Master in Intelligent Interactive Systems
Universitat Pompeu Fabra

Predicting the Use of Car Parks in the Province of Barcelona

David Moreno Esteban

Supervisor: Vicenç Gómez

July 2021

Contents

Dedication

I would like to dedicate this work to every single person that has helped me to achieve this. The master's thesis, the master itself, the previous degree, everything.

Acknowledgement

I would like to express my sincere gratitude to:

- My supervisor, Dr. Vicenç Gómez, who has shown great interest in making the project a success, and has always been available to help me and discuss what I needed.

- My family, who has supported me, and has understood my lack of presence during the busiest weeks.

- My master mates. Each of them has taught me something and has made me learn and improve.

- Judit, who has given me strength and support from the distance.

- Last but not least, Luna, who has brighten my darkest days.

Abstract

Designing and generating wiser policies for urban systems and infrastructures is a challenge of paramount importance. Today, the cities that present the most successful transport strategies are prioritising the movement of people, giving residents and visitors a wider variety of attractive transport options while creating effective ways to switch from private to public transport means. Understanding the use and the impact of public infrastructures that facilitate mobility is crucial.

We consider a dataset of one year of activity in the form of car park occupancy in the province of Barcelona. The data comprises ten different parking facilities located close to train stations. We propose and analyze different and intuitive prediction models based on statistical and mathematical approximations.

First, we analyze the occupancy recordings in different parking locations and show that the activity is strongly coupled with the circadian rhythm, following a 24-hours cyclic pattern. Second, we implement a predictive model to provide the occupancy of a particular parking for an entire future day. We show that for both, statistical and mathematical approximations it performs quite accurately. Third, we implement a predictive model to guess the occupancy of the remaining hours of the day given the occupancy of the previous hours. Finally, a qualitative and quantitative analysis of the parking occupancy during the Covid-19 pandemic has been performed in order to understand how the global situation has influenced the parking usage.

Our results show that, despite the apparent complexity associated to public mobility and use of car parks, very simple models motivated in intuitive principles are sufficient to understand and predict this dynamics. Overall, our results can facilitate the design of public policies to facilitate the mobility within Barcelona and its surroundings, by providing a better understanding of how the citizens switch between private cars and public trains.

Keywords: Parking occupancy; Prediction methods; Temporal series; Covid-19 influence

Chapter 1

Introduction

For the last decades, most cities transport strategies have prioritised cars and traffic speed flows. This urban development model, based on motorized private traffic when considering the citizen mobility from big cities to its surroundings and the other way around, has had multiple social and environmental impacts, which have consequent costs in public economy, society and health [1]. Today, the cities that present the most successful transport strategies are prioritising the movement of people, giving residents and visitors a wide variety of attractive transport options while creating effective ways to switch from private to public transport means. Consequently these last years there has been a shared common movement in many European big cities to integrate a more socially responsible mobility, development and strategies [2].

Nowadays, every single day huge amounts of data are being generated and collected regarding this social mobility and their corresponding traffic flows. All this data can be used to generate better long-term strategies and adapt big cities to the needs of its citizens and not the other way around. For example, in the case of Barcelona, "la Autoritat de Transport Metropolità (ATM)"[1] has installed sensors in different parking locations to understand better their usage patterns and to draw a long-term strategy that allows to improve the interurban mobility from and to Barcelona. One of the key points on this conversion to a greener and more social mobility is allow-

[1]ATM website: `https://www.atm.cat/web/index.php`.

ing a bridge between private and public transportation means when considering the interurban mobility of territories. And here is where the infrastructure of public parking lots located within the Barcelona province plays a key role [2], being most of them related to several train stations and allowing people to park their cars and change directly there to the public transportation system, working as a bridge between both public transports and private transportation means.

Thus, being able to characterize the mobility of citizen within a city is a social phenomena of utter importance. Considering the public park lots located in the outskirt of Barcelona as an effective alternative to ensure this switch from private to public transportation, finding new methods to predict their usage and characterize the evolution of their occupancy in time would be a good tool to allow a better planning of the public transportation means. This is why generating a predictive model of this occupancy would be a challenging and ambitious project.

This master thesis has been developed in collaboration with the ATM who kindly provided the dataset of car park occupancy during the recorded period. The project has been jointly carried out with Josep Ferrer [3]. Both thesis share the same introduction and preliminary analysis, but they differ in different aspects and methods. While [3] is focused on the problem of *modeling* the occupancy behavior, this thesis is focused on the problem of *predicting* the data. For this reason, during the *Preliminary analysis* chapter I use *we*, and for the rest of the chapters I use *I*.

1.1 Motivation and Objectives

This thesis contains several goals to be achieved, most of them complementing each other.

1. **Understand the behavior of the occupancy**: A first direct goal is to understand how the parking occupancy evolves with time, being able to detect

[2] ATM private-public change strategy: `https://territori.gencat.cat/ca/06_territori_i_urbanisme/planejament_urbanistic/pla_director_urbanistic_pdu/en_curs/Catalunya/pdu_parks_and_ride/`.

differences between days such as holidays, weekdays and weekends. A second approach to understand how this occupancy changes depending on the parking that is being used is to be considered as well.

2. **Effective predicting model**: Train an effective model to predict the occupancy behaviour in time. Thus, generating a predictive model based on historic data means and precomputed temporal profiles, to analyze and extract the dynamics and behaviours of the parking occupancy.

3. **Detect external influences on the occupancy behaviour**: Try to detect possible factors that might influence the occupancy behaviour of these parking stations such as the Covid-19 pandemic.

4. **Detect errors in the stored data.**: Detecting errors that are present on data might be a good way to check the effectiveness of the model and the quality of the working data.

5. **Predict empty gaps**: Just in case some sensors stop working, predicting empty gaps using the model would be useful.

6. **Obtaining Real Time Data**: Generating a script that using the API provided by ATM is able to obtain real-time occupancy data from different parking spots.

1.2 Dataset

For this project we use several datasets coming from different parking spots. A total amount of 10 car parks have been used, each of them located in different cities within the province of Barcelona: Cerdanyola (140), Granollers(198), Martorell(139), Mollet (264), Prat del Llobregat (482), Quatrecamins(178), Sant Boi(394), Sant Quirze(410), Sant Sadurni (257) and Vilanova(488).

Their corresponding location can be observed in the following map:

Figure 1: Map of Barcelona province with the parking locations

Data is continuously gathered from physical sensors placed in the entrance and exit of the car parks, where an entrance of a vehicle increases the occupancy of the particular parking, and a departure of a vehicle decreases that counter. In this case, the location where the vehicle parks is irrelevant.

The original data provided by the ATM has the following fields:

- **Datetime:** The day and the hour the measurement took place. This is provided in a synchronously time scale starting at 00:00h and ending at 23.30h of each day, with a resolution of 30 minutes.

- **Free slots:** The number of free slots in the moment of the measurement. It contains decimals due to the 30 minute resolution, as can be observed in Figure 2. The entrance and the departure of vehicles in that period of time is averaged, resulting in decimal slot capacities. As we will see in the following chapter, we create a new field named "Occupancy", which is the difference between the maximal capacity of each parking and the specific number of Free slots for each interval, due to the fact that we want to measure the occupancy.

	DateTime	Parking Granollers Renfe plazas totales
0	06/01/2020 7:00	176
1	06/01/2020 7:30	176,6032144
2	06/01/2020 8:00	176,4983267
3	06/01/2020 8:30	176

Figure 2: Original Dataframe provided by ATM

The simplicity of the data and the fact that is quite new, together with the cyclic pattern of samples, are some of the strengths of the dataset and the project itself, as it eases the process of intuit trends and possible approximations. On the other hand, the lack of robustness in some sensors, as well as the Covid-19 pandemic comprises the weaknesses, as both factors reduce the training/testing sets.

1.3 Related Work

There exists an extensive literature concerning the prediction of free parking slots or occupancy in some specific parking locations. Some few examples can be easily found when checking some keywords such as *parking occupancy prediction* or *parking occupancy modelling* directly in google scholar. Hence, it is quite clear that parking occupancy pattern and free slot behaviour with time have been intensively studied these last years and present an increase trend as shown in Figure 3.

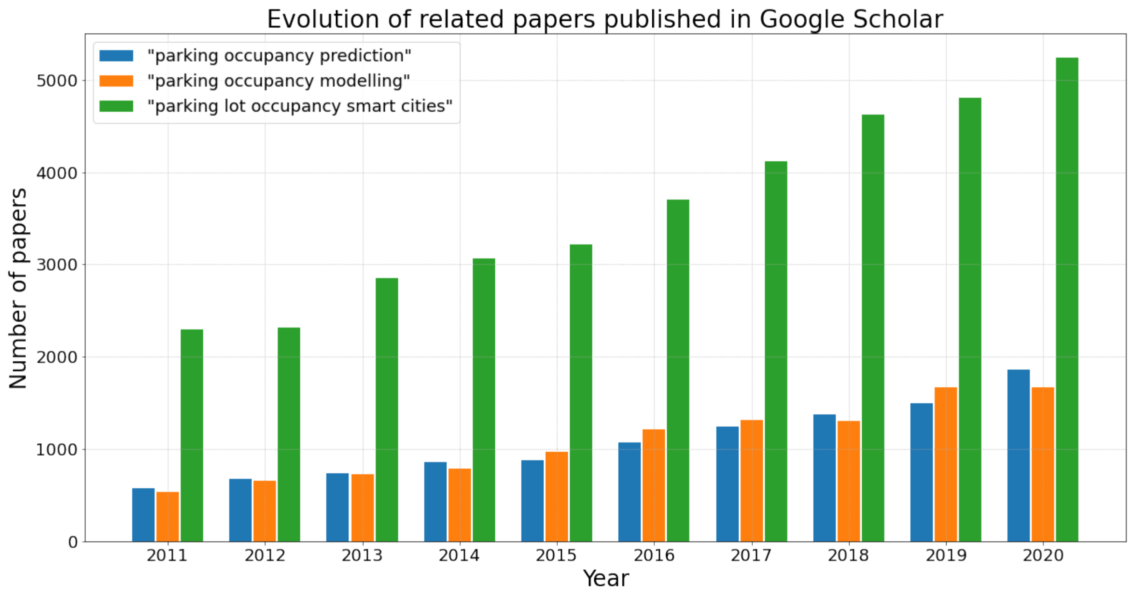

Figure 3: Evoluton of papers published

Several methods have been proposed to model such functions, from linear regression [4] to SVM [4] and different variations of neural networks [4, 5, 6], a wide range of diverse techniques have already been proven as efficient and accurate. However, the existing literature considers very different datasets and, to the best of our knowledge, different settings and tools than the ones we consider.

In our case, we want to demonstrate that using easy and simple methods are sufficient to characterize globally the occupancy behavior. Furthermore, using methods that can be intuitively understood is better to obtain some meaningful explanation of what is happening there. This is why, our two main approaches will focus on a first statistical approach based on prototype-like characterization of our data. Afterwards, a second mathematical approach using a mixture of gamma distributions will be performed as well to further simplify the model and obtain a more intuitive and meaningful model.

1.4 Structure of the Thesis

This thesis is structured as follows:

1. **Preliminary Analysis**

   A first preliminary analysis of the available data is performed to extract some first intuitions about the behaviour of this occupancy. The data s cleaned and filtered (to sort out outliers or mistaken data) to have a better dataset with which train and test the model.

2. **Prediction of parking occupancy.**

   Prediction methodology for both, offline and online data. Algorithm explanation is given for each method, as well as the error measurements.

   - **Approximation of the parking occupancy**: Summary of the methodology chapter provided in a parallel collaborative thesis.
   - **Prediction of an entire day of occupancy**: How to predict a full future day.
   - **Real time prediction**: How to predict the remaining hours of the day using given data about the occupancy of a part of the day.
   - **API connection**: HTTP requests to the original API where we can access in real time to the status of the different car parks.

3. **Covid-19 influence on parking occupancy**

   This chapter aims to provide an understanding on how the Covid-19 has influenced the way the car parks are used.

4. **Conclusions and future work**

   Summary of the results achieved and its possible impact is provided, as well as a path to follow for future improvements and applications of this study.

Chapter 2

Preliminary Analysis

In this section, we perform a preliminary analysis of the data provided to us by the ATM for each different car parking spot. In this analysis, we aim to understand how the occupancy varies over time, and characterize the statistical properties (mean and variance) of each parking. Also, it is interesting to understand the behaviour of the occupancy along the days, in order to extract some hypothesis, and to be able to identify different profiles or prototypes. A more extensive analysis is not possible due to the limited amount of data, as we are highly constrained by the Covid-19 pandemic, which affected dramatically to the data from March 13th of 2020 until now.

It is important to mention that, prior to the first visualizations of means and global behaviours, a preliminary cleaning of the data is carried out, removing from the original dataset wrong data that might bias or affect a little bit the future predicting model. We understand as a wrong data some days with huge lack of data, or some days where the sensor of a particular parking has not worked properly. To work with the available data, both the influence of holiday days and the Covid-19 pandemic days are considered as well, omitting these days from the study and comparing them later with some obtained behavior, as it is explained later on this same chapter.

2.1 Data Description

Our available data comprises the period between January 2020 and mid-March 2020. The working data is discretized with a 30 minutes resolution, having then 48 samples per day. For a practical understanding, this is quite useful, as when working for instance with Granollers, the resolution and intervals are exactly the same as when working with Quatrecamins. Therefore, all park stations follow the same pattern, making it way easier to compare and generalize the observations to any other parking, which eases the visualization and the understanding process.

All procedures and computations shown throughout this thesis are implemented in all parking datasets. For logical coherence and a better comprehension of the reader, we only display results for the Granollers and the Sant Sadurní car parks. In some specific cases other parking can be used to illustrate some meaningful examples. However, all results and conclusions can be generalized to all available parking datasets.

2.2 Preliminary Data Analysis

We start analyzing the average occupancy of each week day for the three first months of 2020. Figure 4 shows this mean occupancy. We observe that occupancy is high and stable form Monday to Thursday, and then decreases during the weekends, suggesting two types of activity profiles, or prototypes.

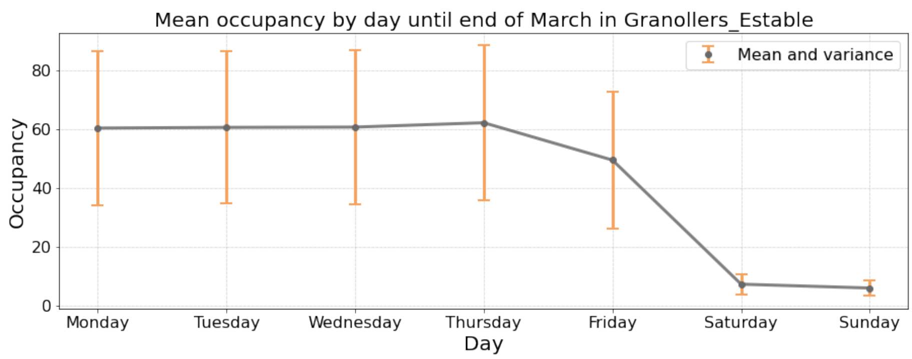

Figure 4: Mean occupancy for each weekday from January to March of 2020 in the parking of Granollers.

However, when analysing it in more detail, it can be observed that from Tuesday to Thursday the mean occupancy remains almost the same. Mondays tend to have a lower occupancy, but still close to them. All these four days do present big variances. Weekends have a low occupancy and present a small variance. When focusing on Fridays, it is noticeable that present a quite particular behavior: It is a mid-step between the weekdays and weekends. Moving forward, analysing the data on a minor scale, we visualized the evolution of the occupancy of each parking along time.

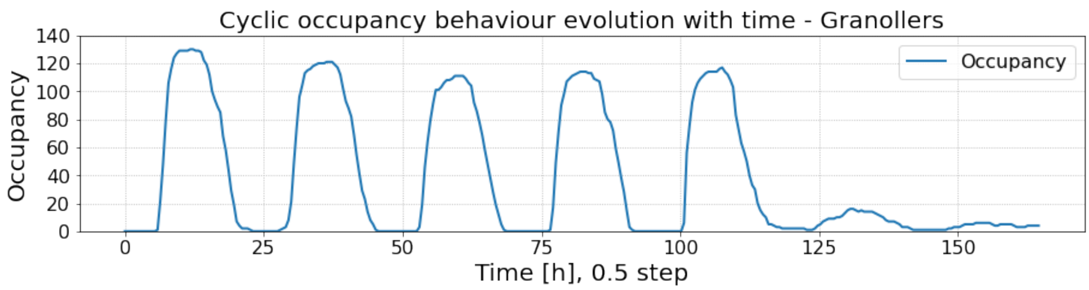

Figure 5: Cyclic behaviour of the Granollers parking from Monday 24th of February to Sunday 2nd of March 2020.

Figure 5 clearly shows how the occupancy behaviour presents a cyclic evolution along time: The same pattern is repeated every single day, having some particularities following the initial guesses that have been already brought up above. To understand better this cyclic behaviour, we plot all the available cyclic profiles corresponding to a given day, in this case for all Tuesdays, and their corresponding historical mean.

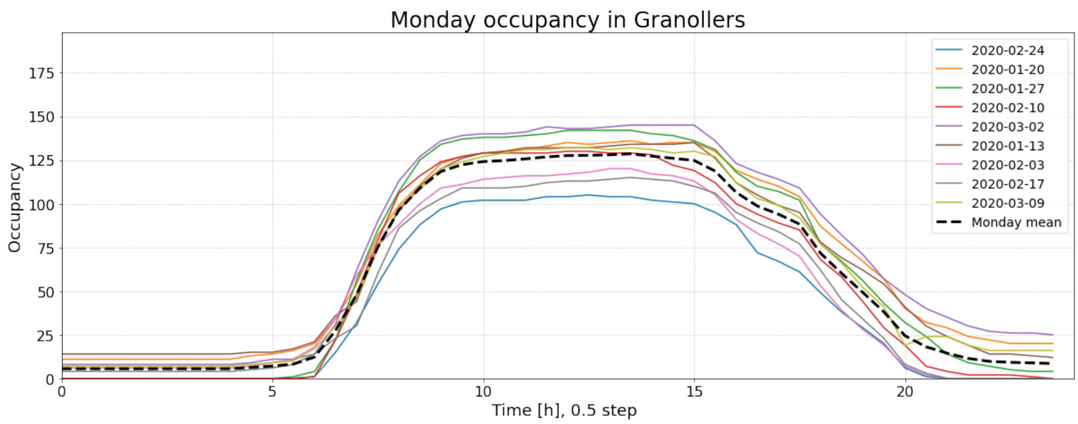

Figure 6: 24-hour occupancy cyclic behaviour for Mondays in the parking of Granollers.

As it can be observed in the Figure 6, all Mondays behave similarly and the historical mean captures properly this behaviour. Even though the total amount of occupied parking slots do vary (the volume of entering cars), the behavior of the cycle is always the same. It does not exist any significant activity or variation during the first 5 hours approximately. The existing offset at the beginning and at the end of the day is directly related to the previous or next day occupation, which means some cars remained parked in the parking throughout the whole night. As both minimal occupancy values coincide both at the first region (from 0 to 5) and at the final one (from 20h on), we can assume the stability at these hours is maintained and thus, almost every day the same amount of cars remained parked.

Then, a clear peak of activity can be spotted, where the occupancy of the parking increase abruptly, from 6 until 9h. The next observation is that this peak is quite constant until 16, where the occupancy starts to decrease dramatically, until it reaches again a stability on the lowest occupancy values from 21:00 to 00:00h. Then the same cyclic pattern should be repeated as well.

2.3 Analysing Data Through Historical Means

At this point, we have achieved some first intuitions and we observe that it makes sense to use historical means to understand better the behaviour of the data, we can move forward visualizing the mean behaviour of each day. Which means, how the occupancy evolves along hours for each type of day, shown in Figure 7.

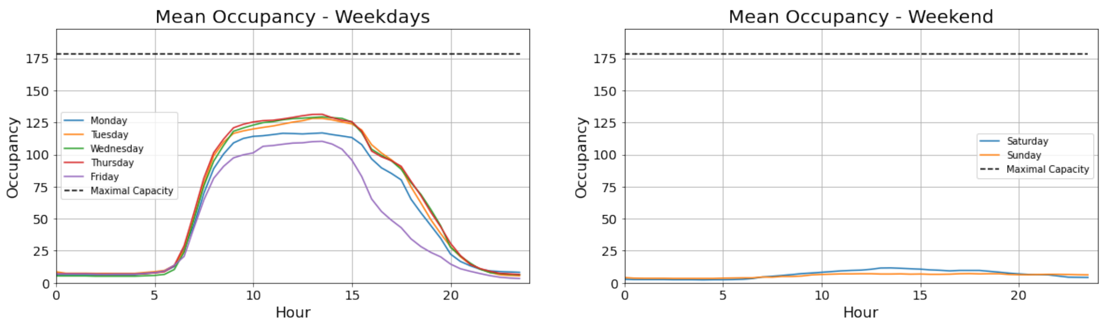

Figure 7: 24-hour occupancy cyclic behaviour from Monday to Sunday in the parking of Granollers.

Now, focusing on the differences between the different type of days we can conclude there are three different behaviours. Whereas weekdays from Monday to Thursday present a high occupancy value, Fridays present a slightly lower one and a narrower stable peak. Finally, weekends are characterized for a significantly low flat occupancy with a reduced variation along the hours. Therefore, we can already assume there might be three different main behaviours: Weekdays from Monday to Tuesday, Fridays and Weekends.

Trying to find an explanation to this, Weekends present a completely different dynamic than the rest of the weekdays. However, when comparing Fridays with the other 4 weekdays, it is understandable that Fridays present their own dynamic, resembling the behaviour that Weekdays have but presenting lower values and an earlier decrease of this occupancy, as many people tend to finish earlier from work or goes back home earlier for then meeting family or friends.

Another remarkable observation from Granollers data is that the occupancy never reaches the maximal capacity of the parking. To show that some parking data can get really closer to a maximal capacity value (even though saturate some particular days), as well as show that the behavior explained so far can be generalized for other parking lots, the same graphic is displayed in Figure 8, but in this case using the Sant Sadurní parking dataset.

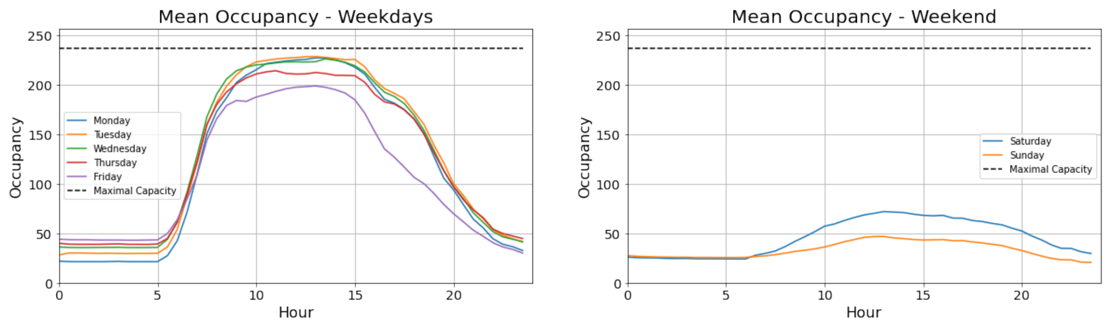

Figure 8: 24-hour occupancy cyclic behaviour from Monday to Sunday in the parking of Sant Sadurní.

As it can be observed again in this second parking (Figure 8), all three behaviors can be differentiated as well. An important appointment is that while weekdays mean behavior from Monday to Thursday are close to saturate (reach the maximum capacity), Fridays in average are quite far from this maximum capacity.

An overall observation for both parking is that, the difference between the four first weekdays and Fridays is not only on the volume of cars, or in the narrower stabilized period, but also in the fact that the decrease of the occupancy is much more smooth and gradual on Fridays than in the rest of the weekdays, where it decreases much more abruptly.

Once observed the mean occupation along the hours for each day, it is important to consider the variance, as it will affect directly to the quality of the predictive model. In the case the data shows high variance, it will be more difficult to achieve a high accuracy on the predictions.

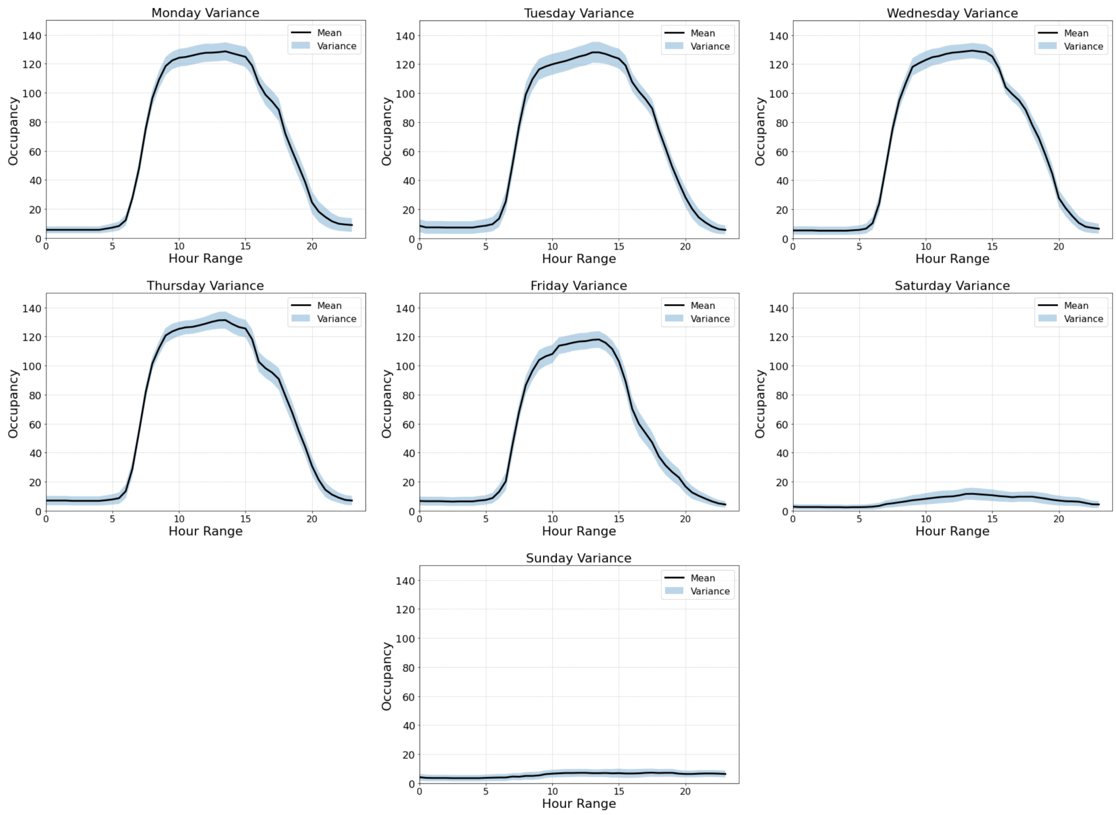

Figure 9: Historical mean for each type of day with its corresponding Variance for the Granollers parking

Analyzing the Figure 9, it can be observed that the higher variance values (semi-transparent blue) are present during the hours the parking is highly occupied, from 8h to 16h approximately, whereas the increment and the decrease periods are less variable. Hence, from this preliminary analysis we can extract the conclusion that the arrivals and departures follow a quite precise pattern, varying on the volume of vehicles, but being quite accurate on the arriving/departure time.

2.4 Holidays and Outliers

All holiday days (specially those weekday holiday, more importantly then the weekend holiday) are omitted during this first preliminary analysis and during the whole model development. This final subsection is aimed to explain the existence of this abnormalities in the dataset and how to treat them. At first, we differentiate between three kinds of abnormalities: The ones that presents an extremely low occupancy because they are holidays, the ones due Covid19 restrictions and finally some unjustifiable abnormal days. In the second case, all of them reproduce the behaviour of a weekend, recovering a similar historical mean.

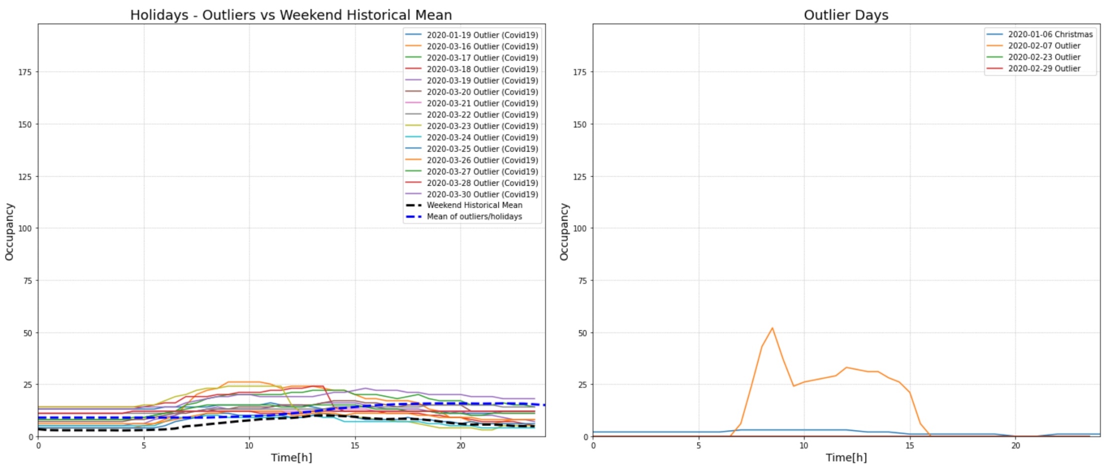

Figure 10: Comparison between the historical mean of Holidays and Outliers to the historical mean of Weekends on the left, and unjustifiable outliers in the right.

Some other additional abnormal days have been detected, as can be observed in the right plot of Figure 10. In this case, 3 of them belong to February, when theoretically there were neither regional nor national holiday. Nevertheless, only one of these February days, in addition to the 6th of January (in the right plot due to the fact it only has 34 sensor values) present real information. The remaining 2 February outliers seems to show that there were some sensors error, as the occupancy is 0 along the whole day.

2.5 Complementary Analysis: Renfe Rodalies Stations

As a parallel task, we want to know how the different park lots we are working with are related to train stations, and furthermore, to observe how incidents in some train stations, or even on other car parks, may affect those. The problem with this is that there was no dataset available representing the different stations and their connections, so we had to build this dataset manually, writing every single connection in every train line. Knowing where the nearest train stations are located, and identifying the closer ones to every single parking has helped to understand the behaviour of the occupancy. One of the main means of transport is trains, and often people get the car to reach the train stations. This direct relation affects the occupancy behaviours, as it depends also on the train timetables, the frequency of trains among others.

In the Figure 11 it can be observed the network of the train stations (orange nodes), visualized with Gephi and using a GeoLayout, which helped to interpret such network and how close the train stations are from each parking that we are working with (green/blue nodes). The names of each train station as well as the location of the parking lots are not shown due difficulty to read them in such small dimension.

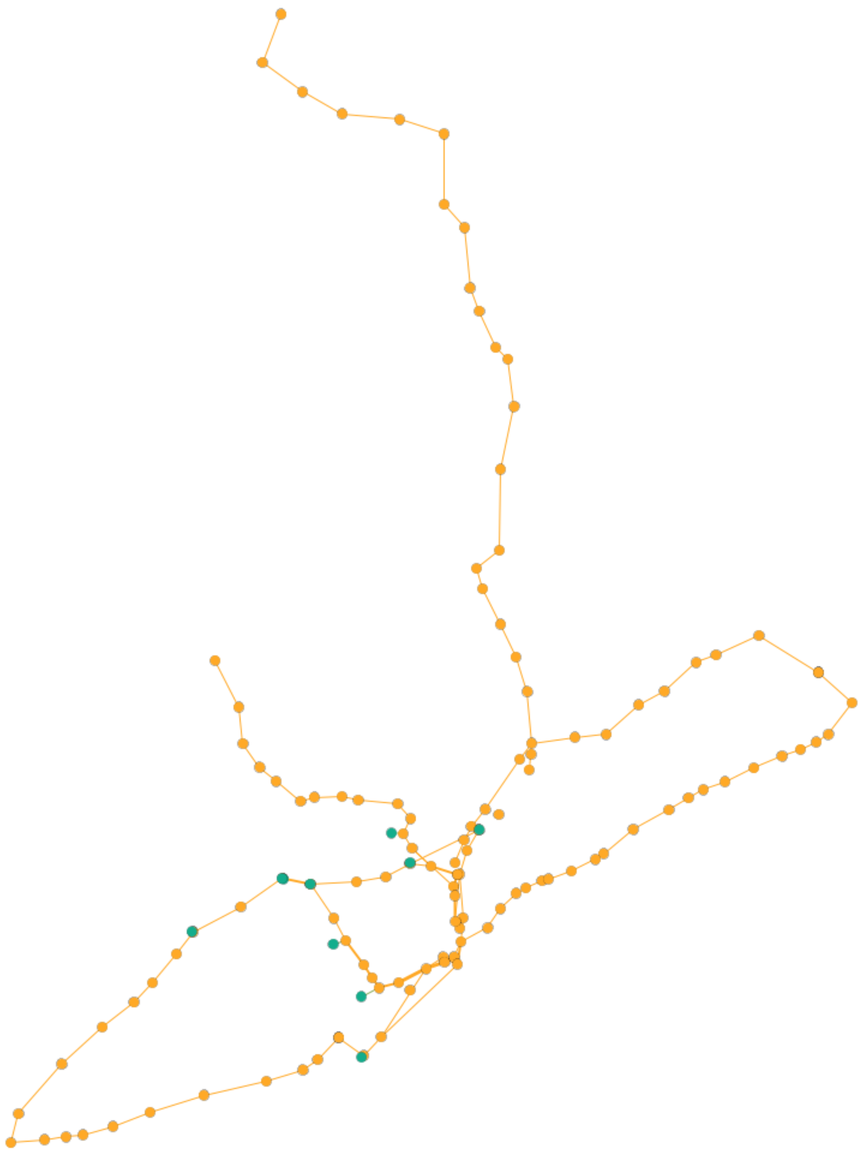

Figure 11: Rodalies Renfe stations and connections in orange, as well as car parks of the study.

Chapter 3

Prediction of Parking Occupancy

In this fourth chapter, the parking occupancy is predicted using statistical and mathematical models. The methodology used to predict aims to be simple, interpretable and transparent in order to demonstrate that it can perform similarly than more complex (and commonly used nowadays) models, such as MLP, Recurrent Neural Networks, LSTM [5, 7], or Linear Regression [4], using really few data. These previous approaches are more data driven, whilst our approach is more model based, using less parameters, despite these parameters tend to be more abstract. In different Neural Networks or Linear Regression models, the parameters tend to be more tangible, such as parking, weather or calendar properties.

I use the data from Granollers parking lots, as it has quite stable sensors, and the variance is relatively low. For more precise information from other parking lots, such as Vilanova, a preliminary and extensive analysis has been carried out in [3]. Moreover, I provide prediction results and accuracy for less precise car parks, in this case Sant Sadurní, which presents a huge variance during the first 6 hours and the 3 last hours of the day. In order to reduce the dimension of the document, I only show graphically Granollers results and some relevant plots of the rest of the car parks, but I provide the rest of the results in an online repository [1], where all the plots and results shown in this chapter can be found for the remaining locations.

[1]Repository containing the results for all the parking lots: `https://drive.google.com/drive/u/1/folders/1JM8CtXTuRbtlIbrhfU35rEDzSRYk5oOM`

First, I provide a brief introduction to the parking occupancy approximations. Then, I present two different types of prediction: entire day prediction (offline) and real time prediction (online) having data about the occupancy of a part of the day. Each of this prediction types are approached by both, the statistical and mathematical model, and compared each other to determine which performs better and which is the optimal trade-off between accuracy and interpretability.

Throughout this chapter I use the word *prototype* quite a lot. So first, a brief definition of what I consider as a Prototype is necessary to understand better why do I need them.

Prototype: *A model that generalizes all common characteristics present in all things of the same kind. In our case, present in all data of the same kind.*

3.1 Approximation of the Parking Occupancy

In [3] there are the details and the mathematical explanation of the methodology and the modeling of the parking occupancy behaviour presented in the previous chapter. To be able to understand the predictive models, I provide first a brief summary of the different approximations:

1. **Statistical Approach using Prototypes:** Similar data is detected to define and generate different prototypes that could capture the main characteristics of each group of similar days. Each prototype is computed as the historical mean of the following similar groups: Weekdays (Monday to Thursday), Fridays and Weekends. Hence, each of this 3 prototypes, represented in the Figure 12 captures the behaviour of the different days.

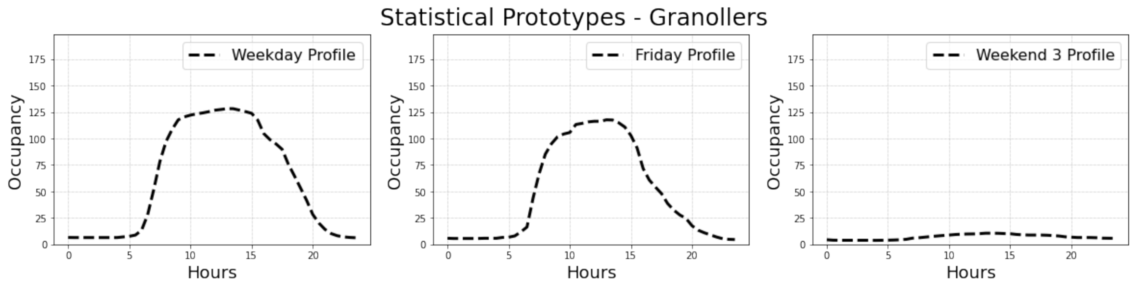

Figure 12: Statistical prototypes for Granollers parking.

2. **Mathematical approach using parametric fitting:** In a second phase, the previously detected prototypes are parametrized using mathematical functions, like Andreas Kaltenbrunner et al. [8]. An aggregation of gammas is used to approximate the behaviours of the Weekdays and the Fridays, and a single gamma for the weekends. The details on this are extensively explained in [3]. Thus, both a mathematical explanation of the behaviour and an easier model can be achieved. Here also 3 different prototypes are obtained, as it can be observed in Figure 13 for the same groups, Weekdays (Monday to Thursday), Fridays and Weekends.

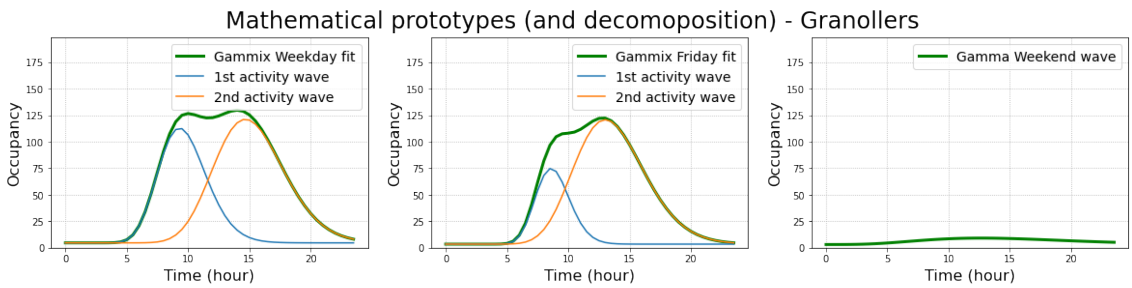

Figure 13: Gamma mixture (Gammix) approximation for each prototype (Weekday, Friday and Weekend from left to right).

From now on, I might also refer to statistic prototypes for the ones obtained using the first approach, and mathematical prototypes for the ones obtained using the second one.

3.2 Prediction of an Entire Day of Occupancy

In few words, one of the approaches that I aim to develop is: given a day, predict the occupancy of a particular parking along all the day, from 00:00h to 23:30h. As there exist two different methodologies, the statistical and the mathematical, I perform and evaluate the predictions for both. I use three weeks of data for testing the predictions. Thus, knowing that I only have three months of data, and that I have previously removed several outliers due to the Covid-19, holidays and to errors in the sensors, I have 5/6 weeks of training data.

To measure the performance of the offline and real time predictions, and hence, measure how good they are, I compute the proportional error for every interval i of 30 minutes (e_i in Equation 3.1), and afterwards I compute the average error (e_d, in Equation 3.2) of the N (48) proportional errors to know the mean proportional error of a day d.

$$e_d[i] = \frac{|(predicted_i - test_i)|}{max\_capacity} \cdot 100 \tag{3.1}$$

$$avg\_e_d = \frac{1}{N} \sum_{i=0}^{N} e_d[i] \tag{3.2}$$

The reason of using proportional error instead of MSE/RMSE is that, for some cases where the real data is 0 and the prediction is 1, the relative error is infinite, whilst in reality, an error of 1 parking slot is considerably low. Or between 1 and 2 (during early hours) provides a 0.5 relative error, unfairly high. Also it is important to put in context the difference: 1 slot in a 45 capacity parking is not the same than 1 slot in a 250 capacity parking. Hence, neither MSE nor MAE seemed an appropriate measure to me to compute the error of this predictions.

3.2.1 Statistical Full Day Prediction

For this approach I directly use the statistical prototypes to determine the occupancy of the parking for a forthcoming day. In this case, the predictive model consist of 48 different parameters, one per each half an hour. Given a day *d*, it is possible to know to which prototype it belongs: Weekday, Friday or Weekend. Once known, I directly get the 48 parameters of the prototype, which represent the parking occupancy for the 24 hours with a 30 minutes resolution.

Disaggregated by day in Figure 14, I obtain the prediction (dashed blue line) for every single day of the week compared to real testing days (different color). It is important to remark that, due to the data cleaning, 3 weeks of testing data might contain different number of weekdays, for instance, 4 Tuesdays and 2 Fridays.

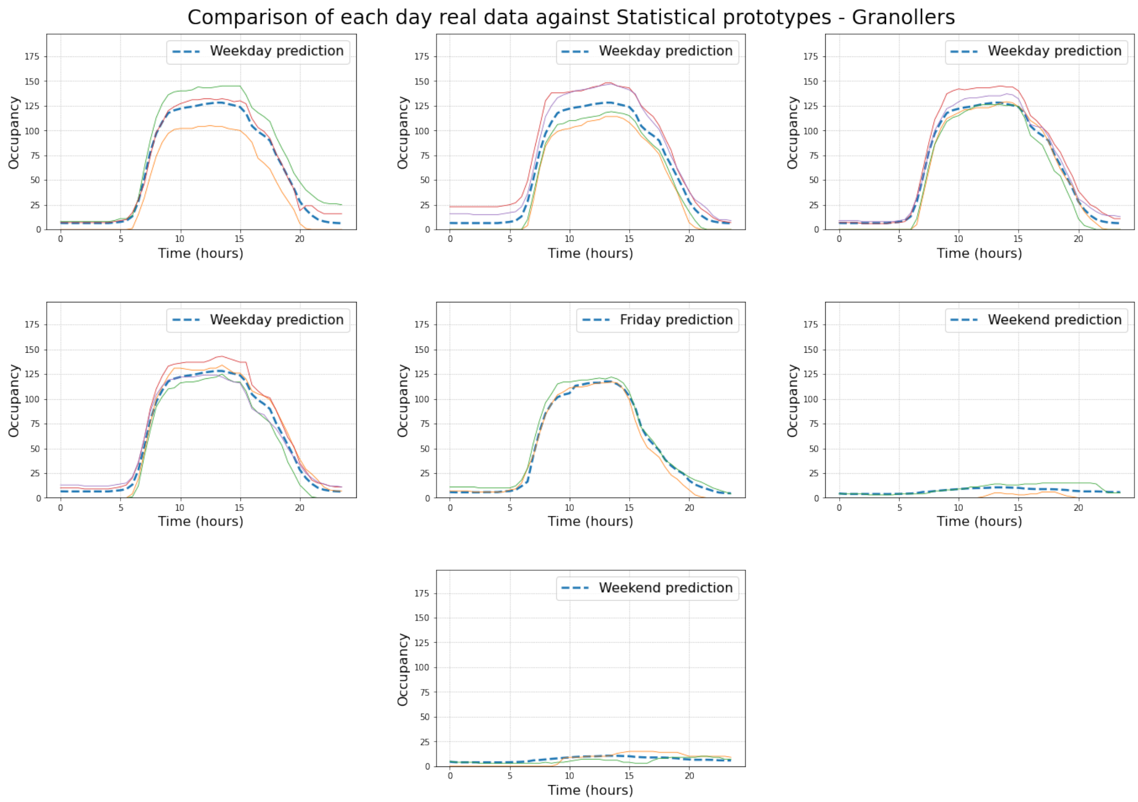

Figure 14: Comparison of the entire day prediction using statistical prototypes with testing days in Granollers from Monday to Sunday.

In the Figure 14, it can be observed that, despite the variance of the data, the prediction behaves quite appropriately. I find the highest differences during two different time periods, depending on the day: During the early stable hours or during the maximum occupancy hours, which in fact, are also quite stable.

Taking the values shown in the previous Figure 14, I compute the proportional error for every single testing day, for each day of the week. Moreover, after computing the proportional error for the complete day with a resolution of 30 minutes, I also computed the mean of the errors by day.

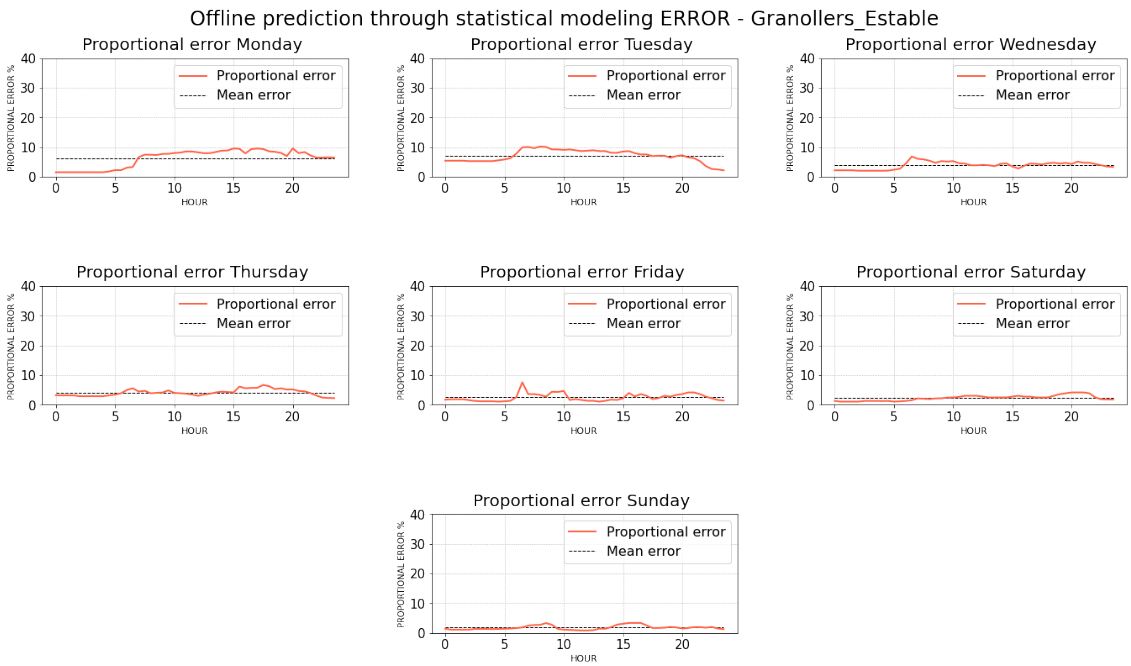

Figure 15: Proportional error of the prediction for each day of the week using statistical prototypes in Granollers.

The results of these computations are quite satisfactory, as can be seen in the Figure 15. It is quite clear that the worst days are the Mondays and the Tuesdays, with an approximated mean error of 7% and a peak proportional error at the end of the day of 10%, whilst the rest of the days do not exceed the 5% and the peaks of proportional error do not reach the 10%. In other words, I obtain a mean proportional accuracy of 93-97% with only 5/6 training weeks. It can also be observed that there is no a pattern on the hours that supply more error to the mean, as the proportional error behaves quite uniformly.

In the case of the statistical approach to predict an entire day, for this particular parking, the variance has not a massive effect on the error and it is hard to find a correlation between the error and the variance observed in the *Preliminary Analysis* chapter. Nevertheless, I can state that during the maximum occupancy hours, where more variance can be found, is when normally I obtain a subtle bigger error.

Analyzing Sant Sadurní, which differs from Granollers in some aspects, it can be observed that some testing days occupancy saturates reaching the maximum capacity of the parking, whilst the prototype does not saturate. This can be due to the fact that the training and testing data belongs to different months, or even different periods of a month (early days vs last days), and this may affect the occupancy.

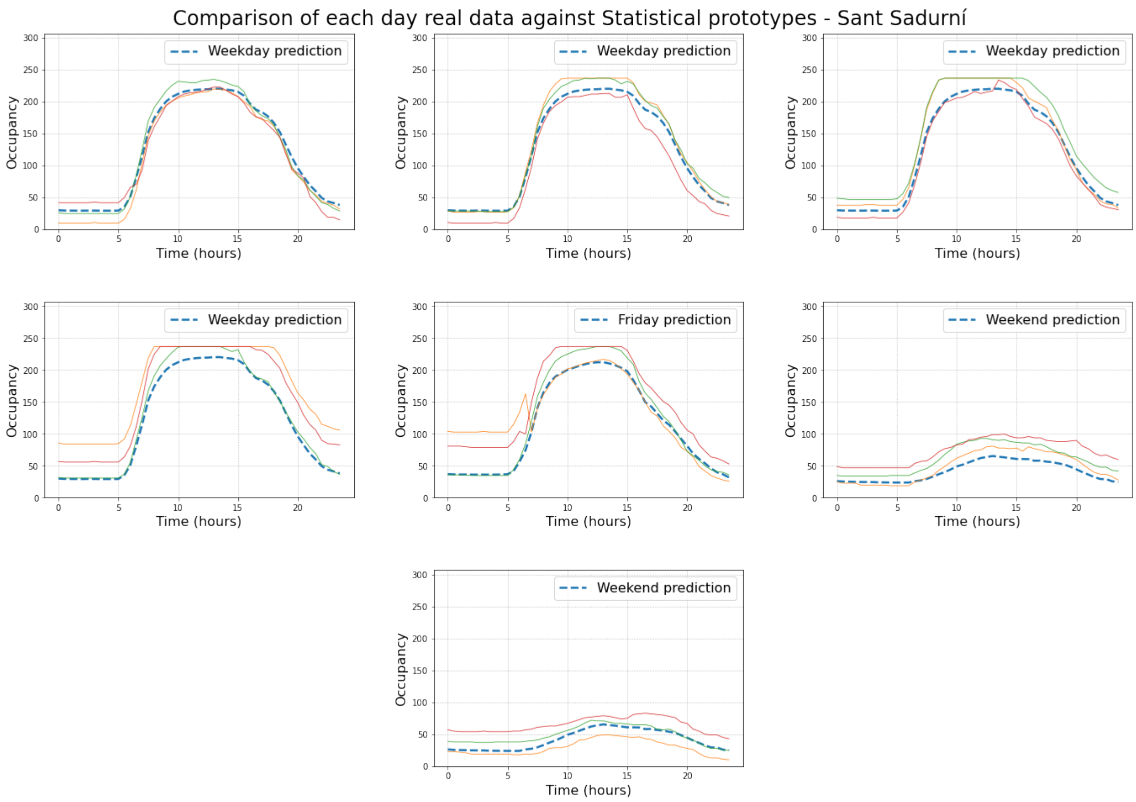

Figure 16: Prototypes compared to testing data in Sant Sadurní from Monday to Sunday.

In Figure 16 Fridays show a huge variance on the first hours of the day, which is directly related to the variance at the end of Thursdays. People leave the car that night quite randomly, that is why there exist such a great variance. Obviously, this will make the prediction of an entire day perform worse.

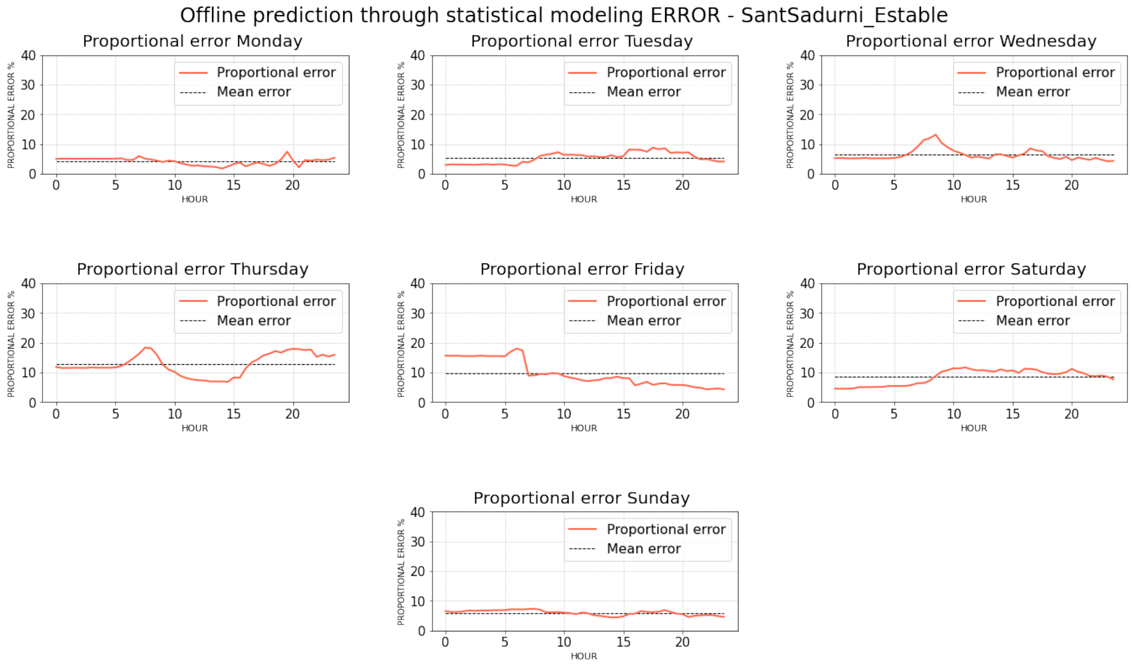

Figure 17: Proportional error of the prediction for each day of the week using statistical prototypes in Sant Sadurní.

Comparing Granollers with Sant Sadurní, which has more variance in particular time period, in the Figure 17 can be seen that the higher mean proportional error happens on Thursday and Friday, being over the 10% in average, and having peaks of error of almost 20%. Also the accuracy during the weekends decreases, while Mondays and Tuesdays are predicted more accurately. Here it is also observable a correlation between variance of the testing data and proportional error, as the periods with higher variance provide the highest error peaks.

3.2.2 Mathematical Full Day Prediction

The second approach is the mathematical one. In this case, I use the parameters extracted to generate the normalized gamma mixtures of each prototype and the mean of the integral of the days belonging to each prototype to re-scale. Hence, with this approach there are as maximum 8 parameters, 40 less than with the statistical. I say maximum 8 due to the fact that the weekends present a much simpler behaviour. Thus, the parametrization of the behaviour for weekends is just a simple gamma, and I only need 5 parameters. Counting, I need the gamma mixture/gamma parameters, the area (the mean of the integral) to denormalize (re-scale) the gamma mixture distribution to parking occupancy values, and be able to perform the proportional error detailed in the previous section, and, consequently, to be able to compare both models performance, and finally (the 8th or the 5th parameter) the mean offset of the prototype, as some car parks never reach 0 occupancy, and I need to return this information after the normalization. Hence, the process is:

1. Detect the prototype of the day I want to predict, in order to know which are the parameters that I need to reconstruct the appropriate gamma mixture, as well as to know which is the correct re-scaling factor (area) and mean offset. Also to verify if I need a gamma mixture or a simple gamma, in case the day d belongs to Weekends.

2. Once known if the day to predict belongs to a Weekday, Friday or Weekend prototype, get the parameters that allows to regenerate the gamma/gamma mixture kernel (3 or 6 parameter for the gamma/gamma mixture, 1 parameter to re-scale and 1 to apply the offset).

3. After obtaining the proper parameters, generate the gamma mixture distribution and re-scale it by the factor of the mean area of the detected prototype (Weekday, Friday or Weekend). Once I have re-scaled, I aggregate the mean offset (48 value array) and I obtain the prediction of the occupancy for a full day d for the current parking with a 30 minutes resolution.

Disaggregated by day, I obtain the following prediction (dashed blue line) compared to real testing days (random color):

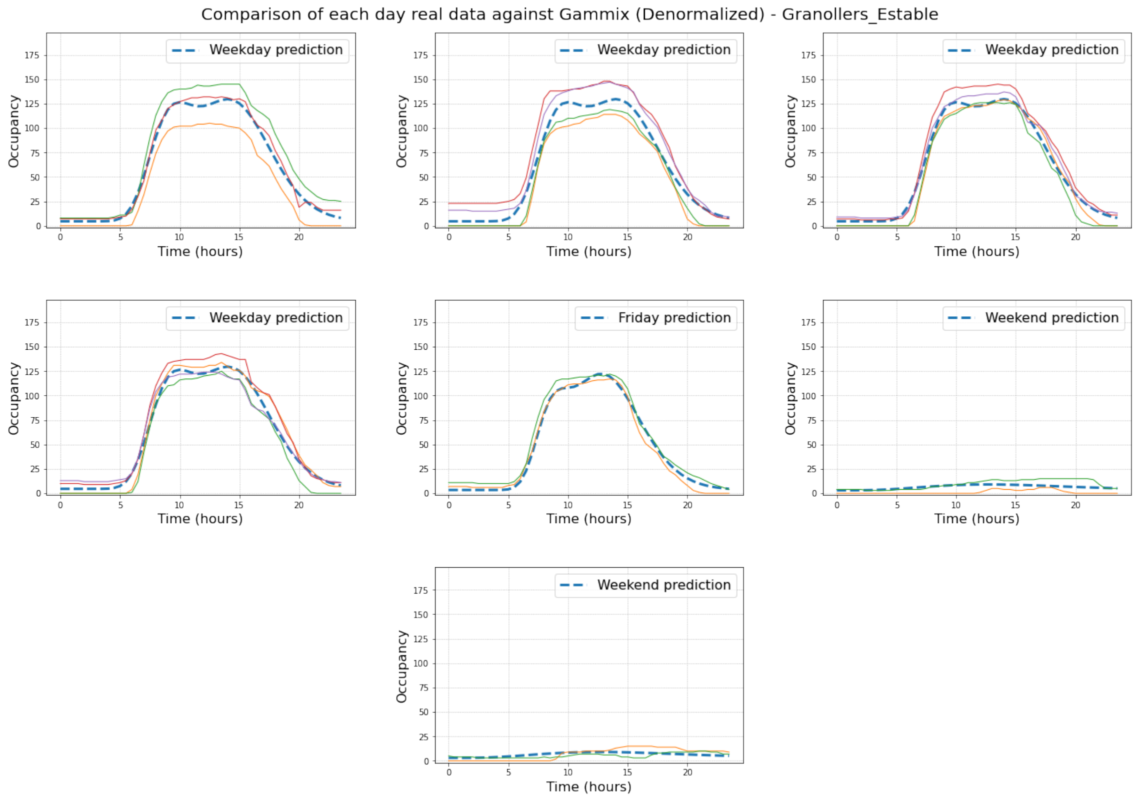

Figure 18: Comparison of the prediction using Gammix prototypes with testing days from Monday to Sunday.

Despite the evident difference in the approximation and methodology, it can be easily seen in the Figure 18 that similar results are obtained. Again, there exist variance in some periods of every day. For instance, this becomes evident in Mondays, in the maximum occupancy hours, and also in the Tuesdays subplot, too, during the early stable hours and also during the maximum occupancy hours. In order to compare both, I compute the proportional error proceeding in the same way than previously done.

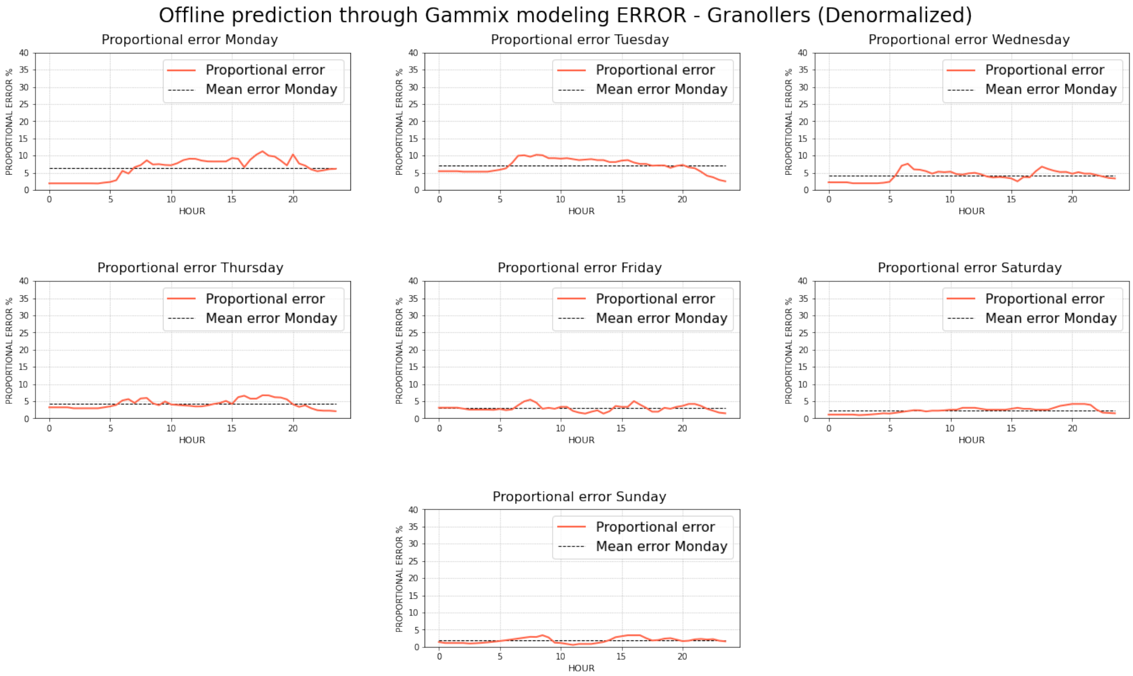

Figure 19: Proportional error of the prediction for each day of the week using Gammix prototypes in Granollers.

In Figure 19, it can be seen that, again, the worst days are the Mondays and the Tuesdays, with an approximated mean error of 7%, whilst the rest of the days do not exceed the 5% and for the Fridays and Weekends it's under 3% of mean proportional error, which means that the model might fail around 5 slots in average in a 140 capacity parking. In other words, I obtain a proportional accuracy of 93-97% with only 5/6 training weeks. Regarding peaks in the proportional error in particular time periods, Monday is the only day that exceeds the 10% at 18:00h/19:00h. Regarding the rest of the days, the error stays quite constant along the hours.

Concerning Sant Sadurní, which has been analyzed in the statistical approach, I can extract the same conclusions that I have previously mentioned: The variance in this parking affects directly to the performance of the predictive model, irrespective of whether the approach is statistical or mathematical. And this is perfectly seen again by observing where the peak proportional errors occur in the Figure 20, Thursday 18:00h - 23:30h and Friday 00:00h - 06:00h.

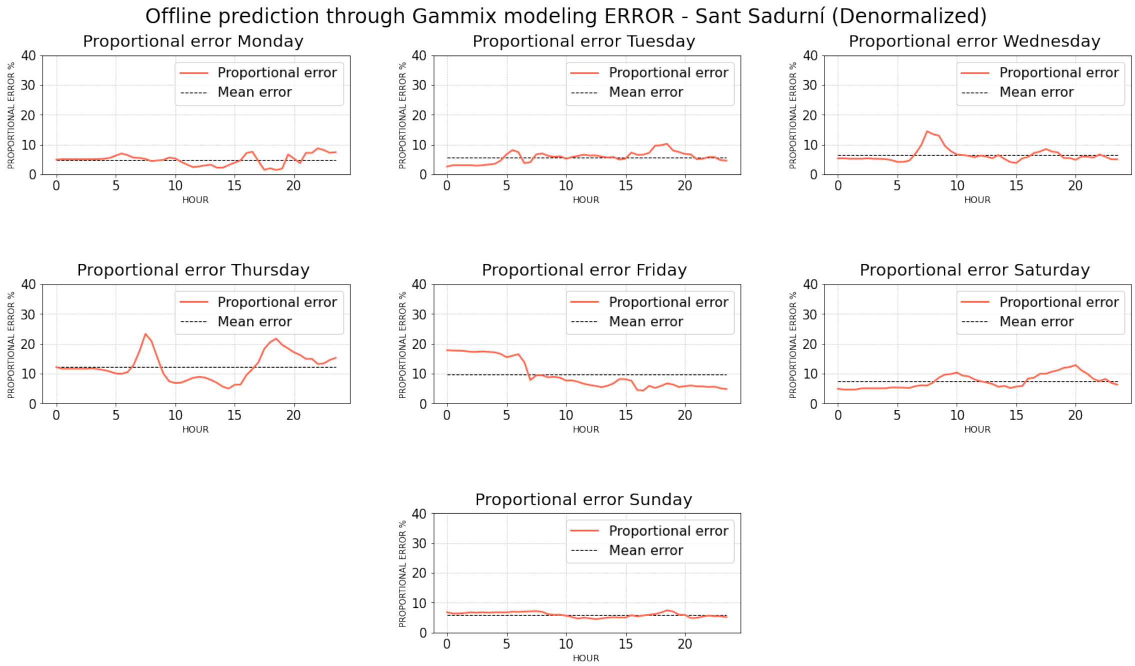

Figure 20: Proportional error of the prediction for each day of the week using Gammix prototypes in Sant Sadurní.

I collect the mean proportional errors and the corresponding standard deviations for the Granollers parking in the Table 1, which supports the conclusions already made: The accuracy of the prediction for entire days performs satisfactorily.

Table 1: Mean (± stdv) complete day prediction proportional error comparison for Granollers parking.

Day	**Statistical prop. error (%)**	**Gammix prop. error (%)**
Monday	6.31 (±2.92)	6.43 (±2.84)
Tuesday	7.13 (±2.03)	7.18 (±1.93)
Wednesday	3.96 (±1.21)	4.19 (±1.48)
Thursday	4.16 (±1.09)	4.20 (±1.28)
Friday	2.55 (±1.28)	2.97 (±0.92)
Saturday	2.35 (±0.89)	2.35 (±0.90)
Sunday	1.79 (±0.71)	1.88 (±0.78)
Mean	**4.04** (±2.03)	**4.17** (±2.00)

3.3 Real Time Prediction

The second problem that I want to solve is the following: at a given time t within the hours of the day d, with a resolution of 30 minutes, I want to predict the occupancy of a particular parking in the following hours until the end of the day, taking into account the occupancy of such parking in the previous N hours, being N the number of hours from 00:00h to t.

In this case, one might believe that directly using the prototypes again would work properly, but that is not completely true, due to the fact that there exist a variance on the volume of vehicles, and that variance can not be perfectly captured with this method. Also, another possibility would be a parametric approximation to fit the known data, but due to the composition of two different gammas, when few hours of data are provided, the second activity peak would have to be approximated from a linear interpolation of the Gammix prototype and the current prediction through the parameters approximation. This has not provided accurate results for the first half of the day. Hence, those are the reasons why I decide to approach this problem using a real-time scaling methodology, inspired by the prediction models used in [8, 9].

The auto-complete approach consists on taking into account the data obtained the N previous hours. To do so, I should perform a minimization of the error between the current known data and the prototype (statistical or mathematical). But that can be well approximated by considering the current occupancy of the parking and the hour of the day. The reason of not computing the minimization is that, in some cases, there can be an unexpected/outlier behaviour in some earlier hours to the last known hour, and that might worsen the current status, that, it could (or not) be following perfectly the prototype. Hence, what the chosen methodology does is:

1. Detect the prototype that day d follows (Weekday, Friday or Weekend) and get the last known value and the time t this value was computed.

2. Get the values of the prototypes (Gammix and statistical) at time t.

3. For both, Gammix and statistical, divide the real value by the prototype value to obtain the scale factor for each approach.

4. Scale the prototypes by the corresponding scale factors.

To showcase this approach, I chose a random Monday of the testing data, and I perform the real time scaling prediction simulating different known windows and limit hours. The results are shown in Figure 21:

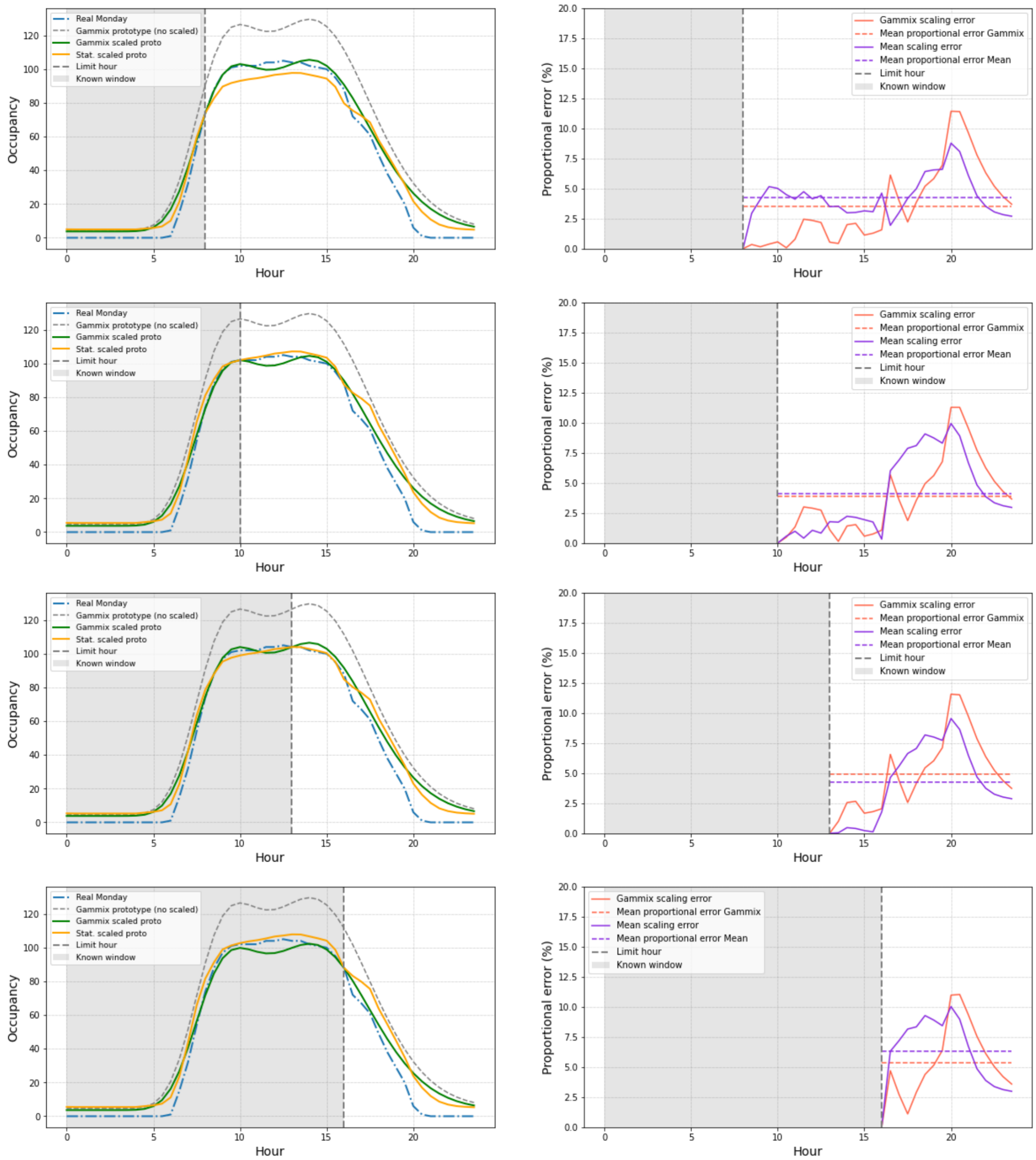

Figure 21: Prediction and proportional error of the remaining hours of a day given different times t=8,10,13,16 from top to bottom of an arbitrary testing Monday.

As it can be observed, in this particular example both prototypes work quite accurately having different known data windows. Nevertheless, there is a time period where both, the statistical and the gamma mixture prototypes, performs worse, and this period belongs to the stabilization of the occupancy at the end of the day. I have already commented on that during the preliminary analysis, and it is also commented in [3], along the *Methodology* chapter. As it is a sudden stabilization of the occupancy, it is really difficult to capture that abrupt change on the behaviour. Despite that, the proportional mean error remains between a reasonable limits. More precisely, the mean proportional errors and the standard deviations are detailed in the Table 2.

Table 2: Testing Monday mean proportional error (± stdv) along different hours using scaling up in statistical and mathematical model in Granollers.

Monday	**08:00h**	**10:00h**	**13:00h**	**16:00h**
Gammix (%)	3.51 (± 3.22)	3.87 (± 3.17)	4.93 (± 3.14)	5.32 (± 3.08)
Stat istical (%)	4.26 (± 1.73)	4.09 (± 3.21)	4.23 (± 3.15)	6.29 (± 2.82)

The mean proportional errors for both approaches (for this particular example) are under the 6.5%, which means that for Granollers parking, which has a maximum capacity of 140 approximately, the highest average error would be around 9 slots, and also, that error would happen during the hours that present a low occupancy which, in practical terms, is less worrying than making a mistake during the maximum occupancy hours.

Comparing both approaches, there are no significant differences. In this particular example, it performs slightly better the mathematical approximation, using the Gammix approach, but in some other cases, it is the opposite. This can be seen in the Appendix, in order to not extend this thesis with the huge amount of graphical and tables results. Apart from predicting a little better than the statistical model, the mathematical one has the advantage that is self explanatory and has an interpretation, and furthermore it require much less parameters to be reconstructed than the statistical one.

I have just provided an example of a testing Monday, which belongs to the Weekday prototype. To provide a complete analysis, the result tables for the same analysis but performed for a Friday and a Saturday can be observed, to cover the three prototypes that I mention along the whole thesis.

Table 3: Testing Friday mean proportional error (± stdv) along different hours using scaling up in statistical and mathematical model in Granollers.

Friday	**08:00h**	**10:00h**	**13:00h**	**16:00h**
Gammix (%)	3.45 (± 2.21)	3.76 (± 2.22)	3.34 (± 1.40)	2.03 (± 1.30)
Statistical (%)	7.32 (±6.69)	8.27 (± 8.24)	10.34 (± 7.91)	4.63 (± 2.33)

Table 4: Testing Saturday mean proportional error (± stdv) along different hours using scaling up in statistical and mathematical model in Granollers.

Saturday	**08:00h**	**10:00h**	**13:00h**	**16:00h**
Gammix (%)	2.44 (± 2.41)	2.49 (± 1.53)	1.48 (± 1.02)	1.61 (± 0.83)
Statistical (%)	3.31 (± 1.75)	3.68 (± 2.29)	3.19 (± 2.48)	3.65 (± 2.32)

The results obtained for this particular case for Friday (Table 3) and Saturday (Table 4), which in fact represent quite well the results of rest of the testing days for Granollers, show a better performance of the real time prediction model than the one for the Monday. The Friday testing day shows a really similar accuracy than the Monday. Regarding the Saturday, it performs even better with both methods, the statistical and mathematical.

Other car parks, such as Vilanova or Quatrecamins, provide really similar results. Nevertheless, I find interesting to compare these results again with Sant Sadurní, which present more unstable training and testing data, due to different factors: the nature of the behaviour in that parking, failures of the sensors, or another different unknown factor.

In the case of Sant Sadurní, I simulate the real time prediction for the same days, in order to be consistent. For Mondays, the statistical approach seems to perform better, as it is shown in Table 5. Nevertheless, the mathematical model provides clearly lower mean proportional errors and standard deviations for Friday and Saturday. The main difference between Sant Sadurní and the before-mentioned car parks is that the weekends real time prediction perform way worse for both approaches than the rest of the days, unlike in Granollers, Vilanova or Quatrecamins.

Table 5: Testing Monday mean proportional error (± stdv) along different hours using scaling up in statistical and mathematical model in Sant Sadurní.

Monday	**08:00**	**10:00**	**13:00**	**16:00**
Gammix (%)	4.62 (± 3.32)	3.40 (± 1.92)	4.71 (± 2.47)	3.28 (± 1.85)
Statistical (%)	1.58 (± 1.48)	1.71 (± 1.46)	2.89 (± 1.90)	2.50 (± 1.75)

Table 6: Testing Friday mean proportional error (± stdv) along different hours using scaling up in statistical and mathematical model in Sant Sadurní.

Friday	**08:00**	**10:00**	**13:00**	**16:00**
Gammix (%)	2.82 (± 2.39)	3.06 (± 2.64)	4.17 (± 2.56)	3.72 (± 2.46)
Statistical (%)	5.82 (± 4.84)	6.81 (± 5.76)	9.79 (± 6.24)	5.52 (± 3.50)

Table 7: Testing Saturday mean proportional error (± stdv) along different hours using scaling up in statistical and mathematical model in Sant Sadurní.

Friday	**08:00**	**10:00**	**13:00**	**16:00**
Gammix (%)	3.98 (± 3.96)	5.35 (± 4.43)	3.16 (± 1.91)	3.03 (± 1.33)
Statistical (%)	10.66 (± 3.59)	6.11 (± 4.70)	5.60 (± 4.65)	6.44 (± 4.53)

In conclusion, the real time prediction model performs really well, despite the not-so-complex architecture. It performs even better than the offline prediction, obtaining proportional mean error rates always under 5.5% in the case of Granollers, acceptable standard deviations, and also a good performance in the rest of the car parks. The highest proportional error periods remain constant in every single day, and in every parking: from 19:00h to 21:00 approximately, the decrease of occupancy, it is really difficult to predict accurately for our real time predictive model.

Regarding the two methodologies, it has been shown that both, statistical and Gammix performs quite similar for the offline and online approaches, providing good results, and alternating which is the best one depending on the day and the particular parking. Nevertheless, due to the fact that the mathematical one provides a more understandable and interpretable explanation, I can consider it as the best method. The results of both predictions (offline and online) for the rest of the parking lots is available in an online repository [2].

[2]Repository containing the results for all the parking lots: `https://drive.google.com/drive/u/1/folders/1JM8CtXTuRbtlIbrhfU35rEDzSRYk5oOM`

3.4 API Connection for Real Time Predictions

The previous sections provide the results for simulations, where I use static data. The following step is to test this predictive model using real time data, obtained directly from the API.

An HTTP connection with the API is implemented, with an automatic refresh of the authentication. Through this connection, I am able to request for the current status of the parking occupancy of all the car parks at the same time. This fact is really positive when it comes to think about a future deployed application. The central system, from Telefonica, is updated every time a car enters or leaves any parking. This would be a huge amount of unnecessary updates, so our proposal is to call this serves every 30 minutes, and the server will provide the last status for every single parking. I obtain the data in Figure 22, among much more parameters that might be useful for a future application:

	Parking	Last_update (UTC)	Last_update (Local)	Occupancy	Parking capacity
0	ElPrat	2021-06-22T15:18:24.247Z	2021-06-22T17:18:24.247Z	60	397
1	SantBoi	2021-06-22T15:18:19.251Z	2021-06-22T17:18:19.251Z	291	81
2	LesFonts	2021-06-22T15:13:48.462Z	2021-06-22T17:13:48.462Z	288	0
3	MartorellCentral	2021-06-22T15:18:15.996Z	2021-06-22T17:18:15.996Z	15	101
4	MolletSantFost	2021-06-22T15:18:24.969Z	2021-06-22T17:18:24.969Z	244	0
5	CerdanyolaUni	2021-06-22T15:18:18.813Z	2021-06-22T17:18:18.813Z	3	116
6	Vilanova	2021-06-22T15:14:08.634Z	2021-06-22T17:14:08.634Z	72	393
7	GranollersCentre	2021-06-22T15:15:17.278Z	2021-06-22T17:15:17.278Z	128	48
8	Castellbisbal	2021-06-22T15:18:14.275Z	2021-06-22T17:18:14.275Z	133	177
9	SantQuirze	2021-06-22T15:18:25.303Z	2021-06-22T17:18:25.303Z	193	178
10	SantSadurni	2020-06-25T17:22:28.000Z	2020-06-25T19:22:28.000Z	45	189
11	QuatreCamins	2021-06-22T15:15:54.946Z	2021-06-22T17:15:54.946Z	65	89

Figure 22: Data gathered using the HTTP request to Urbiotica sensor central system.

Once gathered the real time data, the further action is to apply what has been explained in the *Real Time Prediction* section.

Chapter 4

Covid-19 Influence on the Parking Occupancy

Along the previous chapters of the thesis, it has been visualized, analyzed and predicted the behaviour of the parking occupancy taking into account a global normal situation. But, unfortunately, this last year will be remembered by a worldwide pandemic, the Covid-19. It has been mentioned that the lack of data has been a constraint and a possible reason to be limited to achieve better results. Nevertheless, I have access to the parking occupancy during the pandemic, from April 2020 to March 2021. This chapter aims to provide an understanding on how the Covid-19 influences the way the car parks are used, exemplifying this with Granollers parking.

4.1 Behaviour along Different Lockdown Phases

For the study of the influence of Covid-19, an analysis on how the occupancy was affected in average during the different phases of the pandemic, starting from the strict lockdown to the last phases has been carried out.

In Figure 23 it can be observed how I split the different phases, according to the restrictions imposed in Barcelona and Catalonia.

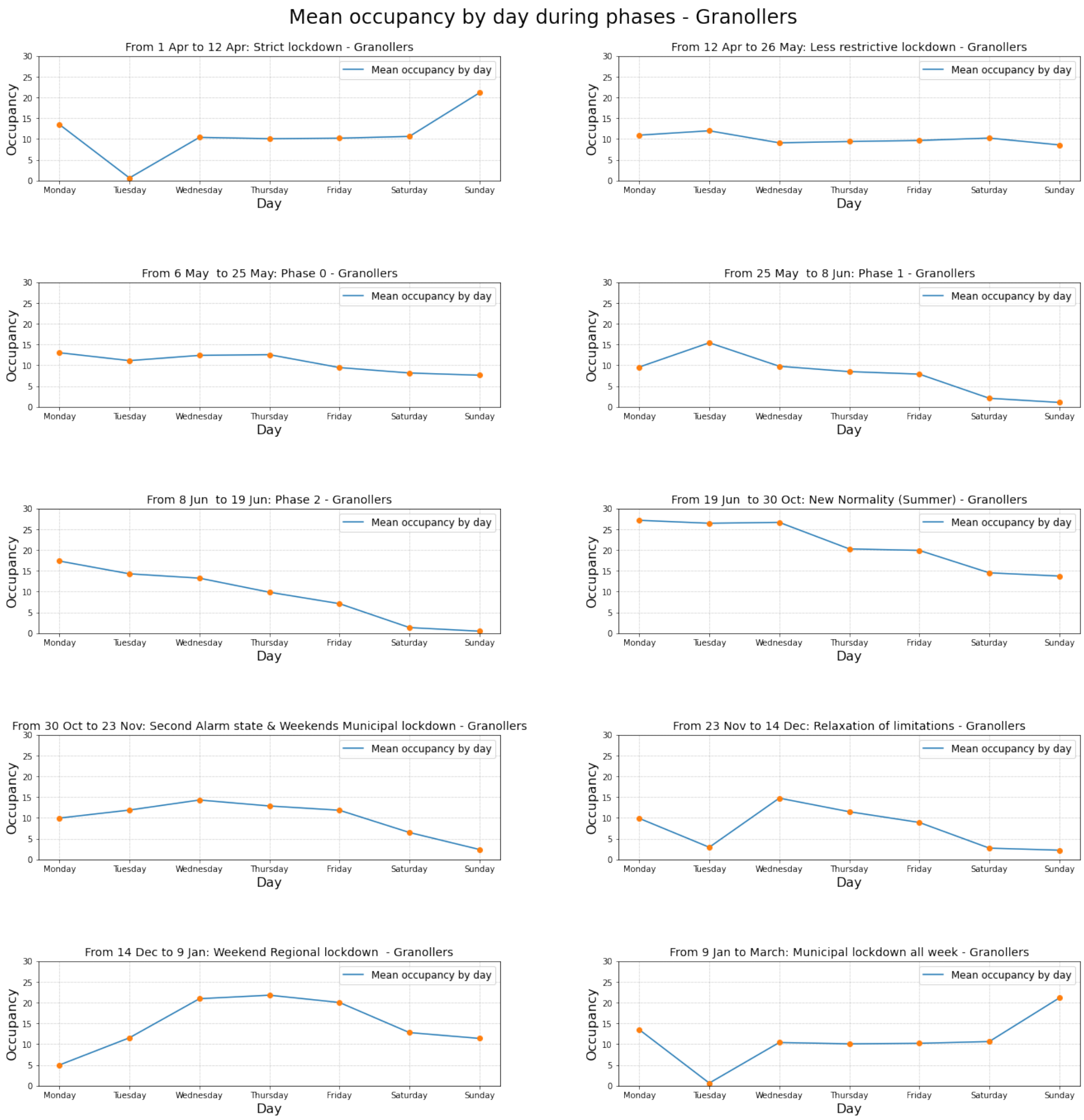

Figure 23: Mean day occupancy during different phases of the Covid-19 pandemic in Barcelona province.

In the top left subplot, as well as in the bottom right one, it can be observed that Tuesday mean occupancy is 0 (or nearly 0). This event might happen due to the fact that the particular phases last only 1 week approximately, and coincidentally that Tuesday the sensors or the system failed. Apart from this particular artifact, it can be also observed that during the summer, when the "new normality" was achieved and people had more freedom of movement, the mean occupancy increased and behaved more similar to a pre-Covid period.

To sum up this first analysis from a full day perspective, it can be concluded that the mean occupancy has decreased considerably compared to the occupancy in the pre-pandemic months. Much more flat graphics, together with the previous statement, determine that the difference in the occupancy between weekdays and weekends has been reduced.

4.2 Behaviour and Volume Variation

Once observed that the mean occupancy along the day changes, there is the need to observe and understand how the occupancy changed along the hours during the pandemic with respect to the pre-covid months. First, a qualitative analysis of the behaviour, or in other word, an change in the behaviour analysis, is performed, and once the difference on the shape is determined and the hours where there is more difference are clearly identified, the difference on the volume of car arrivals and departures is quantified.

At first sight, in the Figure 24 it can be seen that the mean occupancy along the days during the pandemic (right) and during the previous period of 2020 (left).

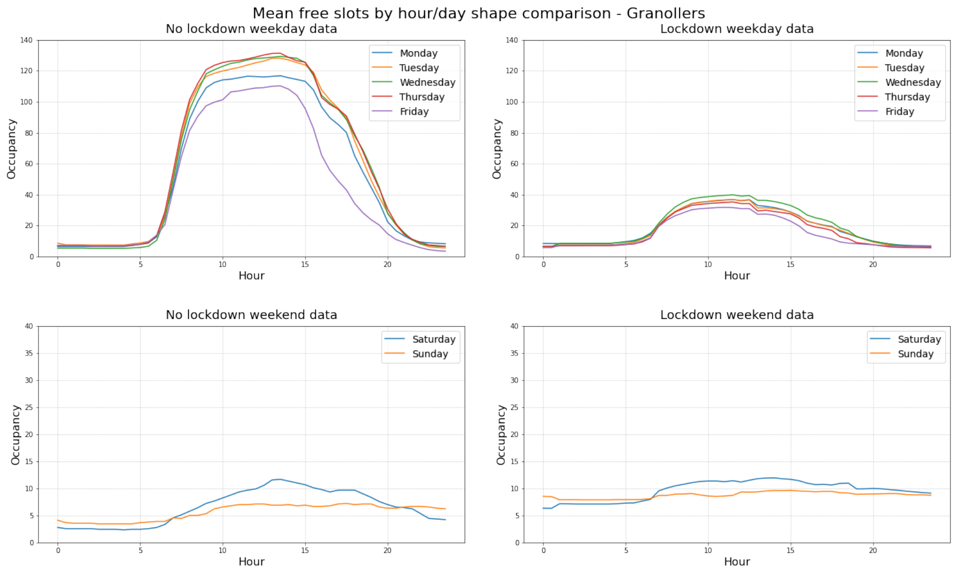

Figure 24: Mean occupancy by day comparison between pre-pandemic and pandemic data.

The shape of both plots, pre-pandemic and during the pandemic, do not show a dramatic difference at first sight. The main observations and hypotheses are:

- The maximum mean occupancy during the pandemic barely reaches 40 car slots occupied, while it reaches a maximum occupancy of almost 135 during the previous months (January, February and half March 2020).

- The departures seem to happen earlier during the pandemic, showing a behaviour more similar to the pre-covid Fridays. Specially, at 13h, there is an abrupt decrease in the occupancy, probably due to the special job schedules during Covid-19.

- Weekends still have a less predictive behaviour, varying the intensity of the increment and decrease of occupancy. Nevertheless, Sundays still present a flatter occupancy than Saturdays. In average, during the pandemic, here were more cars parked along the first and last hours of the weekend than in the pre-pandemic period.

4.2.1 Change in the Occupancy Behaviour

To support these previous statements, I need to compare them in the same conditions. Hence, a normalization between 0 and 1 is performed in order to be able to compare both, covid and pre-covid behaviour, in a fair way. It is true that, for lower values, might be a higher offset in the early hours and also in the later hours despite having the same occupancy, but this is treated in the forthcoming quantitative analysis. Once normalized and having all values between 0 an 1, I compute the absolute difference between every day, from Monday to Sunday, of both periods with a 30 minutes precision, in order to maintain consistency throughout the document. Nonetheless, I treat weekend days separately, as it requires an specific explanation.

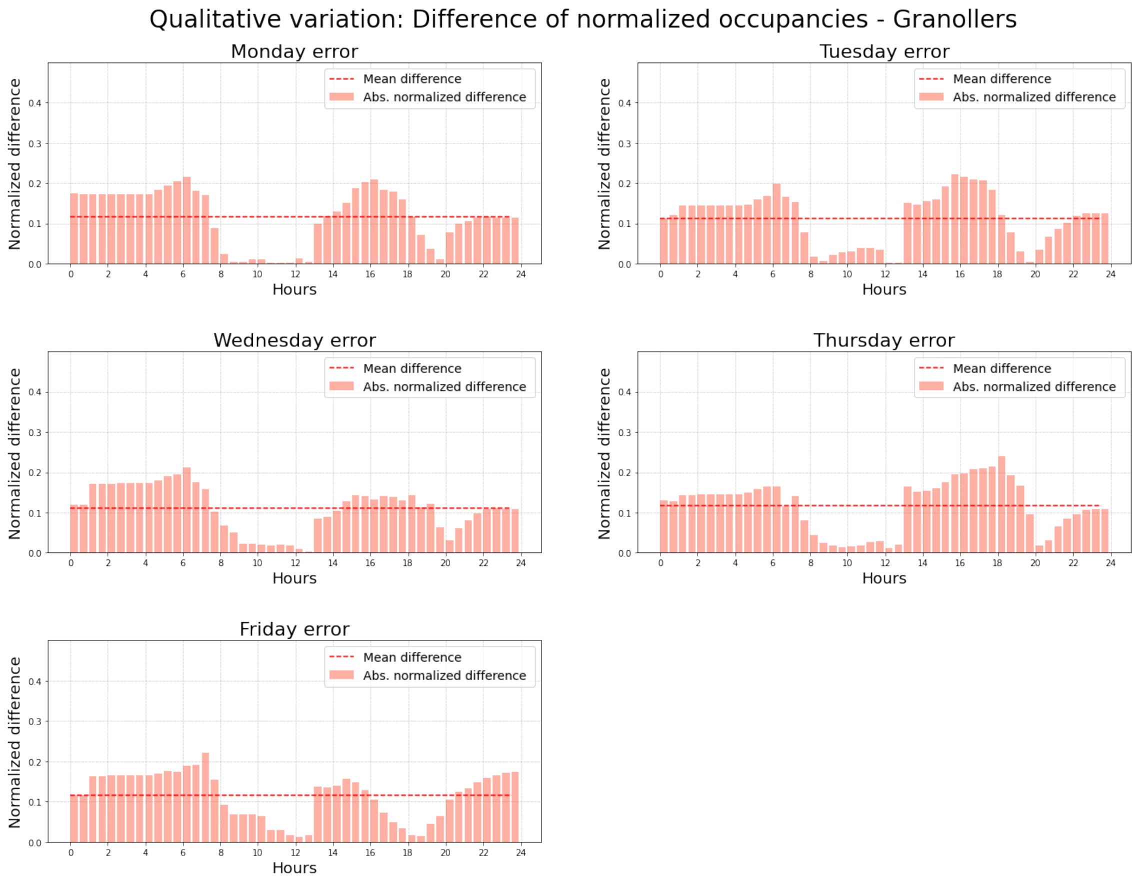

Figure 25: Qualitative difference between normalized data from pre-pandemic and pandemic occupancy.

As the first intuitions show, the lowest qualitative difference occurs during the maximum occupancy hours, while during the increase and decrease phases it reaches the maximum qualitative difference, due to the less abrupt car arrivals and early (and also less abrupt, too) cars departure. From 13:00h to 18:00h there exist the difference caused by the earlier departure of cars during Covid-19, as well as by the slower rhythm of departures. In overall, a pattern on the difference along the hours is observable for all the working days, and a average qualitative normalized difference of 10% approximately allows us to state that the behaviour during the pandemic is qualitatively similar.

As previously mentioned, weekends requires special attention in this analysis:

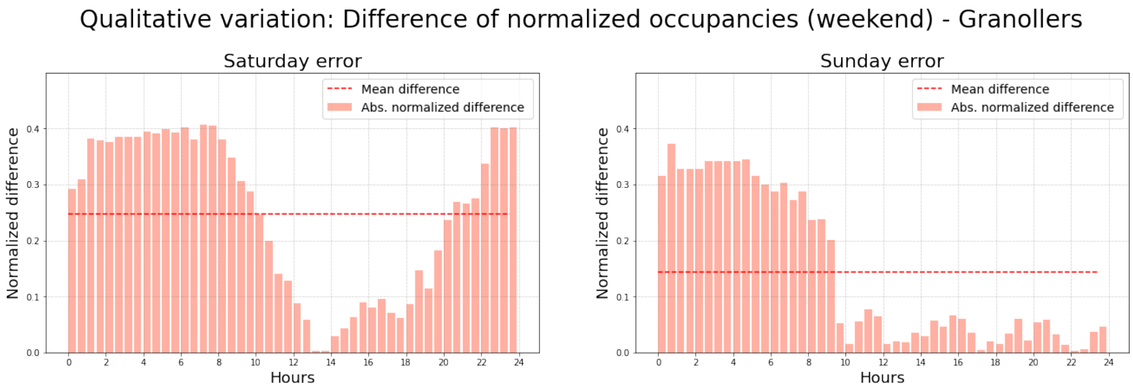

Figure 26: Qualitative difference between normalized data from pre-pandemic and pandemic occupancy during the weekends.

Looking at the weekends in the Figure 26, there is a much higher qualitative difference due to the instability and variability of their behaviour. So much so that Saturdays show the biggest qualitative difference in the whole analysis. Sundays, despite being slightly bigger than the weekdays, show a reasonable qualitative difference that allows as to state that the occupancy behaviour during the Sundays do not vary dramatically between both periods.

4.2.2 Change in the Volume of Vehicles

Once is clear where the main difference lies, it is important to quantify also the increase or decrease on the vehicle volume. To achieve this, I use the pre-lockdown data as a base, and I compute the proportion of the lockdown mean occupancy by hour for every single day. Prior to demonstrate the results, to understand what the data represented means I provide the following clarification: The closer the value is to 1, the less quantitative difference exists. If it's over 1, it means that the occupancy during the Covid-19 during these hours was higher than the pre-pandemic months. On the other hand, if it is under 1, that means that the occupancy during the Covid-19 was lower. Once clarified, the results for the quantitative difference in the parking of Granollers are:

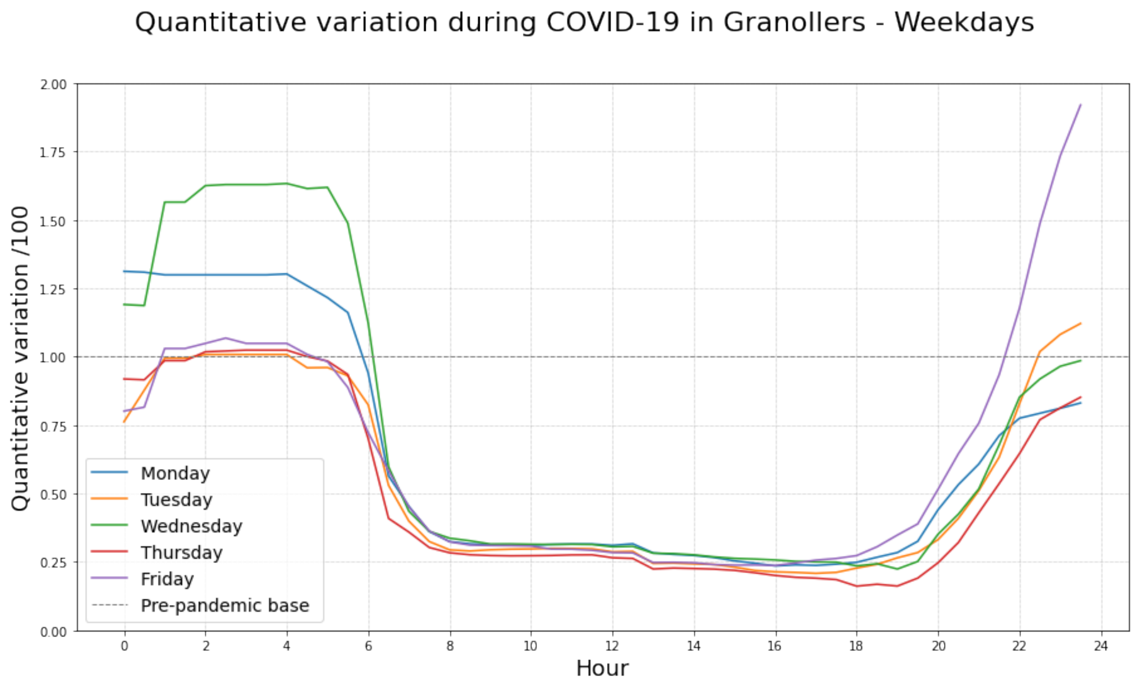

Figure 27: Quantitative difference between normalized data from pre-pandemic and pandemic occupancy during the weekdays.

Some conclusions that I can extract by observing the previous figure are:

- During the hours of increase of the occupancy. maximum capacity and decrease of occupancy (6:00h - 20:00h), we can state that the occupancy during Covid-19 was reduced by a 30% in average.

- Despite the statement of the previous bullet point, during the sleep hours, the occupancy has behaved really similar to the occupancy during the previous months of the pandemic. There are two exceptions, Mondays and Wednesdays, where more cars have stayed parked during the nights in average.

- It is also remarkable the difference of cars parked from 22:00h on Fridays, which is almost twice the number of cars parked during the pre-pandemic period.

Again, the weekends need to be treated separately, as the quantitative variation differs from the weekdays and Fridays.

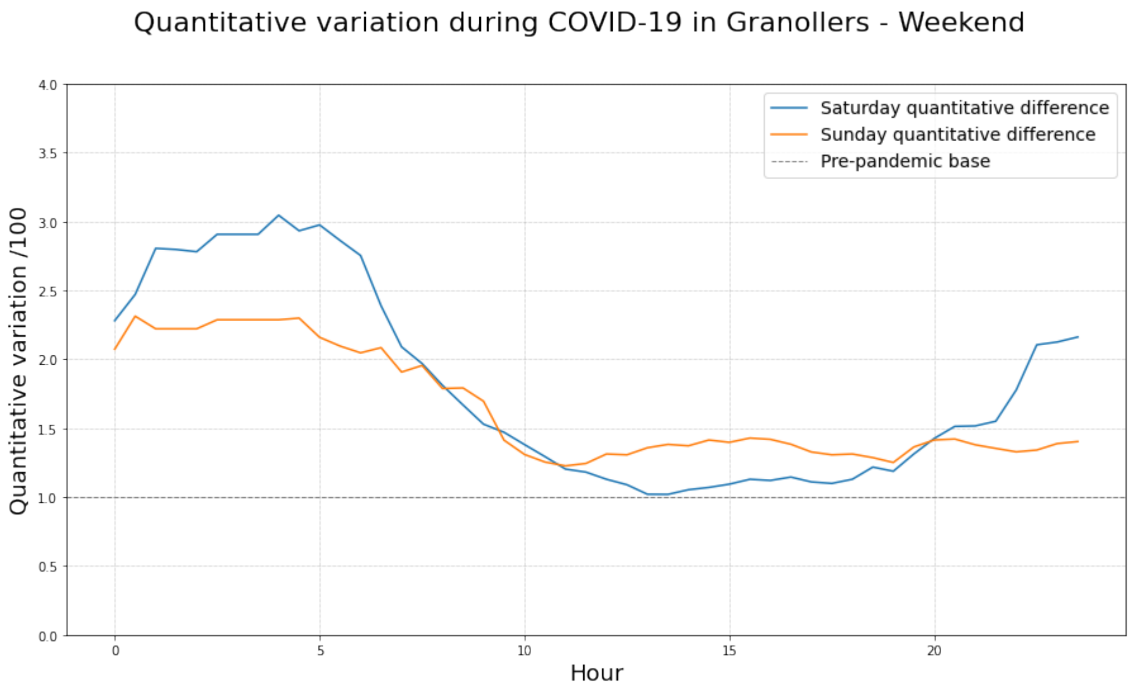

Figure 28: Quantitative difference between normalized data from pre-pandemic and pandemic occupancy during the weekends.

During the weekends of the Covid-19 period, I show in Figure 28 that the occupancy is higher than during the weekends of January, February and March, especially during the night hours.

As a conclusion for this chapter, I can state that the mean occupancy by hours of the different days of the week present differences. Nevertheless, this differences are more quantitative than qualitative. People have used car parks in the same way, but to a lesser extent. This can be due to the fact that some compulsory jobs have not changed the schedule, making the people follow the same routines and using the car parks the same way they were doing previously. This analysis is performed for different car parks, and the conclusion considers them as well.

Chapter 5

Conclusions and Future Work

This last chapter aims to collect and summarize all the results obtained from this study, as well as to provide a path to continue the research in the future.

5.1 Conclusions

The first part of the study consists in understanding human behaviour when it comes to parking usage in the area of Barcelona, using stored data from different car parks, which tend to be place close to a train stations. We observe that the occupancy follows a circadian cycle, really aligned with the average person routine. Parallel to this thesis, the behaviour of the occupancy is successfully approximated through a mathematical approach: an linear mixture of gammas. Using the output of [3], a highly explainable and understandable predictive model is designed and developed, having as its strengths the simplicity (in the good senses of the word), interpretability and transparency of the model and the low quantity of required data to train the model. I demonstrate that simpler models like this one can give results similar to those approaches which use RNN, or DL in general, which have an opaque layer that sometimes makes it difficult to interpret the model and the results.

I show that both, statistical modeling of prototypes using historical means, and mathematical modeling using an linear mixture of gammas, performs successfully, obtaining mean proportional error rates always under the 10%, and in a high number of the cases, under the 5%. Also, I am able to predict the occupancy of the future hours with a surprisingly high accuracy, despite testing the prediction from early hours, as can be seen in the *Real Time Prediction* section of the *Prediction* chapter. In these days, being able to predict parking occupancy is highly relevant in the society, as approximately the 30% of the pollution produced by cars, is generates by the search of free places for parking [10].

This study provides relevant information, as the general behaviour of parking occupancy does not change so much in the rest of the areas and countries, as can be observed in [4]. Parking occupancy behaviour is completely correlated with the circadian cycles.

Another interesting conclusion that I extract from this study is that the Covid-19 has had an impact in the parking occupancy, but that impact has not been as dramatic as one could expect, as we have observed in the fourth chapter.

5.2 Future Work

In the future, there is job to do due to some of the current constraints that we have been mentioning along this study. The pandemic has limited the amount of valid data, fact that has made our tests even more challenging. The optimal study would have consisted of a system which takes into account a whole year of parking occupancy, to better detect different patterns depending on the season of the year, and perform the training of the model every 3 weeks in order to keep the system updated to the latest parking occupancy trend. Thus, training and testing sliding windows of data (3 weeks of training and the last one of testing) and performing cross validation tests (using the 3 first weeks to train, and the last to test, and then change the testing week and the training weeks) are the next challenges.

Moreover, this thesis has been carried out in collaboration with the ATM, which aims to deploy and integrate the predictive model inside Mou-te application in the near future, in order to provide the results to the users. Hence, a better integration with the API and a more formal and robust deployment could be an interesting and useful line of work. Also, a dynamic implementation that allows to slide the training and testing windows easily, just as I have previously mentioned, would improve the system and would keep it more updated.

List of Figures

List of Tables

Bibliography

[1] Khreis, H. & Nieuwenhuijsen, M. Car free cities: Pathway to healthy urban living. *Environment International* **94**, 251–262 (2016).

[2] Comission, E. Mobility strategy | mobility and transport. `https://ec.europa.eu/transport/themes/mobilitystrategy_en`. (Accessed on 06/21/2021).

[3] Sánchez, J. F. Modelling the use of car parks in the province of barcelona. Tech. Rep., Universitat Pompeu Fabra. Master in Intelligent and Interactive Systems (MIIS) (2021).

[4] Ziyao Zhao, Y. Z., Yi Zhang. A comparative study of parking occupancy prediction methods considering parking type and parking scale. *Journal of Advanced Transportation 2020(10):1-12* (2020).

[5] Shao, W., Zhang, Y., Guo, B., Qin, K. K., Chan, J. & Salim, F. Parking availability prediction with long short term memory model. *Li S. (eds) Green, Pervasive, and Cloud Computing. GPC 2018. Lecture Notes in Computer Science, vol 11204* 124–137 (2019).

[6] Zheng, Y., Rajasegarar, S. & Leckie, C. 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)*, 1–6 (2015).

[7] Vlahogianni, E., Kepaptsoglou, K., Tsetsos, V. & Karlaftis, M. A real-time parking prediction system for smart cities. *Journal of Intelligent Transportation Systems* **20**, 00–00 (2015).

[8] Kaltenbrunner, A., Gómez, V. & López, V. Description and prediction of slashdot activity. In *2007 Latin American Web Conference (LA-WEB 2007)*, 57–66 (2007).

[9] Kaltenbrunner, A., Gómez, V., Moghnieh, A., Meza, R., Blat, J. & López, V. Homogeneous temporal activity patterns in a large online communication space. *International Journal on WWW/INTERNET* **6**, 61–76 (2008).

[10] Koumetio Tekouabou, C. S., ABDELLAOUI ALAOUI, E. A., Cherif, W. & Silkan, H. Improving parking availability prediction in smart cities with iot and ensemble-based model. *Journal of King Saud University - Computer and Information Sciences* (2020).