Master in Intelligent Interactive Systems
Universitat Pompeu Fabra

Modelling the Use of Car Parks in the Province of Barcelona

Josep Ferrer Sánchez

Supervisor: Vicenç Gómez

July 2021

Universitat Pompeu Fabra Barcelona

Contents

1 Introduction **1**

1.1 Motivation and Objectives 2

1.2 Dataset 3

1.3 Related Work 5

1.4 Structure of the Thesis 6

2 Preliminary Analysis **8**

2.1 Data Description 9

2.2 Preliminary Data Analysis 9

2.3 Analysing Data through Historical Means 11

2.4 Holidays and Outliers 14

2.5 Preliminary Analysis of the Renfe Rodalies Stations 16

3 Methodology **17**

3.1 Statistical Approach using Prototypes 18

3.1.1 Definition of the Prototypes 18

3.1.2 Prototypes Computation Using a Historical Mean Approach 19

3.1.3 Characterization of the Prototypes 21

3.1.4 Statistical Model Characteristics 25

3.2 Mathematical Approach (Gammix Model) 26

3.2.1 Derivation of the Model 26

3.2.2 Mathematical Characterization of the car park Occupancy 27

3.2.3 Steps to Characterize the Occupancy Behaviour: 29

3.2.4 Meaningful Explanation of the Gammix Modelling 35
3.2.5 Mathematical Model Characteristics 37
3.3 Autocomplete Model 38

4 Weather Influence **40**
4.1 First Intuitions 41
4.2 Influence of the Weather in the Occupancy Behaviour 42
4.3 Quantification of the Variation due the Weather Conditions 44
4.4 Concluded Influence towards the Occupancy 45

5 Conclusions and Future Work **46**
5.1 Conclusions 46
5.2 Future Work 48

List of Figures **49**

List of Tables **51**

Bibliography **52**

Dedication

I would like to dedicate this thesis to my mother, without whom nothing of this would have been possible.

Acknowledgement

I would like to express my sincere gratitude to:

- My supervisor Dr. Vicenç Gómez, who has always helped me out whenever I needed and has ensured that this thesis ended up like a success.
- My mother and my brother, who have always supported and motivated me no matter what.
- All my master team mates, with whom I have shared countless studying days and have always been there for any help I needed. Without them none of this would have been possible.

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 park station facilities located close to train stations. We propose and analyze different models based on statistical and mathematical techniques.

First, we analyze the occupancy recordings in different park station locations and show that the activity is strongly coupled with the circadian rhythm, following a 24-hours cyclic pattern. Second, we perform a statistical characterization based on different activity prototypes for days associated with similar activity. We show that such a simple statistical model is enough to characterize accurately the car park occupancy. Third, we propose a mathematical model that explicitly captures the rise and decay of occupancy using a mixture of two different Gamma kernels. We optimize the parameters for each park station independently and show that the resulting model globally captures better the activity than the previous statistical characterization, in terms of accuracy, simplicity (has less parameters) and interpretability.

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 this dynamics to a large extent. 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: Car Park, Car Park Occupancy, Prediction Model, Gamma Kernel.

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 park station 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 car park 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 David Moreno [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 *predicting* the occupancy behavior, this thesis is focused on the problem of *modeling* the data. For this reason, during the *Preliminary analysis* chapter the *we* form is used as a plural, and for the rest of the chapters the *I* form is used as the first person.

1.1 Motivation and Objectives

The motivation and objectives are

1. **To understand the behavior of car park occupancy**: A first direct goal is to understand how this 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 car park that is being used is to be considered as well.

2. **To design simple and accurate models of car park occupancy**: Train an effective model to predict the occupancy behaviour in time. A first predictive model based on a statistical historical means and precomputed temporal profiles is generated to analyze and extract the dynamics and behaviours of the car park occupancy. A second mathematical method to further simplify the statistical one is to be performed as well. To do so, both testing and training datasets will be used to train these models and test them directly on data.

3. **To detect external influences on the occupancy behaviour**: Try to detect the influence that the weather might produce on the occupancy behaviour of these car park lots.

4. **To detect errors in the stored data.**: Detecting errors that are present on data, finding outliers and specific days that might present some characteristic distortion of the pattern due some known or unknown reason.

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

1.2 Dataset

For this project we use several datasets coming from different car park spots. A total amount of 10 car parks have been used, each of them located in different cities within the province of Barcelona. The location together with the maximal number of car park spots that each of them presents is listed below: 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 locations can be observed in the following map:

Figures/Introduction/park stationLocation.PNG

Figure 1: Map of Barcelona province with the park station 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 park station, 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 mins.
- **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 car park and the specific number of Free slots for each interval, due to the fact that we want to measure the occupancy.

	DateTime	Parking Vilanova Renfe plazas totales
0	01/01/2020 0:00	425,5705639
1	01/01/2020 0:30	425,0122716
2	01/01/2020 1:00	425,566
3	01/01/2020 1:30	424,6528307

Figure 2: Original Dataframe provided by ATM for the Vilanova car park location

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 car park slots or occupancy in some specific car park locations. Some few examples can be easily found when checking some keywords such as *car park occupancy prediction* or *park station occupancy modelling* directly in google scholar. Hence, it is quite clear that park station 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.

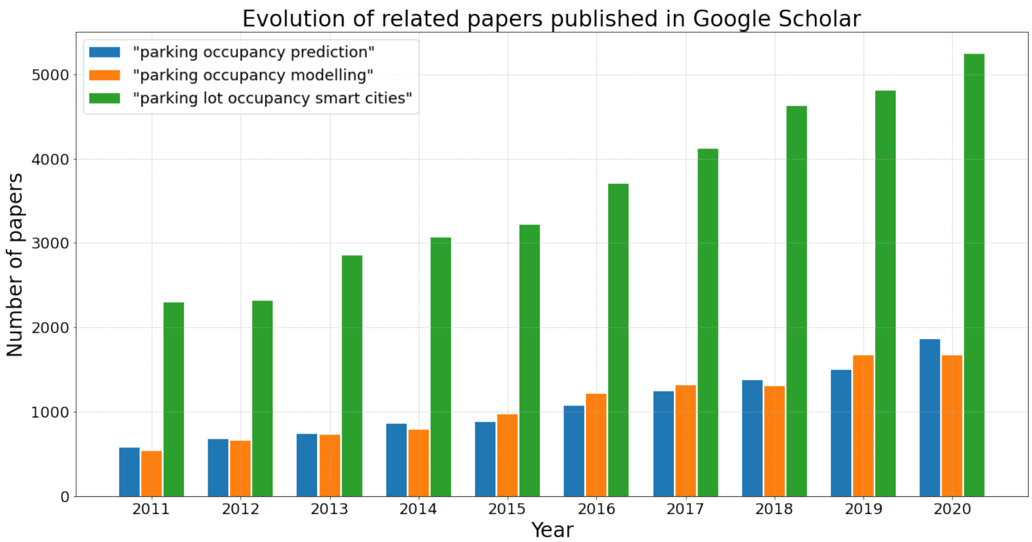

Figure 3: Evolution of the number of papers published related to the keywords.

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 will be 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 is cleaned and filtered (to sort out outliers or mistaken data) to have a better data set to train and test the model.

2. **Methodology**

   Once the data is already cleaned to start working with it, different methods are developed to achieve our main goal: Predict future behaviour. Two final methodologies are used: Historical means of data and a Gamma mixture behaviour fitting.

   - **Statistical approach using historical mean Prototypes:** Similar data is detected to define and generate different prototypes that can capture the main characteristics of each group of similar days.
   - **Mathematical approach:** The previously detected prototypes are parametric using mathematical functions. Thus, both a mathematical explanation of the behaviour and an easier model can be achieved.
   - **Autocomplete Model:** To predict occupancies in the future, we can use some partial already known data to improve even more the predictions.

3. **Influence of weather on occupancy data** Some external factors as weather are considered to understand better what is the influence they have on the occupancy behavior of the different park stations.

4. **Conclusions and Future Work** A final Chapter of this thesis is dedicated to extract the main conclusions of the whole research work and how to proceed in the future to improve the current model.

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 car park spot. In this analysis, we aim to understand how the occupancy varies over time, and characterize the statistical properties (mean and variance) of each car park, as these factors directly affect to the quality of the occupancy prediction for each park station lot. 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 has not been 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 core dataset wrong data that might bias or affect a little bit the future predicting model, understanding as a wrong data some days with huge lack of data, or some days where the sensor of a particular park station lot has not worked properly. To work with the available data, both the influence of holidays days and the Covid-19 pandemic have been considered as well, omitting these days from the study and comparing them later with some obtained behavior, as it will be 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 starting at 00:00h every midnight and ending at 23:30h within the same day. This procedure is repeated for every day cycle and presents a synchronous measurement system for all park station locations.

For a practical understanding, this is quite useful, as when working for instance with Vilanova, the resolution and intervals will be 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 park station, which eases the visualization and the understanding process.

All procedures and computations showed throughout this thesis have been implemented in all car park datasets. For logical coherence and a better comprehension of the reader, within this thesis will be only displayed results for the Vilanova and the Quatrecamins car parks. In some specific cases, other car parks can be used to illustrate some meaningul examples. However, all results and conclusions can be generalized to all available car park 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.

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 as well. Weekends have a really low occupancy and present a small variance. When focusing on Fridays, it is noticeable that present a quite particular behavior: It is a

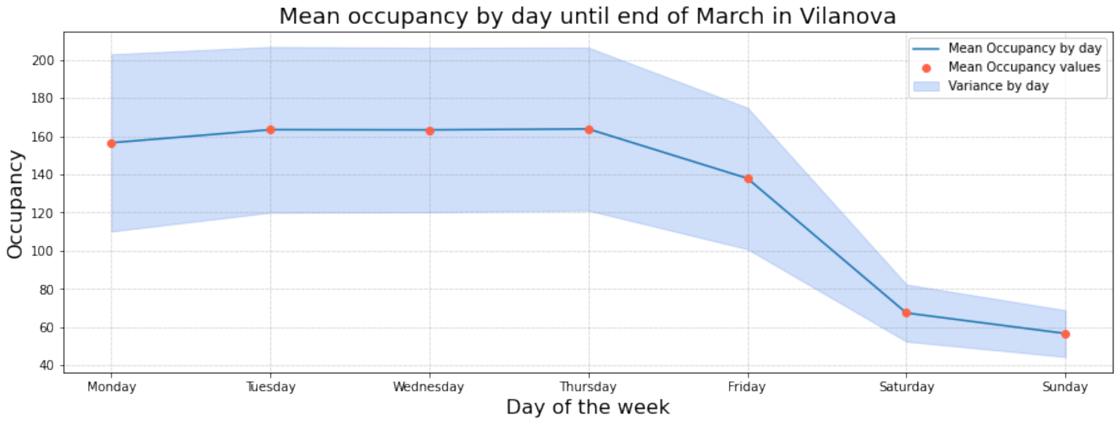

Figure 4: Mean occupancy and variance for each weekday from January to March of 2020 in the car park of Vilanova.

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 car park along time.

Figure 5: Cyclic behaviour of the Vilanova car park for the week starting from Monday 24th of February to Sunday 2nd of March 2020.

Figure 5 clearly shows how the occupancy behaviour presents a cyclic evolution with time: The same pattern is repeated every single day, having some particularities following the initial guesses that were 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.

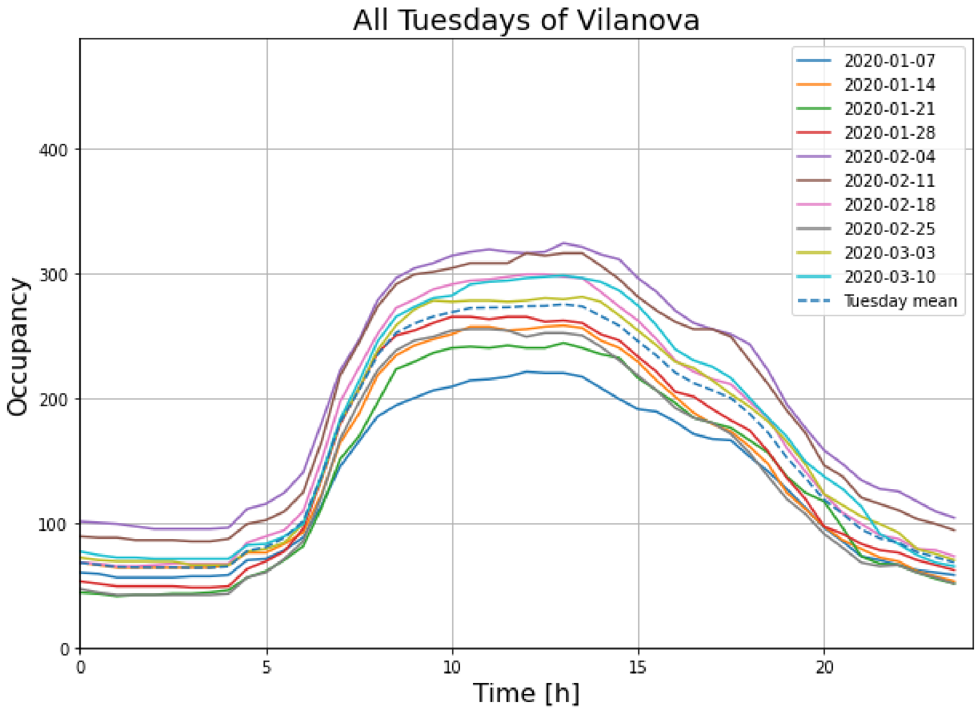

Figure 6: 24-hour occupancy cyclic behaviour for Tuesdays in the car park of Vilanova.

As it can be observed in Figure 6, all Tuesdays do behave similarly and the historical mean captures properly this behaviour. Even though the total amount of occupied car park slots do vary (the volume of entering cars), the behavior of the cycle are always the same.

2.3 Analysing Data through Historical Means

As some first intuitions were already achieved and we have observed that it makes sense to use historical means to understand better the behaviour of the data, we can move forward trying to visualize the mean behaviour of each day. This means we check how the occupancy evolves along hours for each type of day.

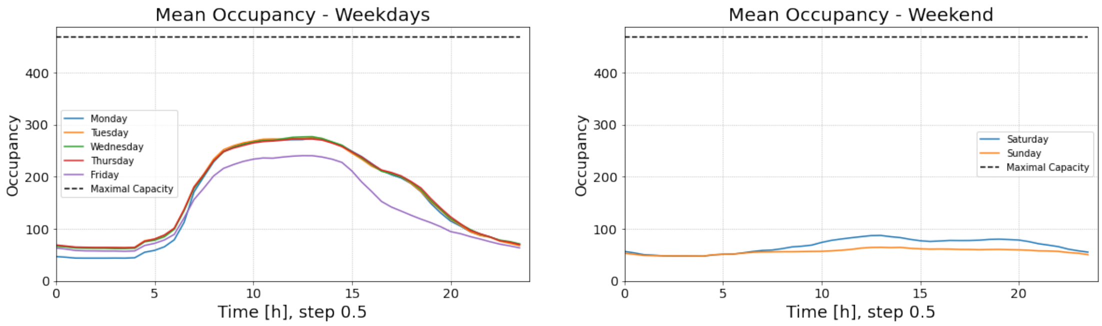

Figure 7: 24-hour occupancy cyclic behaviour from Monday to Sunday in the car park of Vilanova.

As it can be easily observed in Figure 7, where the historical mean for each type of day for the Vilanova park station are displayed by hours, it does not exist any significant activity or variation during the first 5 hours approximately. The existing baseline occupancy 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 park station 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 park station increases 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 is repeated as well.

Now, focusing on the differences between the different type of days we can conclude there are three completely different behaviours. Whereas weekdays from Monday to Thursday present a high occupancy value, Fridays present a lower one and a narrower stable peak. Finally, Weekends are characterized for a low flat occupancy with no much variation. 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, as it was expected, 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 Vilanova data, is that the occupancy never reaches the maximal capacity of the park station. To show that some park station data can saturate to a maximal capacity value and to show that the behavior explained so far can be generalized for other car parks as well, the same graph that can be found in the Figure 7 is displayed in Figure 8 but in this case using the Quatrecamins car park dataset.

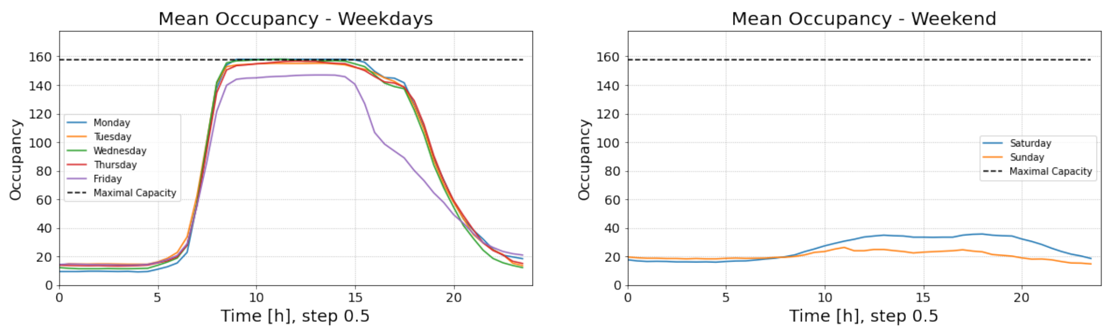

Figure 8: 24-hour occupancy cyclic behaviour from Monday to Sunday in the car park of Quatrecamins.

As it can be observed again in this second park station, all three behaviors can be differentiated as well. An important appointment is that while Weekdays mean behavior from Monday to Thursday saturates, Fridays do not do so.

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.

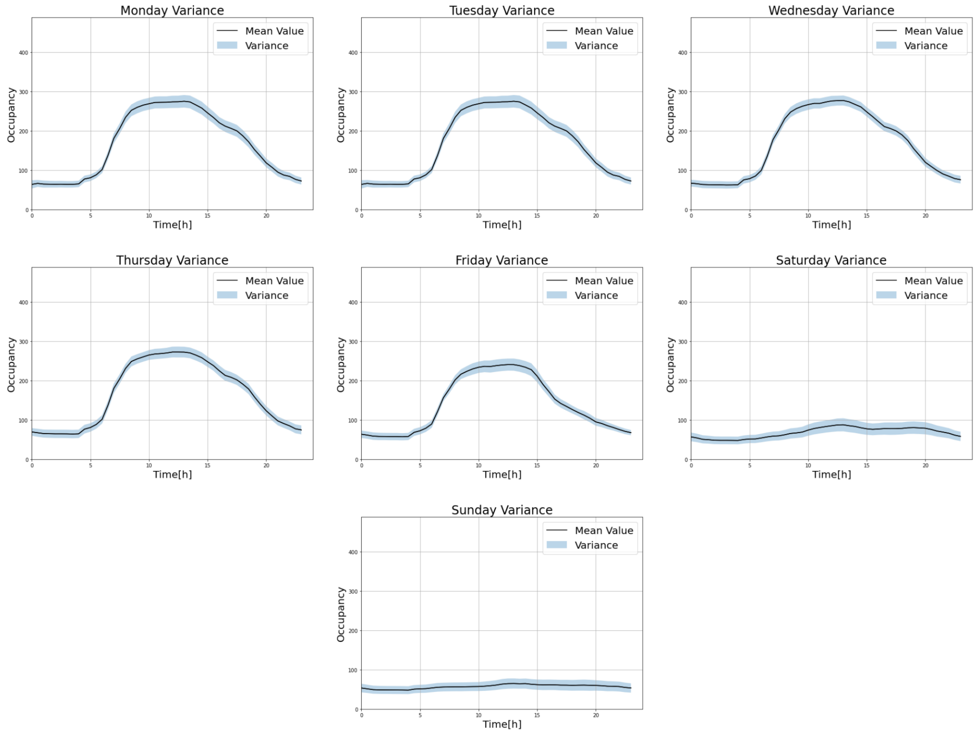

Figure 9: Historical mean for each type of day with its corresponding Variance for the Vilanova park station

Analyzing the Figure 9, it can be observed that the higher variance values (represented in semitransparent blue) are present during the hours the park station is highly occupied, from 8:00h to 16:00h 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 have been ommited during this first preliminary analysis and during the whole model development. This final subsections is aimed to explain the existence of this abnormalities in the dataset and how to treat them. The corresponding graphs attached to illustrate the different types of outliers correspond to the Vilanova dataset.

At first, we differentiated between four different types of abnormalities:

1. Profiles with an extremely low occupancy corresponding to holidays.
2. Profiles with low occupancy due Covid19 restrictions.
3. Profiles with low activity due some external but known phenomena.
4. Profiles with a weird behaviour due unknown factors.

In the first two cases, all of them resemble the behaviour of a weekend, recovering a similar historical mean as it can be observed in the left graph of the Figure 10. In this case, all outliers present a flat occupancy behaviour with a small increase of the occupancy during some central hours.

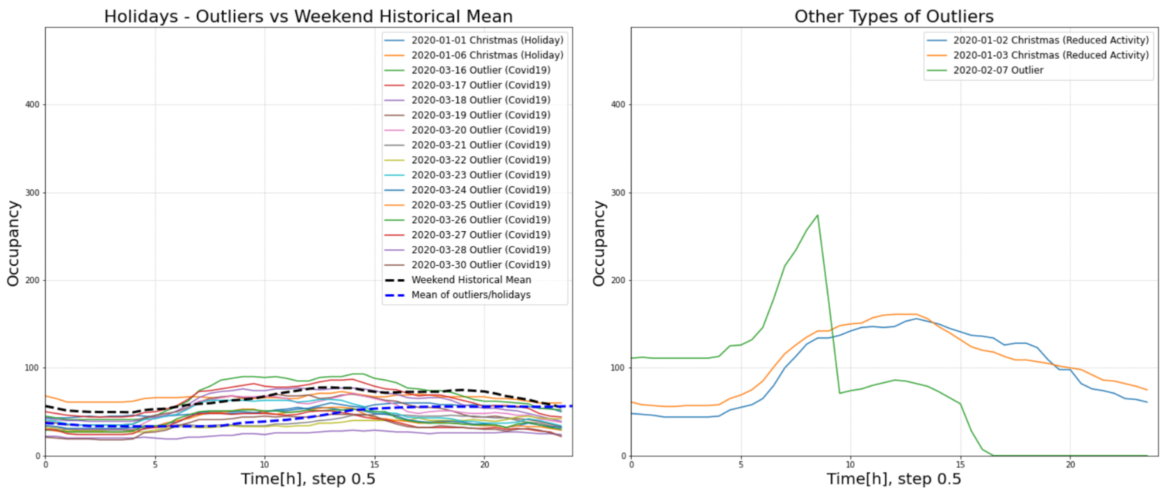

Figure 10: Some other abnormal days not used during the thesis

The last two types of outliers stated before can be observed in the in the right graph of the Figure 10. In this case, both 2nd and 3th of January are not holidays, but as they are amid the Christmas times, present an abnormally low occupancy. However, these both days still do present an activity profile, but with a too low amount of cars to be considered as normal days. A final outlier is found the 7th of February. In this last case, the cause is unknown and I assume it is due some sensors error.

2.5 Preliminary Analysis of the Renfe Rodalies Stations

As a parallel task, we want to know how the park stations we are working with are related to train stations, and furthermore, to observe how incidents in some train stations or even on other park stations may affect those. The main challenge we faced with this task is that there is no dataset available representing the different stations and their connection, so we have built this dataset manually.

In the Figure 11 we can observe the network of the train stations, visualized with Gephi and using a GeoLayout, which helped to interpret such network and how close the train stations are from the car parks that we are working with. The names of each train station as well as the location of the car park lots are not shown due difficulty to read them in such small dimension.

Figure 11: Network of car park lots related to their corresponding train stations

Chapter 3

Methodology

Once the preliminary analysis is ended, and some useful information has been achieved, I design the process to follow to predict the occupancy of the car parks in the near future. A first logical step is trying to model the actual behavior of occupancy through different groups of similar days and try to characterize it as much as possible. Afterwards, all this modelling can be implemented to predict future days.

For this purpose, three different phases have been sequentially performed:

1. **Statistical Approach using Prototypes:** In this first phase, similar data are detected to define and generate different prototypes that could capture the main characteristics of each group of similar days.

2. **Mathematical approach:** In this second phase, the previously detected prototypes are parametrized using mathematical functions. Thus, both a mathematical explanation of the behaviour and an easier model can be achieved.

3. **Autocomplete model for the current day:** To predict occupancies in the future, some partial already known data can be used to improve even more the predictions. This final step is about discussing how to do so.

3.1 Statistical Approach using Prototypes

3.1.1 Definition of the Prototypes

Throughout this thesis the concept of Prototype is used quite often. Hence, a brief definition of what is considered as a Prototype is necessary to understand the model that is developed in the coming sections.

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

Previous preliminary observations have leaded me to determine that the available data should be grouped into 3 different similar behaviours: Weekdays (until Thursday), Fridays and Weekends. The differentiation between the Weekends and the rest of the days is quite obvious, as one can easily observe that there exists a completely different behavior for both Weekdays and Weekends. However, the differentiation between Weekdays and Fridays is slightly more subtle. Fridays behaviour seem to be a mid-step between Weekdays and Weekends.

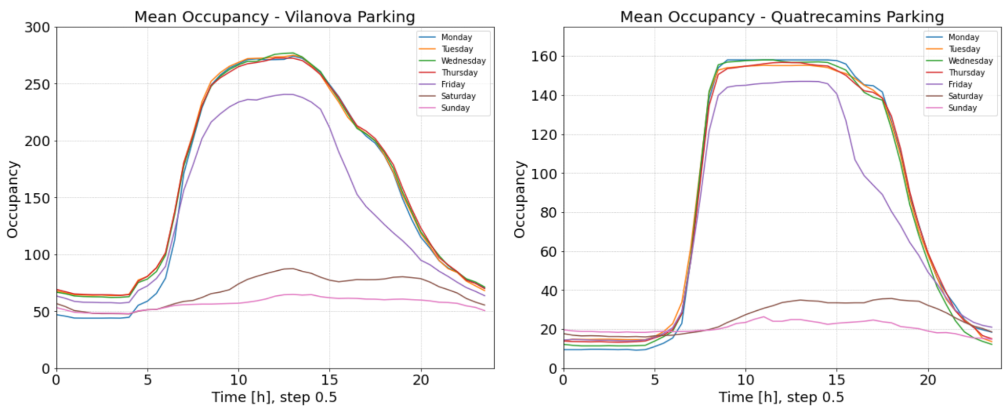

Figure 12: Comparison of the historical mean for each day. Left graph corresponding to Vilanova and right one to Quatrecamins. Notice the scale has been adapted to each car park to observe better different days behaviours.

As it can be observed in Figure 12, when comparing the behavior of different day types in a deeper way, it can be spotted a clear difference between Monday to

Tuesday and Fridays. The following three-group classification is justified as follows:

- **From Monday to Thursday:** Their corresponding occupancy functions coincide, presenting a wider stable region with a high occupancy from 9h to 16 and presenting a first slower decrease around 15h and a second faster and shorter decrease around 16. Their maximal occupancy value coincide as well, presenting all of them almost the same highest occupancy. In the case of Quatrecamins, all four weekdays saturate, presenting the maximal capacity of the car park.

- **Friday:** Even though presents a high occupancy value, it is still lower than the rest of the weekdays. Furthermore, the width of stable region is narrower, presenting a single slower but longer decrease starting around 15h. When focusing on the model displayed in the Quatrecamins car park, unlike the rest of the weekdays, it never reaches the maximal capacity of the car park.

- **Weekends:** The main characteristic is the low occupancy both Sunday and Saturday present. Compared to the weekdays, the occupancy seems quite flat. However, later on we will figure-out on a minor-scale they present some dynamic pattern as well.

After several computations and comparisons with real data, the differential characteristics of Fridays are confirmed. Treating Fridays as a particular prototype improve the accuracy of the characterization of the resting weekdays as well as its own accuracy. This make sense, as Fridays are a special day, where the routine of the people tend to variate slightly, as for instance, the school schedule, or working hours.

3.1.2 Prototypes Computation Using a Historical Mean Approach

The first direct approach to compute a prototype that captures the behaviour of the data is computing the historical mean of the three already defined similar groups:

Weekdays (Until Thursday), Fridays and Weekends. Hence, a historical mean prototype for each group is obtained.

The process to compute the prototypes is the following one:

1. Erase all holidays and outliers from the dataset.
2. Classify each of the available days from January to the 3rd week of February into Weekday, Friday or Weekend.
3. Separate each category and compute the corresponding Historical Mean

To do so, an additional column is generated within the dataset labelling each day within one of the three predefined groups.

After this procedure is performed, a final prototype is obtained for each type of day.

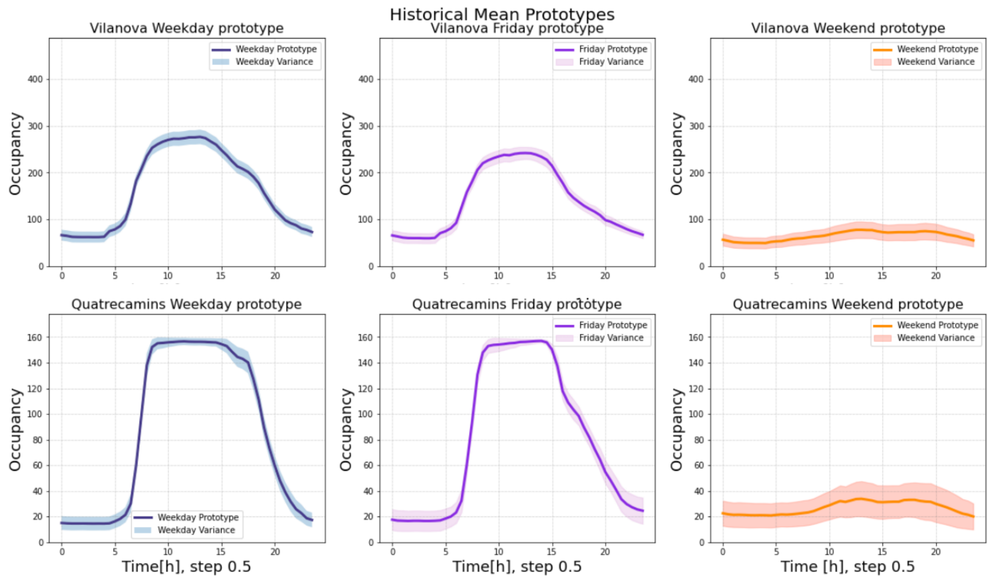

Figure 13: The three different prototypes obtained for both Vilanova and Quatrecamins Datasets.

As it can be observed in the Figure 13, all prototypes resemble their corresponding real days, as their variance is quite low. To ensure that all prototypes are accurate enough, a comparison between real data from the last week of February and the prototype is performed. This comparison can be observed in the Figure 14.

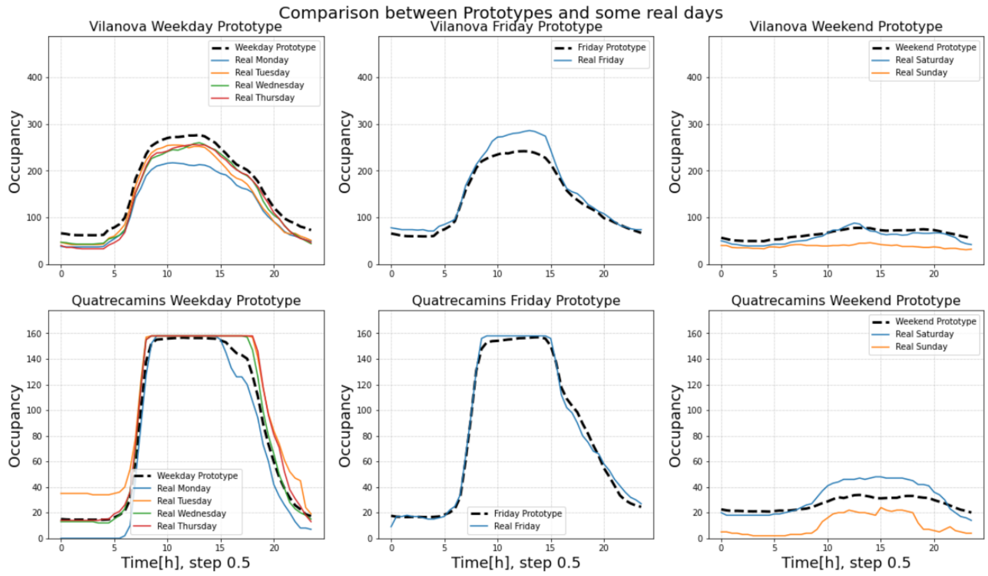

Figure 14: Comparison of each Prototype with real days corresponding to them for both Vilanova and Quatrecamins Datasets.

After the comparison, it can be seen that all prototypes resemble a lot to their corresponding real days. In the case of both real Mondays for Quatrecamins and Vilanova, although they presents a lower stabilization, the behaviour still remains wider than the corresponding Friday Prototype and with a later decrease as well. Thus, observing these results it can be concluded that the three-prototype categorization of all available days is quite accurate and can allow us to obtain good predictions in the long-term.

3.1.3 Characterization of the Prototypes

Once all three prototypes are obtained, the next logical step is to characterize each of the prototypes to understand better how is the behaviour of the occupancy cycle and how to model this same behaviour mathematically following some parametrical function. All three Prototypes present four different regions that are easily characterizable and are found in all their corresponding day [24 hours] cycles.

Weekday Prototype (from Monday to Thursday)

The main characteristics of the Weekday Prototypes are that weekdays usually present the highest maximal values and have a wider cycle with a faster and shorter decrease. All four regions will be explained as follows and can be observed in the Figure 15:

1. **[00-06h]** ***Minimal Occupancy:*** Corresponding to the number of cars that remain parked within the car park during the whole night. The minimal value that the occupancy will take though out the whole cycle. From now on, this minimal value will be referred as the *baseline* occupancy.

2. **[06-09h]** ***Fast Increase:*** A first rapid increase that coincides with the first flow of people entering the car park lot to park their cars. From 6 to 9h, the occupancy raises rapidily until it takes the maximal value of occupancy. In the case of Quatrecamins, this maximal occupancy coincides with the maximal capacity of the car park.

3. **[09-16h]** ***Stabilization Zone:*** After taking the maximal value, the occupancy value stabilizes for a long time-span. From 9 to 16h, the park station occupancy remains stable.

4. **[16-21h]** ***Fast Decrease:*** After 16h, the occupancy starts to decrease rapidly, coinciding with he maximal flow of people who leave the car park.

5. **[21-00h]** ***Minimal Occupancy:*** Once the occupancy has decreased, the occupancy gets back to the minimal value. This value will remain constant until the beginning of the following cycle.

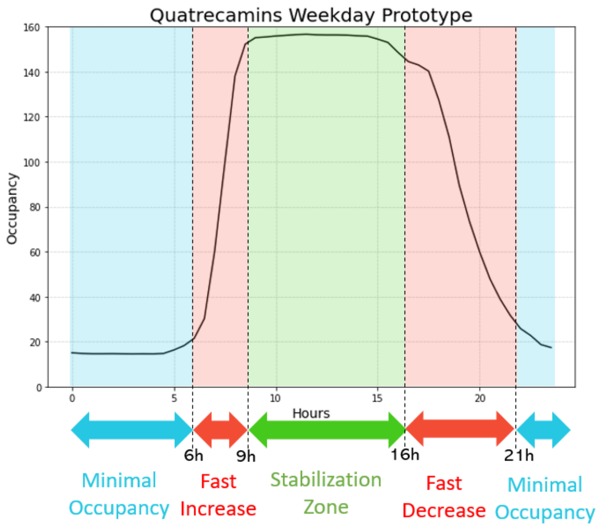

Figure 15: Region-characterization of the Weekday Prototype.

Friday Prototype

The main characteristics of the Friday Prototype are that present a narrower cycle with a slower and longer final decrease that starts earlier as well. The final prototype and its corresponding regions can be observed in the and can be observed in the Figure 16

1. **[00-06h]** ***Minimal Occupancy:*** Corresponds to the baseline occupancy of the car park.

2. **[06-09h]** ***Fast Increase:*** A first rapid increase that coincides with the first flow of cars entering the park station lot. From 6 to 9h, the occupancy raises rapidily until take the maximal value of occupancy. This first increase coincides with the one presented by the rest of the weekdays.

3. **[09-15h]** ***Stabilization Zone:*** After reaching the maximal value, the occupancy value stabilizes for a long time-span.

4. **[15-21h]** ***Fast Decrease:*** After 15h, the occupancy starts to decrease rapidly, coinciding with he maximal flow of people who leave the park station. This

final decrease starts earlier and takes longer to reach back the baseline value of the occupancy.

5. **[21-00h]** ***Minimal Occupancy:*** After the maximal occupancy value, the baseline occupancy value is achieved until the beginning of the following cycle.

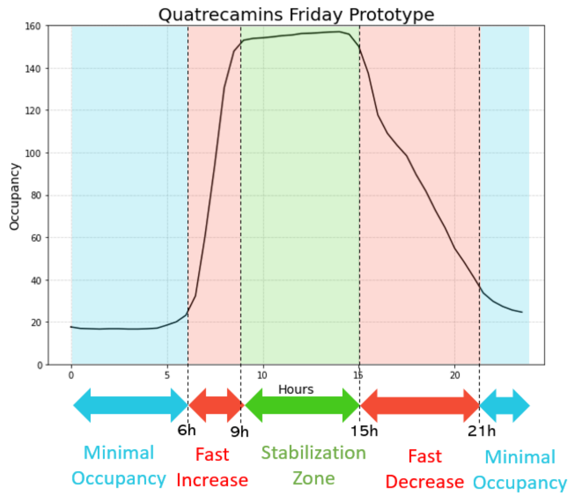

Figure 16: Region-characterization of the Friday Prototype.

Weekend Prototype

The Weekend Prototype presents the most different behavior. The occupancy maximal value is much lower, and the temporal variations are harder to be detected. However, when analysing in deeper detail, the same four regions can be found as well, as it can be observed in the Figure 17.

1. **[00-06h]** ***Minimal Occupancy:*** Corresponds to the baseline occupancy value of the car park.

2. **[06-09h]** ***Fast Increase:*** Corresponds to a first slight increase.

3. **[09-19h]** ***Stabilization Zone:*** After reaching the maximal value, the occupancy value stabilizes on this value for a long time-span. In this case, the stabilization zone presents some litter variations, not being as stable as for the rest of the days.

4. **[19-21h]** ***Fast Decrease:*** After 19h, the decreases again. This final decrease starts way later than the rest of the prototypes ones.

5. **[21-00h]** ***Minimal Occupancy:*** Once the decrease stops, the baseline value is achieved until the beginning of the following cycle.

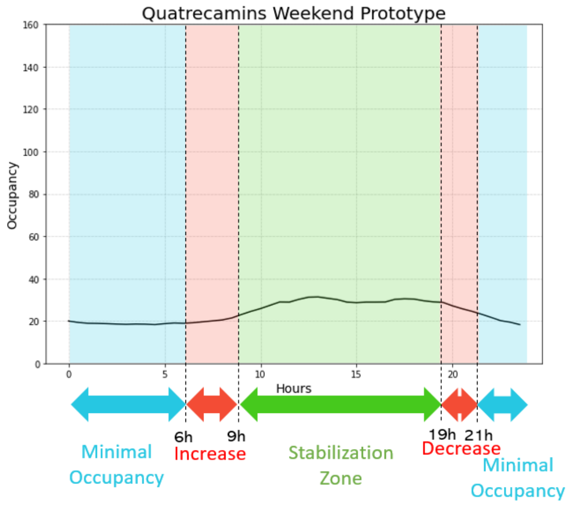

Figure 17: Region-characterization of the Weekend Prototype..

3.1.4 Statistical Model Characteristics

As the first conclusion of this first part of the methodology, it can be stated that the statistical approach is more than enough to characterize and replicate the historical behaviour of the occupancy. The main conclusions are:

1. Prototypes can be grouped to capture "similar" days.
2. Three main Prototypes are enough to explain the available data: Weekday (Monday to Thursday), Friday and Weekend.
3. Real data tend to be similar to their corresponding prototype.
4. It is important to notice that to characterize each of the prototypes, 48 values corresponding to every half an hour within a day cycle need to be stored.

The performance of the statistical model can be checked in David Moreno msc. thesis [3].

3.2 Mathematical Approach (Gammix Model)

Once all data is categorized under similar-behaviour prototypes (Weekdays, Fridays and Weekends), the next logical step is improving and simplifying the model by characterizing it in a mathematical way that requires a smaller amount of stored parameters, but exploiting at the same time the idea that the available data can be easily separated in three differentiated groups.This is why, a second approach is to identify some mathematical explanation to the behaviours previously obtained.

I perform some first attempts trying to find a good parametric approximation that could imitate the behaviour of the available data. Two different profiles are considered: Gamma and Gaussian. The Gamma profile turns to be way better to capture the behavior presented by the occupancy data. This is why, some first approximations are already performed using a single Gamma kernel. However, the usage of a single Gamma kernel is far too simple to approximate most of the available data. A mathematical approach is to be performed to define in a parametric way the behaviour of the available data. In this case, the main approach is inspired by the prediction models used in [7, 8].

3.2.1 Derivation of the Model

Assuming that cars arrive to a specific car park randomly, a specific car remains parked occupying a car park slot for an unknown time span and eventually leaves the car park lot.

Following the previous idea, it can be considered that the occupancy behavior pattern observed in the available data does contain the corresponding probability of a free slot being occupied. It is worth mentioning that the main goal is to predict free slots. This means, I consider as a single static car whenever some car arrives and at the same time some other car leaves, making no change on the occupancy value.

This probability can be achieved when removing the baseline value of the occupancy profile and normalizing its area, obtaining a normalized probability distribution corresponding to the probability of an arrival of a random car (and a consequent increase or decrease of the occupancy value). This probability can then be approximated using a modified Gamma kernel to obtain a mathematical parametric function. This process is explained with deeper details in the coming sections.

3.2.2 Mathematical Characterization of the car park Occupancy

I consider the gamma distribution, parameterized by the shape parameter α and scale parameter β, to characterize the temporal profile of occupancy.

$$\mathrm{Gamma}(x; \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta^{\alpha}). \tag{3.1}$$

Figure 18 shows the gamma distribution function for different values of α and β. The fact that it is an asymmetrical distribution makes it is convenient to model the occupancies over time.

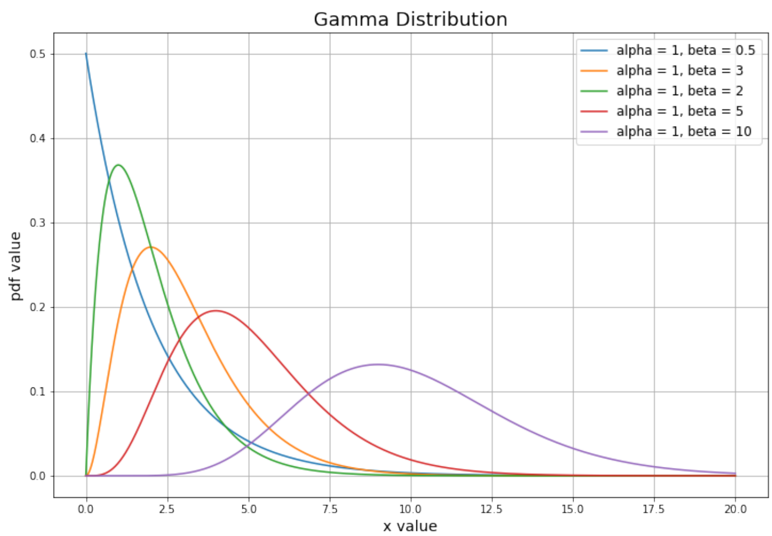

Figure 18: Different gamma kernel behaviours depending on α and β parameters.

Having three different groups of similar days, and after trying different functions and fittings, I end up concluding that the best way to characterize each prototype is following two different approaches. Both of them are referred as the *Gammix model.*

Gammix Model Approach I: A Single Gamma

In the case of Weekends data, I use only a rescaled gamma since their temporal profile is simpler. In this case, an amplitude w is added to the function to modulate the corresponding function shape.

$$\text{Gammix}(x; w, a, b) = w \cdot \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-b^a) \tag{3.2}$$

In this first case, a parameter vector θ is defined containing all the required values to generate the corresponding *Gammix* function for the Weekends.

$$\theta = [w, a, b] \tag{3.3}$$

Gammix Model Approach II: Gamma Mixture for Weekdays and Fridays complex behaviours

In the case of dealing with Weekdays and Fridays, the corresponding occupancy behavior is more complex, presenting 4 completely differentiated regions and having a high variation of free slots throughout the whole 24h cycle.

Hence, a mixture of two different Gamma kernels are to be used to fit the Weekends and Fridays data. Each Gamma kernel presents its corresponding amplitude w as well, to have more parameters with which obtain a better fitting for the function.

$$\begin{aligned} \text{Gammix}(x; w_1, a_1, b_1, w_2, a_2, b_2) \;&=\; w_1 \cdot \frac{b_1^{a_1}}{\Gamma(a_1)} x^{a_1-1} \exp(-b_1^{a_1}) \\ &\quad + w_2 \cdot \frac{b_2^{a_2}}{\Gamma(a_2)} x^{a_2-1} \exp(-b_2^{a_2}). \end{aligned} \tag{3.4}$$

In this second case, a parameter vector θ is defined containing all the required values to generate a corresponding *Gammix* function for each Weekdays and Fridays group of similar days.

$$\theta = [w_1, a_1, b_1, w_2, a_2, b_2] \tag{3.5}$$

3.2.3 Steps to Characterize the Occupancy Behaviour:

In order to convert the occupancy behaviour to a probability and characterize it using the *Gammix function*, some steps are required to be performed.

I. Erasing the baseline of each day:

The first step is erasing the baseline value of the occupancy corresponding to each cycle profile. When having baseline value different than zero, there is no probability that any car park slot can be occupied, as it corresponds to static cars parked within the car park spot during the whole day. In the case that a car substitutes another having no influence on the free slots, they are considered as a static car as well. Hence, this baseline needs to be erased, making every single day occupancy cycle starting at zero in order to maintain the occupancy behavior that contains probability information related to a free slot being occupied.

In the Figure 19, one can observe the original occupancy behavior and the modified one where the baseline value is removed. The baseline value changes for each 24-hour cyclic profile, as it is the minimal occupancy value of each daily profile.

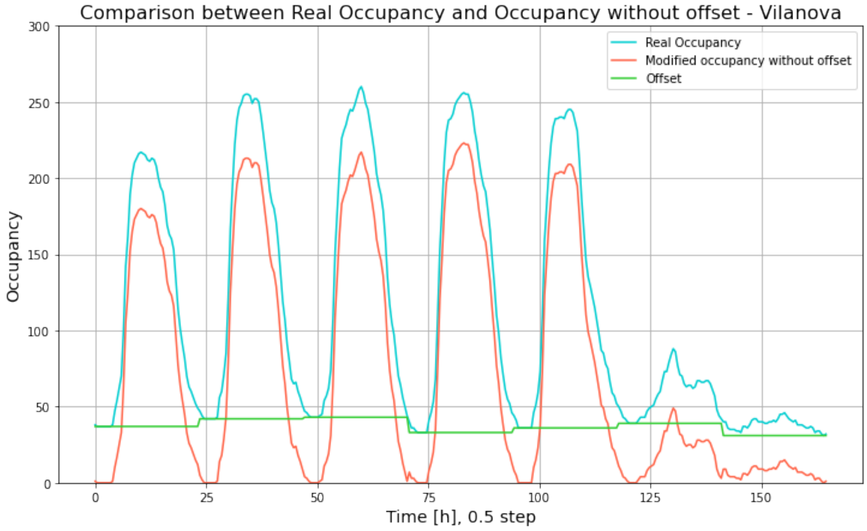

Figure 19: Comparison between real occupancy and modified occupancy where the baseline value corresponding for each cycle has been removed.

In order to keep some approximated baseline value for recovering afterwards the occupancy approximation from the *Gammix* probability prototype output, a mean of all baseline values is performed to be used later. Assuming that a specific group of similar days presents N total individual days, and each cycle day i presents a baseline value $baseline_i$, the corresponding mean for each groups is computed as follows:

$$\text{baseline}_{mean} = \frac{1}{N}\sum_{i=1}^{N}(\text{baseline}_i) \tag{3.6}$$

In the Table 1, the obtained baseline mean values for both Vilanova and Quatre-camins are displayed:

Table 1: Corresponding baseline mean value value for each prototype.

Prototype	**Baseline Vilanova**	**Baseline Quatrecamins**
Weekdays	57.69	9.97
Fridays	53.89	13.2
Weekends	47.05	14.37

II. Normalizing each cycle occupancy data:

Once the baseline value is already removed, the next logical step is normalizing the corresponding profiles by their area, which means their corresponding area equals 1. To do so, the *integrate.simpson* algorithm from the *scipy* library of python is used. This predefined algorithm of python allows to get directly the integration of a given vector summing up all its components. Hence, using the *integrate.simpson* algorithm, each cycle profile can be individually integrated and obtain its corresponding area. Once this procedure is repeated for each 24-hours cycle, each of them can be normalized being divided by its area. In the Figure 20, the obtained normalized occupancy can be observed.

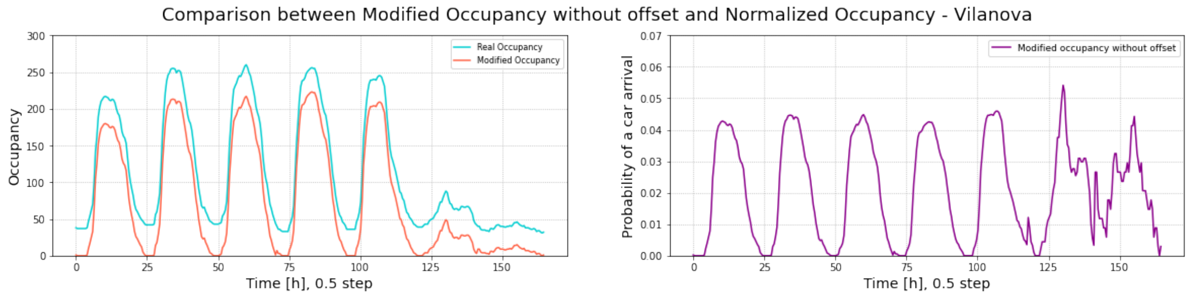

Figure 20: Comparison between modified occupancy and normalized occupancy. Each day cycle has been normalized individually.

In order to keep some approximated area value for recovering afterwards the occupancy approximation from the *Gammix* probability prototype output, a mean of all areas values is performed to be used later as well. Assuming a specific group of similar days presents N total individual days, and each cycle day i presents an area $area_i$, the corresponding mean of the area for each group is computed as follows:

$$\text{Area}_{mean} = \frac{1}{N}\sum_{i=1}^{N}(\text{Area}_i) \tag{3.7}$$

In the Table 2, the obtained mean Area for both Vilanova and Quatrecamins are displayed. In this case, this mean area values are used to convert the Gammix function probability output back to occupancy.

Table 2: Corresponding area mean value for each prototype.

Prototype	**Area Vilanova**	**Area Quatrecamins**
Weekdays	4961.59	3516.66
Fridays	4076.99	3082.10
Weekends	754.40	435.21

III. Model Fitting:

Once all the available profiles are without a baseline value, normalized and categorized into their corresponding group of similar days, the next step is generating the corresponding *Gammix* function model for each group. This is the most important step in order to obtain my goal model.

To do so, three different functions are defined. The same *Gammix* function described in the Equation 3.4 with its input parameters contained in the parameter vector defined in Equation 3.5 is to be used for both Weekdays and Fridays, finding for each of them different optimal parameters. An additional more simple *Gammix* function following the equation described in the Equation 3.2 is to be used for Weekends with its corresponding input parameters contained in the parameter vector described in Equation 3.3. In this third case, its corresponding optimal parameters are to be found as well.

Once all three initial functions and parameter vectors are defined, the optimal values that generate an accurate *Gammix* function output that resembles each group of days needs to be found. To do so, a minimization problem is defined. In this case, I divide all available days from January to the third week of February in their respective prototype. After that, each of the groups has undergone an algorithm consisting of a minimization of the aggregated Mean Squared Error (MSE).

Given a dataset $\mathcal{D} = \{x_{i,j} \mid i = 1, \ldots, N, j = 1, \ldots, 48\}$ that contains the (normalized) occupancy $x_{i,j}$ of day i at time-bin j, I minimize the sum of squared differences between the normalized occupancy as predicted by the model and the actual occupancy for each day. Thus the optimal parameters θ^* are obtained as

$$\begin{aligned}\theta^* &= \arg\min_{\theta} \mathcal{L}(\theta) \\ &= \arg\min_{\theta} \sum_{i=1}^{N} \sum_{j=1}^{48} \left(x_{i,j} - \text{gammix}(x_{i,j};\theta)\right)^2 .\end{aligned} \tag{3.8}$$

The corresponding optimal vectors obtained for Vilanova can be observed in Table 3.

Table 3: Optimal parameter values for every of the prototypes in Vilanova.

Parameter	**Weekday**	**Friday**	**Weekend**
w_1 / w	0.1666	0.04605	0.5089
a_1 / a	2.1434	5.2162	0.3577
a_1 / b	19.7430	42.2427	6.2465
w_2	0.3388	0.4538	-
a_2	1.1897	0.8208	-
b_2	17.8948	10.9309	-

Using the optimal values for Vilanova displayed in Table 3 and the corresponding ones to Quatrecamins, the corresponding *Gammix* functions can be directly plotted using the optimal parameters as input parameters for the Gammix equations. It is worth reminding that both Weekdays and Fridays present a double-gamma kernel whereas the Weekends present one single Gamma kernel. The obtained output probability *Gammix* functions can be observed in Figure 21.

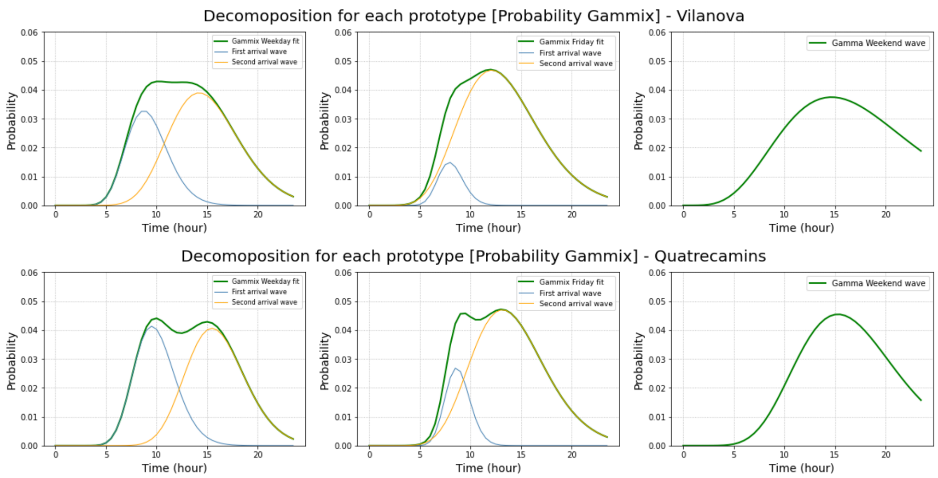

Figure 21: Corresponding Gammix probability prototypes for all Weekday, Friday and Weekend.

IV. Denormalizing and agreggating the baseline value:

Once each Gammix probability prototype is obtained, I perform the denormalization of each of them and the aggregation of their corresponding baseline mean value to get back the real occupancy behaviour. This is why, the previously obtained normalized *Gammix* prototypes are multiplied by their corresponding mean area factor and aggregated to their corresponding baseline mean value. By doing so, the *Gammix* probability prototypes are converted into *Gammix* occupancy prototypes.

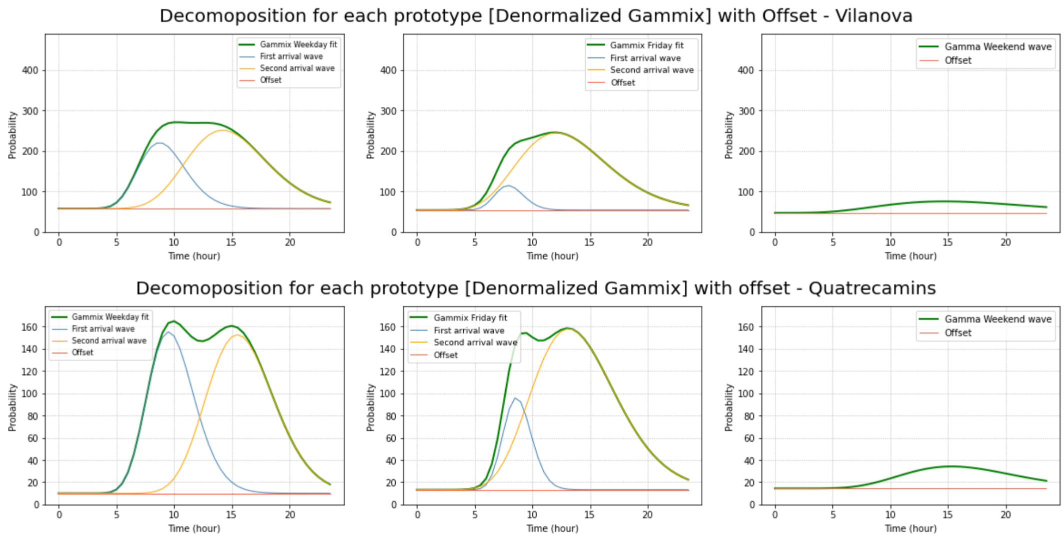

Figure 22: Corresponding *Gammix* denormalized prototypes for all Weekday, Friday and Weekend Prototypes.

The occupancy *Gammix* prototypes are attached in the Figure 22. When comparing the denormalized prototypes with real data, it can be noticed that the output *Gammix* prototypes resemble quite a lot to the real data. This comparison is attached in the Figure 23.

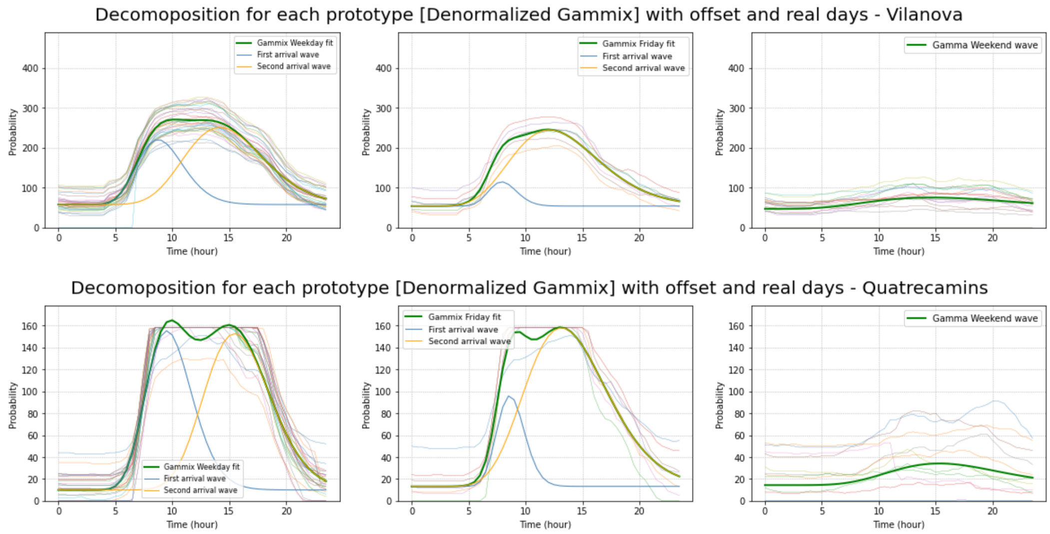

Figure 23: Corresponding denormalized prototypes for all Weekday, Friday and Weekend Prototypes compared with the real data.

3.2.4 Meaningful Explanation of the Gammix Modelling

The corresponding *Gammix* decomposition for Weekdays and Fridays is composed by two different Gamma kernels that present maximal peaks at early hours for the first wave and later hours for the second one. Between both peaks, a stable region shows how the arrivals and departures of cars are almost the same, thus presenting a constant value of occupancy.

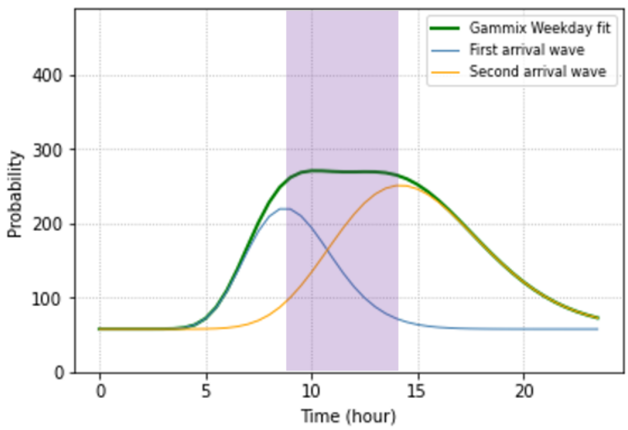

Figure 24: Meaningul explanation of the Wave decomposition.

This can be observed in all Figure 21, Figure 22 and Figure 23. A deeper analysis of the meaning of each wave is performed. In the Figure 24 one can observe a first wave in blue, a second one in orange, a total wave in green and a purple region comprised between both peaks.

Each of the decomposed waves present its own meaning. The First Wave represents the first flux of cars arriving early in the morning. It is characterized by:

- A first rapid increase until 9h.
- A maximal peak representing the maximal probability that a given car enters the car park in early hours.
- A rapid decrease around noon.
- A final zero probability region.

The Second Wave represents the second flux of cars arriving in later hours. It is characterized by:

- A first zero probability region.
- A rapid increase until 15h.
- A maximal peak around 15 representing the maximal probability that a given car arrives the car park in later hour.
- A rapid decrease.

Between the peaks of both secondary probabilities waves, a Stable Region is found. It represents the equilibrium region where it is equally likely that a park station spot either is occupied or set free by a car, making the car park occupancy almost constant.

It is interesting to notice that for Fridays, the first flux of cars arriving at early hours is much lower, showing that the behavior of car users changes completely during Fridays, and most of them start the day later. This phenomena can be observed in almost all car parks. When observing the Figure 21, it is quite clear that both Weekday *Gammix* prototypes present equally weighted decomposed probability waves. However, in the case of Fridays, this behaviour is completely changed to a first tiny probability wave and a second bigger one.

3.2.5 Mathematical Model Characteristics

As some first conclusions for this second part of the methodology, it can be stated that the *Gammix* function recovers the behaviour of the occupancy pattern. As the main conclusions, I have:

- The gamma kernel can be used to approximate our data.
- Complex behaviours for Weekdays and Fridays require two gamma kernels.
- The *Gammix* function presents a meaningful explanation. For the Weekends, the *Gammix* output defines directly the probability of a given occupancy of a free slot. For both Weekdays and Fridays, each *Gammix* wave presents a different meaning:
  - First Wave: Corresponding to the probability that a given car enters in the car park in the first flux of arrivals early in the morning.
  - Second Wave: Corresponds to the probability that a given car enters in the park station in the second flux of arrivals later in in the afternoon.
- **Simplification of the statistical model:**
  - The *Gammix* function for both Weekdays and Friday needs 6 parameters plus two additional Area and Baseline mean value. Therefore, 8 parameters are to be stored.
  - The *Gammix* function for Weekends needs 3 parameters plus two additional Area and Baseline mean values. Therefore, 5 parameters are to be stored.

The performance of the mathematical model can be checked in [3]. Additionally, all the code implementation of both models (mathematical and statistical) and its corresponding application and usage for real-day prediction is available in an online repository [1]

3.3 Autocomplete Model

When facing the real-time approximation problem, there is a differential factor to be considered: *Some data of the day is already known.*

This is a really interesting factor to consider, as there is already some more extra information apart from the historical behaviour or the obtained Gammix prototype of the day to use. Hence, an improvement of the modelling performed so far is considered in order to obtain better results and consequently, better future predictions.

The first approach to generate a more accurate predictor using already existing available data is trying to scale-up the prototype function (either the historical mean or the Gammix one), assuming that the current day will behave quite similar to the historical data, but the specific values themselves would be scalated by a computed factor.

This means, I assume that the behaviour of the occupancy will remain the same, and the only variation is due the total volume of cars entering the park station lot. Thus, scaling the whole prototype (either Historical or Gammix) comparing the known data with the prototype will allow a further improvement of the modelling. In Figure 25, this can be clearly observed:

The autocomplete model improvement will be explained in a deeper detail in David Moreno's thesis. [3].

[1]Repository containing the implementation and results for both models: `https://drive.google.com/drive/u/1/folders/1JM8CtXTuRbtlIbrhfU35rEDzSRYk5oOM`.

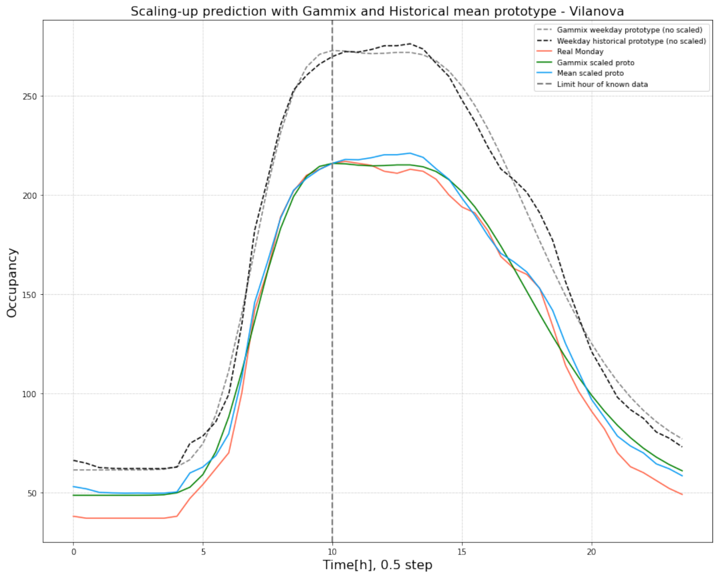

Figure 25: Autocomplete improvement applied to both statistical and mathematical models for Vilanova. Both statistical and mathematical prototypes are multiplied by a factor to adjust better the real data.

Chapter 4

Weather Influence

Some external factors can influence directly on the behaviour of the occupancy of the different car parks, making the users as a whole to behave differently and thus, change the occupancy temporal evolution patterns that have been observed so far. Weather is one of the main factors that can completely change the traffic behavior and the use of transportation means. In order to check if this external factor is significant to be considered as an additional variable, a deep analysis of its influence towards the data is performed. To start analysing this effect, a dataset with the rainy days of Barcelona is found directly in the Open Data Barcelona [1]. In this case the influence of the weather in the grand Barcelona area is analysed in order to see if this influences directly the different studied car park spots, as they are directly related to train stations that users take to get to Barcelona.

Once a historical dataset to know what days were rainy in Barcelona, a simple merging between the occupancy dataset and this later one can be performed to have directly related each day occupancy behavior and its corresponding weather characteristics. A first simple distinction between rainy and sunny days is performed, considering as a rainy day any day that presented an Accumulated daily precipitation (mm) higher than 0.1 in the corresponding day.

[1] Open Data Barcelona: https://opendata-ajuntament.barcelona.cat/data/en/dataset/mesures-estacions-meteorologiques

4.1 First Intuitions

Some first preliminary analysis are performed in this case again, in order to observe how the weather might influence the occupancy temporal pattern. In this case, it is worth saying that there was no rainy Saturday in our study temporal historical data (January to mid March).

First, the mean occupancy by day is plotted for both rainy and not-rainy days in Figure 26, and an additional total mean occupancy by day is plotted as well.

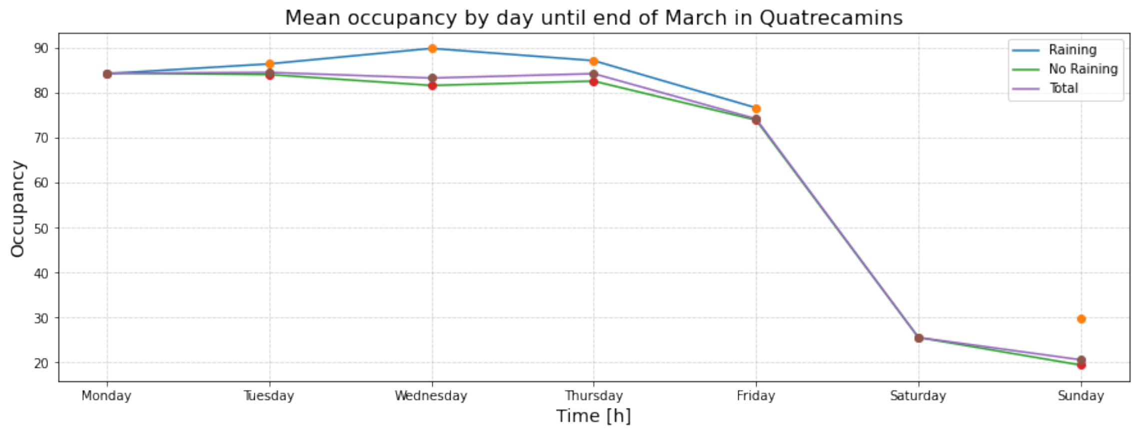

Figure 26: Comparison mean Occupancy by day of Quatrecamins for rainy and not-rainy days.

As it can be observed in the previous figure, when computing the mean for all rainy days, the mean-occupancy values are increased in all cases. When considering Mondays, the value is almost the same. Consequently, when considering the not-rainy days mean, the occupancy values are lower.

Now the previous analysis is repeated again, in this case only considering from Monday to Friday and the whole occupancy pattern for the 24 hours cycle. As it can be observed in the Figure 27, in the case of Quatrecamins, all rainy days occupancy do saturate (including the Fridays) while the not-rainy maintain the same behaviour that has been explained before.

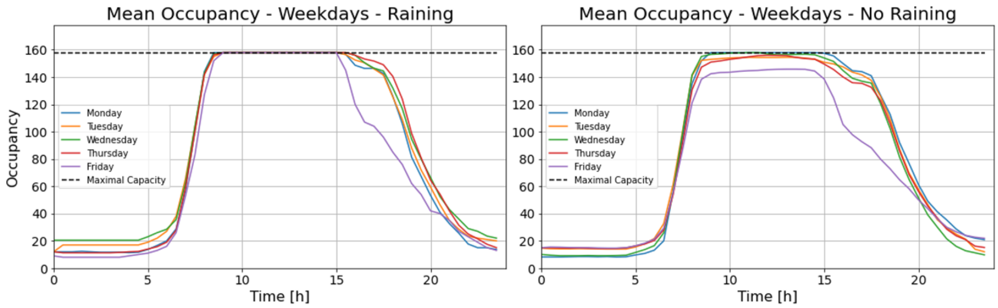

Figure 27: Comparison of the mean Occupancy by weekday of Quatrecamins for rainy and not-rainy.

4.2 Influence of the Weather in the Occupancy Behaviour

In this second analysis, the main idea is focused on detecting if a rainy condition implies an increase or a decrease of the occupancy pattern. Even though in the preliminary analysis of Quatrecamins everything leaded to think of an increase of the occupancy, some car parks do present a different behaviour. Even though most car parks present an increase of the occupancy pattern due rainy weather conditions as expected, a few others do present a decrease as well.

This is why, the same analysis is replicated for different all available car parks. In this case, the mean occupancy from Mondays to Friday and the corresponding variance is computed. It is worth saying that in this case, I consider all days from Monday to Friday as a whole group due the low amount of rainy days that the dataset presents. It is remarkable that Barcelona presents a dry weather where there are not many rainy days. The weekends are omitted for this same reason. There is no a significant amount of rainy Sundays to perform an analysis on their own.

Rainy days implies an increase of Occupancy:

In this case, the figures of both Quatrecamins and Granollers are been attached in the Figure 28. As it can be observed, in both cases the rainy days present a higher occupancy within the park stations. It is worth noticing that all Quatrecamins rainy days do saturate and present no variance during the stable region.

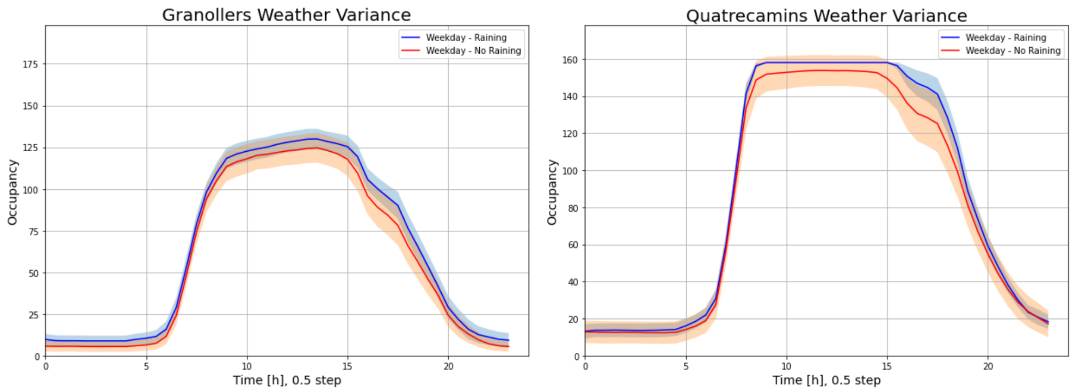

Figure 28: Comparison of the mean occupancy and variance of Granollers (left) and Quatrecamins (right) for Rainy and Not-Rainy days.

Rainy days implied a decrease of Occupancy:

In this second case, the figures of both Mollet and Sant Sadurni are attached in Figure 29. As it can be observed, in both cases the rainy days present a lower occupancy within the park station lots.

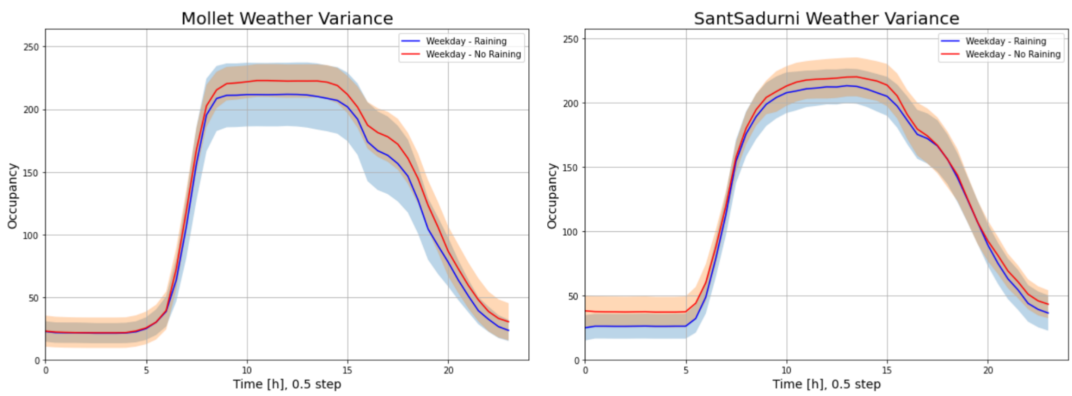

Figure 29: Comparison of the mean occupancy and variance of Mollet (left) and Sant Sadurni (right) for Rainy and Not-Rainy days.

4.3 Quantification of the Variation due the Weather Conditions

The difference between the rainy and non-rainy days means for each park station is computed. Thus, the variation due the rainy weather factor is quantified. As it can be observed in the following Table 4. As it can be observed, 6 days present a higher occupancy for rainy days whilst the resting 3 present lower occupancy values.

Table 4: Difference between rainy and non-rainy days for all available park stations.

park station	Diff (%)	Present Higher Values	Real Slots difference (%)
Cerdanyola	2.83	Rainy	3
El Prat	15.86	Rainy	73
Granollers	3.00	Rainy	6
Mollet	3.32	Non-Rainy	8
Quatrecamins	3.29	Rainy	5
Sant Boi	1.43	Rainy	5
Sant Quirze	8.8	Rainy	33
Sant Sadurni	2.96	Non-Rainy	7
Vilanova	1.12	Non-Rainy	5

When considering the less and most affected park stations, Vilanova presents a difference of 1.12% and 5 real slots difference, while El Prat presents a 15, 86% difference with a 73 slots difference. Their corresponding mean and variance for both rainy and non-rainy days are attached in the Figure 30.

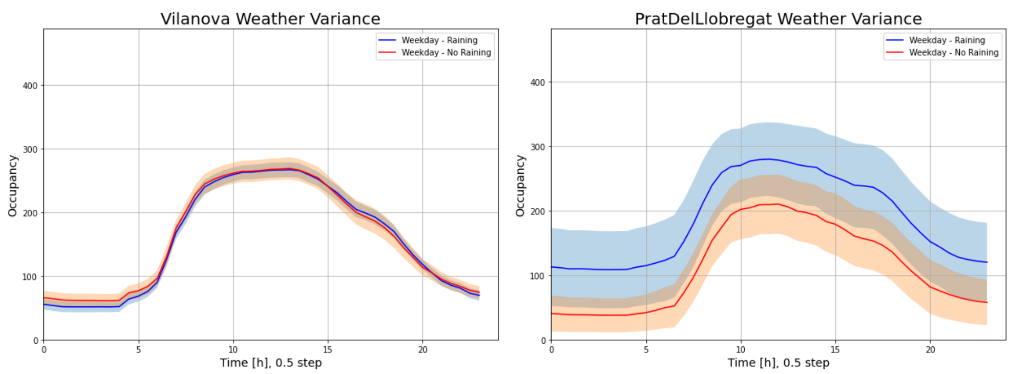

Figure 30: Comparison of the mean occupancy and variance of Vilanova (left) and El Prat (right) for Rainy and Not-Rainy days.

4.4 Concluded Influence towards the Occupancy

The weather influences the main behaviour of the occupancy, but it is not a determinant influence, as it was expected before performing the later analysis. Even though each park station presents its own characteristics, most park stations (6 out of 9) do present an increase due rainy weather. However, the difference in most cases do not overpass the 4%.

Chapter 5

Conclusions and Future Work

During the whole thesis, both models have proved to be useful and accurate. Furthermore, their corresponding results have been highly satisfactory. As it has been shown during the whole thesis, they allow to characterize the occupancy evolution with time and to capture satisfactorily the behaviour of such patterns. In this final chapter, a brief summary of all results obtained throughout this thesis and some guides to improve the current results in the future are collected.

5.1 Conclusions

The main goal of this thesis is the understanding and modelling of the occupancy behaviour of different car park lots located within the province of Barcelona. Additionally, all models to be developed are aimed to be simple, transparent and highly intuitive. All these requirements have been successfully fulfilled. The models have been developed using stored data from different public car park lots that are placed close to train stations.

A first preliminary analysis has proved that occupancy pattern presents a circadian cycle that is highly aligned with the average person routine. This is why a first statistical approach has been considered to demonstrate that only using historical data is more than enough to characterize and predict the future behaviour of car park occupancy. Generating three different prototypes has proved to produce good results with a low error. Furthermore, this simple statistical approach provides us with a simple but effective model to predict long-term occupancy.

Once this first statistical approach has been obtained, the next logical step has been generating a mathematical model that further simplifies the current model without losing prediction and provides us with some meaningful explanations of what is happening. This has been achieved using what I have called the *Gammix function*: A mixture of two Gamma kernels for the case of Weekdays and Fridays and a single Gamma kernel in the case of Weekends. The resulting parametric model of the data behaviour is meaningful and easily interpretable and further simplifies the statistical model, requiring only 8 parameters for both Weekdays and Fridays and 5 ones for Weekends, reducing significantly the 48 parameters required for each of the statistical prototypes.

Additionally, both the statistics and the mathematical models have been proved useful and accurate to predict the future occupancy thoughout David Moreno collaborative msc. thesis [3]. Both models have mostly presented errors under the 5%, showing really satisfactory results. A final improvement of the model called *Autocomplete Model* has showed to be useful as well in order to predict real-time data exploiting the idea that some data of the current day is already known.

It has been clear that one key point of developing wiser policies for urban systems and infrastructures is allowing people to have access to real-time car parks occupancy and thus, make them to take more informed decisions during their everyday lives and improve the traffic derived due interurban mobility [9]. There has been enough evidences showing that using simple statistical and mathematical models are more than enough to characterize and predict these occupancy patterns and provide drivers of real-time occupancy information.

Some final conclusions regarding the weather influence on the occupancy behavior are quite clear: Weather does not affect car park occupancy as much as it has been first expected. However, one can observe some specific effects that can be considered. For instance, making Quatrecamins car park slot saturate for all rainy days or presenting most car parks a different occupancy behaviour when the weather is rainy.

5.2 Future Work

Some factors can be further enhanced to improve the current state of the model. One of our biggest challenges has been characterizing the occupancy pattern with such few data. Considering that the newly implemented sensors system was first installed on the car park lots in the beginning of 2020 January and that the Covid-19 pandemic started at the mid-march of the same year, the total amount of usable data has been significantly reduced. Using more data would allow us to characterize better the change of occupancy throughout a whole year and would end up in more detailed and interesting results, such as occupancy seasonal changes or correlate better both the occupancy patterns with external factors. Another path to be considered in the future is the correlation between close car parks. Even though some car park lots such as Sant Boi and El Prat are close enough to present common trends, considering the amount of available data and that some car park data sets were not accurate enough, this correlation study has not been successful. Another improvement in the future would be correlating the occupancy patterns with the public network of train stations and the incidents that this system presents. As this thesis has been developed under a collaboration with *ATM*, a future project would be to deploy and integrate the model inside the an application to allow users to check real-time car park slots occupancy.

List of Figures

1 Map of Barcelona province with the park station locations 4

2 Original Dataframe provided by ATM for the Vilanova car park location 4

3 Evolution of the number of papers published related to the keywords. 5

4 Mean occupancy and variance for each weekday from January to March of 2020 in the car park of Vilanova. 10

5 Cyclic behaviour of the Vilanova car park for the week starting from Monday 24th of February to Sunday 2nd of March 2020. 10

6 24-hour occupancy cyclic behaviour for Tuesdays in the car park of Vilanova. 11

7 24-hour occupancy cyclic behaviour from Monday to Sunday in the car park of Vilanova. 12

8 24-hour occupancy cyclic behaviour from Monday to Sunday in the car park of Quatrecamins. 13

9 Historical mean for each type of day with its corresponding Variance for the Vilanova park station 14

10 Some other abnormal days not used during the thesis 15

11 Network of car park lots related to their corresponding train stations 16

12 Comparison of the historical mean for each day. Left graph corresponding to Vilanova and right one to Quatrecamins. Notice the scale has been adapted to each car park to observe better different days behaviours. 18

13 The three different prototypes obtained for both Vilanova and Quatrecamins Datasets. 20

14 Comparison of each Prototype with real days corresponding to them for both Vilanova and Quatrecamins Datasets. 21

15 Region-characterization of the Weekday Prototype. 23

16 Region-characterization of the Friday Prototype. 24

17 Region-characterization of the Weekend Prototype.. 25

18 Different gamma kernel behaviours depending on α and β parameters. 27

19 Comparison between real occupancy and modified occupancy where the baseline value corresponding for each cycle has been removed. . . 30

20 Comparison between modified occupancy and normalized occupancy. Each day cycle has been normalized individually. 31

21 Corresponding Gammix probability prototypes for all Weekday, Friday and Weekend. 33

22 Corresponding *Gammix* denormalized prototypes for all Weekday, Friday and Weekend Prototypes. 34

23 Corresponding denormalized prototypes for all Weekday, Friday and Weekend Prototypes compared with the real data. 35

24 Meaningul explanation of the Wave decomposition. 35

25 Autocomplete improvement applied to both statistical and mathematical models for Vilanova. Both statistical and mathematical prototypes are multiplied by a factor to adjust better the real data. . . . 39

26 Comparison mean Occupancy by day of Quatrecamins for rainy and not-rainy days. 41

27 Comparison of the mean Occupancy by weekday of Quatrecamins for rainy and not-rainy. 42

28 Comparison of the mean occupancy and variance of Granollers (left) and Quatrecamins (right) for Rainy and Not-Rainy days. 43

29 Comparison of the mean occupancy and variance of Mollet (left) and Sant Sadurni (right) for Rainy and Not-Rainy days. 43

30 Comparison of the mean occupancy and variance of Vilanova (left) and El Prat (right) for Rainy and Not-Rainy days. 44

List of Tables

1 Corresponding baseline mean value value for each prototype. 30
2 Corresponding area mean value for each prototype. 32
3 Optimal parameter values for every of the prototypes in Vilanova. . . 33
4 Difference between rainy and non-rainy days for all available park stations. 44

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

[8] 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).

[9] Provoost, J. C., Kamilaris, A., Wismans, L. J., van der Drift, S. J. & van Keulen, M. Predicting parking occupancy via machine learning in the web of things. *Internet of Things* **12**, 100301 (2020). URL `https://www.sciencedirect.com/science/article/pii/S2542660520301335`.