1.072
QuatreCamins	0.141	0.149
SantSadurni	0.135	0.459
SantBoi	0.159	0.395

during weekdays and Fridays. The corresponding values of the average loss functions per day of training data are given in Tables 5 and 6. Note that the values of the average loss function are not directly comparable between the TN and the TNL model, as the two models perform different normalisation strategies.

Abbreviations

ARIMA, auto-regressive integrated moving average; ATM, Autoritat del Transport Metropolità; CDF, cumulative distribution function; PDF, probability density function; SVM, support vector machine; TN, Truncated Normal; TNL, Truncated Normal with Limit.

Acknowledgements

The authors would like to thank ATM Barcelona (Autoritat del Transport Metropolità) for providing the data.

Author contributions

AK and VG designed the research project and performed data analysis, model optimization and evaluation. JF and DM performed data analysis, model optimization and evaluation. All authors contributed to the writing and reviewed the manuscript.

Funding information

VG has received funding from "la Caixa" Foundation (ID 100010434), under the agreement LCF/PR/PR16/51110009.

Data availability

All the data used in the paper and the corresponding code are publicly available at: <https://github.com/aig-upf/car-park>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 14 March 2025 Accepted: 9 September 2025 Published online: 10 October 2025

References

1. Lin T, Rivano H, Le Mouél F (2017) A survey of smart parking solutions. *IEEE Trans Intell Transp Syst* 18(12):3229–3253
2. Xiao X, Peng Z, Lin Y, Jin Z, Shao W, Chen R, Cheng N, Mao G (2023) Parking prediction in smart cities: a survey. *IEEE Trans Intell Transp Syst* 24(10):10302–10326
3. Zheng Y, Rajasegarar S, Leckie C (2015) Parking availability prediction for sensor-enabled car parks in smart cities. In: 2015 IEEE tenth international conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pp 1–6
4. Camero A, Toutouh J, Stolfi DH, Alba E (2019) Evolutionary deep learning for car park occupancy prediction in smart cities. In: *Learning and intelligent optimization*. Springer, Cham, pp 386–401
5. Zhao Z, Zhang Y, Zhang Y (2020) A comparative study of parking occupancy prediction methods considering parking type and parking scale. *J Adv Transp*: 1–12
6. Awan FM, Saleem Y, Minerva R, Crespi N (2020) A comparative analysis of machine/deep learning models for parking space availability prediction. *Sensors* 20(1):322
7. Chen X (2014) Parking occupancy prediction and pattern analysis. Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep. CS229-2014
8. Ji Y, Tang D, Blythe P, Guo W, Wang W (2015) Short-term forecasting of available parking space using wavelet neural network model. *IET Intell Transp Syst* 9(2):202–209
9. Chawathe SS (2019) Using historical data to predict parking occupancy. In: 2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON), pp 0534–0540
10. Fokker ES, Koch T, Leeuwen M, Dugundji ER (2022) Short-term forecasting of off-street parking occupancy. *Transp Res Rec* 2676(1):637–654
11. Vakrinou K, Mantouka EG, Vlahogianni EI (2025) Leveraging multisource data for off-street parking occupancy prediction: a twofold analysis. *Int J Transp Sci Technol*
12. Tavaafoghi H, Poolla K, Varaiya P (2019) A queuing approach to parking: modeling, verification, and prediction. *ArXiv*. [arXiv:1908.11479](https://arxiv.org/abs/1908.11479)
13. Daniotti S, Monechi B, Ubaldi E (2023) A maximum entropy approach for the modelling of car-sharing parking dynamics. *Sci Rep* 13(1):2993
14. Schneble M, Kauermann G (2025) Statistical modelling of on-street parking spot occupancy in smart cities. *J R Stat Soc, Ser C, Appl Stat* 017
15. Kaltenbrunner A, Gómez V, Lopez V (2007) Description and prediction of slashdot activity. In: *Web conference, 2007. LA-WEB 2007. Latin American*. IEEE, pp 57–66
16. Szabo G, Huberman BA (2010) Predicting the popularity of online content. *Commun ACM* 53(8):80–88
17. Xue Y, Fan H, Guan H (2019) Commuter departure time choice considering parking space shortage and commuter's bounded rationality. *J Adv Transp* 2019
18. Zhao X, Li Y, Xia H (2017) Behavior decision model for park-and-ride facilities utilization. *Adv Mech Eng* 9
19. Generalitat de Catalunya (1999) PDU for public transport-private vehicle modal interchange car parks within the scope of the ATM integrated tariff system in the Barcelona area. <https://shorturl.at/fep5O>. Accessed 5-Mar-2025

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)