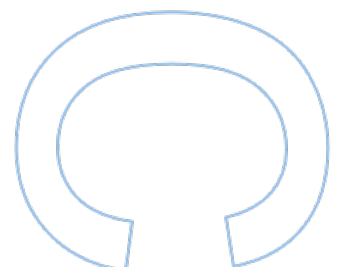
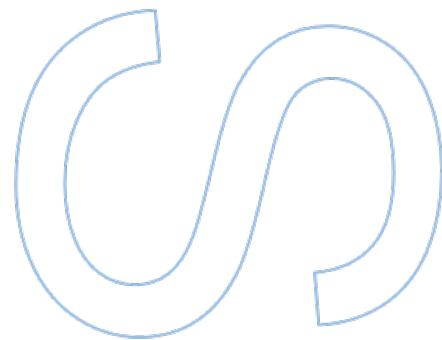
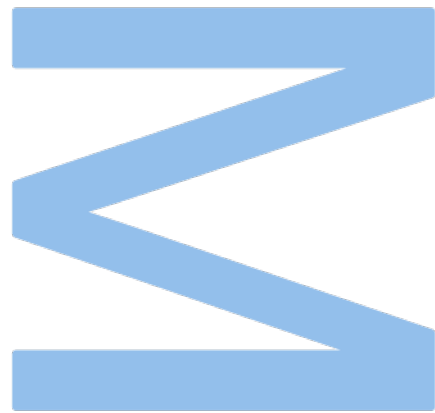


Urban City Dynamics Forecasting using Traffic Data

Sara Raquel Gonçalves de Sá
Master in Network and Information Systems Engineering
Department of Computer Science
Faculty of Sciences of the University of Porto
2024



Urban City Dynamics Forecasting using Traffic Data

Sara Raquel Gonçalves de Sá

Dissertation carried out as part of the Master in Network and
Information Systems Engineering
Department of Computer Science
2024

Supervisor

Fernando M. A. Silva, Full Professor, FCUP

Co-supervisor

Miguel Almeida, Data Scientist, Porto Digital Association



Abstract

Effective traffic control and forecasting in large metropolitan areas is essential for optimizing urban management and the transport system. This thesis aims to develop a forecasting system that makes it possible to predict the intensity and occupancy of road traffic at different time intervals. This work uses traffic data provided by Porto Digital Association. The data was collected from inductive loop detectors placed at critical street intersections within the Municipality of Porto to monitor traffic volume and occupancy.

The methodology adopted applies the Prophet model and its cross-validation method to predict traffic intensity and occupancy. This was intentional to assess this well-known platform, and thus, within this exploratory work, no comparisons were made with other approaches or models currently accepted in this specific area of research. The exploratory analysis carried out with data from January 2022 revealed significant patterns in traffic behavior, particularly during weekdays and peak hours.

The model was initially run for two detectors, one located close to the "Estádio do Dragão" and the other positioned further away. This initial analysis demonstrated the model's ability to capture traffic patterns, and variations in performance were observed between the two locations. The model was then extended to evaluate data from a total of 50 detectors spread across the city of Porto. The evaluation revealed greater accuracy in predicting traffic intensity compared to occupancy, with relative error metrics - specifically, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) - showing consistently lower values for intensity predictions.

Although this work suggests that the approach taken is an important starting point for the development of a more comprehensive traffic forecasting system, it also highlights the need for further refinements and the integration of different data sources. Further research in this field is crucial for the development of adaptive traffic management solutions that can respond to the evolving dynamics of urban transport systems.

Keywords: forecasting time series prediction, prophet model, time series, traffic, urban mobility, inductive loop detectors, prediction models, short-time prediction, long-time prediction

Resumo

O controlo e a previsão eficazes do tráfego nas grandes áreas metropolitanas são essenciais para otimizar a gestão urbana e o sistema de transportes. O principal objetivo desta tese é desenvolver um sistema de previsão que permita prever a intensidade e a ocupação do tráfego rodoviário em diferentes intervalos de tempo. Este trabalho utiliza dados de tráfego fornecidos pela Associação Porto Digital. Os dados foram recolhidos a partir de detetores de laços indutivos colocados em interseções críticas de ruas no Município do Porto para monitorizar o volume de tráfego e a ocupação.

A metodologia utilizada neste trabalho é a aplicação do modelo Prophet e do seu método de validação cruzada para prever a intensidade e ocupação do tráfego. Foi intencional avaliar esta plataforma bem conhecida e, portanto, dentro deste trabalho exploratório, não foram feitas comparações com outras abordagens ou modelos atualmente aceites nesta área específica de investigação. A análise exploratória efetuada com dados de janeiro de 2022 revelou padrões significativos no comportamento do tráfego, particularmente durante os dias úteis e horas de ponta.

O modelo foi inicialmente executado para dois detetores, um localizado perto do Estádio do Dragão e o outro mais afastado. Esta análise inicial demonstrou a capacidade do modelo para captar os padrões de tráfego, tendo sido observadas variações de desempenho entre os dois locais. O modelo foi então alargado para avaliar dados de um total de 50 detetores espalhados pela cidade do Porto. A avaliação revelou uma maior precisão na previsão da intensidade do tráfego em comparação com a ocupação, com métricas de erro relativo - especificamente, o Erro Absoluto Médio (MAE) e o Erro Quadrático Médio (RMSE) - mostrando valores consistentemente mais baixos para as previsões de intensidade.

Embora este trabalho sugira que a abordagem adotada é um ponto de partida importante para o desenvolvimento de um sistema de previsão de tráfego mais abrangente, também salienta a necessidade de mais aperfeiçoamentos e da integração de diferentes fontes de dados. A continuação da investigação neste domínio é crucial para o desenvolvimento de soluções de gestão do tráfego adaptáveis que possam responder à dinâmica evolutiva dos sistemas de transportes urbanos.

Palavras-chave: previsão de séries temporais, modelo prophet, séries temporais, tráfego, mobilidade urbana, detetores de laços indutivos, modelos de previsão, previsão a curto prazo, previsão a longo prazo.

Agradecimentos

Agradeço ao meu orientador, o Professor Doutor Fernando Silva da Faculdade de Ciências da Universidade do Porto, pela orientação prestada ao longo da dissertação. Ao meu co-orientador Doutor Miguel Almeida da Associação Porto Digital por toda a preocupação, disponibilidade e apoio e por tornar possível a realização desta dissertação.

Gostaria também de agradecer à Associação Porto Digital, especialmente ao Doutor João Bastos pela oportunidade de contactar com o mundo do trabalho e ao mesmo tempo desenvolver este trabalho num ambiente profissional, acolhedor, desafiante e em equipa.

Quero também agradecer a todos os membros da equipa da Associação Porto Digital, da qual fiz parte durante o meu estágio, por toda a ajuda e disponibilidade diárias.

Por fim, quero agradecer aos meus pais e à minha irmã Catarina por toda a ajuda, paciência, motivação que sempre acreditaram em mim e que sem eles nada disto era possível nem seria a mesma coisa.

Contents

Abstract	i
Resumo	ii
Agradecimentos	iii
Contents	vi
List of Tables	vii
Acronyms	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	3
1.3 Contributions	4
1.4 Document Structure	4
2 Basic Concepts	5
2.1 Traffic Data	5
2.2 Fundamental Traffic Diagram	6
2.3 Time Series	7
2.3.1 Components of Time Series	8
2.3.2 Stationary Time Series	9

2.3.3	Time Series Models	10
2.4	The Prophet Model	10
2.4.1	Model Fitting	11
2.4.2	Prophet Cross-Validation	17
2.4.3	Model Evaluation Metrics	21
2.5	The Prophet Software	22
3	Related Work	24
3.1	Traffic forecasting with Prophet	24
3.2	Traffic forecasting using time series analysis	25
3.3	General Applications of Time Series Forecasting with Prophet	26
3.4	Limitations	28
4	Design, Implementation and Results	29
4.1	Data Collection	29
4.1.1	Structure of the Data	29
4.1.2	Description of Data Objects	31
4.2	Exploratory Analysis	32
4.2.1	Plots of the Exploratory Analysis	33
4.2.2	Data Availability of the data	38
4.2.3	Detection of Anomalies on the data	39
4.2.4	Scatter Plots that compare intensity with occupancy	40
4.3	Forecasting process using the Prophet model	44
4.3.1	Reading and merging data:	44
4.3.2	Preparing the data for Prophet:	44
4.3.3	Filtering data for the desired period:	45
4.3.4	Incorporation of public holidays and football games:	45
4.3.5	Train the Prophet model:	45
4.3.6	Generating future forecasts:	45

4.3.7	Cross-validation and performance metrics:	45
4.3.8	Cross-validation process in Prophet	46
4.3.9	Evaluating the performance of the forecasting system with 2 detectors - CMP_020-e3 and CMP_070-e2	48
4.3.10	Forecast visualization:	52
5	Conclusions	63
5.1	Contributions	65
5.2	Future work	66
	References	68
A	Results table of CMP_020-e3	70
B	Results table of CMP_070-e2	71
C	Intensity results for 50 detectors for 1 day	72
D	Occupancy results for 50 detectors, period of 1 day	80
E	Intensity results for 50 detectors, period and horizon of 1 hour	86
F	Occupancy results for 50 detectors, period and horizon of 1 hour	88
G	Histograms of Relative Errors of RMSE and MAE for Intensity for 50 Detectors	90
H	Histograms of Relative Errors of RMSE and MAE for Occupancy for 50 Detectors	96

List of Tables

3.1 Comparison of Neural Prophet and Prophet Models	25
---	----

List of Figures

1.1	Inductive loop detectors. Source: Wikipedia.	2
2.1	Flow-Density Diagram [4]	7
2.2	Trend, Seasonality and Irregular Variations of a time series [16]	9
2.3	Plot with the trend component of the intensity value	15
2.4	Plot with the weekly component of the intensity value	15
2.5	Plot with the holidays component of the intensity value.	16
2.6	Plot with the daily component of the intensity value.	16
2.7	Plot visualizing the historical data, the fitted model and the forecasted values for the intensity value	17
2.8	First cycle of the 10-day horizon, forecasts made every 1 day.	18
2.9	Second cycle of the 10-day horizon, forecasts made every 1 day.	18
2.10	First cycle of 10-day horizon and 10-day Period.	19
2.11	Second cycle of 10-day horizon and 10-day Period.	19
2.12	First cycle of 5-day horizon, forecasts made every 10 days	20
2.13	Second cycle for 5-day horizon, forecasts made every 10 days	20
2.14	Cross-validation using initial, period and horizon. In the first iteration, only the Initial is used. In subsequent iterations, Period is added to Initial.	21
4.1	Location of the detectors present in the controller CMP_070 of the crossing of Rua da Alegria and Rua Latino Coelho	30
4.2	Section of the JSON file describing the CMP_070 controller.	32
4.3	Monthly plot for the detector CMP_070-e2 for the intensity	33

4.4	Monthly plot for the detector CMP_070-e2 for the occupancy	33
4.5	Plot of the intensity for January 27th from the detector CMP_070-e2	34
4.6	Plot of the occupancy for January 17th from the detector CMP_070-e2	35
4.7	Weekly plot for the detector CMP_070-e2 for the intensity	35
4.8	Weekly plot for the detector CMP_070-e2 for the occupancy	36
4.9	Daily plot for the detector CMP_070-e2 for the intensity	36
4.10	Daily plot for the detector CMP_070-e2 for the occupancy	37
4.11	Weekend plot for the detector CMP_070-e2 for the intensity	37
4.12	Weekend plot for the detector CMP_070-e2 for the occupancy	38
4.13	Plot that shows the intensity on January 2022 for the detector e1 in the controller CMP_070-e1.	39
4.14	Plot that shows the intensity on January 2022 for the detector e2 in the controller CMP_070-e2.	39
4.15	Weekly plot that shows the occupancy values of the detector CMP_131-e1 between January 10 and 17.	40
4.16	Scatter plot of January 2022 between the intensity and occupancy of the detector CMP_070-e2.	41
4.17	Scatter plot of the Weekly Analysis between the intensity and occupancy of the detector CMP_070-e2.	42
4.18	Scatter plot of the Weekday (Monday) Analysis between the intensity and occupancy of the detector CMP_070-e2.	43
4.19	Scatter plot of the Weekend (Sunday) Analysis between the intensity and occupancy of the detector CMP_070-e2.	43
4.20	Example of a Prophet Cross-validation process	47
4.21	Plot that compares the values of RRMAE for CMP_020-e3	49
4.22	Plot that compares the values of RRMSE for CMP_020-e3	50
4.23	Plot that compares the values of RRMAE for CMP_070-e2	51
4.24	Plot that compares the values of RRMSE for CMP_070-e2	51
4.25	Trend Component of the detector CMP_020-e3 for intensity.	53
4.26	Trend Component of the detector CMP_070-e2 for intensity.	53

4.27	Trend Component of the detector CMP_020-e3 for occupancy	54
4.28	Trend Component of the detector CMP_070-e2 for occupancy	54
4.29	Weekly Component of the detector CMP_020-e3 for intensity.	54
4.30	Weekly Component of the detector CMP_070-e2 for intensity.	55
4.31	Weekly Component of the detector CMP_020-e3 for occupancy.	55
4.32	Weekly Component of the detector CMP_070-e2 for occupancy.	56
4.33	Daily Component of the detector CMP_020-e3 for intensity.	56
4.34	Daily Component of the detector CMP_070-e2 for intensity.	56
4.35	Daily Component of the detector CMP_020-e3 for occupancy	57
4.36	Daily Component of the detector CMP_070-e2 for occupancy	57
4.37	Holiday Component for Intensity Prediction at Detector CMP_020-e3	58
4.38	Holiday Component for Intensity Prediction at Detector CMP_070-e2	59
4.39	Holiday Component for Occupancy Prediction at Detector CMP_020-e3	59
4.40	Holiday Component for Occupancy Prediction at Detector CMP_070-e2	60
4.41	Variation of the RRMSE and RMAE for intensity and occupancy in different time horizons	61
A.1	Results table of CMP_020-e3	70
B.1	Results table of CMP_070-e2	71
C.1	Results table of the 50 detectors for Intensity	73
C.2	Results table of the 50 detectors for Intensity	74
C.3	Results table of the 50 detectors for Intensity	75
C.4	Results table of the 50 detectors for Intensity	76
C.5	Results table of the 50 detectors for Intensity	77
C.6	Results table of the 50 detectors for Intensity	78
C.7	Results table of the 50 detectors for Intensity	79
C.8	Results table of the 50 detectors for Intensity	79

D.1	Results table of the 50 detectors for Occupancy	81
D.2	Results table of the 50 detectors for Occupancy	81
D.3	Results table of the 50 detectors for Occupancy	82
D.4	Results table of the 50 detectors for Occupancy	82
D.5	Results table of the 50 detectors for Occupancy	83
D.6	Results table of the 50 detectors for Occupancy	84
D.7	Results table of the 50 detectors for Occupancy	84
D.8	Results table of the 50 detectors for Occupancy	85
E.1	Results table of the 50 detectors for the period and horizon of 1 hour for Intensity	87
E.2	Results table of the 50 detectors for the period and horizon of 1 hour for Intensity	87
F.1	Results table of the 50 detectors for the period and horizon of 1 hour for Occupancy	89
F.2	Results table of the 50 detectors for the period and horizon of 1 hour for Occupancy	89
G.1	Histogram of RRMSE for Intensity - 1 Hour Horizon	91
G.2	Histogram of RRMSE for Intensity - 1 Day Horizon	91
G.3	Histogram of RRMSE for Intensity - 7 Days Horizon	92
G.4	Histogram of RRMSE for Intensity - 14 Days Horizon	92
G.5	Histogram of RRMSE for Intensity - 28 Days Horizon	93
G.6	Histogram of RMAE for Intensity - 1 Hour Horizon	93
G.7	Histogram of RMAE for Intensity - 1 Day Horizon	94
G.8	Histogram of RMAE for Intensity - 7 Days Horizon	94
G.9	Histogram of RMAE for Intensity - 14 Days Horizon	95
G.10	Histogram of RMAE for Intensity - 28 Days Horizon	95
H.1	Histogram of RRMSE for Occupancy - 1 Hour Horizon	97
H.2	Histogram of RRMSE for Occupancy - 1 Day Horizon	97
H.3	Histogram of RRMSE for Occupancy - 7 Days Horizon	98
H.4	Histogram of RRMSE for Occupancy - 14 Days Horizon	98

H.5	Histogram of RRMSE for Occupancy - 28 Days Horizon	99
H.6	Histogram of RMAE for Occupancy - 1 Hour Horizon	99
H.7	Histogram of RMAE for Occupancy - 1 Day Horizon	100
H.8	Histogram of RMAE for Occupancy - 7 Days Horizon	100
H.9	Histogram of RMAE for Occupancy - 14 Days Horizon	101
H.10	Histogram of RMAE for Occupancy - 28 Days Horizon	101

Acronyms

DCC Departamento de Ciência de
Computadores

FCUP Faculdade de Ciências da
Universidade do Porto

MAE Mean Absolute Error

RMSE Root Mean Absolute Error

RMAE Relative Mean Absolute Error

RRMSE Relative Root Mean Squared Error

Chapter 1

Introduction

1.1 Motivation

The continuous growth of cities globally has led to a more intricate administration of urban transportation. Along with the increase in the number of cars, the accelerated increase in urban population has resulted in a substantial increase in traffic congestion, pollution and overall pressure on urban infrastructures.

Ensuring efficient traffic management is vital to mitigating environmental effects, improving the quality of life of residents and allowing a smooth flow of vehicles. Urban mobility has a substantial influence on the efficiency, economic activity, service accessibility, and overall social well-being of a city. The municipality of Porto, similar to other major metropolitan areas, is naturally faced with these issues.

To address the growing demands on its traffic infrastructure, Porto has used advanced technologies to improve traffic monitoring and management. An essential element of the urban traffic management system is the deliberate positioning of inductive loop detectors on the exposed road surface, whose main use is to select the timing of traffic light cycles.

The detectors detect changes in the magnetic field caused by contact with heavy metal objects such as cars, buses and trucks. As vehicles pass over them, they collect information that is essential for understanding and controlling traffic flows in the city. The data acquired by these detectors is distinguished by its consistency and precision, thereby providing a useful understanding of two main indicators:

Intensity : refers to the measurement of the number of cars moving through a specific location within one hour. It is a fundamental indication of traffic flow and provides important information on the level of utilization.

Occupancy : the percentage of time that a metallic object was detected over the inductive loop.

Accurate, real-time information is crucial for city planners and traffic managers, as they heavily depend on these indicators to make well-informed judgments and adjustments. Through the examination of intensity and occupancy data, we can effectively address developing traffic scenarios, such as unexpected rises in vehicle volume during busy periods or unexpected obstructions. For instance, when there is a high level of intensity, it may be necessary to make changes to traffic lights or implement extra traffic management systems to avoid congestion. Likewise, if a road segment consistently has a large number of vehicles occupying it for a lengthy period of time, it can indicate congestion although this may not be the only cause. These are examples requiring quick measures to reduce the traffic congestion.

However, such data points have a deeper meaning that goes beyond the immediate reaction. In the contemporary field of data-driven urban management, forecasting future traffic conditions has emerged as an extremely useful tool. Predictive analytics allows cities to accurately anticipate and prepare for future traffic scenarios, in addition to responding to current traffic conditions. This proactive approach is particularly advantageous in a dynamic urban environment, where traffic patterns can be influenced by many elements, including normal travel cycles and unforeseen incidents such as accidents or sudden road closures.

In this context, the main objective of this thesis is to develop a traffic forecasting system that can optimize urban management in Porto. To this end, time series forecasting models using the Prophet model will be tested and evaluated, applied to data collected from inductive loops. The study will initially cover an exploratory analysis of the data to detect anomalies and identify patterns, followed by the application of the model to different traffic scenarios, including assessing the impact of local events and public holidays.

Figure 1.1 shows an image of inductive loop detectors placed in a road:



Figure 1.1: Inductive loop detectors. Source: Wikipedia.

1.2 Objectives

This thesis aims to provide city managers with a timely understanding of city traffic by creating a specialized forecasting tool for the traffic conditions in the city of Porto. The specific objective is to forecast the level of traffic and the number of vehicles occupying a given sensor in the city's network for short and long periods of time. These projections are really helpful for municipal leaders who require strategies for different time periods, guaranteeing that suitable steps are implemented to efficiently handle traffic.

In order to accomplish this, the study will utilize the Prophet framework, which is a powerful tool for predicting time series data. The Prophet model has been selected for its adaptability and precision, especially in managing seasonal variations and the influence of holidays, which are frequently observed in urban traffic data. The framework's ability to integrate and adapt to various variables makes it a very suitable option for this study, especially given the intricate nature of traffic patterns in a vibrant municipality such as Porto.

Furthermore, this study will investigate the model's ability to adjust to diverse circumstances, such as fluctuating traffic patterns on weekdays compared to weekends, as well as during special events that attract substantial crowds and consequently cause considerable changes in regular traffic patterns. Weekdays generally have high traffic volumes during the morning and evening rush hours, however, weekends may have more unpredictable traffic patterns influenced by shopping or tourism. The importance of this study resides in providing further decision tools to assist the administration of urban traffic in Porto.

In summary, the main objectives of this dissertation are :

- to accomplish an exploratory analysis using traffic data gathered from inductive loop detectors to gain a better understanding of the data and to detect possible anomalies, identifying patterns and the main problems that may exist in the data;
- to test and evaluate the applicability of the Prophet platform as a time series forecasting model for this type of traffic data. A first approach is to be followed based on data from two detectors, one close to the Estádio do Dragão and the other far from it, to analyze the impact of football games and public holidays on both detectors. In addition, the system is tested over various periods and horizons using the cross-validation method, thus allowing the intensity and occupancy forecasts to be evaluated.
- to extend the Prophet analysis and forecasting model to a larger number of detectors, namely 50 randomly selected detectors, to assess the accuracy of its performance in predicting intensity and occupancy over different periods and horizons;
- to evaluate the performance for both scenarios 2 and 50 detectors, by calculating the values of the relative errors of the MAE and RMSE performance metrics obtained through the cross-validation method, so that we can better understand whether the prediction of this system is as expected or not.

1.3 Contributions

The main contributions of this dissertation correspond to the main objectives detailed in the previous section, namely:

- exploratory analysis revealed missing data for January 2022;
- purposely, we selected detectors with reduced missing data to explore the Prophet platform to predict traffic intensity and occupancy in two detectors located near the Estádio do Dragão;
- expanded our study to 50 randomly selected detectors to confirm or even surpass the predication accuracy we saw for 2 detectors;
- introduced information on public holidays and our results show that this information significantly improves the model's ability to predict traffic dynamics, thus improving traffic forecasting.
- our results also show that including external events, such as football games, improves the model and provides a better understanding of traffic flow.

These contributions were useful for a better understanding of the traffic patterns in the Municipality of Porto since we could analyze that both intensity and occupancy follow cyclical patterns over time, especially on weekdays during peak hours. The exploratory analysis helped us understand the data and detect some gaps in the data. Also, we could verify that using the Prophet model on this data, the system had more difficulty in predicting occupancy values than intensity and that it was more accurate over shorter than longer horizons.

1.4 Document Structure

This document is organized into four additional chapters. Chapter 2 includes the fundamental concepts necessary for the understanding of the subjects covered in this work. It provides a theoretical explanation of the data used and the methodologies used to develop this study. Chapter 3 includes a section on related work, which discusses relevant and helpful research for this study. Chapter 4, entitled “Design, Implementation and Results”, explains the exploratory analysis carried out on the data, the implementation of the forecasting system and the discussion of the results obtained. Chapter 5 presents the conclusions drawn from this work, its limitations, and proposes possibilities for improving the system for future work.

Chapter 2

Basic Concepts

This chapter provides the fundamental concepts that form the theoretical basis of this dissertation, mainly in the modeling and prediction of traffic data. Firstly, traffic data is presented, which is obtained through inductive loop detectors that offer two main metrics: intensity and occupancy. Analyzing this data is fundamental to understanding traffic patterns over time, detecting peak periods and helping to develop traffic management measures. Subsequently, the concepts of time series are discussed, which provide the mathematical basis for modeling and predicting intensity and occupancy data. Finally, the Prophet model is presented, which is frequently used for time series forecasting, with an emphasis on traffic data. This model is significantly useful due to its ability to detect complex, non-linear trends. Prophet's growth models are detailed, some of the methods it uses, the configurations that are needed to improve its performance and the evaluation metrics that assess its performance.

2.1 Traffic Data

The ability to monitor traffic in an urban city is, nowadays, critical to improving city management and mobility. For the city of Porto, it is important to anticipate traffic intensity so that measures can be taken that may avoid, for example, road congestion.

Therefore, the municipality of Porto uses inductive loop detectors to collect traffic data, namely the traffic volume and occupancy at each detector, measured with an interval of 10 minutes, to follow the TrafficFlowObserved data model from Fiware.

As noted by Cherrett et al. [5], inductive loop detectors provide an effective way of collecting essential parameters such as vehicle count (or intensity), speed and occupancy. These parameters are fundamental to support real-time traffic management, especially in urban areas where congestion is a significant concern.

The inductive loop detectors are large coils placed under the pavement which can detect large metallic objects on top of them, such as cars. This happens because the impedance of the loop is

modified by the amount of metal in the vehicle. The variation in the loop's inductance depends on the impact the vehicle has on the magnetic field as it moves [5].

An illustration of these inductive loops is given in the image 1.1.

In the Municipality of Porto, each detector sends data to a controller, a physical device housed in a streetside telecoms cabinet. A controller typically gathers data from each detector at a single intersection. A single controller does, however, occasionally collect data from two or three nearby intersections.

The inductive loop detectors provide two primary measurements at 10-minute intervals:

Intensity: represents the total number of cars that crossed the detector within a 10-minute interval. For example, if the detector records a total of 30 vehicles within this period, this value can be transformed into an hourly rate by multiplying it by 6. Consequently, in this particular instance, the traffic intensity would be quantified as 180 cars per hour. This conversion is essential for conducting traffic analysis which helps in understanding periods of high traffic.

Occupancy: measures the proportion of time during which metallic objects, such as cars, are detected by the inductive loop. For instance, if the detector detects the existence of vehicles for 4 minutes during a 10-minute time frame, the occupancy rate would be computed as 40%. Assessing the level of congestion on a roadway is crucial and can provide valuable information for developing traffic management measures.

Monitoring both traffic volume and occupancy is fundamental for congestion detection since a high vehicle occupancy rate associated with a low vehicle count can indicate slow or stationary traffic, which signals the occurrence of congestion [5].

Traffic data based on the two measures just presented is gathered over time and thus constitutes an observed time series, one for intensity and another for occupancy.

2.2 Fundamental Traffic Diagram

The fundamental traffic diagram illustrates the traffic flow under optimal conditions, characterized by uniform cars and drivers, as well as balanced traffic.

The following diagram implies that flow (Q), velocity (V), and density (D) are interconnected using the subsequent formula [21]:

$$Q = D \cdot V$$

The flow (Q) refers to the quantity of vehicles per unit of time, often articulated in vehicles per hour. Density (D) denotes the number of vehicles per unit length of roadway, often articulated in vehicles per kilometre. Ultimately, vehicle velocity (V) is articulated in kilometres per hour.

The flow-density diagram, based on empirical data, presents a more accurate representation of reality. The diverse cars and drivers provide a more detailed understanding of traffic behavior through their interrelationships.

The next image illustrates the flow-density diagram that relates the traffic flow and the density.

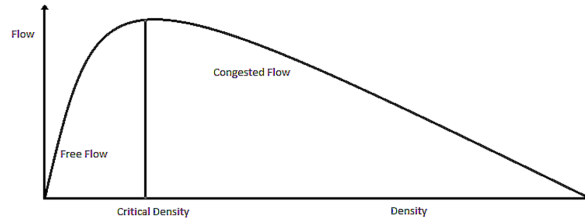


Figure 2.1: Flow-Density Diagram [4]

The image above illustrates three essential aspects of these diagrams [7]:

- **Free Flow Phase:** In this phase, the density values are low, suggesting that there are few vehicles on the road and consequently they can circulate freely. As density grows, traffic flow also escalates.
- **Critical Density:** During this phase, the maximum flow value is attained, resulting in achieving its full capacity.
- **Constrained Flow:** Upon achieving the critical density limit, the flow values start a decline, indicating that an increase in occupancy values correlates with a decrease in flow values. The rise in density values results in a reduction in vehicle speed, which leads to a decrease in the number of vehicles crossing a certain segment of the road during that period.

For a more detailed explanation of these diagrams and their dynamics, we suggest the book [21].

This diagram will be useful to analyse the patterns of the intensity and occupancy over time on different time scales in our data in the chapter 4.

2.3 Time Series

A stochastic process is a family of random variables indexed in a set $\mathbb{T}$, $\{x_t\}_{t \in \mathbb{T}}$. A time series in discrete time is a set of observations indexed by time x_t , $t = 1, \dots, T$. As an example, we can consider the number of cars passing over a bridge at different times of the day [18].

Time series may be univariate in which case the observations are univariate at each time stamp t , e.g. the number of cars passing over a bridge at time t , or multivariate in which case several variables are observed at time t , e.g., the number of cars passing over a bridge and air humidity at time t .

2.3.1 Components of Time Series

Time series consists of 4 components: trend, seasonality, cyclical variations and irregular variations. Next, we will describe them in more detail [1]:

- **Trend Component** represents the long-term trajectory or orientation of the data series. It indicates the general direction, whether it is increasing, decreasing or remaining constant, over a long period of time.
- **Seasonality** is the occurrence of regular, short-term variations in a data series over a given period of time, usually a year. The patterns observed are induced by external influences, such as variations in weather conditions, the occurrence of holidays and cultural events, resulting in anticipated and consistent variations in the data.
- **Cyclical variations** Cyclical patterns refer to changes in the series that occur cyclically and repeatedly over long periods, usually influenced by economic or business cycles.
- **Irregular variations** commonly referred to as residual or random variations, consists of unpredictable and random fluctuations in the data series. These variations cannot be attributed to trend, seasonal or cyclical factors and are often caused by unplanned occurrences or random disturbances.

The next image illustrates common components of a time series:

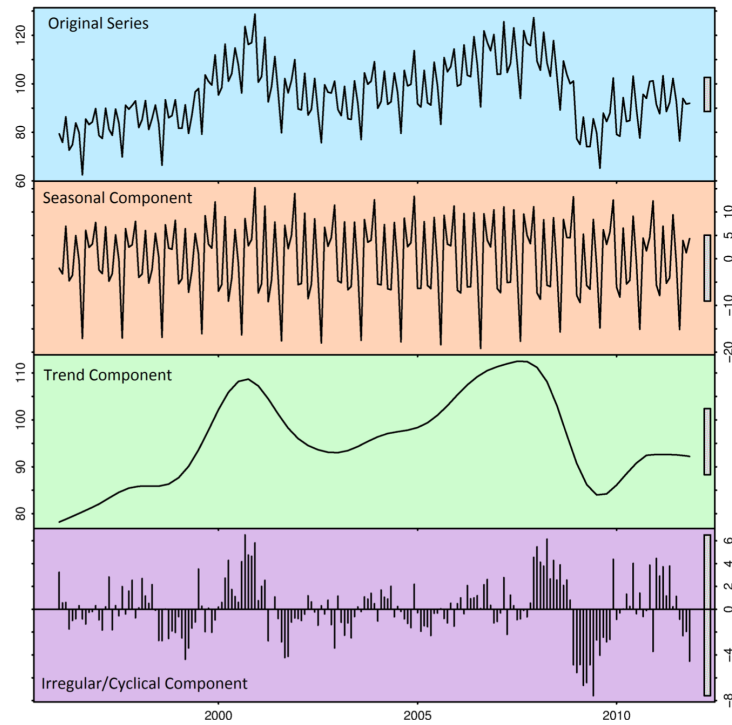


Figure 2.2: Trend, Seasonality and Irregular Variations of a time series [16]

2.3.2 Stationary Time Series

In time series analysis, it is common to observe a certain regularity in the behavior of a series over time, i.e. stationarity.

A time series can be either strictly stationary or second-order stationary. In the case of a strictly stationary series, the probabilistic behavior of any set of values and their displacements remains the same over time [18].

Therefore, the sets $\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\} \in \{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$ are identical for all $k = 1, 2, \dots$, all instants $t_1, t_2, \dots, t_k$, and for any deviations $h = 0, \pm 1, \pm 2, \dots$

A stationary second-order time series requires a process with finite variance, in which [18]:

- (i) the mean function, μ_t , is constant and does not vary with time t ;
- (ii) the autocovariance function, $\gamma(s, t)$, depends only on the difference between s and t , i.e. the distance $|s - t|$.

Stationarity requires that both the mean and autocorrelation functions have a certain regularity so that these quantities can at least be estimated using their means. A strictly stationary time series with finite variance is also stationary. However, the opposite is not always true. A relevant case in which stationarity implies strict stationarity occurs when the time series

is Gaussian, i.e. when all finite subsets of the series follow a Gaussian distribution [18].

2.3.3 Time Series Models

Time series modeling is a crucial component of time series analysis, covering diverse techniques and strategies to comprehend and forecast future values using past data. The main models employed in time series analysis include autoregressive models (AR), moving average models (MA), autoregressive moving average models (ARMA), autoregressive integrated moving average models (ARIMA) and seasonal autoregressive integrated moving average models (SARIMA). Every model possesses distinct attributes and uses, offering a comprehensive set of tools for analysts and researchers. For a more detailed discussion on each of these models, the book by Davies and Blockwell [2] provides a comprehensive introduction to time series analysis, offering theoretical and practical insights into the application of these models.

2.4 The Prophet Model

Prophet is an open-source tool for time series forecasting created by Facebook to address forecasting goals. It is characterized by being a model that can evaluate data on a daily, weekly or annual basis as well as its trend over time. It takes into account specific events that may occur in the data as well as holidays. It can also deal with patterns that may exist in the data, such as missing values [20].

Structure of the Decomposable Additive Model Prophet is considered to be one of the decomposable models, that is, it separates the time series into different components: trend, seasonality and events. The model does this in an additive way, adding the components to obtain the final forecast. Therefore, we can characterize Prophet's additive model as the mathematical form that allows for the decomposition of data [20].

With the Prophet decomposable additive model, we can obtain a clear analysis of how each component behaves over time. As a result, the decomposition of the time series makes it easier to understand and adjust forecasts, allowing detailed visualization of seasonal patterns, long-term trends and the impact of specific events. This model results in the following equation [20]:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (2.1)$$

where, $s(t)$ represents daily, weekly and annual seasonality, $g(t)$ represents the trend, $h(t)$ refers to holidays and specific, non-periodic events and ϵ_t is the error term that covers changes that the model cannot identify.

The $s(t)$ in the equation can simultaneously represent the seasonality components - daily, weekly and annual, depending on how the model was developed. Generally, each of these

components is treated separately and then added together to obtain the total $s(t)$. This allows the model to capture various periodicities within the same term.

2.4.1 Model Fitting

Once the Prophet model has been defined, it then needs to be fitted to the historical data that has already been obtained. This adjustment allows the Prophet model to gather seasonal patterns, trends or other events, adjusting them to the data that has been observed. The historical data supplied to Prophet must always contain two columns: “ds” and “y”. The “ds” column must be in date format, preferably YYYY-MM-DD or YYYY-MM-DD HH:MM, and the ‘y’ column must be numeric and represent the value to predict in the training set, which consists of the historical data [15]. This model uses the historical data to make forecasting and this data will be used to train the model, so we can consider that this historical data is the training set.

After the data has been formatted accordingly, Prophet categorizes the time series into three primary components: trend, seasonality, and holidays/events.

Trend Component The trend component is modeled by fitting a segmented linear curve to the trend or non-periodic part of the time series, ensuring that the trend is not overly affected by outliers or gaps in the data. This approach provides a more robust representation of the evolution of data over time [17].

Prophet contains two growth models in the growth parameter in the fit function: the logistic model (growth = 'logistic') and the linear model (growth = 'linear'). The main difference between the two is whether the carrying capacity is limited or not [17]. In the logistic model, this capacity is limited (for example, population growth up to a maximum limit), while in the linear model, there is no such limitation, allowing indefinite growth [17].

Logistic growth The logistic growth model is most appropriate for scenarios where growth is constrained by a maximum limit (carrying capacity). It is useful for predicting time series that use saturation, such as population growth in an environment where resources are limited [17].

The equation for logistic growth with non-linear saturation in its primary form is given by [20]:

$$g(t) = \frac{C}{1 + e^{-k(t-m)}}$$

Where:

- C is the carrying capacity;
- k is the growth rate;

- m is the shift parameter;
- t is the time;

If the data demonstrates variable capacity over time and an inconsistent growth rate, the model must be capable of capturing these variations and adapting to the historical data. The piecewise growth model integrates the changes regarding the trend by establishing changepoints, and its equation may be articulated as follows [20]:

$$g(t) = \frac{C(t)}{1 + \exp(-(k + \mathbf{a}(t)^\top \delta)(t - (m + \mathbf{a}(t)^\top \gamma)))}.$$

where:

- C is the maximum capacity variable over time;
- k is the growth rate;
- a is a vector of indicator functions for each changepoints ;
- δ is the vector of changes in the growth rate at each changepoint;
- t is the time;
- m is the shift parameter;
- γ is the adjustment of the shift;
- T represents the total horizon;

Piecewise Segmented Linear Growth In the piecewise linear model, the trend is made up of several linear segments with a constant growth rate linked together, but where their slope can change at each changepoint, offering greater flexibility to capture abrupt changes in the data [20]. Changepoints may be established explicitly, or Prophet may utilize default options, as demonstrated in this dissertation.

The expression for the piecewise segmented linear growth is [20]:

$$g(t) = (k(t) + \mathbf{a}(t)^\top \delta)t + (m(t) + \mathbf{a}(t)^\top \gamma)$$

Where:

- $k(t)$ represents the growth rate in time t ;
- $a(t)$ is the vector of indicator functions for the changepoints in time t ;

- $m(t)$ indicates the shift parameter in time t ;
- δ is the adjustment to the growth rate;
- γ incorporates s and δ and is adjust with $-s_j\delta_j$ to ensure that the function is continuous;
- T represents the total horizon;

Changes in the trend over time are affected by changepoints, i.e. moments when the trajectory of the time series changes significantly. Prophet detects these changepoints automatically and adjusts the model to reflect these changes [17]. Identifies them by designating several potential points along the time series historical data where the growth rate may vary. It then applies sparse regularization (equivalent to L1 regularization) to reduce the number of points used, keeping only those that cause the most significant changes. This allows the model to consider several possible locations for trend changes, but at the same time restrict detection to only the most relevant changes, avoiding excessive use of changepoints [8]. For a better understanding of the L1 regularization see [22].

Upon selecting the suitable changepoints, the model establishes a cost function that minimizes the difference between the observed and forecast values of the historical data, therefore adjusting the trend. By implementing the L1 penalty on these changepoints, the model provides a more continuous trend while reducing abrupt alterations in slope [20]. The model employs an L-BFGS optimization technique to refine the cost function that englobes all of the components of the model (trend, seasonality and holidays/events) and repeatedly modify the trend parameters until convergence is achieved. Therefore, we can assume that this algorithms is used to optimize all of the components. For a more detailed explanation of the L-BFGS algorithm, refer to [3].

The algorithms detect changepoints in historical data, letting the model to adjust to temporal trend differences and formulate a trend function for future forecasts. The segmented linear model with automated changepoint detection provides adaptability in addressing various trend patterns [17].

Seasonality In the context of business, time series frequently display seasonal patterns that repeat regularly, representing cyclical behaviors like consumption cycles, seasons, or particular days of the week. Periodic functions are used frequently to model these seasonal trends because they capture the changes observed in the data [20].

The *Prophet* model is a Bayesian model that uses a Fourier series-based decomposition to reflect seasonality. By using this method, seasonal cycles that rely on the time series regular period can be described. For instance, the period would be $P = 365.25$ days for annual seasonality and $P = 7$ days for weekly trends. The sum of the sine and cosine terms can be used to indicate the seasonal component as follows [20]:

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right)$$

Here, N corresponds to the number of harmonics used to describe the seasonality, and the coefficients a_n and b_n define the amplitude of the cycles. These coefficients form the vector β , represented by [20]:

$$\beta = [a_1, b_1, \dots, a_N, b_N]$$

The coefficients are determined by simple linear regression, which minimizes the discrepancy between the observed data values and those predicted by the seasonality function. The linear regression will be adjusted for each component of the Fourier series.

The matrix $X(t)$, which includes the sine and cosine functions corresponding to the seasonal cycles and applies to both historical data and future forecasts, must be combined to determine the seasonal effect at any time point t . For example, the matrix $X(t)$ would look like this to represent annual seasonality with $N = 10$ [20]:

$$X(t) = \left[\cos\left(\frac{2\pi t}{365.25}\right), \dots, \sin\left(\frac{2\pi(10)t}{365.25}\right) \right]$$

Therefore, the seasonal component of the series can be expressed as [20]:

$$s(t) = X(t)\beta$$

The parameters β are modeled as random variables in the Bayesian framework of *Prophet*, according to an a priori normal distribution $N(0, \sigma^2)$. This introduces smoothing into the seasonal patterns, avoiding sudden variations. By restricting the number of terms N , the model acts as a low-frequency filter, capturing smoother patterns. However, increasing N can make it possible to capture faster seasonal variations, but at the risk of overfitting the model [20].

In practice, weekly patterns often require $N = 3$, whereas annual seasonal patterns typically require a figure of $N = 10$. Furthermore, the model's complexity and the quality of the fit to the data can be balanced by automating the selection of these values using metrics like the Akaike Information Criterion (AIC) [20].

Holidays and Events Holidays and events introduce predictable disruptions into many time series, but they do not always follow regular patterns, which makes them difficult to model. Thus, it is essential to include holidays in forecasts to improve the overall accuracy of the predictions. Prophet allows adding lists of holidays, either a predefined global list or a customized list that includes specific events [20]. The model applies an indicator function to determine if a certain date aligns with a holiday or event, assigning a value that quantifies its influence on

the estimations. The event is assigned a value of 1 if it occurs on that specific day and a value of 0 if it does not occur. Each data is assigned a parameter, which is then modified according to previous data, to capture its impact on the data. To determine the effect of the event on historical data, the model adjusts a coefficient for each event using linear regression. These coefficients signify their magnitude and the positive or negative direction of their influence on the time series. For a more detailed explanation, see [20]. The effects of these holidays are handled independently by Prophet, allowing them to provide a set of past and future dates for each holiday. In addition, Prophet also allows the incorporation of a range of days around the holidays, such as the Christmas season, making these days part of the holiday as well. The impact of these holidays is adjusted with parameters and the effects can be analyzed after the forecast, allowing for a detailed assessment of their influence on the time series [20].

After fitting the model, Prophet allows us to visualize the main components that have an impact on the time series, using the `plot.components()` method, which generates separate plots for each of the following: trend component, weekly component and yearly component [15].

An illustration of these plots can be seen below for specific detector data:

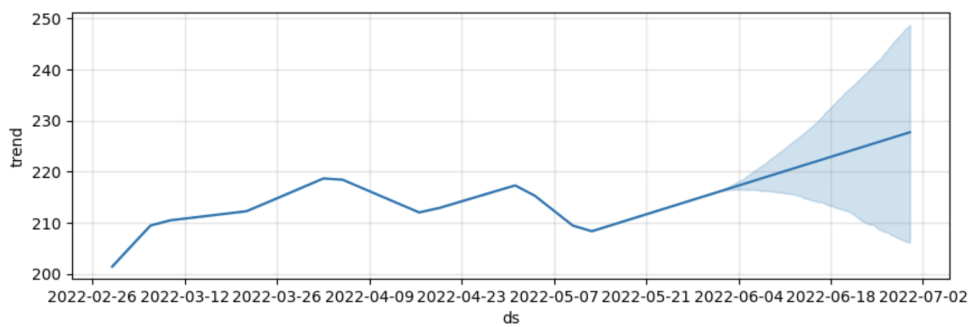


Figure 2.3: Plot with the trend component of the intensity value

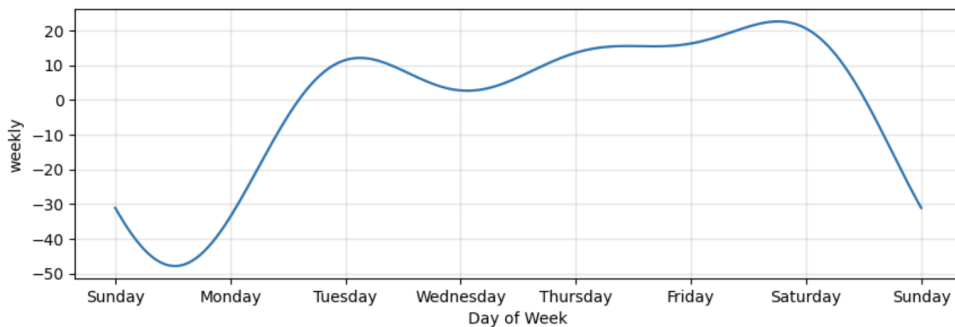


Figure 2.4: Plot with the weekly component of the intensity value

The plots generated illustrate the forecasts as a blue line, while the light blue band represents the uncertainty interval. By looking at these plots, we can understand how changes in trend or

seasonality affect the time series. If there is a sudden change in the trend (a change point), this will be reflected in the trend plot, with the line sloping up or down.

If we consider holidays or special events in the model, Prophet can provide a plot for the component holidays [9]. The Figure 2.5 shows one such example.

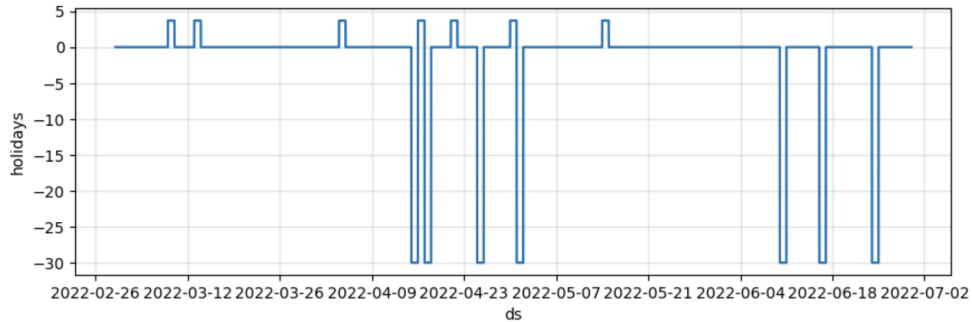


Figure 2.5: Plot with the holidays component of the intensity value.

The data utilized in this study had high resolution (10-minute intervals), allowing Prophet to identify patterns that repeat within the same day, which may be represented as daily components. Prophet dynamically modifies daily seasonality based on the data, effectively capturing intra-day trends. This is illustrated in Figure 2.6 by showing the daily component of a detector for the intensity.

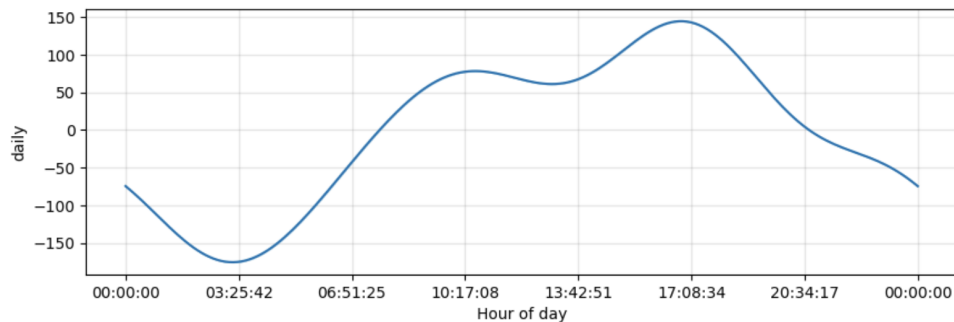


Figure 2.6: Plot with the daily component of the intensity value.

Subsequently, we show the plot illustrating the predicted intensity for a single detector 2.7. The black dots signify the actual values, the blue line depicts the model's predictions, and the blue shaded area represents the confidence interval, indicating the margin of error in the forecasts.

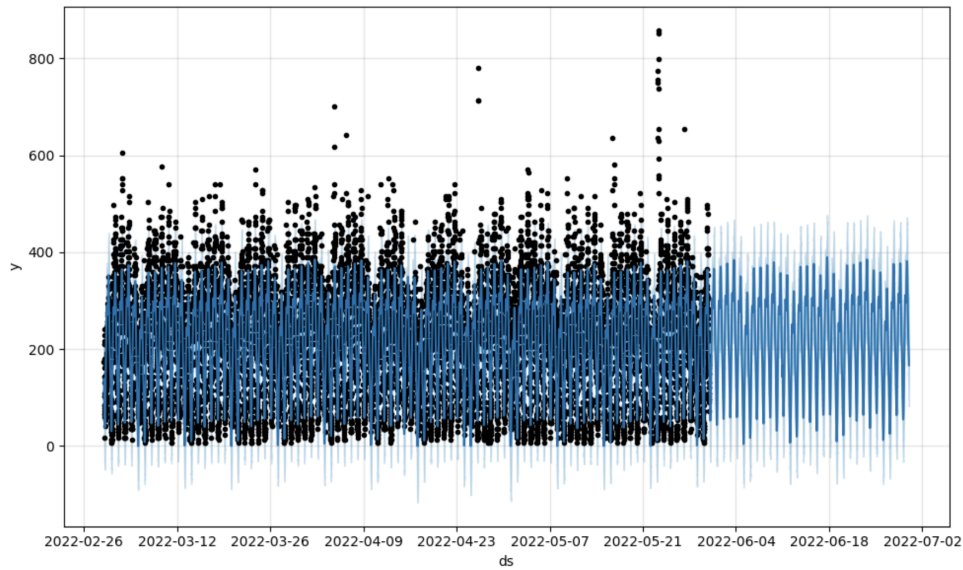


Figure 2.7: Plot visualizing the historical data, the fitted model and the forecasted values for the intensity value

2.4.2 Prophet Cross-Validation

Cross-validation is a method available in Prophet, that allows the evaluation of the model's performance on historical data. This process is useful for identifying possible overfitting or underfitting problems, as well as checking the model's ability to generalize the new data. Prophet's cross-validation method takes a set of historical data as input and creates a table of forecast results based on this cross-validation so that the model's performance can be assessed [23].

This process involves defining important parameters, such as defined in [23]:

- **Initial:** The number of data points that will be used to train the model.
- **Period:** The time interval between consecutive evaluations of the model. For more detailed and frequent forecasts, we chose a shorter period and for more punctual forecasts over long intervals, we opt for a longer period.
- **Horizon:** The entire future duration over which the model is required to generate forecasts.

Various metrics can be calculated to evaluate the model effectively [23].

The horizon denotes the future time frame for which the model will forecast, while the period indicates the time interval between each forecast within that horizon. We can examine, for example the instance of a meteorological institute, the IPMA - Instituto Português do Mar e da Atmosfera that generates daily predictions, including data on weather conditions for the subsequent 10 days in several regions of Portugal. Therefore, the horizon corresponds to 10 days and the period to 1 day [11].

Consequently, according to the IPMA, the following scenarios can be described [11]:

Horizon higher than the period : This scenario is the most frequent and efficient. The application predicts the subsequent 10 days daily as follows:

On the initial day, forecasts are generated for days 1 through 10 as illustrated in figure 2.8, representing the first cycle of forecasts. On the subsequent day, new forecasts are produced for days 2 through 11 representing the second cycle and it is illustrated in the image 2.9; on the third day, predictions are formed for days 3 through 12, and this pattern continues accordingly.

The predictions continuously "shift", consistently including the subsequent 10 days. This results in an overlap in the predictions, since each new day adjusts the predictions for the subsequent 10 days depending on the data that becomes available.

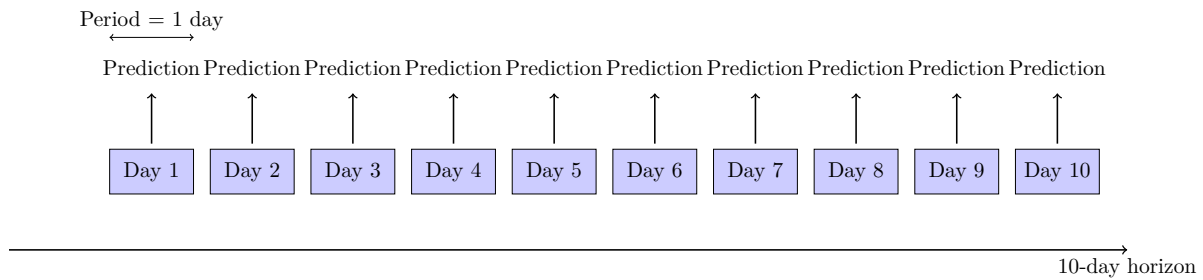


Figure 2.8: First cycle of the 10-day horizon, forecasts made every 1 day.

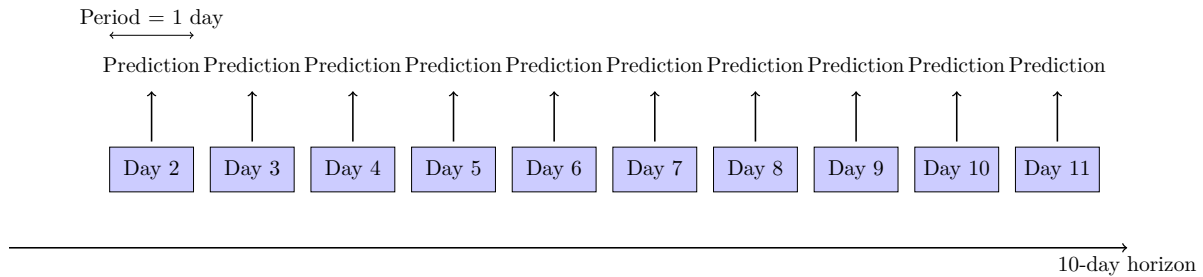


Figure 2.9: Second cycle of the 10-day horizon, forecasts made every 1 day.

Horizon equal to Period: If, in this scenario, the period and horizon are the same, namely both spanning 10 days, the model would provide predictions just every 10 days, so utilizing the entirety of the periods. So, in the first cycle, the predictions are made from Day 1 to Day 10, as illustrated in Figure 2.10. In the second cycle, the predictions are made from Day 11 to Day 20, as illustrated in figure 2.11, and this pattern continues accordingly.

The forecast horizon coincides precisely with the period, resulting in non-overlapping predictions that follow a 10-day interval pattern.

This scenario is plausible; however, it may not be optimal for forecasting objectives due to its decreasing accuracy over time, particularly under quickly changing situations. It is illustrated in the following image:

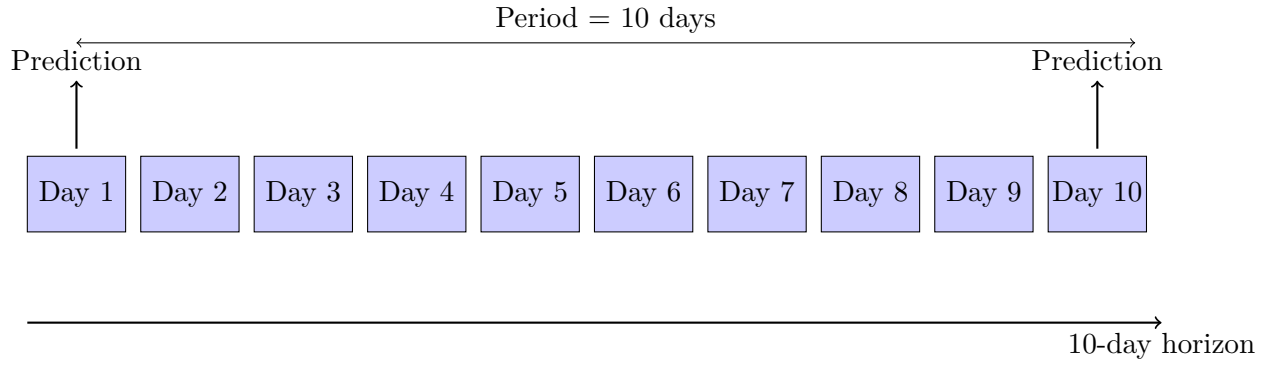


Figure 2.10: First cycle of 10-day horizon and 10-day Period.

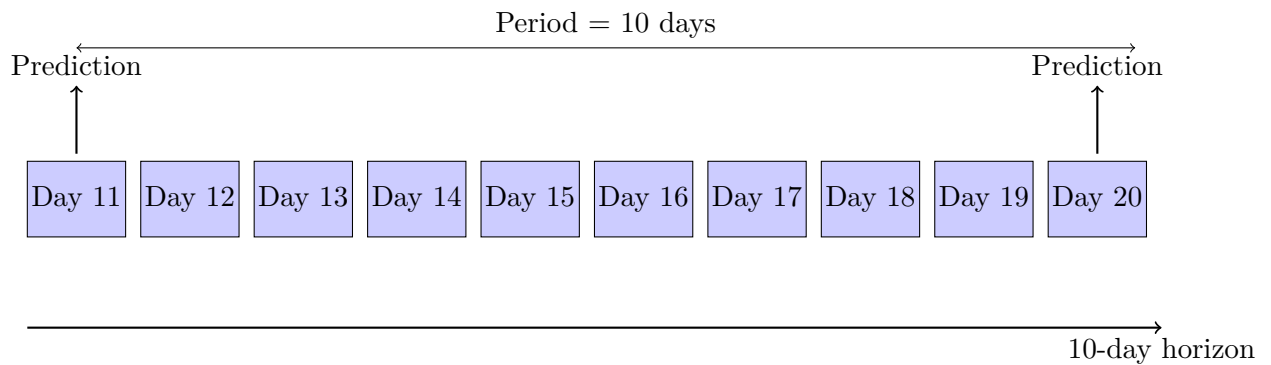


Figure 2.11: Second cycle of 10-day horizon and 10-day Period.

Horizon shorter than Period : If the horizon is shorter than the period, for instance, a 5-day horizon and 10 days, the application would generate forecasts every 10 days, but these would only encompass a 5-day interval ahead. On the first cycle, the model would provide predictions for days 1 through 5 as it is illustrated in the image 2.12, with subsequent forecasts (second cycle) beginning only on day 11 until day 15 represented in the image 2.13, resulting in a lack of predictions from days 6 to 10, as the gap between forecasts exceeds the forecast horizon. This would repeat for the following cycles.

This last scenario is not effective, since it generates prolonged periods absent of forecasts, potentially rendering the system outdated and compromising the quality of the predictions. It is illustrated in the following image:

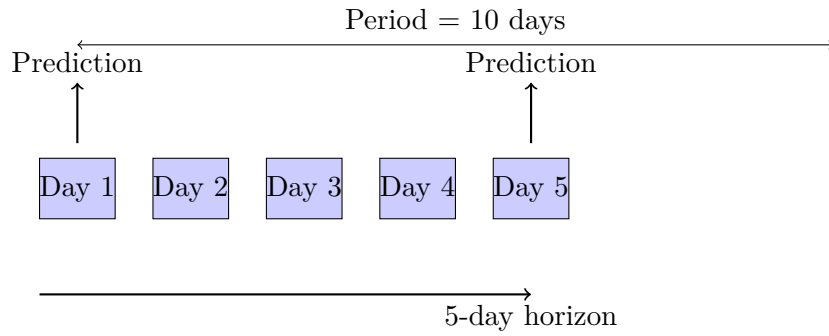


Figure 2.12: First cycle of 5-day horizon, forecasts made every 10 days

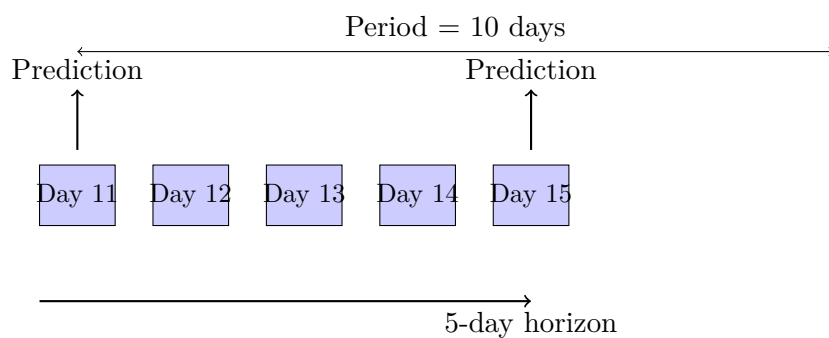


Figure 2.13: Second cycle for 5-day horizon, forecasts made every 10 days

Traditional cross-validation involves dividing the data into random subsets for training and testing, which cannot be applied directly to time series. This means that time series data has an order dependency, i.e. the data runs in the same direction over time. The model can not be prepared on future data and used to predict the past. Therefore, the most appropriate solution for time series is the cross-validation method, backtesting [10]. This process begins by specifying an initial training period and a forecast horizon. The model is then trained on this training data and tested on a subsequent set of data, called test data [10]. For the next iteration (or fold), the training set is expanded and the test set moves forward in time [10]. This method can then be adjusted with a sliding window approach, where the size of the training set remains fixed and only the most recent data is used [10]. This process is repeated several times (several folds), and the final error score is the average of the error scores in each of the iterations [10]. To model real-time forecasts, which replicate the process of continuously adding new data and reevaluating the model, the previously described expanding window technique can be helpful [10].

The next image shows a representation of this process:

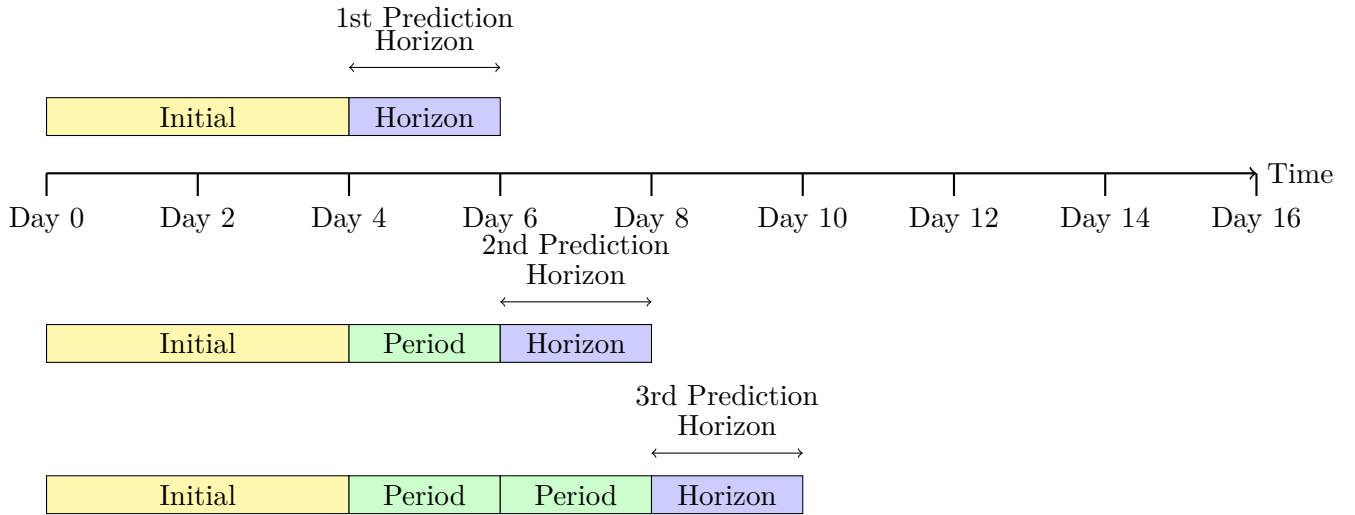


Figure 2.14: Cross-validation using initial, period and horizon. In the first iteration, only the Initial is used. In subsequent iterations, Period is added to Initial.

To assess the quality of the model's prediction, a variety of performance metrics can be used, namely the mean absolute error (MAE), the root mean square error (RMSE), or other metrics. These metrics provide a quantitative view of the use of the model and help to understand how well the model fits the data [10].

In time series, it is important to ensure that the model can predict the future well based on past data. Backtesting and sliding window techniques help assess the model's ability to deal with new and unknown data, which allows behavior to be observed over time. In addition, using robust metrics helps to avoid challenges such as overfitting the training data, ensuring that the model can be generalized to different scenarios [10].

2.4.3 Model Evaluation Metrics

There are various measures of prediction error resulting from the application of Prophet's cross-validation method, in this study we will consider the MAE (Mean Absolute Error) and the RMSE (Root Mean Square Error) [19].

The MAE is a statistical metric that evaluates the difference between pairs of measurements that describe the same phenomenon, for example, comparing two different measurement methods or comparing predicted values with actual values, as is the case in this study [19]. The MAE is obtained by calculating the average of the absolute differences between the measurements and the formula is as follows [1]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| = \frac{1}{n} \sum_{i=1}^n |e_i|$$

where:

- y_i is the observed value
- $\hat{y}_i$ is the predict value
- n is the total number of observations

The RMSE is the root mean square error which is normally used to assess the accuracy of a model [19]. It calculates how large, on average, the prediction errors are compared to the actual values. To do this, it adds up the squares of the differences between the predicted and observed values, averages them and then takes the square root of the result, as shown in the following formula [19]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

- y_i is the observed value
- $\hat{y}_i$ is the predict value
- n is the total number of observations

Comparing the two formulas used to measure forecast errors, the simplest to calculate and understand is the MAE because minimizing it usually leads to predicting the median of the values, while minimizing the RMSE leads to predicting the mean. Therefore, despite being more complex to interpret, the RMSE is the most commonly used [19].

Furthermore, we may highlight MAPE, which is referenced in several articles within the Related Work chapter. MAPE (Mean Absolute Percentage Error) is expressed as a percentage and quantifies the mean absolute percentage error, specifically assessing the magnitude of the discrepancy between projected and actual values. The formula is as follows [1]:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{y_t} \right| \times 100$$

- y_t is the observed value in the time t
- $e_t = y_t - \hat{y}_t$ it is the error at time t
- n is the total number of observations

2.5 The Prophet Software

The Prophet model described in section 2.4 is implemented in the Prophet software, which facilitates the fitting of time series models using the Python language. It has some advantages that are beneficial for forecasting time series, some of which are mentioned below [17]:

- can handle seasonality in data well, allowing it to be of different time scales, depending on the nature of the data and the periods it observes.
- handles the data in a way that is simple and understandable for the user.
- is able to incorporate special events and holidays into its model, meaning that if the data contains such days, it can easily capture them and make its forecast.
- can cope well with the lack of information in the data, allowing it to deal with irregularity.
- imposes upper and lower limits on the data, to understand its tendency to best fit the model to the data.

In conclusion, we explored the essential concepts for traffic analysis and forecasting, in particular the data collected by induction loop detectors, such as intensity and occupancy, which are fundamental to understanding traffic flow. Next, we analyzed time series, which are essential for modeling variations in data over time. Their components were discussed, as well as the need for stationary to guarantee more accurate forecasting models.

We then presented the Prophet model, which is useful for capturing complex traffic patterns, and discussed the necessary adjustments and evaluation metrics to improve its performance. With this, we have consolidated a theoretical basis for the next chapters, where these concepts are applied in practice.

Chapter 3

Related Work

Time series forecasting is commonly used in many areas including traffic, sales and finance. The Prophet model is noted for its ability to operate with time series covering seasonality, non-linear trends, holidays and events. This chapter covers work that has adopted Prophet in traffic forecasting, time series analysis and the impact of external events. In addition, studies that have used the model in areas such as forecasting supermarket sales, road traffic volumes and crypto market rates will be discussed. The current work expands this knowledge by applying Prophet with the cross-validation method on data from the Municipality of Porto, evaluating its performance through the relative error of MAE and RMSE and taking into account events such as football games and holidays.

3.1 Traffic forecasting with Prophet

Traffic forecasting is inherently challenging due to the vast amount of data, seasonal variations, and unforeseen occurrences that might cause complications. Nevertheless, Prophet has demonstrated its utility not just in traffic forecasting but also in several other application domains.

The paper by ChikkaKrishna [6], shows a study in which a time series prediction model was developed using Prophet and Neural-Prophet on a highway in India. The study aimed to forecast traffic volume(vehicles/hour) over 7 days to aid traffic management and planning and to predict possible congestion. These forecasts were evaluated using metrics such as mean absolute error (MAE), root mean square error (RMSE) and mean percentage error (MAPE). MAE and RMSE were expressed in vehicles per hour and MAPE is expressed as a percentage.

The results obtained were as follows:

Metric	Neural Prophet (Train)	Prophet (Train)	Neural Prophet (Test)	Prophet (Test)
MAE	31.095 veh/hr	34.677 veh/hr	49.438 veh/hr	61.232 veh/hr
RMSE	39.349 veh/hr	43.084 veh/hr	56.445 veh/hr	67.944 veh/hr
MAPE (%)	6.513 %	7.236 %	6.513 %	7.236 %

Table 3.1: Comparison of Neural Prophet and Prophet Models

Therefore, the study concluded that Neural-Prophet obtained better MAE, RMSE and MAPE results on both the training set and the test set, outperforming Prophet in terms of accuracy.

This study did not address the impact of one-off events or public holidays, which can create variations in city traffic. This dissertation differs in that it looks at holidays and football games in traffic data from the Municipality of Porto and applies the Prophet model to these events, aiming to explore its performance in capturing them and how it adapts to them.

3.2 Traffic forecasting using time series analysis

The article [19] portrays traffic forecasting using some time series models and aims to distinguish which model is most effective for predicting traffic volume. This study is carried out in Dhaka, Bangladesh, characterized by significant congestion, making this forecast essential in order to reduce travel time and the flow of vehicles on the roads. The data used recorded the number of vehicles per hour over 7 consecutive days. The models applied were as follows: ARIMA, ETS (Exponential Smoothing), SNAIVE (Seasonal Naive) and Prophet and in the end the results of these were combined into a mixed model in order to try and improve the robustness of the forecasts. The evaluation metrics MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) were used to compare the performance of the models and these metrics were expressed in vehicles/hour.

Concerning the results obtained, we can highlight the following:

- Prophet reflected the best performance overall, having low RMSE and MAE values, proving to be the most suitable in 42 of the 76 streets analyzed, particularly in streets with high seasonal patterns.
- In streets with less seasonal patterns, the SNAIVE model demonstrated better performance.
- The mixed model performed consistently in the different scenarios, but was unable to outperform Prophet or SNAIVE in the scenarios in which they performed best.
- The ARIMA and ETS models did not perform as well in this context, proving to be inappropriate for dealing with the seasonal patterns of Dhaka's traffic data.

This study concluded that Prophet is the best-performing model for forecasting, as it was the model that obtained the best results, especially on streets with high daily variation in traffic

volume and greater seasonality, and the one that obtained the best accuracy results. The authors stress the importance of choosing the model to use according to the characteristics of the data and suggest that densely populated cities such as Dhaka benefit from a tool capable of predicting traffic on their streets.

3.3 General Applications of Time Series Forecasting with Prophet

In this dissertation, Prophet is used to model traffic data in specific and sporadic event scenarios, such as football games, where seasonality and trends are interrupted by one-off events that cause significant variations in traffic. As well as being applied in areas such as traffic, Prophet has shown some facility in other areas.

The study [12] applied Prophet to sales forecasting in supermarkets. Its objective was to forecast future sales and optimize management and operational activities. The following models were compared: ARIMA, Holt-Winter's (Triple Exponential) and FB-Prophet.

The MAPE is presented as a percentage, while the RMSE metric was assessed utilizing furniture supermarket sales.

Comparing the models, Prophet showed an RMSE of 65.8 furniture supermarket sales and a MAPE of 8.3%. ARIMA had an RMSE of 152.6 furniture supermarket sales and a MAPE of 14.3%. Holt-Winter's had an RMSE of 85.7 furniture supermarket sales and a MAPE of 11%. Taking these values into account, we can see that FB-Prophet was the best-performing model, being able to forecast sales with the greatest accuracy compared to the other models.

It was concluded that the forecast made by Prophet is close to the real values and performs well for time series. However, they point out that its performance can be further improved by combining Prophet with other techniques. They also pointed out a difficulty in scalability, especially when analyzing large amounts of data, and to combat this, they suggested integrating the model with transfer learning. Finally, they stressed that the accuracy of real-time forecasting depends significantly on the quality of the training and validation of the models to be used.

The article by Yuvarani et. al. [23] analyzed the use of the Prophet model in forecasting the daily closing price of the cryptocurrency called Ethereum. The analysis was carried out using historical data from January 2020 to December 2021 and aimed to evaluate the model's ability to capture seasonal variations and unpredictable events.

The results showed that Prophet can detect trends and seasonality in the Ethereum price, in USD, adapting well to abrupt and non-linear changes in the data. The separation of the time series into trend and seasonality components was crucial in the model, allowing it to capture both short-term variations and recurring seasonal patterns.

To reinforce and guarantee the robustness of the model, the cross-validation method was used, which made it possible to assess the model's performance over different time horizons. The

model's accuracy was also measured using metrics such as MSE, RMSE, MAE and MAPE. The metrics MSE, RMSE were expressed in USD and the metric MAPE was expressed in percentage. These metrics showed satisfactory performance, considering that the data was volatile. Fb Prophet showed MAPE values between 0.24015% and 0.25353% and was compared with the LSTM (Long Short-Term Memory) model, which showed a MAPE value of 43.88%, indicating that LSTM outperformed Fb Prophet in terms of percentage error.

Although the stock market and the financial sector are possible fields for applying these models, they can be adapted to a variety of forecasts and analyses. In this study, the model relies exclusively on historical data from cryptocurrencies. In this way, the study concluded that by using a quality set of historical data, it is possible to build robust models for analysis and forecasting.

The article [14] suggests a hybrid approach to time series forecasting, combining the Prophet model with an improved version of the LSTM (Long Short-Memory) neural network, called P-gLSTNet. The main objective of this study is to improve the accuracy of predicting time series data, particularly the prediction of pollutants from high-emission vehicles. In this way, the study also explores the use of ARIMA, LSTM, Prophet and P-gLSTNet models in the air emissions scenario, using vehicle data in two Chinese cities and compares their performance through metrics such as mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE). The proposed hybrid model combines Prophet's ability to detect seasonality and events with LSTM's ability to model long-term dependencies in time series. The results showed that of the four models, the worst performer was ARIMA and the best performer was the P-gLSTNet hybrid model, thus demonstrating that the ARIMA model is not suitable for forecasting time series in the fields of meteorology and environmental energy. It can also be noted that in all data sets, P-gLSTNet was the most accurate model, followed by LSTM and Prophet.

In conclusion, the P-gLSTNet hybrid model proved to be more effective than traditional models such as ARIMA, LSTM and Prophet at capturing both seasonal trends and complex temporal dependencies. Therefore, this approach is significantly useful in applications involving high-dimensional multivariate time series such as urban traffic forecasting or pollutant emissions.

The article [13] analyzes the use of the SARIMA model in short-term traffic flow prediction, which is fundamental for management and control in advanced transport systems, such as Advanced Traffic Management Systems (ATMS). This study differs from others in that it uses a minimal amount of input data, i.e. three days of observations to predict the next day's traffic flow, unlike other similar studies which usually use large databases. The results showed that the SARIMA model achieved high accuracy in predicting traffic flow, being able to predict peak and off-peak traffic volumes with results close to the real values.

Consequently, it was possible to conclude that the SARIMA model is effective in predicting short-term traffic flow when using a limited amount of historical data, making it suitable for scenarios where the availability of large databases is restricted. This model proved capable of

capturing daily seasonal traffic patterns, outperforming traditional models such as the historical average and the naive method. In addition, it showed that when historical data is combined with data obtained in real-time, the error is even smaller, highlighting the potential of this approach for greater forecasting accuracy.

3.4 Limitations

Although Prophet is a model widely used in different areas, it is possible to identify some limitations. The model has difficulties in capturing abrupt or unpredictable events, such as the volatility of cryptocurrency prices, just as in sales forecasts, it has shown obstacles in predicting sudden changes in consumer behavior. When forecasting road traffic, it performed well in scenarios with clear seasonality but showed some limitations in predicting more complex variations and unexpected emergencies.

The SARIMA model also showed limitations in predicting irregular traffic, particularly when there is little historical data available.

In addition, the model's performance depends on the quality and consistency of the historical data. In scenarios where patterns have a high level of volatility, such as financial markets or unexpected traffic events, it is likely that the model will need to be adjusted or combined with other models for its performance to be effective.

In conclusion, the studies described throughout this chapter have demonstrated the potential and limitations of the Prophet model in different areas, such as road traffic, sales and finance. Although Prophet shows good results in scenarios with well-defined seasonal patterns, it has difficulties in capturing unpredictable events and wide variations.

It is important to highlight that complementary models such as Neural-Prophet and LSTM have been proposed in the literature to deal with these difficulties, to offer better results in more volatile contexts. However, there is still progress to be made to adapt Prophet and other time series models to scenarios that are more complex and dynamic. Future studies should focus on improving the unpredictable variations that can occur in these contexts to improve forecasts and make them more robust and accurate.

Chapter 4

Design, Implementation and Results

This chapter presents the structure of the data used in this dissertation through to the results obtained in the traffic analysis and forecasting process, exploring how the data was collected, structured, processed and analyzed to obtain valuable information about its patterns in the city of Porto. Firstly, the data provided by the Porto Digital Association, obtained from inductive loop detectors, is described in detail. After this, an exploratory analysis was carried out and then a forecasting system was implemented using the Prophet model to predict intensity and occupancy. The model was enriched with external variables such as football games at the Estádio do Dragão and national holidays to evaluate its performance. Finally, the results of the cross-validation method in different scenarios are discussed, both in short-term and long-term forecasts, as well as the relative errors of MAE and RMSE both in the initial phase with the detector located near and far from the Estádio do Dragão and in the fifty detectors spread throughout the municipality of Porto.

4.1 Data Collection

In this section, it will be presented the structure of the data and a detailed description of the objects that it contains. This data comes from inductive loops spread throughout the municipality of Porto on the streets with traffic lights and they are used to collect traffic data which is then sent to the controllers, which are physical devices housed in street-side telecoms cabinets. They measure the intensity (number of vehicles per unit of time) and occupancy (percentage of time that the metallic object was detected over the loop).

4.1.1 Structure of the Data

The data provided by Porto Digital Association is stored in a compressed file organized chronologically by year, with separate directories reserved for each month. Inside each monthly directory are files in JSON format, each corresponding to the readings from a specific controller, taken at 10-

minute intervals. The files are designated following a specific format, for example `CMP_070.json`. The acronym CMP stands for “Câmara Municipal do Porto” and the designation 070 denotes the distinct identification number of the controller associated with each detector. The detector is then characterized by a unique id present within each controller’s JSON file, for example, `CMP_070-e1.json`. Each controller can have more than one detector.

The dataset used in this thesis corresponds to the period from January 2022 to December 2022. The data results from 61 controllers and 695 detectors located in the city. It contains detailed information on traffic intensity and occupancy across various areas of the city.

There are city maps that illustrate the precise positions of controllers placed within the city. A controller, typically, gathers data from all detectors at a single intersection. A single controller does, however, occasionally gather data from two or three nearby intersections as it can be seen in Figure 4.1. This figure illustrates one controller located in the streets "Rua da Alegria" and "Rua Latino Coelho" gathering data from detectors CMP_070-e1 to CMP_070-e4. In the ID of the controller, we use the letter “e”, to refer to “espira”, numerated from 1 to the number of detectors that the controller has. "Espira" is the name in Portuguese that we use to refer to the inductive loop.

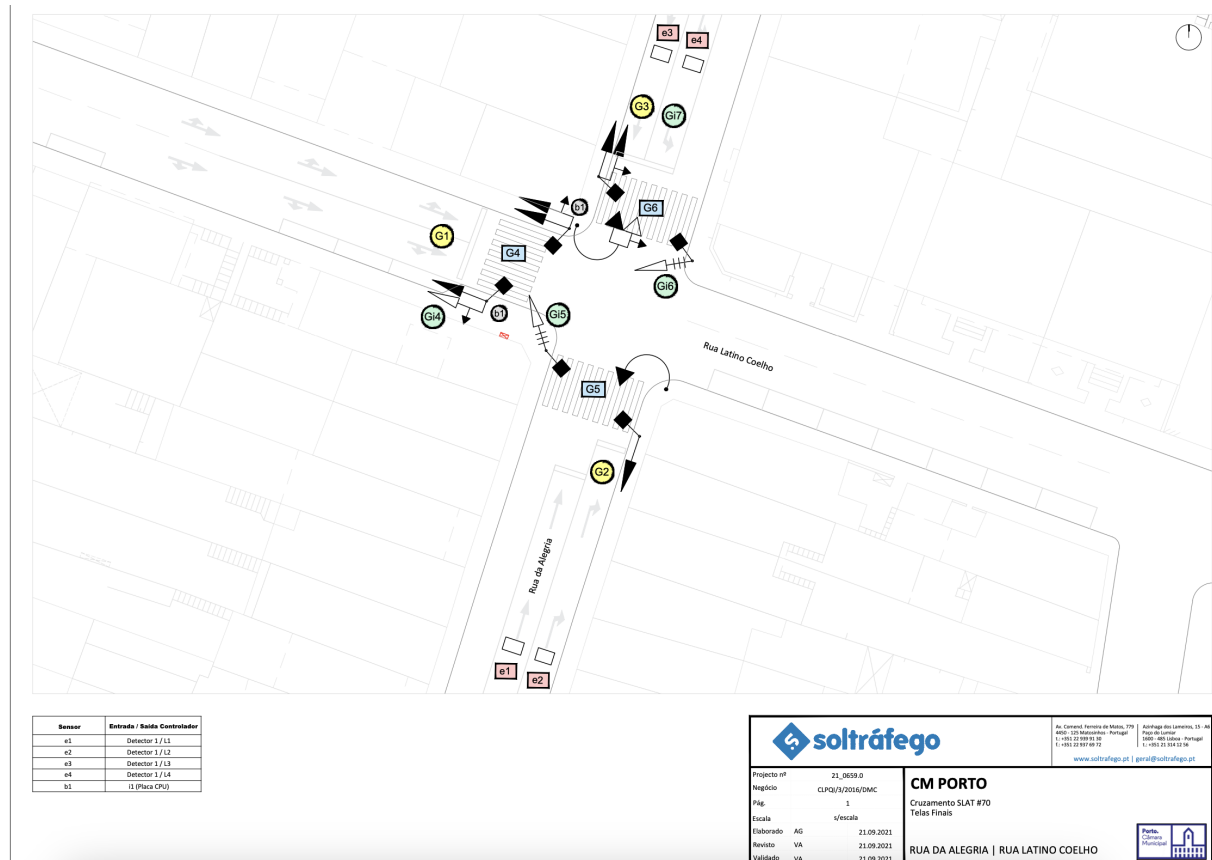


Figure 4.1: Location of the detectors present in the controller CMP_070 of the crossing of Rua da Alegria and Rua Latino Coelho

4.1.2 Description of Data Objects

The list of relevant objects that allow us to interpret the recorded readings from the road detectors is provided next. For each object, we give its name, a description of what it does, the type of data we need to access and an example of the data value.

dateObservedFrom - a discrete numerical value whose datatype is a `Datetime`, specifying the precise day and time at which the observation period begins. An example value is `2022-02-01T00:10:00+00:00`.

dateObservedTo - a discrete numerical value whose datatype is a `Datetime`, indicating the termination point, in terms of day and time, of the observation period. An example value is `2022-02-01T00:10:00+00:00`.

Id - the traffic detector's unique identification number. The format is structured as `CMP_XXX-YYY`, where `xxx` and `yyy` represent the controller and detector identifications, respectively. An example value is as follows: `urn:ngsi-ld:TrafficFlowObserved:porto:traffic:CMP\_070-e1`

location:value_coordinates - represents the estimated position of the controller, not the inductive loop, given by the geographical coordinates representing a position point where the detector is typically located. Its datatype is a `Geolocation` with an example format given by `(41.345676,-8.90076)`.

Intensity - represents the traffic flow observed per hour, corresponding to the number of cars observed in that period. Observations are made every 10 minutes and the number of cars observed in the last 10 minutes is multiplied by 6 to indicate the expected observation per hour. Its datatype is an `int` and if the value is 24, it means that 4 vehicles were observed in the last 10 minutes.

Occupancy - this integer represents the occupancy of traffic flow represented as a percentage of the observation period during which a metallic object was over the loop. Its datatype is an `int` and, for example, a value 4 indicates that the loop had a metal object over it 4 percent of the 10-minute interval.

The following image shows how these data objects are represented in the `CMP_070` controller JSON file present in the January 2022 folder:

```
[
  {
    "dateObservedFrom": {
      "value": "2022-01-01T00:10:00+00:00"
    },
    "dateObservedTo": {
      "value": "2022-01-01T00:20:00+00:00"
    },
    "id": "urn:ngsi-ld:TrafficFlowObserved:porto:traffic:CMP_070-e2",
    "location": {
      "value": {
        "coordinates": [
          41.159431,
          -8.601064
        ]
      }
    },
    "intensity": {
      "value": 144
    },
    "occupancy": {
      "value": 7
    }
  },
  {
    "dateObservedFrom": {
      "value": "2022-01-01T00:20:00+00:00"
    },
    "dateObservedTo": {
      "value": "2022-01-01T00:30:00+00:00"
    },
    "id": "urn:ngsi-ld:TrafficFlowObserved:porto:traffic:CMP_070-e2",
    "location": {
      "value": {
        "coordinates": [
          41.159431,
          -8.601064
        ]
      }
    },
    "intensity": {
      "value": 252
    },
    "occupancy": {
      "value": 24
    }
  }
]
```

Figure 4.2: Section of the JSON file describing the CMP_070 controller.

4.2 Exploratory Analysis

Initially, the data-gathering procedure for the exploratory study began with the traffic controllers employed by the Municipality of Porto. The dataset was provided by the Porto Digital Association.

To streamline the manipulation and processing of data for the implementation of the forecasting system, a Python script was developed to systematically import each JSON file from the designated root directory. Since the dataset provided contained some empty JSON files for certain controllers, the script needed to handle these empty files to ensure that each controller's file ultimately contained usable data. This was necessary to maintain data integrity and completeness across all monthly folders.

Before selecting the detectors for our initial study, we implemented a careful evaluation procedure to guarantee the inclusion of those detectors that had accessible data. The temporal scope employed was January 2022, with particular emphasis on the detectors CMP_070-e1, CMP_070-e2, CMP_131-e2, and CMP_262-e2. The selection of these detectors was done randomly from a sample of detectors that had been verified to possess data files that were not empty.

The timeline of January 2022 was chosen for the exploratory analysis because it would take too long to analyze the entire data set in Google Colab, which is where the code was implemented, so it was analyzed over a shorter period of time.

4.2.1 Plots of the Exploratory Analysis

Figures 4.3 and 4.4 illustrate the plot of the time series corresponding to the data for intensity and occupancy for detector CMP_070-e2.

- **Monthly Plot:** Displays the intensity and occupancy for the full month of January 2022. In the following example, we can see that there is a time in the month with missing data probably due to an error in the collection of the data from the controller CMP_070 in that period.

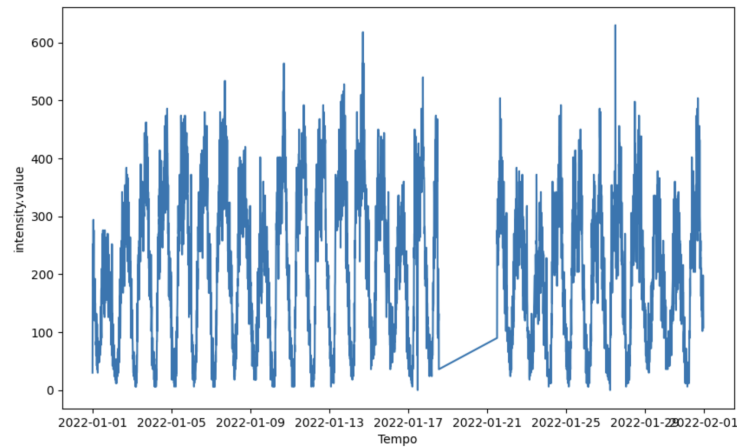


Figure 4.3: Monthly plot for the detector CMP_070-e2 for the intensity

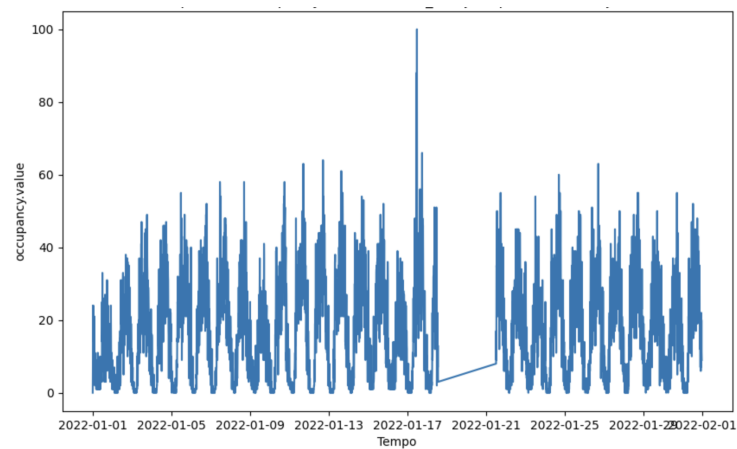


Figure 4.4: Monthly plot for the detector CMP_070-e2 for the occupancy

Analysis of the data reveals a regular trend in the intensity, which represents the number

of cars present during the month. The data demonstrates a cyclical pattern and consistent during the whole month. Furthermore, the occupancy statistics exhibit a similar cyclical pattern, but with a decreased percentage and continuously lower values for the whole month.

The monthly data highlights these cyclical patterns demonstrating the persistent nature of traffic behavior, characterized by significant rises in both occupancy and intensity on weekdays. The cyclical pattern described is characteristic in municipal environments when traffic maximum values are systematically seen at certain periods of the day.

For instance, if we observe the figures 4.3 and 4.4, and we compare the 1st of January 2022, which was a Saturday, to the 5th of January which was a Wednesday, we can see in both plots that the intensity and occupancy values are lower on the 1st compared to the 5th.

The intensity plot shows a regular cyclical pattern, which probably corresponds to the daily behavior of the traffic. However, there is a large decrease around January 18th and 19th, which suggests a lack of data from the controller for this detector on these days. This could be related to the fact that the controllers were relatively new at the time and there was still the possibility of technical failures reflecting momentary problems with the controller.

A particularly interesting observation is the sharp peak in intensity on January 27th, where a maximum of 630 vehicles was recorded. This anomaly stands out from the regular pattern seen on other days. After investigating local events in the city of Porto on that day and finding that there was no specific event that could justify such a significant increase in traffic flow on that day, some hypotheses emerged that could be considered. As this increase occurred between 12pm and 3pm, which is different from the traditional peak hours, it could be related to unexpected changes in traffic or an increase in commercial vehicles, or it could even be a technical fault on the part of the controller. The next plot shows the intensity values on that day.

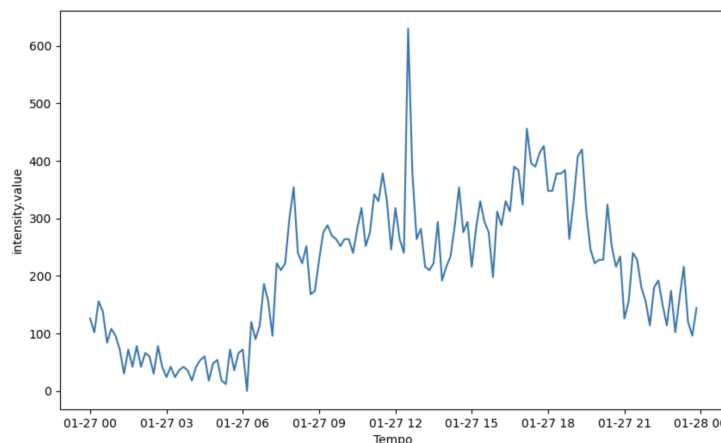


Figure 4.5: Plot of the intensity for January 27th from the detector CMP_070-e2

In the occupancy plot, we observe a pattern similar to that of intensity. However, there is a specific moment on January 17th, at approximately 12 o'clock, when the occupancy value

reaches 100%. This anomaly could be due to a failure in the controller's data logging or it might indicate an instance of atypical congestion, potentially caused by a road accident or other unforeseen traffic event. The next plot illustrates the occupancy values on that day.

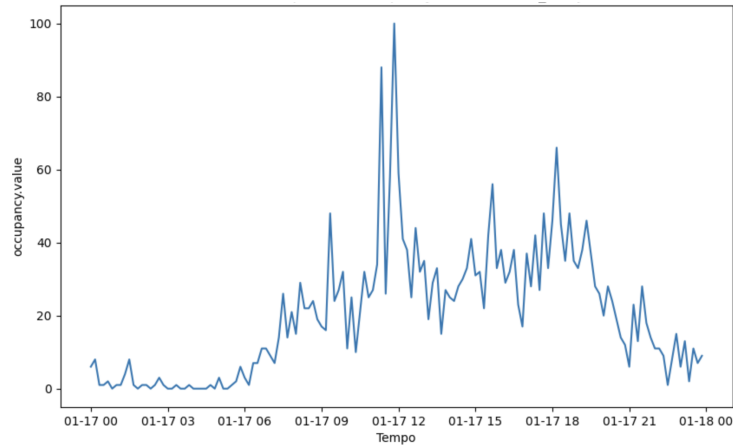


Figure 4.6: Plot of the occupancy for January 17th from the detector CMP_070-e2

Despite this atypical value, the general pattern of occupancy throughout the month assumes peaks between 0% and 60%, which seem normal and correspond to times of day with more and less traffic, while the higher values may correspond to normal peak hours.

- **Weekly Plot:** Shows the flow of traffic in a normal week from 24th of January until 31 of January 2022.

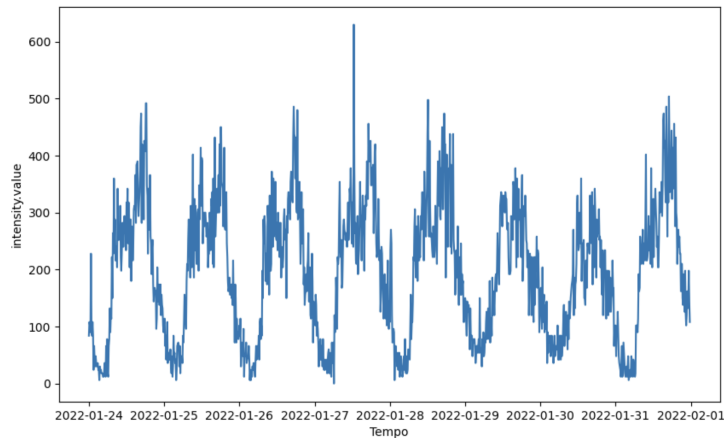


Figure 4.7: Weekly plot for the detector CMP_070-e2 for the intensity

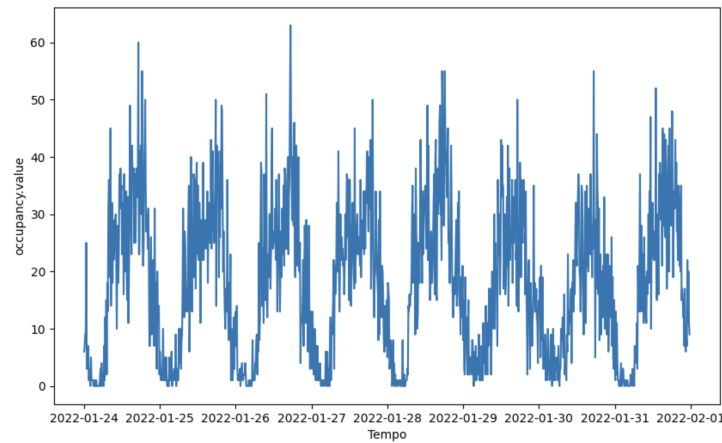


Figure 4.8: Weekly plot for the detector CMP_070-e2 for the occupancy

On weekdays, the data exhibits several distinct peaks in both occupancy and intensity, probably aligning with the morning rush hours (about 7-9 AM) and evening rush hours (approximately 5-7 PM).

On January 29th and 30th, a decrease in traffic intensity values was observed. This reduction is expected, as these dates correspond to a Saturday and Sunday, respectively. Weekend days typically experience less frequent and irregular travel patterns, as they are not dominated by routine commutes to work or school. Consequently, traffic tends to be less predictable and consistent, often resulting in lower intensity levels on these days.

The occupancy values show a different pattern than the intensity values. While the intensity values show sharp peaks during rush hours, the variation in occupancy is less pronounced. and through the whole week, we can see that there is a minimal variation and the variations remain more moderate throughout the week.

- **Daily Plot:** Describes the changes in occupancy and intensity during a given workday. In this case, it was the 25th of January 2022 corresponding to a Tuesday.

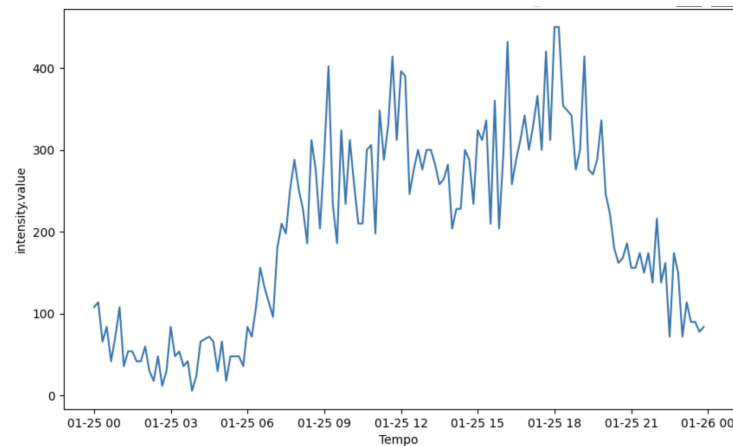


Figure 4.9: Daily plot for the detector CMP_070-e2 for the intensity

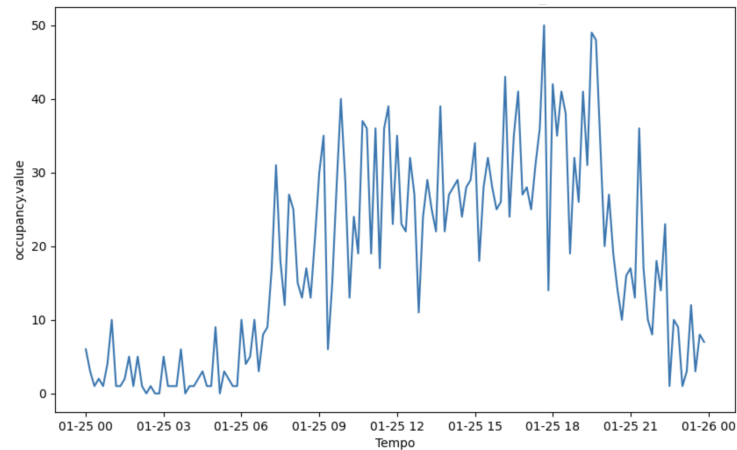


Figure 4.10: Daily plot for the detector CMP_070-e2 for the occupancy

The daily plots provide an in-depth analysis of traffic variations within a single day. In the early morning, both occupancy and intensity experience substantial increases, which are then followed by a slight decrease in the mid-morning, before the late afternoon.

As anticipated, nighttime traffic experienced a substantial decline, as both occupancy and intensity indices reached their lowest points because most people are at home and there are fewer vehicles on the road.

- **Weekend Plot:** Examines the differences in traffic on a weekend day. In this case, it was the 23rd of January 2022 corresponding to a Sunday.

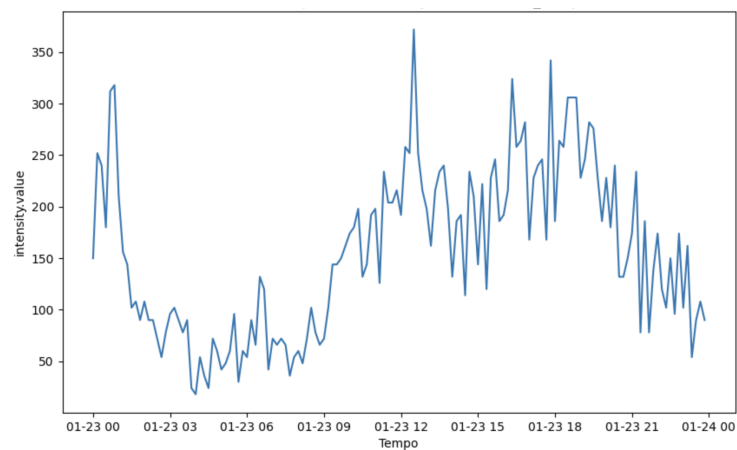


Figure 4.11: Weekend plot for the detector CMP_070-e2 for the intensity

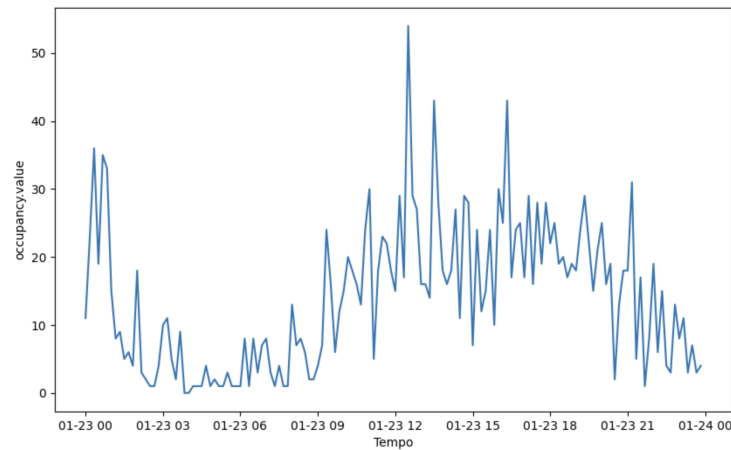


Figure 4.12: Weekend plot for the detector CMP_070-e2 for the occupancy

The weekend plots show that the intensity values are high between 12 AM and 9 PM, which is probably accurate since on Sundays, people tend to leave the house more in the afternoon than in the morning when the values are lower.

As for the occupancy values, we can see that they are higher between 12 AM and 6 PM, probably due to second-row parking or inappropriate parking on the streets, or it could be just an interval during which there is more occupancy on the streets.

4.2.2 Data Availability of the data

We noticed that several detectors did not provide data for every day of the month during the visual analysis. In particular, we observed that in the detector CMP_070-e1, the data for intensity did not begin at the beginning of the month, but rather on a specific day in January 2022:

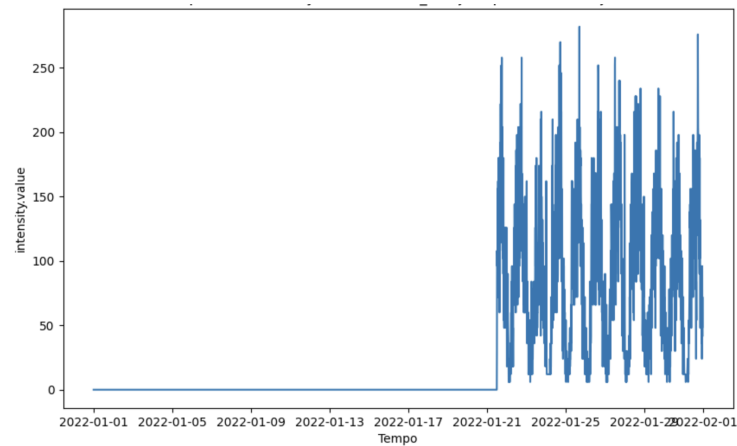


Figure 4.13: Plot that shows the intensity on January 2022 for the detector e1 in the controller CMP_070-e1.

4.2.3 Detection of Anomalies on the data

There are a few detectors that exhibit pauses in the data collection. For example, CMP_070-e2 shows gaps in the data for January 2022, as illustrated in the following image:

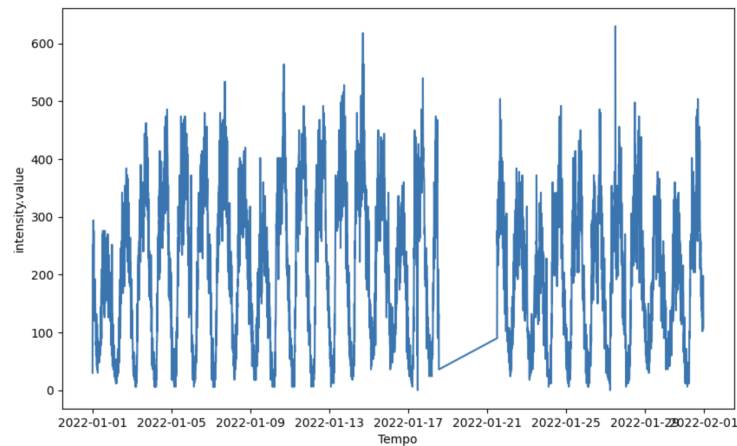


Figure 4.14: Plot that shows the intensity on January 2022 for the detector e2 in the controller CMP_070-e2.

The occupancy statistics recorded in CMP_131.json and the detector CMP_131-e1 for the week of January 10 to 16, 2022, reveal atypical occupancy values. Consistently reaching a maximum of 100 on all days, these values may indicate improper parking, for example.

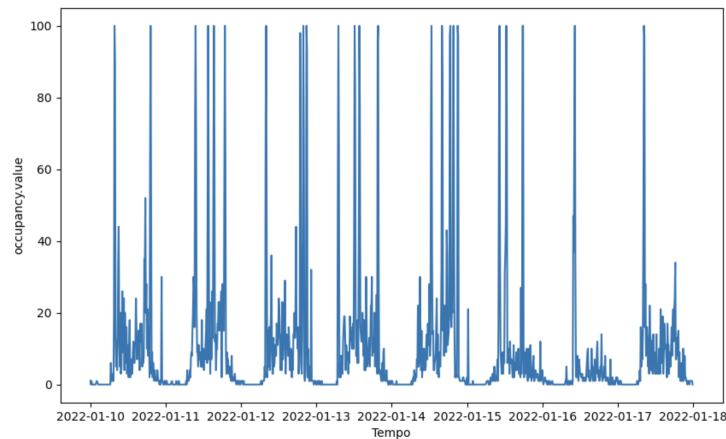


Figure 4.15: Weekly plot that shows the occupancy values of the detector CMP_131-e1 between January 10 and 17.

Conclusions and Implications

- **Predictability:** The data indicates that traffic patterns exhibit some degree of predictability, especially on weekdays. This predictability is crucial for the implementation of a time series forecasting system.
- **Weekend vs Week days:** A comparison of weekend and weekday traffic shows that the differences in traffic patterns align with expectations. Weekends display lower and more variable levels of traffic density compared to weekdays.

4.2.4 Scatter Plots that compare intensity with occupancy

Scatter plots were performed to analyse the intensity and occupancy of the detectors over time on different time scales. We were able to determine relationships between occupancy and intensity using the scatter plots and we can compare them to the fundamental diagram of flow described in the section 2.2 of chapter 2.

Therefore, one can consider that in the scatter plots used for this study, intensity corresponds to the flow, as it represents the number of vehicles per unit of time, and traffic density, although it is not the same as occupancy, it functions as a proxy for density since if vehicles are close together, occupancy will be higher than if they are spaced further apart, which is similar to density (number of vehicles on a given stretch of road).

Below is an in-depth analysis of the observed patterns.

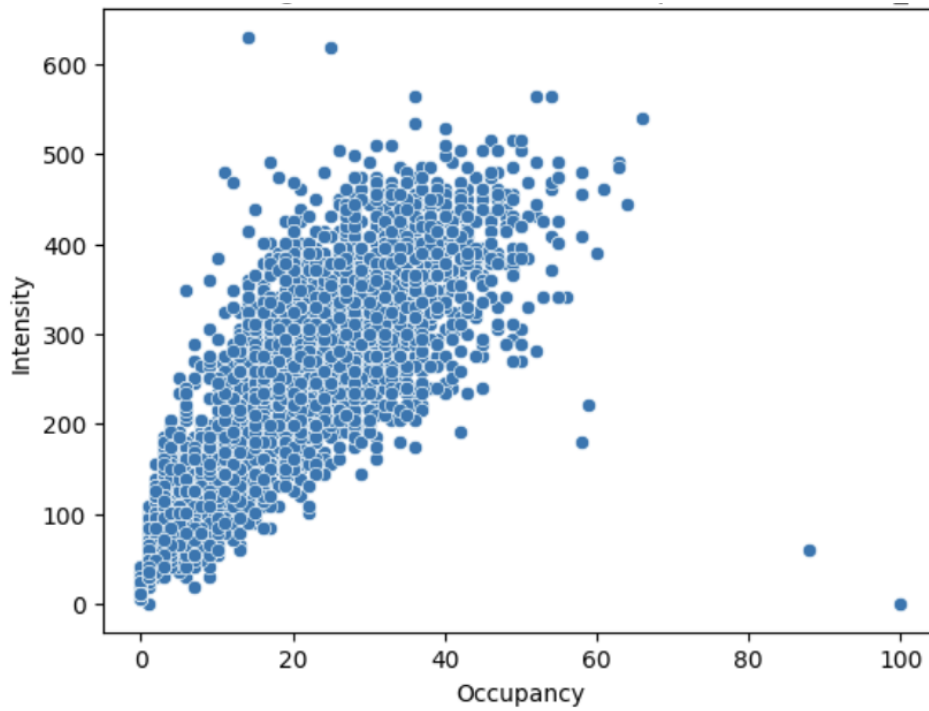


Figure 4.16: Scatter plot of January 2022 between the intensity and occupancy of the detector CMP_070-e2.

Monthly Analysis: An analysis of the scatterplot 4.16 for January 2022 reveals the diagram relating flow to density. In this plot it can be observed that as the intensity increases the occupancy also increases, showing that at lower occupancy values (0-40) the intensity increases linearly (free-flow). When occupancy reaches approximately 40, the plot shows a different behavior, with a negative slope and a dispersion of intensity values and the traffic is in the constrained flow phase, since traffic is unstable but there is no indication that it has reached the congestion phase. The congestion phase would be reached if intensity and occupancy stabilized at high levels.

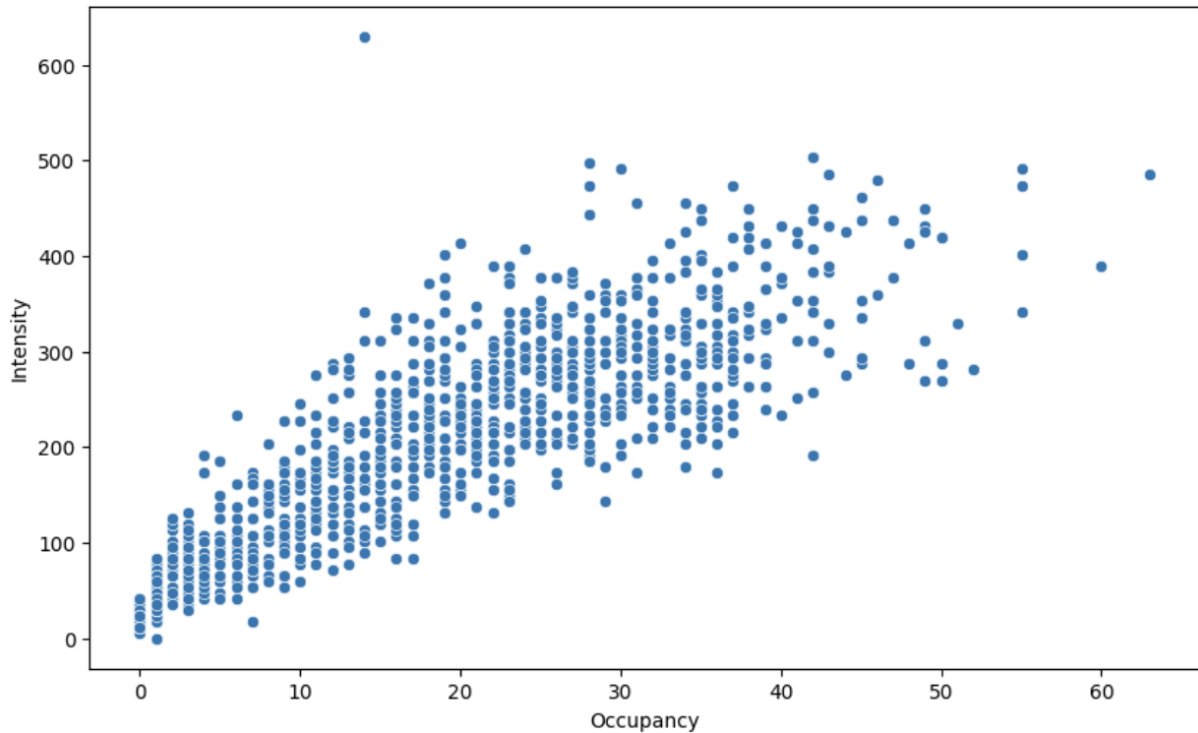


Figure 4.17: Scatter plot of the Weekly Analysis between the intensity and occupancy of the detector CMP_070-e2.

Weekly Analysis - 24th January 2022 to 31th January 2022 The weekly scatterplot 4.17 also demonstrates a positive correlation between intensity and occupancy but with lower values of intensity compared to the monthly scatterplot. In this plot, the intensity concentrates between the range of 0 to 400 cars per hour and the occupancy values up to 60%. Despite the stronger correlation between occupancy and intensity, the plot still shows some outlier points that may indicate instances of unusually high traffic flow.

It is relevant that this plot shows similar behavior to that described in the monthly plot: an initial growth in intensity proportional to occupancy (free-flow) and a constrained flow phase, with no signs of congestion.

Daily Analysis - 24th January 2022 - Monday

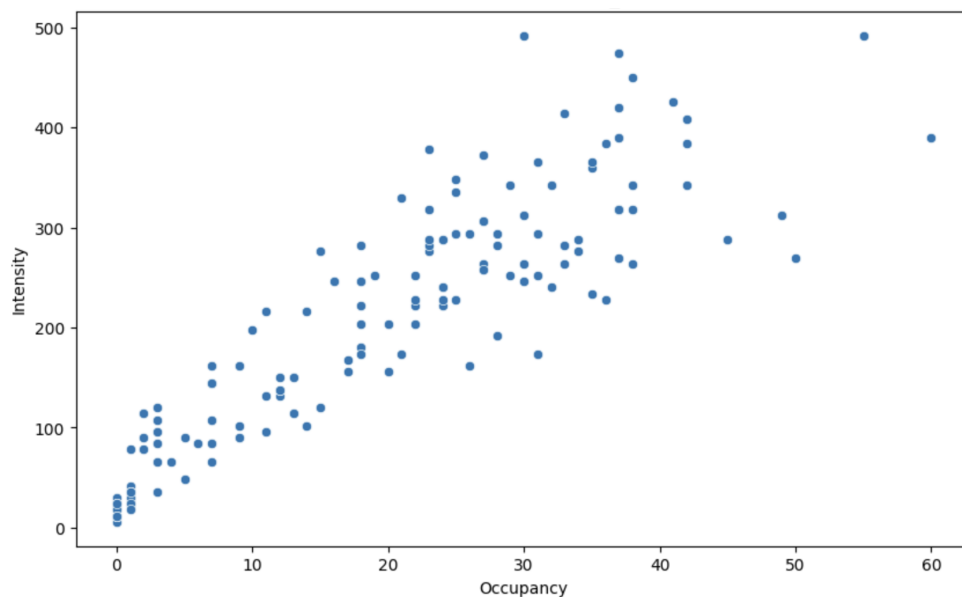


Figure 4.18: Scatter plot of the Weekday (Monday) Analysis between the intensity and occupancy of the detector CMP_070-e2.

Daily Analysis - 23th January 2022 - Sunday

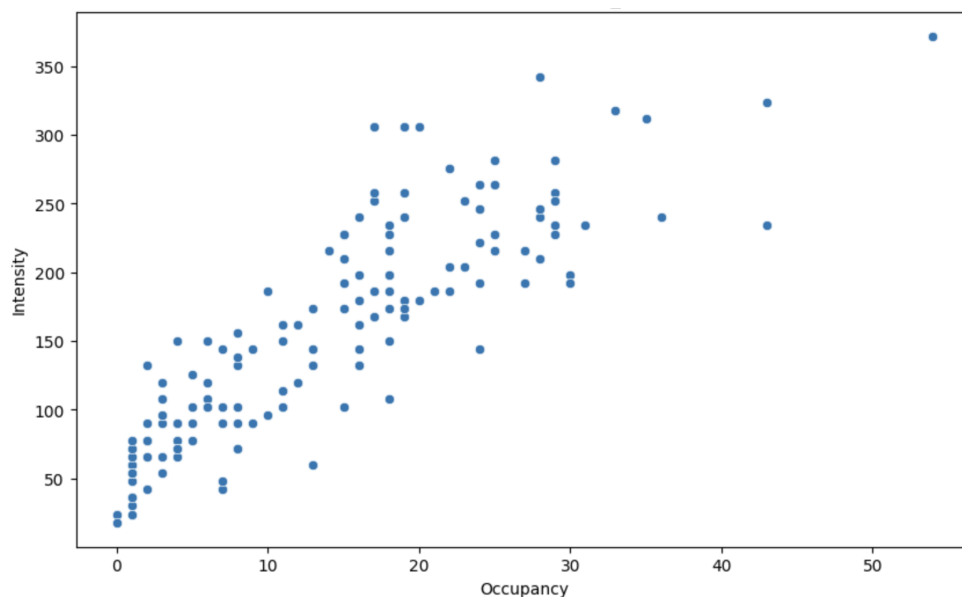


Figure 4.19: Scatter plot of the Weekend (Sunday) Analysis between the intensity and occupancy of the detector CMP_070-e2.

The figures 4.18 and 4.19 compare traffic on a Monday (January 24) and a Sunday (January 23) in 2022, respectively. Both plots show that intensity and occupancy increase proportionally.

It is, therefore, possible to compare this behavior with the initial phase of the traffic flow diagram, where flow increases directly proportionally with density (free flow).

Despite this, it can be seen that on Sunday, the intensity values are lower than on Monday, which is This is to be expected since there isn't as much traffic during the weekend as on a weekday.

It is relevant that these plots show similar behavior to that described in the monthly plot: they show an initial growth in intensity proportional to occupancy (free-flow), a constrained flow phase, with no signs of congestion.

4.3 Forecasting process using the Prophet model

The process of predicting the intensity and occupancy returned by the inductive loops in the Municipality of Porto was carried out using the Prophet model. For a more detailed explanation, see Chapter 2, section 2.4. The steps followed to carry out this forecast are described below.

4.3.1 Reading and merging data:

The first step consisted of iteratively reading the JSON files corresponding to each month's data. This data was filtered based on the ID of the target detector. The Prophet model was applied to two different detectors: detector e2 of controller CMP_070, which is located between Rua da Alegria and Rua Latino Coelho. There is a map of this controller in the Data collection section of this chapter. And then detector e3 on controller CMP_020, which is located between Avenida Fernão Magalhães, Rua Fernando Bulhões and Rua Santo António das Antas. We chose these 2 detectors to test the model to evaluate the traffic behavior at a detector that is located at a greater distance (CMP_070-e2) and another closer (CMP_020-e3) to the Estádio do Dragão, since we incorporated a list of football games played in that stadium in the year 2022. After this, the Prophet model was applied to more detectors, more precisely, fifty detectors in different areas of the Municipality of Porto as well in different periods and horizons. To evaluate the performance of all the detectors, the relative errors of RMSE and MAE were calculated.

4.3.2 Preparing the data for Prophet:

In order to use the data in Prophet, the `dateObservedFrom` column has been converted to `datetime` format, removing the time zone information, as required by the model. Next, we separate the intensity and occupancy data into 2 data frames:

- **Intensity DataFrame:** This DataFrame included the `ds` column (timestamps) and the `y` representing intensity values.

- **Occupancy DataFrame:** This DataFrame contained the `ds` column and the `y` column representing occupancy values.

4.3.3 Filtering data for the desired period:

To focus the forecast on a specific period, the data was filtered to include only the period between March 1, 2022 and May 31, 2022. The resulting data was stored in the `filtered_dataframe_intensity` and `filtered_dataframe_occupancy` data frames, which represent the historical data used by Prophet.

4.3.4 Incorporation of public holidays and football games:

In addition to the historical data, we added a list of national holidays taking place in 2022 (`pt_holidays_df`) and a list of football games taking place at the Estádio do Dragão called (`porto_football_games`) given by the Porto Digital Association. Both were combined into a single data frame called `holidays`, which was then passed to Prophet during initialization in order to help improve the predictive capacity in relation to specific events.

4.3.5 Train the Prophet model:

Two different Prophet models were trained since the objective was to predict two different variables, intensity and occupancy. The model that predicts the intensity was initialized with a Prophet object and the dataframes containing the football games and the holidays were added. After that, the model uses the data frame with the intensity values and corresponding dates of the period we want to analyse to make the prediction. The same was done to the occupancy prediction model.

4.3.6 Generating future forecasts:

After Prophet trained each of the models (for intensity and occupancy), predictions were made for the following 30 days based on the historical data. To do this, it was necessary to calculate the number of 10-minute intervals in the 30 days, knowing that each day has 24 hours and that each hour has 6 10-minute intervals ($30 \times 24 \times 6$). These time intervals were stored in a dataframe which allows the model to predict the intensity and occupancy values in each of them.

4.3.7 Cross-validation and performance metrics:

To assess the model's performance, we conducted cross-validation (explained in more detail in chapter 2 in 2.4.2) using Prophet's cross-validation method. This process tested the forecasting capacity in 3 different scenarios, with the following parameters:

- Scenario 1 : initial = 30days, period= 10min and horizon= 10min
- Scenario 2: initial = 30days, period= 1hour and horizon= 1hour
- Scenario 3: initial= 30days, period= 1day and horizons=[1day,7days,14days,28days]

During cross-validation, Prophet trains the model up to a certain point in time (cutoff) and makes predictions for the defined horizon, moving the cutoff according to the specified period. The predictions are then compared with the actual values and metrics such as MAE (mean absolute error) and RMSE (root mean square error) are calculated.

4.3.8 Cross-validation process in Prophet

To illustrate the cross-validation process used in the Prophet model, we defined the following configuration: an initial training period of 30 days, a prediction horizon of 1 day, and a cutoff period of 1 day. The model iterates through the available data, updating and generating predictions at each step.

The parameters are as follows:

- Initial training period: 30 days of historical data.
- Prediction horizon: 1 day of data after each cutoff.
- Cutoff period: Every 1 day.

How Prophet iterates:

- **First iteration (Cutoff 1):** The model is trained with the first 30 days of data, makes predictions for the next 1 day, compares the predictions with the actual data, and calculates metrics such as MAE and RMSE.
- **Second iteration (Cutoff 2):** The model moves forward by 1 day, retrains with the data available until the second cutoff, and makes predictions for the next 1 day. The performance metrics are recalculated.
- **Subsequent iterations:** This process repeats, moving the cutoff forward 1 day at a time and generating predictions for the next 1 day until all the data has been processed.

Next, it illustrated an example for a better understating of the process of cross-validation:

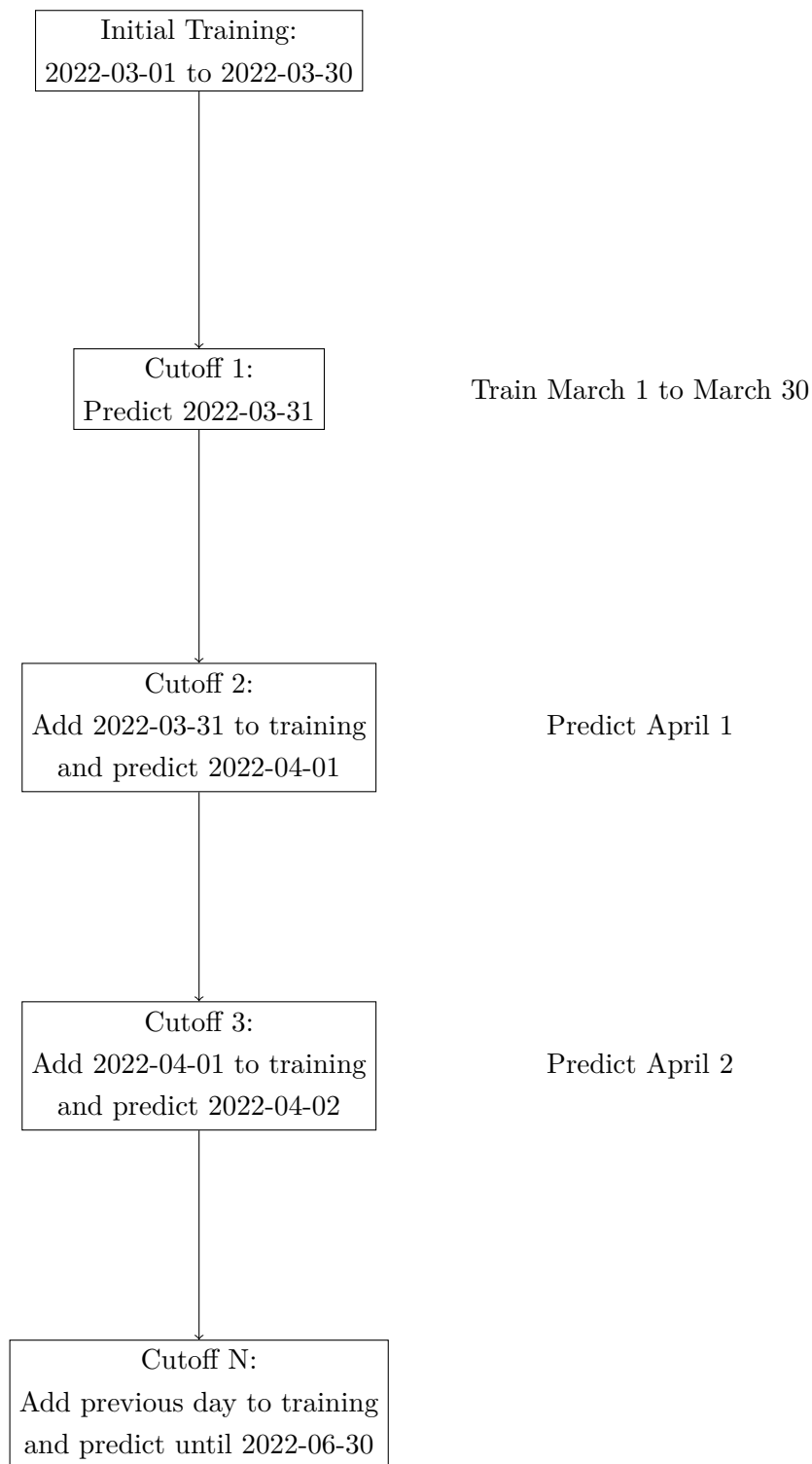


Figure 4.20: Example of a Prophet Cross-validation process

After running the cross-validation, the performance metrics are obtained using the following code: `performance_metrics_intensity = performance_metrics(cv_results)`

These metrics include several metrics but in this study, we focus on the following, these

metrics were defined in the Chapter 2 subsection 2.4.3:

- MAE (Mean Absolute Error): Mean absolute error between actual and predicted values.
- RMSE (Root Mean Square Error): Square root of the mean squared error.

These metrics, obtained from cross-validation, were used to assess the accuracy of the forecasts.

4.3.9 Evaluating the performance of the forecasting system with 2 detectors - CMP__020-e3 and CMP__070-e2

To evaluate the performance of Prophet's cross-validation method, we chose the following parameters for the start date, end date and initial period:

- **Start Date:** '2022-03-01'
- **End Date:** '2022-05-31'
- **Initial Period:** '30 days'

The Start Date, End Date and Initial period were equal to all of the experiences we did but the period and the horizon of the cross-validation would change depending on the horizon since the value of the horizon is higher or equal to the value of the period. We therefore carried out the cross-validation method in the following scenarios:

- 10-minute period and 10-minute horizon
- 1-hour period and 1-hour horizon
- 1-day period and 1-day, 7-days, 14-days and 28-days horizons

For the 2 detectors - CMP__070-e2 and CMP__020-e3, cross-validation was carried out on all the scenarios mentioned above, for both intensity and occupancy.

After obtaining these cross-validation metric results, to be able to assess whether the error values are good or bad, we chose 2 of them for this purpose: the Mean Absolute Error (MAE) and the Root Mean Absolute Error (RMSE) and calculated the relative error of them using the MAE and RMSE values of each scenario(respective period and horizon values) and the average of the y values(intensity or occupancy)for each of the detectors.

The relative MAE (RMAE) and relative RMSE (RRMSE) are calculated as follows:

$$\text{RMAE} = \left(\frac{\text{MAE}}{\bar{y}} \right) \times 100$$

$$\text{RRMSE} = \left(\frac{\text{RMSE}}{\bar{y}} \right) \times 100$$

where $\bar{y}$ represents the mean of the observed values (intensity or occupancy), providing a normalized error relative to the scale of the data.

The acronyms for these relative errors are: Relative Mean Absolute Error (**RMAE**) and Relative Root Mean Squared Error (**RRMSE**).

In the Appendix A is the table of the results of the detector CMP_020-e3 with the respective values of intensity and occupancy as well as the mean and the values of relative RMAE and RRMSE in the periods and horizons considered for this prediction.

Next, the following plots represent the values of intensity and occupancy for CMP_020-e3 in the respective periods and horizons.

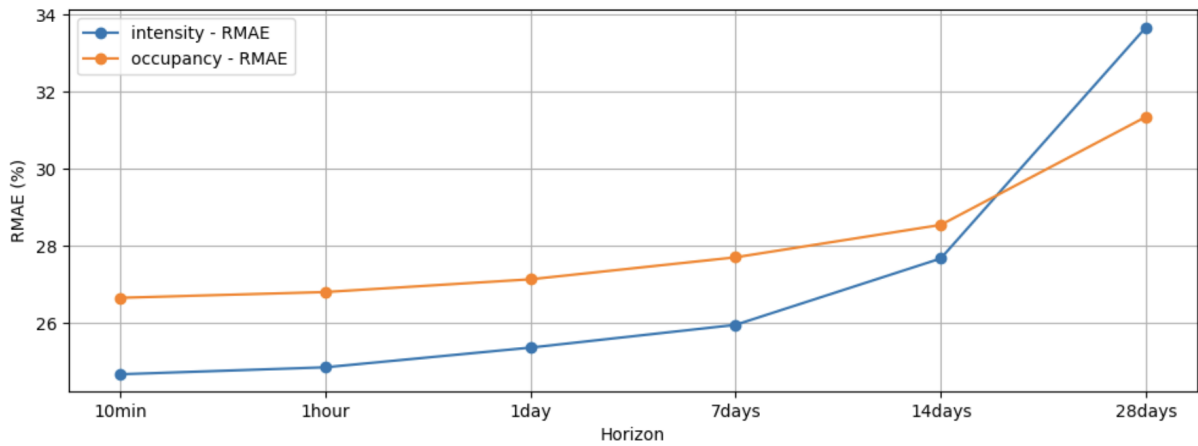


Figure 4.21: Plot that compares the values of RRMAE for CMP_020-e3

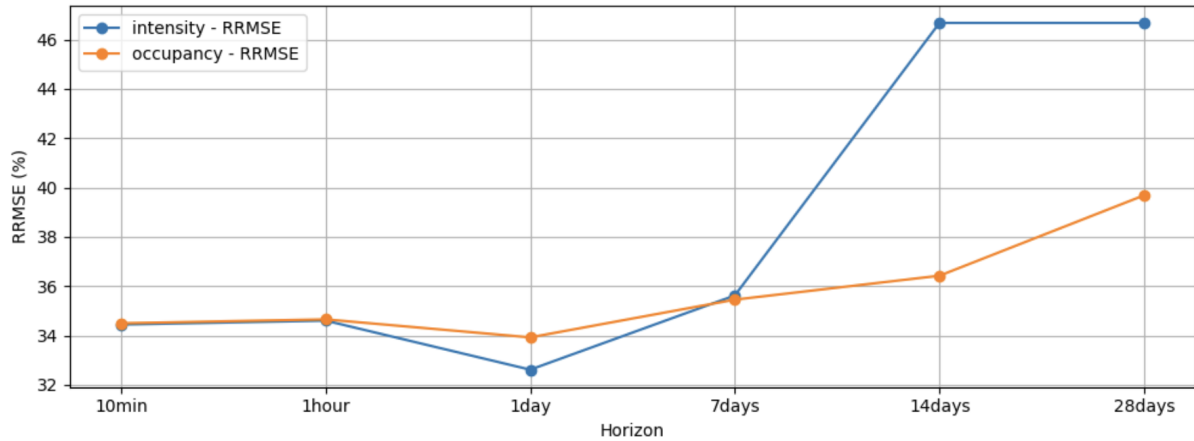


Figure 4.22: Plot that compares the values of RRMSE for CMP_020-e3

Intensity results for CMP_020-e3: The RMAE values range from 24.68% for the 10-minute horizon to 33.77% for the 28-day horizon and we can see that as the horizon increases, the RMAE value also increases. As for the RRMSE values, they start at 34.44% for the 10-minute horizon, which is higher than those of CMP_070-e2. We can suggest that the system consequently indicates that there is greater variation in the error when predicting intensity close to the Estádio do Dragão. As the horizon increases, the RRMSE increases to 46.67% for 14 days and 28 days, reflecting a significant increase when compared to shorter horizons, despite being the same for these 2 longer horizons. Therefore, the system's accuracy is greater at shorter than longer intervals.

Occupancy results for CMP_020-e3: For short periods and horizons (10 minutes and 1 hour), the RMAE is between 26.66% and 26.81%. As with CMP_070-e2, the error increases as the forecast horizon lengthens. For 28-day horizons, the RMAE rises to 31.35%, suggesting that the forecast for long-term occupancy loses accuracy. The RRMSE follows a similar trend, rising from 34.49% for short horizons to 39.68% for long horizons.

As shown in Appendix B, the results for the detector CMP_070-e2 will be interpreted as follows:

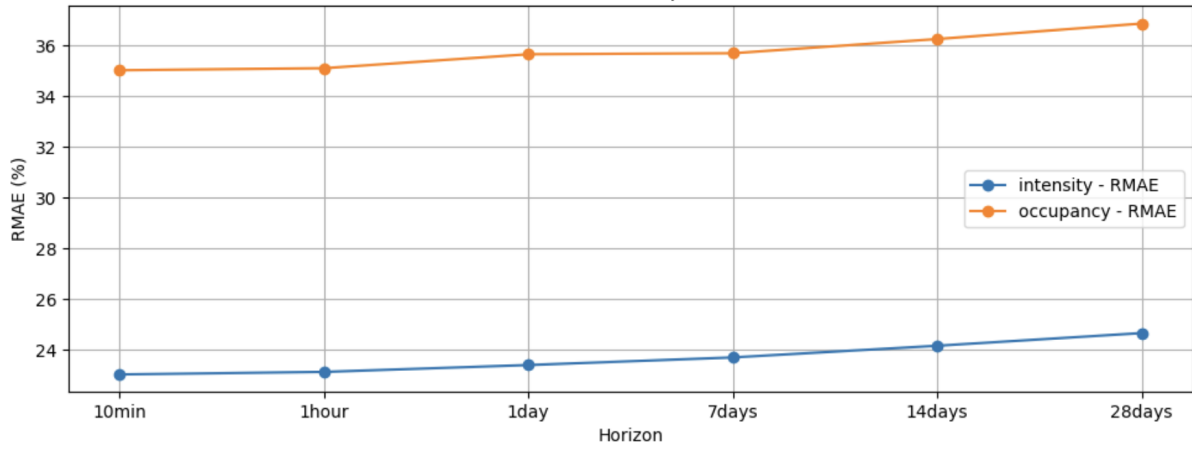


Figure 4.23: Plot that compares the values of RRMAE for CMP_070-e2

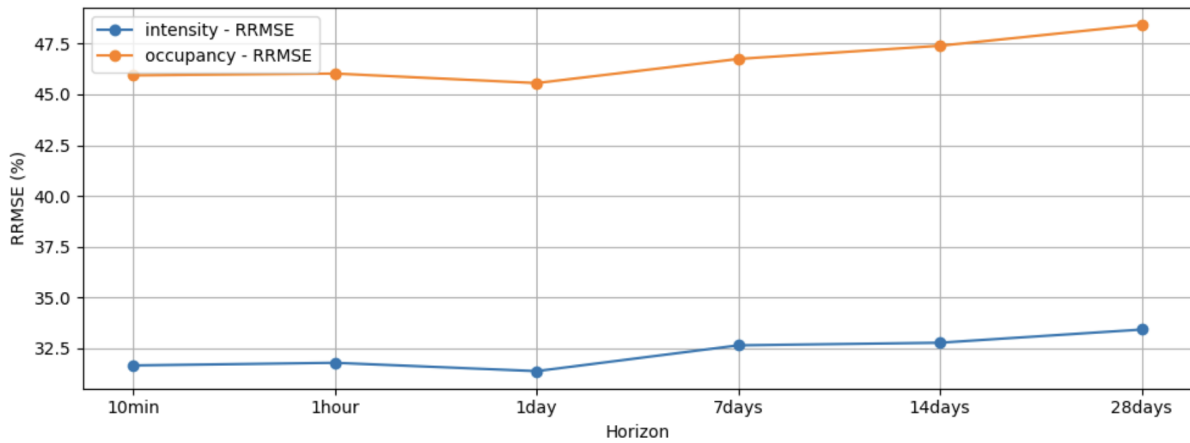


Figure 4.24: Plot that compares the values of RRMSE for CMP_070-e2

Intensity results from CMP_070-e2: We can see that the RRMAE varies from 23.04% to 24.67% and the RRMSE varies from 31.64% to 33.41% for different horizons. In the 10-minute period and horizon, the RMAE is the lowest (23.04%) and as the horizon increases to 28 days, the RMAE rises to 24.67%, indicating that it can more easily predict the intensity value in shorter horizons and that its accuracy decreases slightly in longer horizons. RRMSE starts at 31.64% for the 10-minutes period and 10-minute horizon, suggesting a higher error value compared to RMAE, which may suggest that the predicted values vary more when measured by RRMSE. As the horizon increases, the RRMSE is higher, showing that the error also increases considerably at longer horizons, especially between 14 and 28 days.

Occupancy results for CMP_070-e2: As for the occupancy values, the RMAE varies from 35.03% to 36.87% and the RRMSE varies from 45.94% to 48.43%. It can be seen that from the 1-hour horizon to the 1-day horizon, the RRMSE values decrease slightly, specifically from 46.03% to 45.56% and then continue to increase until the 28-day horizon. As the horizon increases, the RMAE and RRMSE values increase significantly, but we can see that the increase is slightly smaller.

Comparative analysis between CMP_070-e2 and CMP_020-e3: In general, the CMP_070-e2 detector shows slightly lower RMAE and RRMSE values than the CMP_020-e3 detector in almost all scenarios. This may indicate that CMP_070-e2 has a slightly more accurate prediction for traffic intensity compared to CMP_020-e3. The difference between the two detectors may be influenced by location: CMP_020-e3, near the Estádio do Dragão, may experience greater traffic variation on match days or public holidays.

Also, we can observe that both detectors perform better in short-term forecasts, with the RMAE and RRMSE increasing as the forecast horizon expands to 28 days.

Observing the plots, we can analyse that the detector CMP_070-e2 has more consistent values of RMAE and RRMSE in intensity and occupancy in comparison with the detector CMP_020-e3.

4.3.10 Forecast visualization:

The predicted values can be visualized using the Prophet plot method which provides a graphical representation of the predicted intensity and occupancy values over time. In addition, we use a Prophet method to visualize the individual components of the forecast such as the trend (general direction of the time series over time), weekly component (recurring patterns throughout the week), daily component (recurring patterns throughout the day) and holiday effects (impact of specific events such as holidays and football games).

Next, the plots corresponding to the trend, weekly and daily components of each detector for intensity and occupancy will be analyzed, as well as the comparison between them. After that, we will analyse the holiday component for the two detectors.

Trend Component of the detectors CMP_020-e3 and CMP_070-e2 for intensity

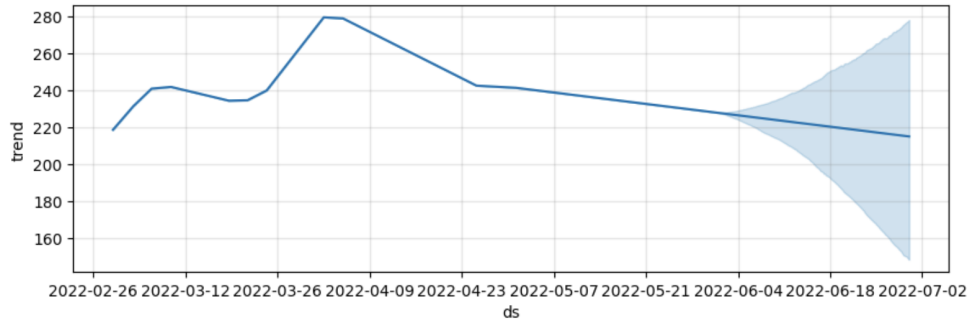


Figure 4.25: Trend Component of the detector CMP_020-e3 for intensity.

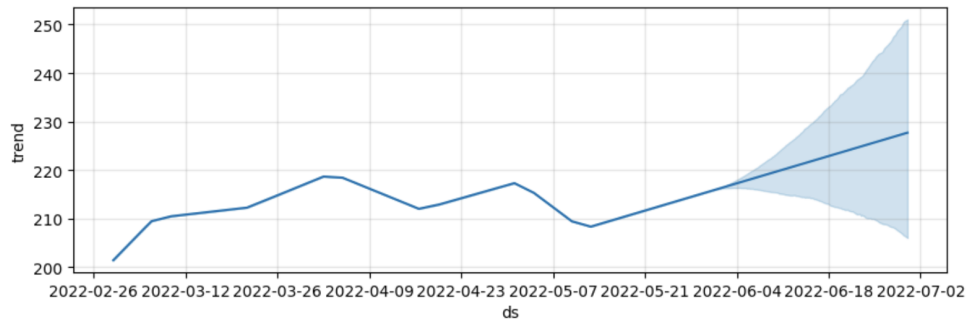


Figure 4.26: Trend Component of the detector CMP_070-e2 for intensity.

The analysis of the trend component in figure 4.25 and figure 4.26 shows that the plots of the two detectors behave differently. The CMP_020-e3 detector shows a peak in intensity around April and then a sharp decrease until the beginning of June. Detector CMP_070-e2 shows relatively consistent values until the beginning of June. The intensity values in the CMP_070-e2 detector are lower and the variations are more moderate compared to the CMP_020-e3 detector, where the intensity values are higher and less consistent. Both plots show an increase in uncertainty, represented by the blue-shaded area, thus suggesting that the behavior of traffic over the forecasted period may vary considerably.

Trend Component of the detectors CMP_020-e3 and CMP_070-e2 for occupancy

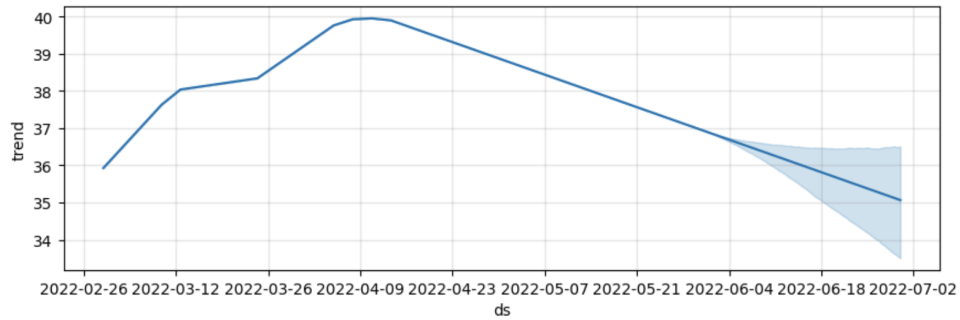


Figure 4.27: Trend Component of the detector CMP_020-e3 for occupancy

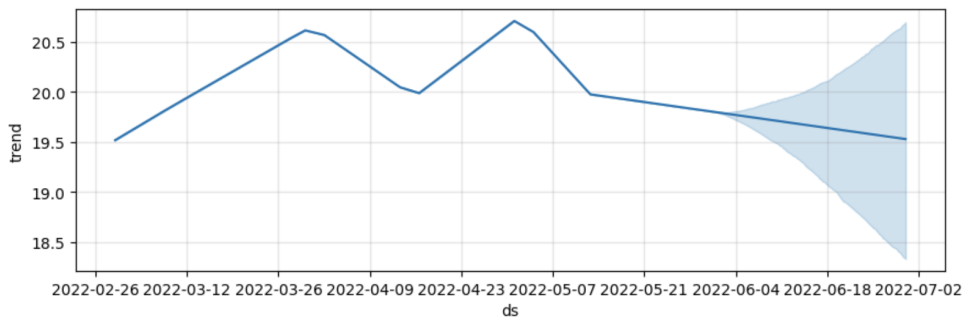


Figure 4.28: Trend Component of the detector CMP_070-e2 for occupancy

By analyzing the figures 4.27 and 4.28 we can see that both detectors show an increase in occupancy values until April, but the CMP_020-e3 detector reaches higher values than the CMP_070-e2 detector. It can be seen that over time, the CMP_020-e3 detector shows a more constant trend and the CMP_070-e2 detector shows sharper variations. Both show an increase in forecast uncertainty from the beginning of June, which is the month we want to forecast.

Weekly Component of the detectors CMP_020-e3 and CMP_070-e2 for intensity

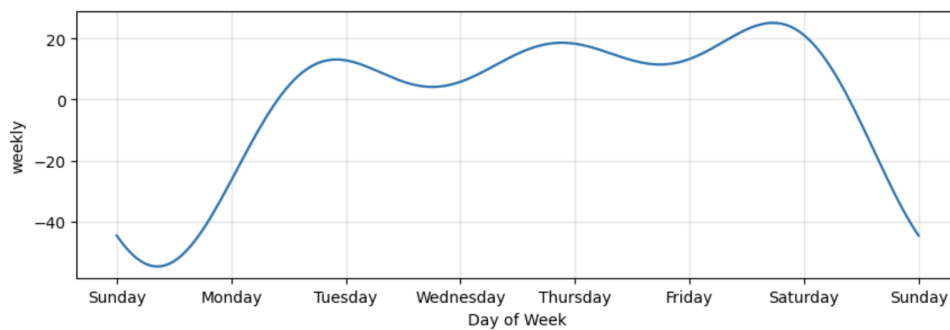


Figure 4.29: Weekly Component of the detector CMP_020-e3 for intensity.

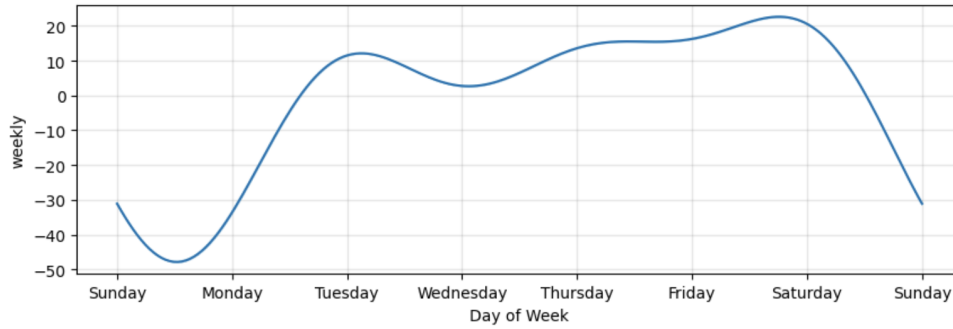


Figure 4.30: Weekly Component of the detector CMP_070-e2 for intensity.

By analyzing the figures 4.29 and 4.30 we can see that the intensity shows a similar pattern for both detectors. The intensity assumes high values from Monday to Tuesday and decreases significantly until Friday. On Saturday it shows an increase, this being the highest value during the week and then on Sunday it shows a sharp decline. The CMP_020-e3 detector has a higher maximum intensity value than the CMP_070-e2 detector.

Weekly Component of the detectors CMP_020-e3 and CMP_070-e2 for occupancy

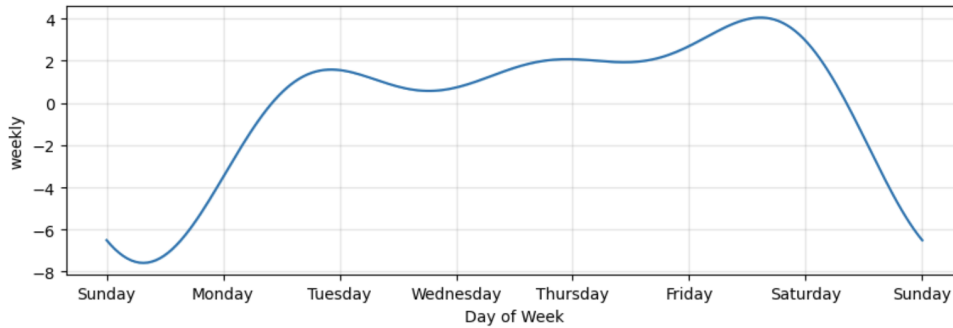


Figure 4.31: Weekly Component of the detector CMP_020-e3 for occupancy.

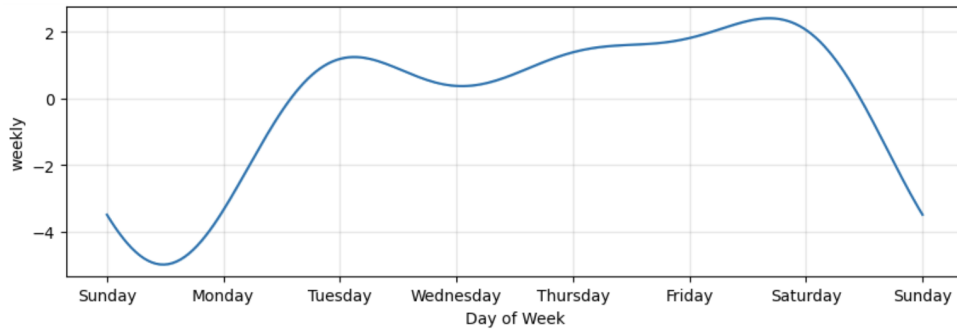


Figure 4.32: Weekly Component of the detector CMP_070-e2 for occupancy.

The figures 4.31 and 4.32 show the same behavior. On Sunday, occupancy is lower and gradually increases from Monday to Tuesday, remaining relatively constant from Wednesday to Friday. On Saturday it rises considerably, before falling sharply on Sunday.

Daily Component of the detectors CMP_020-e3 and CMP_070-e2 for intensity

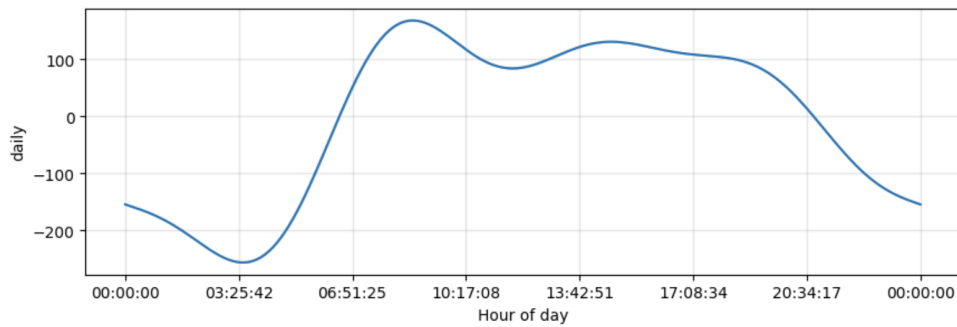


Figure 4.33: Daily Component of the detector CMP_020-e3 for intensity.

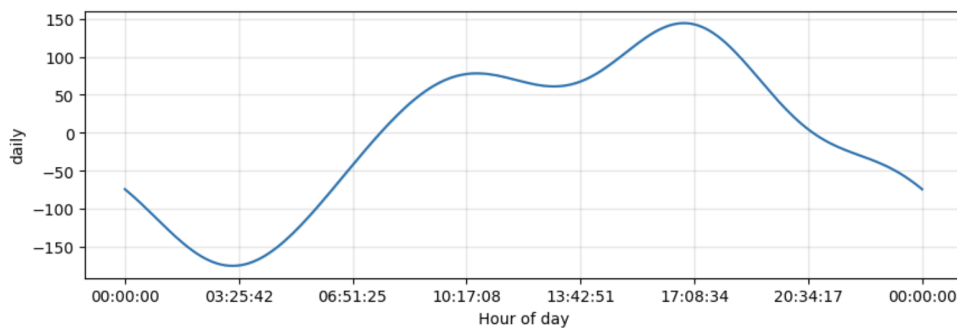


Figure 4.34: Daily Component of the detector CMP_070-e2 for intensity.

In the figures 4.33 and 4.34 we can observe that both detectors have a similar behavior

throughout the day. Both of the detectors perform an intensity peak in the morning around 7 am, which we can consider the beginning of rush hour, then the intensity values decrease and become relatively constant and finally show a significant decrease around 5 pm. Although both detectors show relatively constant values between mid-morning and mid-afternoon, it can be seen that the CMP_070-e2 detector shows more variation than the CMP_020-e3 detector.

Daily Component of the detectors CMP_020-e3 and CMP_070-e2 for occupancy

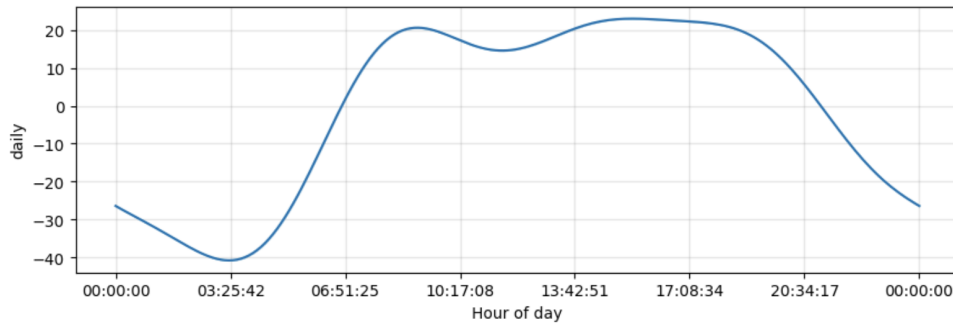


Figure 4.35: Daily Component of the detector CMP_020-e3 for occupancy

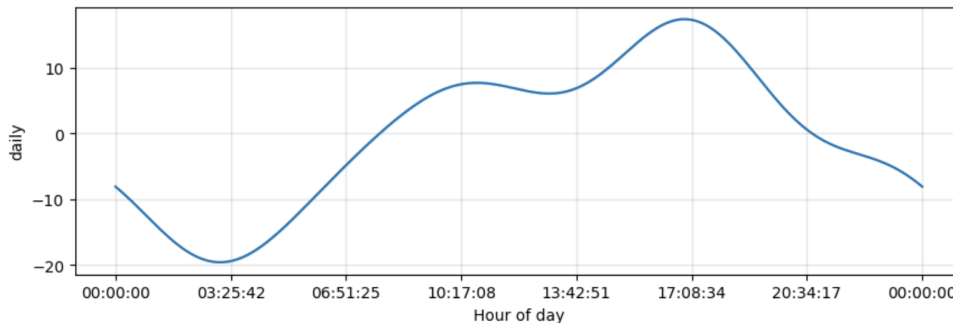


Figure 4.36: Daily Component of the detector CMP_070-e2 for occupancy

In the figure 4.35 we can see that the CMP_020-e3 detector shows a peak in occupancy in the morning between 6 am and 10 am, which corresponds to rush hour, and that at the end of the day, around 5 pm, it shows a sharp decrease until midnight. Between these two peaks, occupancy figures remain relatively consistent.

In figure 4.36 we can see that there is also a high peak in the morning between 6 am and 10 am and that at the end of the day, around 5 pm, there is a decrease in occupancy, just as with the CMP_020-e3 detector.

Although both detectors show similarities in the typical pattern during morning and late afternoon, the CMP_020-e3 detector shows a greater range of variability.

Influence of football games and national holidays on the detectors in the intensity variable

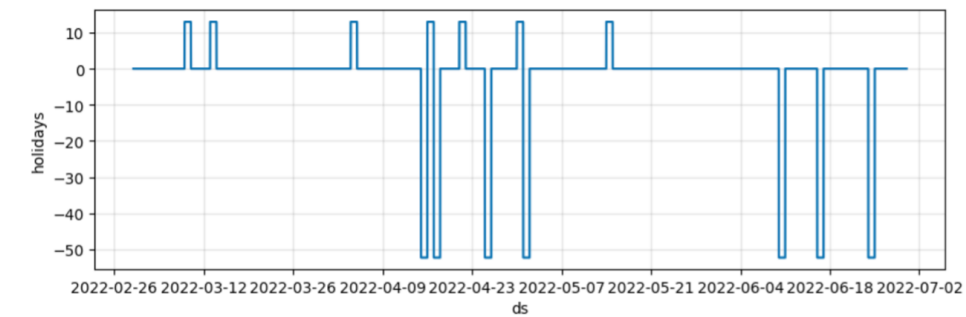


Figure 4.37: Holiday Component for Intensity Prediction at Detector CMP_020-e3

The holidays component, shown in the plot above 4.37, illustrates the impact of football games on traffic at detector CMP_020-e3, which is located near the Estádio do Dragão. Using this plot and the list of football games at the stadium, we can see how the Prophet model considers their impact on this detector. In the plot, each peak with a positive value in the “holidays” component represents the influence of the holiday or football game on the traffic forecast.

For the football games at Estádio do Dragão, the following dates were considered:

February 27, 2022; March 9, 2022; March 13, 2022; April 4, 2022; April 16, 2022; April 21, 2022; April 30, 2022; and May 14, 2022.

We can observe from the figure 4.37 that positive, slightly accentuated peaks align with the dates of these matches, indicating that the model has identified these events as relevant. The traffic intensity on these days is higher than usual, as expected, due to the increased number of people travelling to the stadium to watch the games.

With regard to national holidays, the following dates were considered:

January 1, 2022; April 15, 2022; April 17, 2022; April 25, 2022; May 1, 2022; June 10, 2022; June 16, 2022; June 24, 2022; August 15, 2022; October 5, 2022; November 1, 2022; December 1, 2022; December 8, 2022; and December 25, 2022.

We can see from the plot that the peaks corresponding to these dates are negative, suggesting a reduction in traffic in this region, and we can also see that the reduction is quite high.

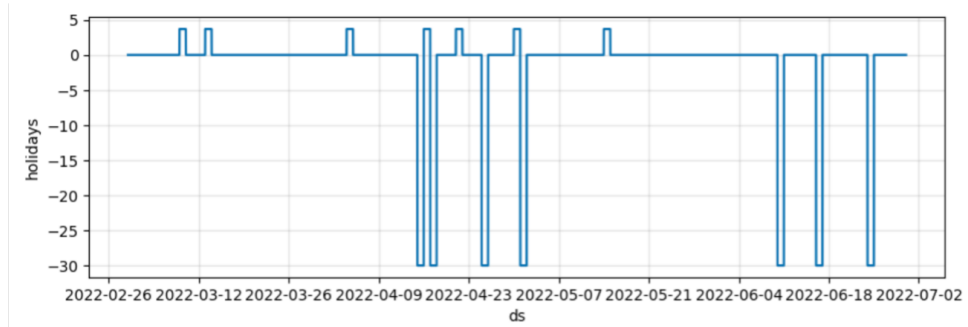


Figure 4.38: Holiday Component for Intensity Prediction at Detector CMP_070-e2

The holidays component for detector CMP_070-e2 represented in the figure 4.38, which is located further away from the Estádio do Dragão, we can see that the peaks corresponding to the dates of football games are much lower than for detector CMP_020-e3, which is plausible, since traffic is more concentrated in the streets near the Dragão stadium and even people who live in these streets may have travelled to watch the match. Concerning the dates of the national holidays, we can see that the peaks are negative, suggesting a reduction in traffic, just as in CMP_020-e3. However, in this detector, the reduction is more pronounced.

Influence of football games and national holidays on the detectors in the occupancy variable

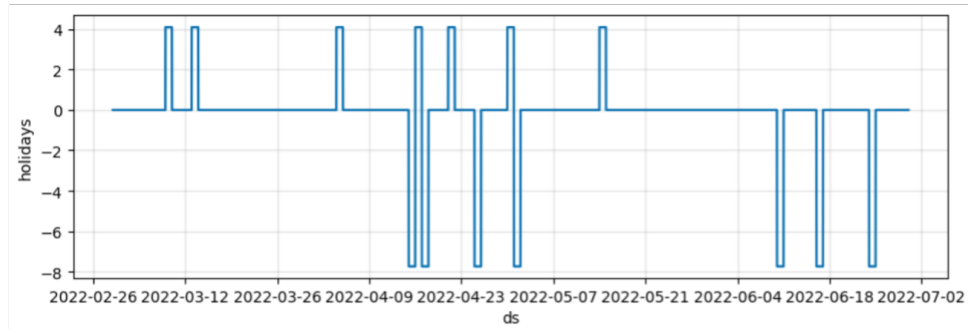


Figure 4.39: Holiday Component for Occupancy Prediction at Detector CMP_020-e3

Looking at the plot in figure 4.39, we can see that the positive peaks refer to the dates of football games, indicating that occupancy is higher on these days. We can infer that with more vehicles in transit around the stadium, the detector registers higher occupancy, meaning that vehicles are occupying the inductive loops for longer periods. This directly correlates with higher occupancy rates, suggesting that traffic is denser and vehicles are moving more slowly or remaining in the surrounding area for longer. In contrast, on public holidays, there is a noticeable reduction in occupancy, likely due to decreased daily activity, such as fewer work-related journeys and general movement in the city.

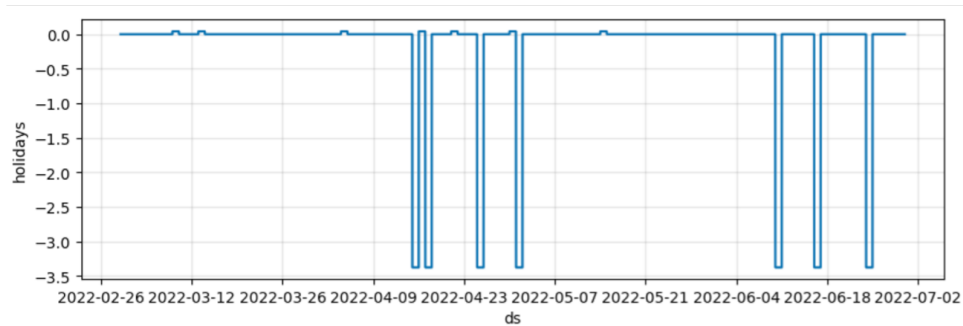


Figure 4.40: Holiday Component for Occupancy Prediction at Detector CMP_070-e2

Compared to the CMP_020-e3 detector, this detector shows in Figure 4.40 a much lower impact on football game days, suggesting that traffic is lower near this location, although not absent on these days. This may suggest that vehicles do not spend as much time on the inductive loops, resulting in relatively low occupancy. During public holidays, there is a pattern of decreasing occupancy values, around 3.5% which may suggest that traffic on these days is lower, probably due to people staying at home with their families.

Cross validation for the 50 detectors After these initial tests on 2 detectors, we decided to expand the study to include 50 detectors distributed across different areas of the Municipality of Porto. This expansion aimed to evaluate traffic behavior in various regions of the city and the increase in the number of detectors allows for a more robust and detailed analysis of traffic patterns and the sensitivity of the system in making these intensity and occupancy predictions. For these 50 detectors, cross-validation was only carried out in 2 scenarios, since, as this work was developed in a Google Colab context, the 10-minutes period and horizon scenario required very high computational resources. Therefore, the two scenarios carried out include the following parameters:

- **Scenario 1:** initial 30 days, period of 1 hour and horizon of 1 hour
- **Scenario 2:** initial 30 days, period of 1 day and horizons of 1 day, 7 days, 14 days and 28 days.

As with the 2 initial detectors, in order to assess the system's performance, the relative errors of MAE and RMSE (RMAE and RRMSE) were calculated. These values can be found in Appendix C to Appendix F, as well as histograms for the intensity in Appendix G and histograms for the occupancy H showing the distribution of these values in each of the scenarios with different periods and horizons.

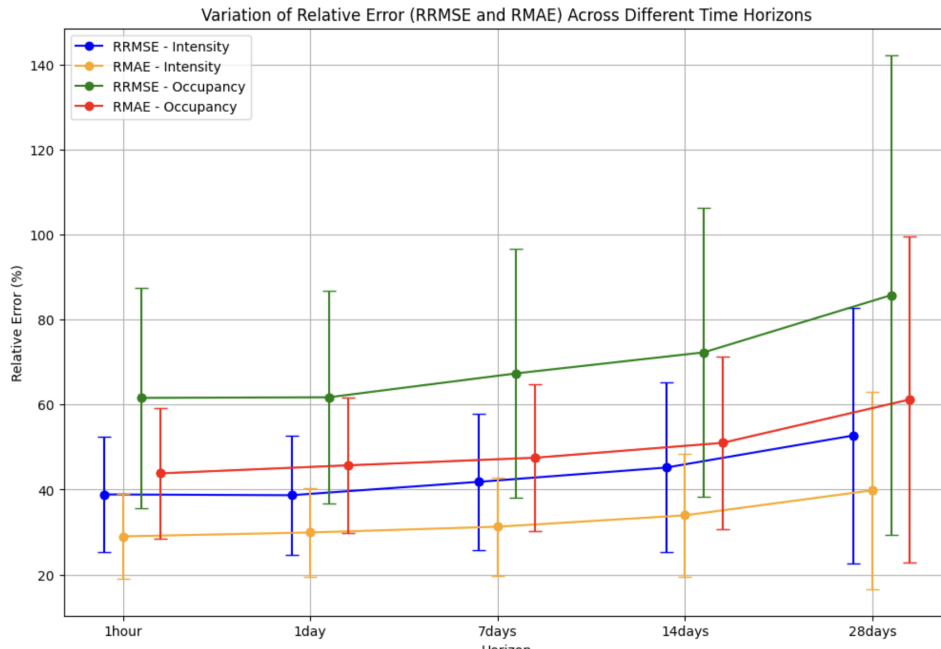


Figure 4.41: Variation of the RRMSE and RMAE for intensity and occupancy in different time horizons

When plotting the RMAE and RRMSE metrics for intensity and occupancy over the different forecast horizons, it was observed that the data tended to overlap on the x-axis. This is because the points for RRMSE and RMAE for intensity and occupancy refer to the same forecast horizon, which would result in the points on the plot overlapping. To avoid this overlap, an offset was used to separate the horizontal lines while maintaining the order of the forecast horizons. In this way, all the curves are visibly different and comparable.

The line plot was created to represent the variation in the RRMSE and RMAE over the different time horizons for the intensity and occupancy forecasts. This plot also includes error bars indicating the standard deviation of each error point.

Therefore, the y-axis represents the value of the relative MAE and RMSE error in percentage and the x-axis represents the different forecast horizons. The blue and yellow lines represent the RRMSE and RMAE for intensity. The green and red lines represent the RRMSE and RMAE about occupancy. The error bars along the lines represent the standard deviation, expressing the variability of the error values over the horizons. The longer the error bar, the greater the uncertainty or variation in the values. Shorter bars indicate that the error values in that horizon are more consistent.

Analysis of RMAE and RRMSE values for Intensity The Relative Mean Absolute Error (RMAE) values for intensity showed the lowest rates at all forecast horizons, varying approximately between 20% and 35%. The Relative Root Mean Square Error (RRMSE) values

are higher and increase proportionally as the forecast horizons get longer, especially between the 14 and 28-day periods. This suggests that long-term forecasts show greater variability compared to short-term forecasts.

Analysis of RMAE and RRMSE Values for Occupancy The RMAE values for occupancy initiate at a value slightly above 40% in the 1-hour forecasts and gradually increase, reaching around 60% when the forecast horizon reaches 28 days. With regard to RRMSE, it can be seen that the values for occupancy are the highest of all the relative error metrics, demonstrating a high degree of variability which also increases proportionally as the forecast horizons increase.

Comparison between Intensity and Occupancy When comparing the RMAE and RRMSE values between intensity and occupancy, we observe that the intensity forecasts are more accurate, consistent and stable. This is evidenced by the lower RMAE and RRMSE values, particularly in the case of RMAE, and by the lower uncertainty, as indicated by the smaller standard deviation bars. Conversely, the occupancy forecasts demonstrate higher RMAE and RRMSE at all horizons, as well as greater variability, as evidenced by the larger standard deviation bars, especially for the RRMSE. These results indicate that the model has greater difficulty in predicting vehicle occupancy, especially over longer forecast horizons.

The model is more accurate at predicting intensity, especially at shorter horizons (1 hour and 1 day), with lower absolute error (MAE) and controlled variation. However, for longer horizons (14 to 28 days), the uncertainty increases significantly. Occupancy forecasts are less reliable, with higher RMAE and RRMSE and variability, particularly in the case of RMSE. This suggests that the model has greater difficulty in consistently capturing vehicle occupancy patterns, especially over longer forecast horizons.

Chapter 5

Conclusions

Although traffic prediction using the prophet system shows some potential, for inductive loop detectors data, it is important to recognize that its performance is highly dependent on the quality and volume of the data used, as well as the system's ability to adapt to the nuances of specific traffic behaviors. From the results observed, we can assert that the system still has limitations and improvements are needed before it can demonstrate enough reliability for real-world applications.

This study has shown that the Prophet system has a higher degree of accuracy in short-term forecasts compared to long-term forecasts. This is because, over shorter horizons, traffic conditions tend to be less volatile and more predictable. Additionally, the inclusion of patterns related to holidays and special events allowed the model to more accurately capture variations in traffic, especially during busy periods such as football games at the Estádio do Dragão.

Based on the analysis of 50 detectors, it was possible to observe that the values of the relative errors of MAE and RMSE were considerably lower for intensity predictions than for occupancy. Therefore, we can conclude that occupancy is more difficult to predict and shows greater variability, especially as the horizon increases.

Although this forecasting system has shown some potential, there is still room for improvement. The relative errors observed indicate that the system is not quite ready to be implemented in broad traffic management operations. The limitations of the model highlight the need for further adjustments and improvements, especially to calibrate the model according to the different traffic conditions and behaviors that occur in various parts of the city. The inclusion of additional data sources, such as weather conditions and construction activities, can significantly improve the model's predictive capacity. Furthermore, expanding the network of detectors and increasing the granularity of the sensors can help capture local variations in traffic more accurately, resulting in greater robustness of the system.

Exploratory data analysis offered valuable insights into traffic behavior, providing a detailed view of urban dynamics. It revealed that both vehicle intensity and occupancy follow cyclical patterns over time, especially on weekdays, particularly during peak hours (morning and

afternoon), when traffic intensity reaches its highest values. This predictable behavior suggests that the forecasting system has the potential to offer more accurate forecasts over short horizons since traffic during these periods tends to be less volatile and more regular. The predictability observed reinforces the suitability of using time series models in routine traffic scenarios, especially in urban areas with strong demand during specific hours.

However, the analysis also highlighted marked differences between traffic patterns on weekdays and weekends. On weekends, traffic showed greater variability, with lower levels of intensity and occupancy, as well as fluctuations that were more difficult to predict. One possible justification for this could be that weekend traffic is more influenced by external and non-routine factors, such as events, leisure and occasional trips, rather than consistent work- or school-related journeys. Such variations indicate that the forecasting model needs to be adjusted to better deal with these irregular traffic conditions, and the inclusion of additional variables, such as local events or weather conditions, can significantly improve the accuracy of forecasts for these periods.

Another important point identified in the exploratory analysis was the detection of anomalies in the data, such as gaps on certain days and peaks in traffic intensity. We can suppose that this is because some controllers only start collecting information on detectors after a certain period of time. Anomalous traffic peaks, such as the one observed on January 27, represent another challenge, as they reflect unexpected fluctuations in the number of vehicles. The occurrence of these anomalies reinforces the need for a more robust data collection infrastructure that can minimize errors and guarantee the continuity of data transmission in real-time. This is crucial for increasing the reliability of the forecasting system and reducing the impact of atypical events on the results.

In addition, the greater variability observed in occupancy measurements compared to intensity measurements suggests that the percentage of time that a metallic object was detected over the inductive loop is difficult to model accurately. Occupancy reflects the length of time that vehicles remain on the sensors, and this parameter can be influenced by factors such as congestion, irregular parking or the slow movement of waiting vehicles. Unlike intensity, which refers only to the number of vehicles passing a point in a given time interval, occupancy is directly linked to the dynamics of vehicle flow, and its variability makes the forecasting process more challenging. Identifying these unpredictable patterns in occupancy points to the need to continue monitoring both parameters (intensity and occupancy) to gain a more comprehensive understanding of traffic conditions.

The system could benefit from the implementation of specialized outlier detection and correction mechanisms that could help the Prophet model reach even better performance and would allow inaccurate data to be filtered out before feeding it into the forecasting models. This would significantly increase the accuracy and practical usefulness of the system in operational traffic management contexts.

Therefore, based on the exploratory analysis, it was possible to identify the need for improvements in traffic controllers and data collection infrastructure to ensure better data

quality and consistency. These improvements are essential to minimize forecast errors and increase the system's reliability in different traffic scenarios. It also became clear that the inclusion of contextual variables and the proper treatment of anomalous data are important steps to improve the system's ability to deal with different traffic behaviors, especially during periods of high variabilities, such as weekends and special events.

In conclusion, this study serves as an important starting point for the further development of a more comprehensive and accurate traffic forecasting system. Although the current system is not yet fully optimized for operational use, the results obtained provide insights that can be used to inform future iterations of the system. By continuing to refine and expand the model, there is the potential to create a highly effective tool for mitigating traffic congestion and improving urban mobility in general. Continuing this work will play a key role in the development of more accurate and efficient traffic management solutions, contributing to long-term sustainability and quality of life in urban areas. As cities like Porto continue to grow and evolve, the implementation of robust systems to manage road traffic will become increasingly crucial, making this line of research essential to meet the transportation challenges of the future.

5.1 Contributions

The contributions of this work are related to the exploratory analysis and development of a traffic forecasting system using the Prophet model. In particular, the following:

- **Exploratory data analysis:** Data was analyzed for January 2022 to identify patterns, seasonalities, anomalies or potential problems that may exist in the data.
- **Testing the Prophet model on two detectors:** The Prophet model was tested on two specific detectors to reduce the complexity of the initial process and also to evaluate its performance on a detector that is close to the Estádio do Dragão and one that is further to understand the impact of football games and holidays on both detectors. By focusing on these 2 detectors, it is possible to better explore the patterns in the data before expanding the system to a larger number of detectors. The system was also tested over various periods and horizons to evaluate the intensity and occupancy predictions.
- **Evaluating the effect of public holidays and football games:** The impact of public holidays, a component of Prophet, was evaluated to understand their impact on intensity and occupancy forecasts. Including external events such as football games helped us to better understand traffic flow.
- **Expansion of the number of detectors:** To test the model's performance in different areas of the municipality of Porto, we expanded the number of detectors to 50, which were randomly generated, assessing their accuracy of intensity and occupancy in different periods and horizons.

- **Calculation of the relative errors of MAE and RMSE:** To help evaluate the system's performance for both the 2 detectors and the 50, the relative error values in percentage of the MAE and RMSE metrics obtained by the cross-validation method were calculated to better understand whether the system's prediction was as expected or not.
- **Suggestions for improvement and future work:** Based on the results obtained, we suggest incorporating more external factors, such as weather data or traffic incidents, which could further refine the model. In addition, the use of a more sophisticated model or a hybrid approach that could complement the Prophet model could be of added value in predicting this data.

5.2 Future work

To improve the performance and accuracy of the forecasting system, several areas can be explored in future work. One of the main directions will be to include more external variables in the system, such as weather data, which can significantly influence traffic. In addition, it will be important to integrate information specific to each intersection, area or zone in the municipality of Porto, since local characteristics can affect traffic behavior differently.

Another approach that could be investigated is traffic forecasting at shorter intervals, such as every 10 minutes. This would allow a faster and more accurate response to variations in traffic, especially during peak periods or special events.

In addition to Prophet, the application and comparison of other time series models, such as ARIMA (AutoRegressive Integrated Moving Average) and SARIMA (Seasonal ARIMA), could be explored to see if they can capture seasonal and trend patterns more efficiently in certain scenarios.

It is also proposed to use all the detectors available in the traffic controllers in the municipality of Porto, thus increasing the scope and representativeness of the forecasts. In addition, the use of a larger and more recent set of historical data could provide the system with a more robust database, which possibly could be analysed to contribute to greater forecast accuracy.

Another possibility would be to adopt hybrid or more advanced models, such as recurrent neural networks (RNNs) or deep learning models, which can capture complex temporal patterns and interactions between variables more efficiently. These models could complement the current system, providing more accurate predictions in more complex traffic scenarios.

In addition, the integration of real-time data, such as information from IoT sensors and camera monitoring systems, could improve the system's ability to adjust to instantaneous changes in traffic conditions. Combining this data with real-time forecasting systems would allow for more dynamic and proactive traffic management.

Finally, it would be interesting to develop intuitive and interactive visual interfaces to present

forecast results in a clear and accessible way to traffic managers, facilitating informed and efficient decision-making.

Implementing these improvements could contribute to significantly increasing the effectiveness of the system, providing better tools for traffic management in the municipality of Porto, as well as enabling more informed and efficient decision-making.

References

- [1] R. Adhikari and R. K. Agrawal. [An introductory study on time series modeling and forecasting](https://doi.org/10.48550/arXiv.1302.6613). 2013. doi:<https://doi.org/10.48550/arXiv.1302.6613>.
- [2] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, 2016.
- [3] Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyu Zhu. [A limited memory algorithm for bound constrained optimization](#). *SIAM Journal on Scientific and Statistical Computing*, 16(5):1190–1208, 1995. doi:10.1137/0916069.
- [4] C. Champion and C. Champion. [Agent-based implementation of particle hopping traffic model with stochastic and queuing elements](#). 2018. Accessed: 24-10-2024.
- [5] T. Cherrett, H. Bell, and M. McDonald. Traffic management parameters from single inductive loop detectors. In *Proceedings of the 7th World Congress on Intelligent Transport Systems*, pages 243–251, 2000.
- [6] N. K. ChikkaKrishna, P. Rachakonda, and T. Tallam. Short-term traffic prediction using fb-prophet and neural-prophet. *Journal of Forecasting*, 41(2):275–290, 2022.
- [7] Xiaoli Deng and Yao Hu. [Macroscopic fundamental diagram estimation considering traffic flow condition of road network](#). *Promet – Traffic & Transportation*, 35(5):681–697, 2023. doi:10.7307/ptt.v35i5.107.
- [8] Prophet Official Documentation. Prophet: Forecasting at scale. <https://facebook.github.io/prophet/>, 2018.
- [9] Facebook. [Modeling holidays and special events](#), 2024. Accessed: 2024-10-23.
- [10] R. Garg, S. Barpanda, N. S. G. R. Salanke, and S. Ramya. Machine learning algorithms for time series analysis and forecasting. In *Proceedings of the MCCS 2021 Conference*, pages 1–7, 2021.
- [11] Instituto Português do Mar e da Atmosfera. [Previsão horária por localidade - 10 dias](#), n.d. Acesso em: 2 nov. 2024.

- [12] B. K. Jha and S. Pande. [Time series forecasting model for supermarket sales using fb-prophet](#). In *Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC)*, pages 547–554. IEEE, 2021. doi:10.1109/ICCMC51019.2021.9418033.
- [13] S. V. Kumar and L. Vanajakshi. [Short-term traffic flow prediction using seasonal arima model with limited input data](#). *European Transport Research Review*, 7:21, 2015. doi:10.1007/s12544-015-0170-8.
- [14] W. Liu, X. Yu, Q. Zhao, G. Cheng, X. Hou, and S. He. [Time series forecasting fusion network model based on prophet and improved lstm](#). *Computers, Materials & Continua*, 74(2):3200–3217, 2023. doi:10.32604/cmc.2023.032595.
- [15] Meta. [Prophet: Forecasting at scale](#), 2024. Accessed: 2024-09-27.
- [16] Quantdare. [Using decomposition to improve time series prediction](#). *Quantdare*, 2014. Accessed: 24-10-2024.
- [17] S. Sathyan and J. S. N. Jagadeesha. Forecasting traffic volume using statistical and artificial intelligence tools: Prophet, fbprophet, and neural prophet. *International Journal of Intelligent Systems and Applications in Engineering*, 12(4):1421–1430, 2024.
- [18] R. Shumway and D. S. Stoffer. *Time Series Analysis Using the R Statistical Package*. Free Dog Publishing, 2018.
- [19] M. A. R. Shuvo, M. Zubair, A. T. Purnota, S. Hossain, and M. I. Hossain. [Traffic forecasting using time-series analysis](#). In *Proceedings of the Sixth International Conference on Inventive Computation Technologies (ICICT 2021)*, pages 269–274, 2021. doi:10.1109/ICICT50816.2021.9358682.
- [20] S. J. Taylor and B. Letham. [Forecasting at scale](#). *The American Statistician*, 72(1):37–45, 2018. doi:10.1080/00031305.2017.1380080.
- [21] M. Treiber and A. Kesting. *Traffic Flow Dynamics: Data, Models and Simulation*. Springer, Heidelberg, New York, Dordrecht, London, 2013. ISBN: 978-3-642-32459-8. doi:10.1007/978-3-642-32460-4.
- [22] Wikipedia contributors. [Regularization \(mathematics\) — wikipedia, the free encyclopedia](#), 2023. [Online; accessed 30-October-2024].
- [23] P. Yuvarani, P. Dinesh, P. Bharani, and B. Dharun. [Time series forecasting of ethereum price by fb-prophet](#). In *Proceedings of the 4th International Conference on Signal Processing and Communication (ICSPC 2023)*, pages 272–277, 2023. doi:10.1109/ICSPC57692.2023.10125661.

Appendix A

Results table of CMP_020-e3

The following table presents the results obtained for the CMP_020-e3 detector. The table includes the following columns:

- **Variable y:** The traffic variable being predicted, such as intensity or occupancy.
- **Period:** The time period over which the prediction is evaluated.
- **Horizon:** The forecast horizon (e.g., 10 minutes, 1 hour, or more).
- **MAE:** Mean Absolute Error
- **RMSE:** Root Mean Square Error
- **Mean_y:** The average of the actual observed values.
- **Relative MAE:** The MAE as a percentage of the mean observed values.
- **Relative RMSE:** The RMSE as a percentage of the mean observed values.

variable y	period	horizon	mae	rmse	mean_y	relative mae	relative rmse
intensity	10 minutes	10 minutes	61,28783327	85,51815741	248,2924	24,68	34,44
intensity	1 hour	1 hour	61,71112746	85,93173995	248,2924	24,86	34,6
intensity	1 day	1 day	62,98002092	80,97007046	248,2924855	25,37	32,61
intensity	1 day	7 days	64,10510097	87,94426128	246,9189922	25,96	35,62
intensity	1 day	14 days	67,95608288	93,53367549	245,5119705	27,68	46,67
intensity	1 day	28 days	82,44657868	113,9447896	244,137078	33,77	46,67
occupancy	10 minutes	10 minutes	10,42699686	13,48704918	39,10381	26,66	34,49
occupancy	1 hour	1 hour	10,4819315	13,55086895	39,10381503	26,81	34,65
occupancy	1 day	1 day	10,61458987	13,26370543	39,10381503	27,14	33,92
occupancy	1 day	7 days	10,8304081	13,85653992	39,08421438	27,71	35,45
occupancy	1 day	14 days	11,14045526	14,21364328	39,025032	28,55	36,42
occupancy	1 day	28 days	12,2130336	15,45838739	38,96088563	31,35	39,68

Figure A.1: Results table of CMP_020-e3

Appendix B

Results table of CMP_070-e2

The following table presents the results obtained for the CMP_070-e2 detector. The table includes the following columns:

- **Variable y:** The traffic variable being predicted, such as intensity or occupancy.
- **Period:** The time period over which the prediction is evaluated.
- **Horizon:** The forecast horizon (e.g., 10 minutes, 1 hour, or more).
- **MAE:** Mean Absolute Error
- **RMSE:** Root Mean Square Error
- **Mean_y:** The average of the actual observed values.
- **Relative MAE:** The MAE as a percentage of the mean observed values.
- **Relative RMSE:** The RMSE as a percentage of the mean observed values.

variable y	period	horizon	mae	rmse	mean_y	relative mae	relative rmse
intensity	10 minutes	10 minutes	48,91139325	67,17695586	212,3451429	23,04	31,64
intensity	1 hour	1 hour	49,15918984	67,47637087	212,3451429	23,14	31,77
intensity	1 day	1 day	49,70962067	66,59069686	212,3451429	23,41	31,36
intensity	1 day	7 days	50,27402803	69,19538491	212,062011	23,71	32,63
intensity	1 day	14 days	51,06476264	69,2095249	211,2638195	24,17	32,76
intensity	1 day	28 days	52,09472108	70,55692919	211,2043311	24,67	33,41
occupancy	10 minutes	10 minutes	6,993575704	9,174588139	19,97142857	35,03	45,94
occupancy	1 hour	1 hour	7,011554942	9,193091922	19,97142857	35,11	46,03
occupancy	1 day	1 day	7,122698784	9,098035697	19,97142857	35,66	45,56
occupancy	1 day	7 days	7,115615709	9,319375561	19,93371487	35,7	46,75
occupancy	1 day	14 days	7,213278899	9,429061991	19,89136297	36,26	47,4
occupancy	1 day	28 days	7,361064621	9,668492162	19,96467582	36,87	48,43

Figure B.1: Results table of CMP_070-e2

Appendix C

Intensity results for 50 detectors for 1 day

The following table presents the results obtained for the 50 detectors, evaluated over 1 day, with forecast horizons of 1 day, 7 days, 14 days, and 28 days for the intensity predictions. The table includes the following columns:

- **Detector:** The unique identification number or code assigned to each traffic detector.
- **Horizon:** The forecast horizon, which refers to the time period for which the prediction is made (e.g., 1 day, 7 days, 14 days, 28 days).
- **Mean_y_intensity:** The average observed intensity value over the specified horizon.
- **Mean_RMSE:** The average Root Mean Square Error (RMSE) for intensity predictions across the specified horizon.
- **Mean_MAE:** The average Mean Absolute Error (MAE) for intensity predictions across the specified horizon.
- **Relative Error RMSE Intensity:** The RMSE expressed as a percentage of the average observed intensity value.
- **Relative Error MAE Intensity:** The MAE expressed as a percentage of the average observed intensity value.

Detector	Horizon	mean_y_intensity	mean_rmse	mean_mae	Relative Error RMSE Intensity	Relative Error MAE Intensity
CMP_020-e6	1day	109,3845043	73,70081745	60,09642575	67,38	54,94
CMP_020-e3	1day	248,2924855	81,09083268	63,07468461	32,66	25,4
CMP_020-e5	1day	412,4526064	89,95872239	69,60858103	21,81	16,88
CMP_070-e2	1day	212,3451429	66,61241125	49,73117436	31,37	23,42
CMP_070-e4	1day	78,31567515	41,81535744	30,57627739	53,39	39,04
CMP_074-e1	1day	253,839396	75,42484599	59,20958813	29,71	23,33
CMP_074-e2	1day	273,7694079	121,8715716	87,80526261	44,52	32,07
CMP_075-e2	1day	139,0653254	72,93347753	59,76454836	52,45	42,98
CMP_070-e3	1day	283,2388247	93,22128377	74,61356659	32,91	26,34
CMP_070-e1	1day	97,8069907	39,52533802	30,58831498	40,41	31,27
CMP_020-e1	1day	583,7748888	113,5750754	87,95845179	19,46	15,07
CMP_075-e3	1day	54,65941748	40,41279229	31,0237353	73,94	56,76
CMP_020-e7	1day	228,988972	67,68135349	53,88980374	29,56	23,53
CMP_075-e4	1day	54,27024891	33,23307711	26,73578986	61,24	49,26
CMP_075-e1	1day	413,2329953	104,7205674	82,70738761	25,34	20,01
CMP_082-e3	1day	87,88462891	38,32087184	30,15118865	43,6	34,31
CMP_088-e1	1day	444,8971579	152,6876995	122,0842781	34,32	27,44
CMP_131-e1	1day	102,8479078	47,15998468	36,54914248	45,85	35,54
CMP_250-e1	1day	138,1990741	60,18775291	47,10080668	43,55	34,08
CMP_082-e2	1day	480,1345935	99,53451509	78,81087091	20,73	16,41
CMP_103-e3	1day	210,0083199	97,73878132	73,9018973	46,54	35,19
CMP_103-e5	1day	98,68689492	60,93015216	43,60423108	61,74	44,18
CMP_209-e3	1day	123,2391081	50,6801236	39,40035848	41,12	31,97
CMP_103-e4	1day	137,5011876	63,68253789	49,16399661	46,31	35,76
CMP_225-e2	1day	288,7037975	64,28823848	50,98833766	22,27	17,66
CMP_088-e2	1day	111,8263761	59,80201314	46,38021317	53,48	41,48
CMP_250-e2	1day	230,7802107	91,65369948	72,47944733	39,71	31,41

Figure C.1: Results table of the 50 detectors for Intensity

CMP_225-e1, 1day	245,8245019	58,6839908	45,57514021	23,87	18,54
CMP_081-e6, 1day	295,1502165	76,57921892	57,64897659	25,95	19,53
CMP_101-e1, 1day	643,9943065	140,8743637	112,2691127	21,88	17,43
CMP_081-e2, 1day	313,1395723	91,11422983	72,31480293	29,1	23,09
CMP_128-e3, 1day	204,2596325	73,12019613	56,16143754	35,8	27,5
CMP_103-e2, 1day	280,6907074	128,6312773	90,96850709	45,83	32,41
CMP_081-e7, 1day	256,3802412	84,67318442	66,83119748	33,03	26,07
CMP_081-e4, 1day	112,1989557	66,7320004	44,92224697	59,48	40,04
CMP_209-e1, 1day	16,72502806	8,377886858	6,333654633	50,09	37,87
CMP_137-e1, 1day	115,7100977	45,43023721	36,1600084	39,26	31,25
CMP_131-e2, 1day	88,93802714	37,14055545	29,26596231	41,76	32,91
CMP_264-e1, 1day	596,4818836	174,7824476	137,7855924	29,3	23,1
CMP_081-e3, 1day	83,57580175	38,00575025	30,21047893	45,47	36,15
CMP_128-e1, 1day	865,9656115	160,4766602	125,1654722	18,53	14,45
CMP_081-e1, 1day	305,4307517	64,61348742	51,24185967	21,15	16,78
CMP_101-e2, 1day	507,9592348	119,986762	95,30758238	23,62	18,76
CMP_121-e2, 1day	169,5117295	57,16495442	44,69491341	33,72	26,37
CMP_082-e5, 1day	275,8834751	96,23766839	76,17970197	34,88	27,61
CMP_225-e4, 1day	150,8733303	65,60060839	52,25396567	43,48	34,63
CMP_101-e3, 1day	237,7560917	72,43918773	57,8931215	30,47	24,35
CMP_075-e7, 1day	149,1864407	60,564881	47,92606976	40,6	32,12
CMP_128-e2, 1day	676,4616788	150,771491	117,329514	22,29	17,34
CMP_103-e1, 1day	188,4673286	130,0764422	96,86024676	69,02	51,39
CMP_020-e6, 7days	113,3033828	85,00325694	67,20287975	75,02	59,31
CMP_020-e3, 7days	246,9189922	87,9906965	64,14329463	35,64	25,98
CMP_020-e5, 7days	411,615703	94,86115688	71,88237042	23,05	17,46
CMP_070-e2, 7days	212,062011	69,20902059	50,28267688	32,64	23,71
CMP_070-e4, 7days	78,57772928	43,35146131	30,92282401	55,17	39,35

Figure C.2: Results table of the 50 detectors for Intensity

CMP_074-e1, 7days	253,7740474	78,37796795	58,89634844	30,88	23,21
CMP_074-e2, 7days	272,7965906	143,6639717	102,9973537	52,66	37,76
CMP_075-e2, 7days	139,4787855	80,01731985	60,33475198	57,37	43,26
CMP_070-e3, 7days	282,5301502	96,36070714	74,14645126	34,11	26,24
CMP_070-e1, 7days	97,82346906	40,4166342	30,63711321	41,32	31,32
CMP_020-e1, 7days	583,2187348	122,2540649	91,1708894	20,96	15,63
CMP_075-e3, 7days	54,71936353	41,75319562	30,63306227	76,3	55,98
CMP_020-e7, 7days	230,6930346	70,51885675	54,81473677	30,57	23,76
CMP_075-e4, 7days	54,41810841	34,32877598	26,4000146	63,08	48,51
CMP_075-e1, 7days	413,2283564	111,9852264	84,07640726	27,1	20,35
CMP_082-e3, 7days	87,78208772	40,18326121	29,68707643	45,78	33,82
CMP_088-e1, 7days	454,3512825	186,0173005	144,5957152	40,94	31,82
CMP_131-e1, 7days	103,0329727	48,71022888	36,29507293	47,28	35,23
CMP_250-e1, 7days	138,1792485	61,97533352	46,47093825	44,85	33,63
CMP_082-e2, 7days	480,1997041	107,6490801	83,33771952	22,42	17,35
CMP_103-e3, 7days	207,9550076	115,1193169	84,47349985	55,36	40,62
CMP_103-e5, 7days	97,63980599	73,5649044	53,0469746	75,34	54,33
CMP_209-e3, 7days	122,8182187	55,48417839	41,44921022	45,18	33,75
CMP_103-e4, 7days	137,8703386	75,42868725	56,69865094	54,71	41,12
CMP_225-e2, 7days	288,3375332	66,31921857	51,1399908	23	17,74
CMP_088-e2, 7days	112,0509331	63,75958508	47,02049003	56,9	41,96
CMP_250-e2, 7days	230,952195	95,86856838	73,05325596	41,51	31,63
CMP_225-e1, 7days	246,1776659	61,15604976	47,07453043	24,84	19,12
CMP_081-e6, 7days	296,1432053	78,97633905	59,42780512	26,67	20,07
CMP_101-e1, 7days	642,8101818	153,8405101	115,3882314	23,93	17,95
CMP_081-e2, 7days	312,4509397	99,27809956	75,4437478	31,77	24,15
CMP_128-e3, 7days	204,3845952	77,88932932	56,7546615	38,11	27,77
CMP_103-e2, 7days	287,033076	169,1594069	119,5647979	58,93	41,66

Figure C.3: Results table of the 50 detectors for Intensity

CMP_081-e7, 7days	256,0238134	88,57834361	67,0733003	34,6	26,2
CMP_081-e4, 7days	111,1801959	71,26174244	47,91911781	64,1	43,1
CMP_209-e1, 7days	16,80315431	8,443505804	6,321745761	50,25	37,62
CMP_137-e1, 7days	116,0194988	46,8210414	36,48787032	40,36	31,45
CMP_131-e2, 7days	88,95191991	37,74174893	29,17327996	42,43	32,8
CMP_264-e1, 7days	614,2400521	196,5409622	149,6395559	32	24,36
CMP_081-e3, 7days	83,48968625	39,49744461	30,00381414	47,31	35,94
CMP_128-e1, 7days	865,4019339	175,4405523	130,384256	20,27	15,07
CMP_081-e1, 7days	304,8402952	70,40697011	53,38355676	23,1	17,51
CMP_101-e2, 7days	507,6343989	130,011485	97,8451078	25,61	19,27
CMP_121-e2, 7days	169,6746119	58,42719308	44,75414155	34,43	26,38
CMP_082-e5, 7days	276,0518908	104,4756377	78,46761764	37,85	28,42
CMP_225-e4, 7days	151,3822509	68,56624954	51,71325354	45,29	34,16
CMP_101-e3, 7days	237,9758037	75,54112124	58,86801811	31,74	24,74
CMP_075-e7, 7days	151,1020142	61,92837754	48,30378956	40,98	31,97
CMP_128-e2, 7days	675,1733835	160,7624572	121,6000962	23,81	18,01
CMP_103-e1, 7days	195,6169045	164,5793254	118,7211821	84,13	60,69
CMP_020-e6, 14days	118,7996706	105,0816346	80,24230988	88,45	67,54
CMP_020-e3, 14days	245,5119705	93,46815847	67,91005123	38,07	27,66
CMP_020-e5, 14days	411,0077012	98,41668785	74,90010718	23,95	18,22
CMP_070-e2, 14days	211,2638195	69,19879762	51,02962534	32,75	24,15
CMP_070-e4, 14days	78,71246301	45,10050511	32,09867206	57,3	40,78
CMP_074-e1, 14days	253,0532657	80,10928854	60,18944222	31,66	23,79
CMP_074-e2, 14days	270,0521894	163,6244488	121,3221503	60,59	44,93
CMP_075-e2, 14days	139,0469046	81,92934428	61,92309624	58,92	44,53
CMP_070-e3, 14days	281,6799909	98,01645329	75,62951614	34,8	26,85
CMP_070-e1, 14days	97,86135482	40,63717313	30,8242687	41,53	31,5
CMP_020-e1, 14days	582,8893259	126,3887812	94,93107546	21,68	16,29

Figure C.4: Results table of the 50 detectors for Intensity

CMP_075-e3_14days	54,63444922	43,6990297	31,96225916	79,98	58,5
CMP_020-e7_14days	230,6869829	74,06859228	57,68325233	32,11	25
CMP_075-e4_14days	54,24996405	34,87742268	26,77267832	64,29	49,35
CMP_075-e1_14days	412,3941145	115,5320811	87,34396984	28,01	21,18
CMP_082-e3_14days	87,60651306	41,06476124	30,34396727	46,87	34,64
CMP_088-e1_14days	462,3575834	217,4257991	168,1977481	47,03	36,38
CMP_131-e1_14days	102,7222495	49,16738162	36,61285404	47,86	35,64
CMP_250-e1_14days	137,6947375	64,98049091	48,79224732	47,19	35,44
CMP_082-e2_14days	480,0701994	114,1754069	89,31866362	23,78	18,61
CMP_103-e3_14days	204,9294808	140,5287173	105,1837209	68,57	51,33
CMP_103-e5_14days	96,34781301	91,64040459	67,32988575	95,11	69,88
CMP_209-e3_14days	122,869362	64,05634904	46,17950032	52,13	37,58
CMP_103-e4_14days	137,7786274	89,7904543	67,53562829	65,17	49,02
CMP_225-e2_14days	287,6365174	71,51691031	55,93305407	24,86	19,45
CMP_088-e2_14days	111,4492249	65,09371385	48,35989654	58,41	43,39
CMP_250-e2_14days	230,196746	99,77077769	76,45006321	43,34	33,21
CMP_225-e1_14days	245,1639516	66,71935967	52,08513061	27,21	21,25
CMP_081-e6_14days	295,6068049	82,51562738	63,24810608	27,91	21,4
CMP_101-e1_14days	641,1244923	160,6531498	121,021835	25,06	18,88
CMP_081-e2_14days	311,4083969	105,4858545	80,8610659	33,87	25,97
CMP_128-e3_14days	204,4266988	80,4403028	58,41473625	39,35	28,57
CMP_103-e2_14days	295,944463	228,6096286	167,4828821	77,25	56,59
CMP_081-e7_14days	254,735456	90,95580594	69,08007111	35,71	27,12
CMP_081-e4_14days	109,3011191	73,74342484	50,82901607	67,47	46,5
CMP_209-e1_14days	16,83118761	8,471510785	6,30272136	50,33	37,45
CMP_137-e1_14days	115,4595314	47,64887447	37,00277197	41,27	32,05
CMP_131-e2_14days	88,58809046	37,95893363	29,41972293	42,85	33,21
CMP_264-e1_14days	628,0931624	177,6027067	138,8799406	28,28	22,11

Figure C.5: Results table of the 50 detectors for Intensity

CMP_081-e3	14days	83,01687927	40,16838783	30,49193326	48,39	36,73
CMP_128-e1	14days	863,8444699	181,8303373	135,7181219	21,05	15,71
CMP_081-e1	14days	304,481439	73,57361523	56,46163689	24,16	18,54
CMP_101-e2	14days	506,3506868	135,268282	102,6219235	26,71	20,27
CMP_121-e2	14days	169,6533125	59,62250207	45,88066481	35,14	27,04
CMP_082-e5	14days	275,5553654	108,647359	82,25529998	39,43	29,85
CMP_225-e4	14days	151,511998	72,50993794	54,5026437	47,86	35,97
CMP_101-e3	14days	237,4123842	77,11707264	60,36836804	32,48	25,43
CMP_075-e7	14days	151,9191791	64,96408224	51,28522132	42,76	33,76
CMP_128-e2	14days	673,4439442	166,0322266	125,862311	24,65	18,69
CMP_103-e1	14days	205,4737195	217,9215194	162,2770769	106,06	78,98
CMP_020-e6	28days	125,9844686	147,0358129	101,1348348	116,71	80,28
CMP_020-e3	28days	244,137078	113,4671623	82,20062767	46,48	33,67
CMP_020-e5	28days	411,7302557	104,7705145	78,87164483	25,45	19,16
CMP_070-e2	28days	211,2043311	70,44352609	51,9943126	33,35	24,62
CMP_070-e4	28days	79,49114349	48,85890889	35,02006519	61,46	44,06
CMP_074-e1	28days	253,5629368	82,35795329	61,39045332	32,48	24,21
CMP_074-e2	28days	268,7934202	189,4349075	139,3216395	70,48	51,83
CMP_075-e2	28days	139,9303304	88,99171106	66,6782145	63,6	47,65
CMP_070-e3	28days	281,9296753	100,199765	76,58572745	35,54	27,16
CMP_070-e1	28days	97,70630986	42,13932963	32,01866163	43,13	32,77
CMP_020-e1	28days	584,9289648	131,119448	97,89684304	22,42	16,74
CMP_075-e3	28days	54,90758163	47,73424355	35,08104644	86,94	63,89
CMP_020-e7	28days	230,6439562	85,19394026	66,66860517	36,94	28,91
CMP_075-e4	28days	54,20692984	36,40630948	28,05978684	67,16	51,76
CMP_075-e1	28days	413,9619893	122,0784413	91,14036303	29,49	22,02
CMP_082-e3	28days	87,97844734	42,33112917	31,01218987	48,12	35,25
CMP_088-e1	28days	468,2438419	314,3111193	239,1719454	67,13	51,08

Figure C.6: Results table of the 50 detectors for Intensity

CMP_131-e1	28days	102,7955898	50,46848187	37,71558311	49,1	36,69
CMP_250-e1	28days	138,8796915	70,71150775	52,65498233	50,92	37,91
CMP_082-e2	28days	483,5792323	118,267834	91,47917552	24,46	18,92
CMP_103-e3	28days	203,3275985	198,1213501	149,9020477	97,44	73,72
CMP_103-e5	28days	95,58712481	122,9177892	93,48781278	128,59	97,8
CMP_209-e3	28days	123,6879772	91,48184113	61,57590677	73,96	49,78
CMP_103-e4	28days	139,9551841	125,0683848	93,67406949	89,36	66,93
CMP_225-e2	28days	288,5087654	94,84968968	75,03075146	32,88	26,01
CMP_088-e2	28days	111,8295644	70,44220729	53,65330356	62,99	47,98
CMP_250-e2	28days	231,7242148	107,5850647	81,90226984	46,43	35,34
CMP_225-e1	28days	246,3872902	92,60634774	71,88051789	37,59	29,17
CMP_081-e6	28days	296,9434599	88,89557853	67,28740489	29,94	22,66
CMP_101-e1	28days	642,8029535	165,8105929	124,1434705	25,79	19,31
CMP_081-e2	28days	313,4900259	104,5192681	78,21880039	33,34	24,95
CMP_128-e3	28days	205,3385585	87,68701744	63,57715623	42,7	30,96
CMP_103-e2	28days	317,195961	390,3213382	315,186662	123,05	99,37
CMP_081-e7	28days	255,4093597	93,16841054	69,85844151	36,48	27,35
CMP_081-e4	28days	107,0867911	80,95040312	58,12562549	75,59	54,28
CMP_209-e1	28days	16,78166707	8,357780099	6,308658165	49,8	37,59
CMP_137-e1	28days	115,1789296	49,87476397	38,70200238	43,3	33,6
CMP_131-e2	28days	88,69887778	38,76292661	30,24481072	43,7	34,1
CMP_264-e1	28days	634,1278545	173,6055632	135,7585063	27,38	21,41
CMP_081-e3	28days	83,12747721	41,02591185	30,88460307	49,35	37,15
CMP_128-e1	28days	865,2666808	188,342391	139,2562274	21,77	16,09
CMP_081-e1	28days	306,2605954	72,94503556	55,30708292	23,82	18,06
CMP_101-e2	28days	508,0280755	139,678945	104,8371202	27,49	20,64
CMP_121-e2	28days	170,5058639	62,23504868	47,97070701	36,5	28,13
CMP_082-e5	28days	277,4177832	113,9715006	85,67699871	41,08	30,88

Figure C.7: Results table of the 50 detectors for Intensity

CMP_225-e4	28days	151,4667332	89,52429069	69,92137287	59,1	46,16
CMP_101-e3	28days	237,8659532	79,32258642	61,5627237	33,35	25,88
CMP_075-e7	28days	151,9875279	74,14152157	59,54385143	48,78	39,18
CMP_128-e2	28days	673,4439909	180,0239955	136,577172	26,73	20,28
CMP_103-e1	28days	227,3049303	358,7249071	291,420725	157,82	128,21

Figure C.8: Results table of the 50 detectors for Intensity

Appendix D

Occupancy results for 50 detectors, period of 1 day

The following table presents the results obtained for the 50 detectors, evaluated over 1 day, with forecast horizons of 1 day, 7 days, 14 days, and 28 days for the occupancy predictions. The table includes the following columns:

- **Detector:** The unique identification number or code assigned to each traffic detector.
- **Horizon:** The forecast horizon, refers to the time period for which the prediction is made (e.g., 1 day, 7 days, 14 days, 28 days).
- **Mean_y_occupancy:** The average observed occupancy value over the specified horizon.
- **Mean_RMSE:** The average Root Mean Square Error (RMSE) for occupancy predictions across the specified horizon.
- **Mean_MAE:** The average Mean Absolute Error (MAE) for occupancy predictions across the specified horizon.
- **Relative Error RMSE Occupancy:** The RMSE is expressed as a percentage of the average observed occupancy value.
- **Relative Error MAE Occupancy:** The MAE is expressed as a percentage of the average observed occupancy value.

Detector	Horizon	mean_y_occupancy	mean_rmse	mean_mae	Relative Error RMSE Occupancy	Relative Error MAE Occupancy
CMP_020-e6	1day	46,34612608	21,36286375	15,80774268	46,09	34,11
CMP_020-e3	1day	39,10381503	13,27742885	10,6241225	33,95	27,17
CMP_020-e5	1day	26,14200981	10,42363728	8,014206246	39,87	30,66
CMP_070-e2	1day	19,97142857	9,095494521	7,120151822	45,54	35,65
CMP_070-e4	1day	7,632557017	7,181817793	5,465480747	94,09	71,61
CMP_074-e1	1day	25,50148707	11,57357608	9,218105335	45,38	36,15
CMP_074-e2	1day	33,51497087	22,16477943	16,04056476	66,13	47,86
CMP_075-e2	1day	11,19027997	7,812774374	6,221292699	69,82	55,6
CMP_070-e3	1day	29,53332571	11,45012973	9,192067103	38,77	31,12
CMP_070-e1	1day	7,350829705	6,140903597	4,375558621	83,54	59,52
CMP_020-e1	1day	34,39274718	11,47675306	9,046276796	33,37	26,3
CMP_075-e3	1day	10,90679612	8,521440135	6,675642887	78,13	61,21
CMP_020-e7	1day	35,27589602	11,65582847	9,156674039	33,04	25,96
CMP_075-e4	1day	22,10865271	19,67472087	14,6456274	88,99	66,24
CMP_075-e1	1day	23,62082716	11,65656091	9,068961719	49,35	38,39
CMP_082-e3	1day	12,79962481	8,83987672	6,866796021	69,06	53,65
CMP_088-e1	1day	10,43602329	7,566133953	5,532308975	72,5	53,01
CMP_131-e1	1day	7,465296918	10,58859932	6,059405561	141,84	81,17
CMP_250-e1	1day	2,593364198	2,503602658	1,495645667	96,54	57,67
CMP_082-e2	1day	41,55033022	10,41048414	8,342058661	25,06	20,08
CMP_103-e3	1day	22,47411602	20,1527884	14,03814154	89,67	62,46
CMP_103-e5	1day	21,90318772	19,90332499	13,68633204	90,87	62,49
CMP_209-e3	1day	14,58482829	10,86977481	8,303922744	74,53	56,94
CMP_103-e4	1day	25,67969121	21,19495862	15,2906422	82,54	59,54
CMP_225-e2	1day	17,53417722	9,770374075	7,630374783	55,72	43,52
CMP_088-e2	1day	7,901613102	6,864233749	4,470394604	86,87	56,58
CMP_250-e2	1day	6,542549414	5,60975735	4,305118977	85,74	65,8

Figure D.1: Results table of the 50 detectors for Occupancy

CMP_225-e1	1day	15,37431115	8,229625266	6,388935616	53,53	41,56
CMP_081-e6	1day	42,34294506	12,63649257	10,01063588	29,84	23,64
CMP_101-e1	1day	24,61944887	9,812588596	7,961653724	39,86	32,34
CMP_081-e2	1day	36,96343987	12,73102261	10,24000568	34,44	27,7
CMP_128-e3	1day	31,77854179	13,77387198	10,74692249	43,34	33,82
CMP_103-e2	1day	23,16042534	20,28309995	13,88094876	87,58	59,93
CMP_081-e7	1day	23,18782309	9,903149505	7,90544072	42,71	34,09
CMP_081-e4	1day	33,67077253	22,80969465	16,40466407	67,74	48,72
CMP_209-e1	1day	7,306771418	8,096389304	5,936450785	110,81	81,25
CMP_137-e1	1day	9,080055124	6,903718668	5,261187207	76,03	57,94
CMP_131-e2	1day	11,73071752	8,399688166	6,594803762	71,6	56,22
CMP_264-e1	1day	16,83118071	10,29400616	7,98486478	61,16	47,44
CMP_081-e3	1day	24,82786686	14,50443592	11,80260847	58,42	47,54
CMP_128-e1	1day	21,94898656	10,49316515	8,214848502	47,81	37,43
CMP_081-e1	1day	41,8594533	12,15110746	9,794925141	29,03	23,4
CMP_101-e2	1day	20,7994762	8,596338059	6,945357628	41,33	33,39
CMP_121-e2	1day	19,82313707	10,98874822	8,466836303	55,43	42,71
CMP_082-e5	1day	30,14433464	11,65080513	9,287574783	38,65	30,81
CMP_225-e4	1day	11,71556886	11,26903565	8,181614967	96,19	69,84
CMP_101-e3	1day	19,08265647	8,717520829	6,959498178	45,68	36,47
CMP_075-e7	1day	34,01398716	11,62758158	9,311150972	34,18	27,37
CMP_128-e2	1day	15,73140967	9,687510314	6,873822623	61,58	43,69
CMP_103-e1	1day	46,64892634	19,86624927	12,9380436	42,59	27,73
CMP_020-e6	7days	43,90582428	25,72481513	17,74495136	58,59	40,42
CMP_020-e3	7days	39,08421438	13,86125248	10,8358002	35,47	27,72
CMP_020-e5	7days	26,12409854	10,92753772	8,13995003	41,83	31,16
CMP_070-e2	7days	19,93371487	9,319919877	7,114562429	46,75	35,69
CMP_070-e4	7days	7,676678344	7,558344702	5,371287563	98,46	69,97

Figure D.2: Results table of the 50 detectors for Occupancy

CMP_074-e1 7days	25,45872158	11,81006588	9,214255584	46,39	36,19
CMP_074-e2 7days	34,07465627	25,63839191	18,61034047	75,24	54,62
CMP_075-e2 7days	11,23326287	8,117497692	6,209717938	72,26	55,28
CMP_070-e3 7days	29,47358125	11,66121607	9,138692814	39,56	31,01
CMP_070-e1 7days	7,347652534	6,401275035	4,313047387	87,12	58,7
CMP_020-e1 7days	34,2614698	12,07989781	9,098880832	35,26	26,56
CMP_075-e3 7days	10,91491068	8,820509639	6,5891518	80,81	60,37
CMP_020-e7 7days	35,5449514	11,8402773	9,141143393	33,31	25,72
CMP_075-e4 7days	22,03470639	19,58466466	14,39114859	88,88	65,31
CMP_075-e1 7days	23,62643253	12,08225548	8,888872263	51,14	37,62
CMP_082-e3 7days	12,76997529	9,121372753	6,751184177	71,43	52,87
CMP_088-e1 7days	10,82070748	8,451351157	6,163655689	78,1	56,96
CMP_131-e1 7days	7,456765572	11,28282301	6,003226531	151,31	80,51
CMP_250-e1 7days	2,589227166	2,856932889	1,479729146	110,34	57,15
CMP_082-e2 7days	41,59169794	10,98222888	8,75468147	26,4	21,05
CMP_103-e3 7days	22,36919047	25,78813842	17,44392717	115,28	77,98
CMP_103-e5 7days	21,98511208	26,07034933	17,31688327	118,58	78,77
CMP_209-e3 7days	14,55594363	13,66451861	9,037993965	93,88	62,09
CMP_103-e4 7days	26,23108635	26,22088038	17,98337357	99,96	68,56
CMP_225-e2 7days	17,57132331	10,14090375	7,42206127	57,71	42,24
CMP_088-e2 7days	8,022190589	7,55885723	4,688789161	94,22	58,45
CMP_250-e2 7days	6,578220312	5,813923581	4,276707351	88,38	65,01
CMP_225-e1 7days	15,45900729	8,625703405	6,310177594	55,8	40,82
CMP_081-e6 7days	42,25403961	12,90158769	10,14114462	30,53	24
CMP_101-e1 7days	24,64780271	10,14495836	7,892590285	41,16	32,02
CMP_081-e2 7days	36,9567854	13,23032894	10,39444559	35,8	28,13
CMP_128-e3 7days	31,78506481	14,27791301	10,5771024	44,92	33,28
CMP_103-e2 7days	23,53739344	27,53609036	17,68781331	116,99	75,15

Figure D.3: Results table of the 50 detectors for Occupancy

CMP_081-e7 7days	23,18841371	10,10460398	7,858544966	43,58	33,89
CMP_081-e4 7days	33,96386811	24,8601261	18,0430086	73,2	53,12
CMP_209-e1 7days	7,398590119	8,9830579	6,202847027	121,42	83,84
CMP_137-e1 7days	9,08889639	6,970284448	5,186323804	76,69	57,06
CMP_131-e2 7days	11,74579323	8,578594803	6,498376993	73,04	55,33
CMP_264-e1 7days	16,65389139	10,55366128	7,69697729	63,37	46,23
CMP_081-e3 7days	24,85075403	14,87530463	11,7820545	59,86	47,41
CMP_128-e1 7days	21,94429211	10,98277081	8,089284568	50,05	36,86
CMP_081-e1 7days	41,86227242	12,43624937	9,927442236	29,71	23,71
CMP_101-e2 7days	20,82977342	8,814295372	6,863824337	42,32	32,95
CMP_121-e2 7days	19,87745098	10,91938352	8,13115458	54,93	40,91
CMP_082-e5 7days	30,17624675	12,11782246	9,452356457	40,16	31,32
CMP_225-e4 7days	11,77910448	11,65410354	7,884883521	98,94	66,94
CMP_101-e3 7days	19,13874058	8,934501877	6,976853215	46,68	36,45
CMP_075-e7 7days	34,31643985	11,84946041	9,333985943	34,53	27,2
CMP_128-e2 7days	15,79300249	10,28012744	6,903841719	65,09	43,71
CMP_103-e1 7days	45,1463959	32,70054896	21,06772657	72,43	46,67
CMP_020-e6 14days	40,93945982	33,26771256	22,5347077	81,26	55,04
CMP_020-e3 14days	39,025032	14,19769785	11,12536902	36,38	28,51
CMP_020-e5 14days	26,10103718	11,17683525	8,337834485	42,82	31,94
CMP_070-e2 14days	19,89136297	9,431296476	7,21288323	47,41	36,26
CMP_070-e4 14days	7,672289817	7,631231584	5,418610775	99,46	70,63
CMP_074-e1 14days	25,34343319	12,00317848	9,331749767	47,36	36,82
CMP_074-e2 14days	34,6902074	28,5330305	20,97238234	82,25	60,46
CMP_075-e2 14days	11,18846526	8,192103736	6,315694265	73,22	56,45
CMP_070-e3 14days	29,41344109	11,7830045	9,238422951	40,06	31,41
CMP_070-e1 14days	7,359415241	6,411373055	4,316188017	87,12	58,65
CMP_020-e1 14days	34,14483679	12,29478269	9,220939878	36,01	27,01

Figure D.4: Results table of the 50 detectors for Occupancy

CMP_075-e3 14days	10,83766939	8,891230124	6,651432191	82,04	61,37
CMP_020-e7 14days	35,62040134	12,17610387	9,423630618	34,18	26,46
CMP_075-e4 14days	21,77555715	19,6649919	14,57265351	90,31	66,92
CMP_075-e1 14days	23,53649458	12,41932881	9,181932567	52,77	39,01
CMP_082-e3 14days	12,74251051	9,1923139	6,756784434	72,14	53,03
CMP_088-e1 14days	11,09588956	9,207021558	6,981421591	82,98	62,92
CMP_131-e1 14days	7,394570498	11,30580957	6,054565445	152,89	81,88
CMP_250-e1 14days	2,578926511	3,016895944	1,534889922	116,98	59,52
CMP_082-e2 14days	41,57736711	11,49556498	9,163460693	27,65	22,04
CMP_103-e3 14days	22,28658991	31,25999034	21,78596278	140,26	97,75
CMP_103-e5 14days	22,14846594	31,95747791	21,6007467	144,29	97,53
CMP_209-e3 14days	14,5741096	17,97232451	10,2278071	123,32	70,18
CMP_103-e4 14days	26,67432719	31,31354032	21,35747613	117,39	80,07
CMP_225-e2 14days	17,52162344	10,46246472	7,758347735	59,71	44,28
CMP_088-e2 14days	7,981205567	7,472364955	4,822058836	93,62	60,42
CMP_250-e2 14days	6,595593217	6,101344082	4,522891982	92,51	68,57
CMP_225-e1 14days	15,36345693	8,940776321	6,674203657	58,2	43,44
CMP_081-e6 14days	42,25520451	13,30363487	10,50259173	31,48	24,86
CMP_101-e1 14days	24,58803345	10,38085119	8,096027539	42,22	32,93
CMP_081-e2 14days	36,90372502	13,68009046	10,78141979	37,07	29,21
CMP_128-e3 14days	31,75030672	14,57328683	10,76599816	45,9	33,91
CMP_103-e2 14days	24,18718088	34,06709121	22,05714414	140,85	91,19
CMP_081-e7 14days	23,15345719	10,25848336	8,005735373	44,31	34,58
CMP_081-e4 14days	34,49796141	26,41258991	19,32830477	76,56	56,03
CMP_209-e1 14days	7,424165232	9,194534638	6,35454754	123,85	85,59
CMP_137-e1 14days	9,017466404	6,995674968	5,186708602	77,58	57,52
CMP_131-e2 14days	11,67422335	8,574015475	6,491692084	73,44	55,61
CMP_264-e1 14days	16,38540545	10,48436021	7,684165554	63,99	46,9

Figure D.5: Results table of the 50 detectors for Occupancy

CMP_081-e3	14days	24,80322591	15,03887624	11,93812402	60,63	48,13
CMP_128-e1	14days	21,92661726	11,18974154	8,248097121	51,03	37,62
CMP_081-e1	14days	41,89198325	12,78589912	10,23114723	30,52	24,42
CMP_101-e2	14days	20,77467365	9,041933227	7,063684842	43,52	34
CMP_121-e2	14days	19,79636648	10,93171815	8,129748207	55,22	41,07
CMP_082-e5	14days	30,15329313	12,43694675	9,728528536	41,25	32,26
CMP_225-e4	14days	11,72566951	11,61774205	7,902912573	99,08	67,4
CMP_101-e3	14days	19,09024324	9,102990638	7,118984545	47,68	37,29
CMP_075-e7	14days	34,50835071	12,22918268	9,656181245	35,44	27,98
CMP_128-e2	14days	15,80026287	10,45751348	7,082590182	66,19	44,83
CMP_103-e1	14days	42,96617487	48,43207181	33,52974523	112,72	78,04
CMP_020-e6	28days	37,46179346	43,42178958	27,34162752	115,91	72,99
CMP_020-e3	28days	38,96088563	15,37416951	12,1529447	39,46	31,19
CMP_020-e5	28days	26,17025566	11,57427974	8,612342843	44,23	32,91
CMP_070-e2	28days	19,96467582	9,663597553	7,355130839	48,4	36,84
CMP_070-e4	28days	7,762317472	7,94729606	5,705056522	102,38	73,5
CMP_074-e1	28days	25,30574767	12,48426345	9,723973746	49,33	38,43
CMP_074-e2	28days	35,34379132	25,83297972	18,34734985	73,09	51,91
CMP_075-e2	28days	11,2707563	8,719501093	6,665613375	77,36	59,14
CMP_070-e3	28days	29,46009737	11,95498696	9,344520441	40,58	31,72
CMP_070-e1	28days	7,383119382	6,722903435	4,503892133	91,06	61
CMP_020-e1	28days	34,03033933	12,77369248	9,642482588	37,54	28,33
CMP_075-e3	28days	10,87902266	9,16210107	6,837314808	84,22	62,85
CMP_020-e7	28days	35,57017075	12,89036409	10,14839714	36,24	28,53
CMP_075-e4	28days	21,57394706	20,7251819	15,61170921	96,07	72,36
CMP_075-e1	28days	23,72548061	13,06401419	9,540538858	55,06	40,21
CMP_082-e3	28days	12,71552287	9,397487303	6,861159735	73,91	53,96
CMP_088-e1	28days	11,32253	12,1039154	9,30619963	106,9	82,19

Figure D.6: Results table of the 50 detectors for Occupancy

CMP_131-e1	28days	7,306106329	11,58021911	6,703728468	158,5	91,76
CMP_250-e1	28days	2,597192527	2,856679079	1,617197148	109,99	62,27
CMP_082-e2	28days	41,87458686	11,65289732	9,228861463	27,83	22,04
CMP_103-e3	28days	21,51747894	48,41686798	35,9166497	225,01	166,92
CMP_103-e5	28days	21,71680354	49,60918225	35,13482324	228,44	161,79
CMP_209-e3	28days	14,68057116	31,80134567	14,54561135	216,62	99,08
CMP_103-e4	28days	26,79213445	47,03305835	32,41254161	175,55	120,98
CMP_225-e2	28days	17,57594596	12,422113	10,1378166	70,68	57,68
CMP_088-e2	28days	8,019877339	7,773300925	5,231466941	96,93	65,23
CMP_250-e2	28days	6,733166081	6,705031313	5,007059971	99,58	74,36
CMP_225-e1	28days	15,46807979	11,73842394	9,255891916	75,89	59,84
CMP_081-e6	28days	42,49023358	14,12276484	11,12528586	33,24	26,18
CMP_101-e1	28days	24,71700509	10,75793318	8,331120331	43,52	33,71
CMP_081-e2	28days	37,16028451	13,70598352	10,71102183	36,88	28,82
CMP_128-e3	28days	31,86453409	15,45898569	11,3809835	48,51	35,72
CMP_103-e2	28days	24,80126973	54,57085706	36,6592939	220,03	147,81
CMP_081-e7	28days	23,28766992	10,45474883	8,130645914	44,89	34,91
CMP_081-e4	28days	34,98874104	28,22112464	20,32960683	80,66	58,1
CMP_209-e1	28days	7,417550721	9,634435908	6,736411724	129,89	90,82
CMP_137-e1	28days	8,989697846	7,23587899	5,30384929	80,49	59
CMP_131-e2	28days	11,69855692	8,649408955	6,545291671	73,94	55,95
CMP_264-e1	28days	16,29779385	10,67281458	7,963769401	65,49	48,86
CMP_081-e3	28days	24,9530691	15,27206775	12,11228675	61,2	48,54
CMP_128-e1	28days	22,00366266	11,68517583	8,609345113	53,11	39,13
CMP_081-e1	28days	42,14425715	12,61817763	10,07941159	29,94	23,92
CMP_101-e2	28days	20,89609584	9,350260106	7,21106838	44,75	34,51
CMP_121-e2	28days	19,76197286	11,05804102	8,241237797	55,96	41,7
CMP_082-e5	28days	30,3783746	12,77856791	9,936270996	42,06	32,71

Figure D.7: Results table of the 50 detectors for Occupancy

CMP_225-e4	28days	11,78289265	12,01816952	8,845244199	102	75,07
CMP_101-e3	28days	19,16477225	9,328912245	7,207956097	48,68	37,61
CMP_075-e7	28days	34,55479337	12,86443587	10,11295251	37,23	29,27
CMP_128-e2	28days	15,8203647	10,83417285	7,564783027	68,48	47,82
CMP_103-e1	28days	38,13511173	88,11354995	72,9119636	231,06	191,19

Figure D.8: Results table of the 50 detectors for Occupancy

Appendix E

Intensity results for 50 detectors, period and horizon of 1 hour

The following table presents the results obtained for the 50 detectors, evaluated over a period and horizon of 1 hour for the intensity predictions. The table includes the following columns:

- **Detector:** The unique identification number or code assigned to each traffic detector.
- **Horizon:** The forecast horizon, refers to the time period for which the prediction is made, in this case, 1 hour.
- **Mean_y_intensity:** The average observed intensity value over the specified horizon of 1 hour.
- **Mean_RMSE:** The average Root Mean Square Error (RMSE) for intensity predictions within the 1-hour horizon.
- **Mean_MAE:** The average Mean Absolute Error (MAE) for intensity predictions within the 1-hour horizon.
- **Relative Error RMSE Intensity:** The RMSE is expressed as a percentage of the average observed intensity value.
- **Relative Error MAE Intensity:** The MAE is expressed as a percentage of the average observed intensity value.

Detector	Horizon	mean_y_intensity	mean_y_intensity	mean_rmse	mean_mae	Relative Error RMSE Intensity	Relative Error MAE Intensity
CMP_020-e1	1hour	583,7748888	583,7748888	119,0506995	88,1059657	20,39	15,09
CMP_020-e3	1hour	248,2924855	248,2924855	85,99940497	61,76587104	34,64	24,88
CMP_020-e5	1hour	412,4526064	412,4526064	92,6623489	69,79700426	22,47	16,92
CMP_020-e6	1hour	109,3845043	109,3845043	72,05912874	58,57021858	65,88	53,55
CMP_020-e7	1hour	228,4951	228,4951	68,58061493	53,03281899	30,01	23,21
CMP_070-e1	1hour	97,8170904	97,8170904	39,87265171	30,23517015	40,76	30,91
CMP_070-e2	1hour	212,3451429	212,3451429	67,48754984	49,18111682	31,78	23,16
CMP_070-e3	1hour	283,2388247	283,2388247	94,44196369	71,93175161	33,34	25,4
CMP_070-e4	1hour	78,31567515	78,31567515	41,94667565	29,80952131	53,56	38,06
CMP_074-e1	1hour	253,839396	253,839396	76,4460236	57,04469428	30,12	22,47
CMP_074-e2	1hour	274,7830829	274,7830829	114,0528418	82,65353771	41,51	30,08
CMP_075-e1	1hour	413,2329953	413,2329953	108,1603959	80,29138778	26,17	19,43
CMP_075-e2	1hour	139,0653254	139,0653254	77,63686493	57,99602503	55,83	41,7
CMP_075-e3	1hour	54,76872134	54,76872134	41,41929886	30,30785125	75,63	55,34
CMP_075-e4	1hour	54,28375708	54,28375708	34,38686059	25,98429851	63,35	47,87
CMP_075-e7	1hour	148,8037261	148,8037261	59,56261202	46,30731422	40,03	31,12
CMP_081-e1	1hour	305,4307517	305,4307517	66,837874	49,88045935	21,88	16,33
CMP_081-e2	1hour	313,1395723	313,1395723	92,21751319	69,1790767	29,45	22,09
CMP_081-e3	1hour	83,57580175	83,57580175	39,05744398	29,50027935	46,73	35,3
CMP_081-e4	1hour	112,5017808	112,5017808	65,09120559	42,54551755	57,86	37,82
CMP_081-e6	1hour	295,1502165	295,1502165	78,89646239	56,94899327	26,73	19,29
CMP_081-e7	1hour	256,3802412	256,3802412	85,60233997	64,72251272	33,39	25,24
CMP_082-e2	1hour	480,1345935	480,1345935	101,1042515	77,52985761	21,06	16,15
CMP_082-e3	1hour	87,88462891	87,88462891	39,92838409	29,08219664	45,43	33,09
CMP_082-e5	1hour	275,8834751	275,8834751	101,7056134	74,68767386	36,87	27,07
CMP_088-e1	1hour	444,8971579	444,8971579	153,0845127	116,6481458	34,41	26,22
CMP_088-e2	1hour	111,8263761	111,8263761	61,60256142	45,18149169	55,09	40,4
CMP_101-e1	1hour	643,9943065	643,9943065	147,8990295	109,0990843	22,97	16,94
CMP_101-e2	1hour	507,9592348	507,9592348	125,4631239	93,18809626	24,7	18,35

Figure E.1: Results table of the 50 detectors for the period and horizon of 1 hour for Intensity

CMP_101-e3	1hour	237,8723557	237,8723557	73,67821789	57,1879157	30,97	24,04
CMP_103-e1	1hour	188,5761811	188,5761811	126,1679957	93,61152536	66,91	49,64
CMP_103-e2	1hour	280,6907074	280,6907074	121,7127087	87,01693991	43,36	31
CMP_103-e3	1hour	210,0381415	210,0381415	90,55245634	67,57619171	43,11	32,17
CMP_103-e4	1hour	137,5011876	137,5011876	59,90517656	46,12005857	43,57	33,54
CMP_103-e5	1hour	98,6962788	98,6962788	56,72421964	40,42295706	57,47	40,96
CMP_121-e2	1hour	169,5117295	169,5117295	56,82538772	43,40172614	33,52	25,6
CMP_128-e1	1hour	865,9656115	865,9656115	170,5066423	125,2146219	19,69	14,46
CMP_128-e2	1hour	676,6930625	676,6930625	156,5995436	117,2741074	23,14	17,33
CMP_128-e3	1hour	204,3651246	204,3651246	77,03788689	55,82506715	37,7	27,32
CMP_131-e1	1hour	103,2953789	103,2953789	48,09107178	35,64750028	46,56	34,51
CMP_131-e2	1hour	88,93802714	88,93802714	37,10121496	28,58592858	41,72	32,14
CMP_137-e1	1hour	115,7415716	115,7415716	45,8574092	35,53628471	39,62	30,7
CMP_209-e1	1hour	16,73975637	16,73975637	8,19814087	6,211236671	48,97	37,1
CMP_209-e3	1hour	123,2818869	123,2818869	51,20447762	38,32314682	41,53	31,09
CMP_225-e1	1hour	246,9895877	246,9895877	57,67675893	44,10639417	23,35	17,86
CMP_225-e2	1hour	290,1206468	290,1206468	63,62811104	49,26499762	21,93	16,98
CMP_225-e4	1hour	151,8098894	151,8098894	67,69959197	50,85216581	44,59	33,5
CMP_250-e1	1hour	138,2153055	138,2153055	61,46156083	45,50787925	44,47	32,93
CMP_250-e2	1hour	230,7802107	230,7802107	93,31949094	70,48489674	40,44	30,54
CMP_264-e1	1hour	596,4818836	596,4818836	169,6405711	133,5691276	28,44	22,39

Figure E.2: Results table of the 50 detectors for the period and horizon of 1 hour for Intensity

Appendix F

Occupancy results for 50 detectors, period and horizon of 1 hour

The following table presents the results obtained for the 50 detectors, evaluated over a period and horizon of 1 hour for the occupancy predictions. The table includes the following columns:

- **Detector:** The unique identification number or code assigned to each traffic detector.
- **Horizon:** The forecast horizon, which refers to the time period for which the prediction is made, in this case, 1 hour.
- **Mean_y_occupancy:** The average observed occupancy value over the specified horizon of 1 hour.
- **Mean_RMSE:** The average Root Mean Square Error (RMSE) for occupancy predictions within the 1-hour horizon.
- **Mean_MAE:** The average Mean Absolute Error (MAE) for occupancy predictions within the 1-hour horizon.
- **Relative Error RMSE Occupancy:** The RMSE expressed as a percentage of the average observed occupancy value.
- **Relative Error MAE Occupancy:** The MAE expressed as a percentage of the average observed occupancy value.

Detector	Horizon	mean_y_occupancy	mean_y_occupancy	mean_rmse	mean_mae	Relative Error RMSE Occupancy	Relative Error MAE Occupancy
CMP_020-e1_1hour		34,39274718	34,39274718	11,70627122	8,78502711	34,04	25,54
CMP_020-e3_1hour		39,10381503	39,10381503	13,56734975	10,4942899	34,7	26,84
CMP_020-e5_1hour		26,14200981	26,14200981	10,56502073	7,795480929	40,41	29,82
CMP_020-e6_1hour		46,34612608	46,34612608	20,31043207	15,09499022	43,82	32,57
CMP_020-e7_1hour		35,20893767	35,20893767	11,71810576	9,004718673	33,28	25,58
CMP_070-e1_1hour		7,351694915	7,351694915	6,349959975	4,248197017	86,37	57,79
CMP_070-e2_1hour		19,97142857	19,97142857	9,192330695	7,010654731	46,03	35,1
CMP_070-e3_1hour		29,53332571	29,53332571	11,49556269	8,981925233	38,92	30,41
CMP_070-e4_1hour		7,632557017	7,632557017	7,434271603	5,250218929	97,4	68,79
CMP_074-e1_1hour		25,50148707	25,50148707	11,58579807	8,997583863	45,43	35,28
CMP_074-e2_1hour		33,55020614	33,55020614	19,87218326	14,78343871	59,23	44,06
CMP_075-e1_1hour		23,62082716	23,62082716	11,76680928	8,595963448	49,82	36,39
CMP_075-e2_1hour		11,19027997	11,19027997	7,88760552	5,98276523	70,49	53,46
CMP_075-e3_1hour		10,93008151	10,93008151	8,733052799	6,470996487	79,9	59,2
CMP_075-e4_1hour		22,09656172	22,09656172	19,46891733	14,22828212	88,11	64,39
CMP_075-e7_1hour		33,97728387	33,97728387	11,65453543	9,118709199	34,3	26,84
CMP_081-e1_1hour		41,8594533	41,8594533	12,06713639	9,616418969	28,83	22,97
CMP_081-e2_1hour		36,96343987	36,96343987	12,68934117	9,930068358	34,33	26,86
CMP_081-e3_1hour		24,82786686	24,82786686	14,67701571	11,54655566	59,12	46,51
CMP_081-e4_1hour		33,56931803	33,56931803	21,82415153	15,70485763	65,01	46,78
CMP_081-e6_1hour		42,34294506	42,34294506	12,80978887	9,982240551	30,25	23,57
CMP_081-e7_1hour		23,18782309	23,18782309	9,947934318	7,716947552	42,9	33,28
CMP_082-e2_1hour		41,55033022	41,55033022	10,4289753	8,3060526	25,1	19,99
CMP_082-e3_1hour		12,79962481	12,79962481	9,108013681	6,701897752	71,16	52,36
CMP_082-e5_1hour		30,14433464	30,14433464	11,83297738	9,180044657	39,25	30,45
CMP_088-e1_1hour		10,43602329	10,43602329	7,514541392	5,236145594	72,01	50,17
CMP_088-e2_1hour		7,901613102	7,901613102	7,173910014	4,438932368	90,79	56,18
CMP_101-e1_1hour		24,61944887	24,61944887	9,877975945	7,655244257	40,12	31,09
CMP_101-e2_1hour		20,7994762	20,7994762	8,575829686	6,668001677	41,23	32,06
CMP_101-e3_1hour		19,09346241	19,09346241	8,729818876	6,790723387	45,72	35,57
CMP_103-e1_1hour		46,61811251	46,61811251	17,44058198	11,41339615	37,41	24,48
CMP_103-e2_1hour		23,16042534	23,16042534	17,72789897	12,42595235	76,54	53,65
CMP_103-e3_1hour		22,47792418	22,47792418	17,8082466	12,46728858	79,23	55,46
CMP_103-e4_1hour		25,67969121	25,67969121	19,00284259	13,9010375	74	54,13

Figure F.1: Results table of the 50 detectors for the period and horizon of 1 hour for Occupancy

CMP_103-e5_1hour	21,91080921	21,91080921	17,64489026	12,32999842	80,53	56,27
CMP_121-e2_1hour	19,82313707	19,82313707	10,70006375	7,977092341	53,98	40,24
CMP_128-e1_1hour	21,94898656	21,94898656	10,6777678	7,795609479	48,65	35,52
CMP_128-e2_1hour	15,73801917	15,73801917	10,02183465	6,595240466	63,68	41,91
CMP_128-e3_1hour	31,7970344	31,7970344	14,15360913	10,42017156	44,51	32,77
CMP_131-e1_1hour	7,506715958	7,506715958	11,10691155	5,905194766	147,96	78,67
CMP_131-e2_1hour	11,73071752	11,73071752	8,470091323	6,393455287	72,2	54,5
CMP_137-e1_1hour	9,083218448	9,083218448	6,954357278	5,090756655	76,56	56,05
CMP_209-e1_1hour	7,318936877	7,318936877	8,775063487	6,059167765	119,9	82,79
CMP_209-e3_1hour	14,59030893	14,59030893	10,45430923	7,678372543	71,65	52,63
CMP_225-e1_1hour	15,49625156	15,49625156	8,237456437	6,078039091	53,16	39,22
CMP_225-e2_1hour	17,69154229	17,69154229	10,0106936	7,313403493	56,58	41,34
CMP_225-e4_1hour	11,82388801	11,82388801	11,67883571	7,999422899	98,77	67,65
CMP_250-e1_1hour	2,593697749	2,593697749	2,799008604	1,456144863	107,92	56,14
CMP_250-e2_1hour	6,542549414	6,542549414	5,65942395	4,12385505	86,5	63,03
CMP_264-e1_1hour	16,83118071	16,83118071	10,32608858	7,521419266	61,35	44,69

Figure F.2: Results table of the 50 detectors for the period and horizon of 1 hour for Occupancy

Appendix G

Histograms of Relative Errors of RMSE and MAE for Intensity for 50 Detectors

This section presents histograms that illustrate the distribution of relative errors for Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for intensity forecasts at various forecast horizons over the 50 detectors. The following horizons were considered: 1 hour, 1 day, 7 days, 14 days, and 28 days. The histograms provide a visual representation of the error distribution, helping to identify the variability and performance of the forecasting model at each horizon.

- **1-hour horizon:** The forecast horizon of 1 hour captures short-term traffic patterns, which tend to show lower error variability compared to longer horizons.
- **1-day horizon:** At this horizon, the system captures daily variations in traffic intensity, providing insight into the model's performance over a full day.
- **7-day horizon:** Weekly trends in traffic intensity are captured at this horizon, with errors likely increasing as the forecast period extends.
- **14-day horizon:** The two-week horizon tests the model's ability to forecast over medium-term periods, where traffic patterns may be influenced by external factors such as holidays or events.
- **28-day horizon:** Long-term forecasts, such as the 28-day horizon, present the highest level of uncertainty, as traffic conditions may change significantly over a month.

The histograms for the relative error of RRMSE at each horizon are presented below:

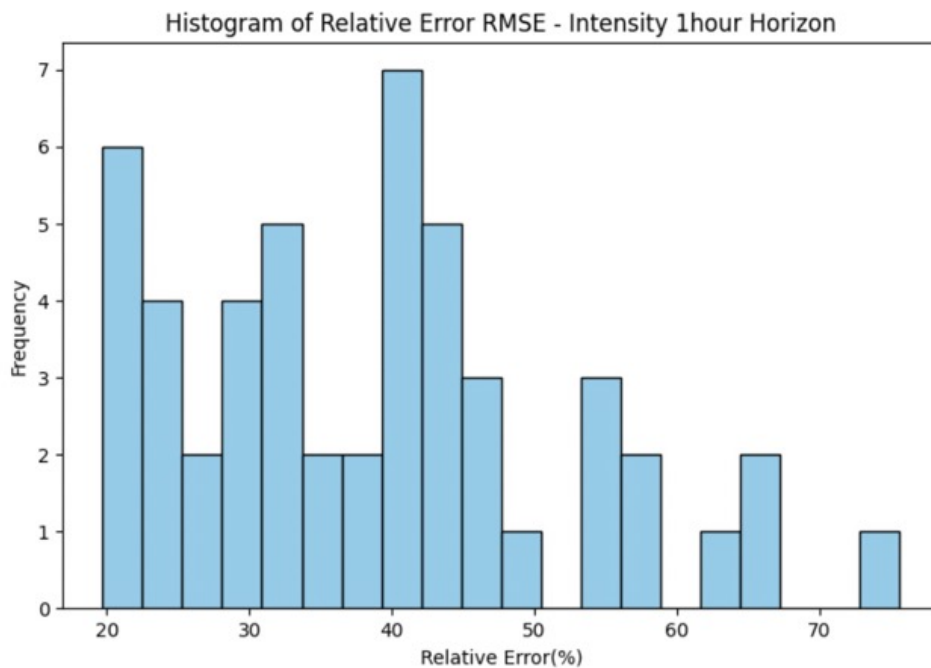


Figure G.1: Histogram of RRMSE for Intensity - 1 Hour Horizon

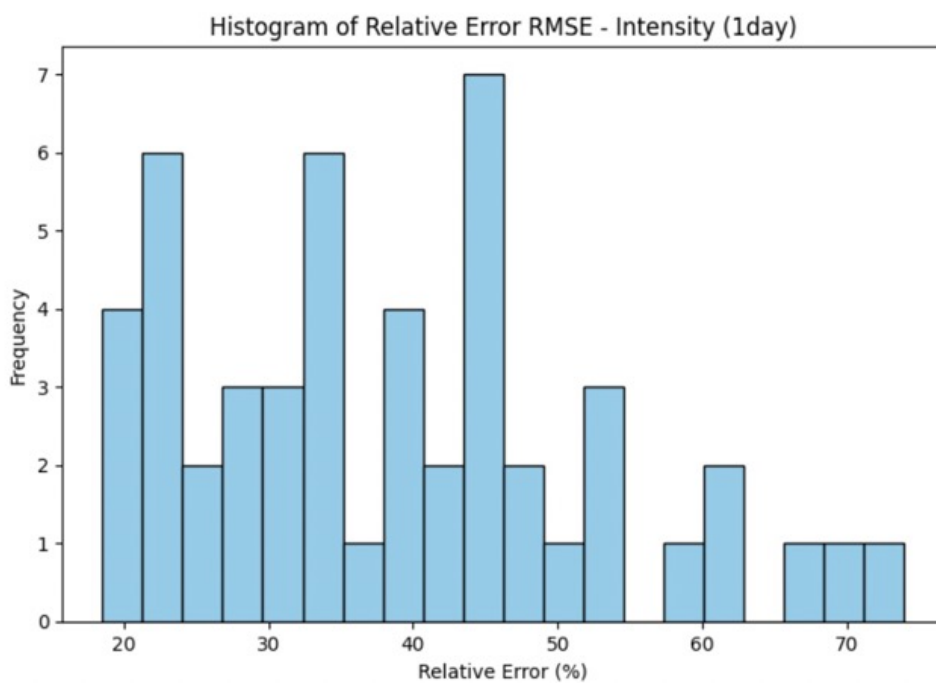


Figure G.2: Histogram of RRMSE for Intensity - 1 Day Horizon

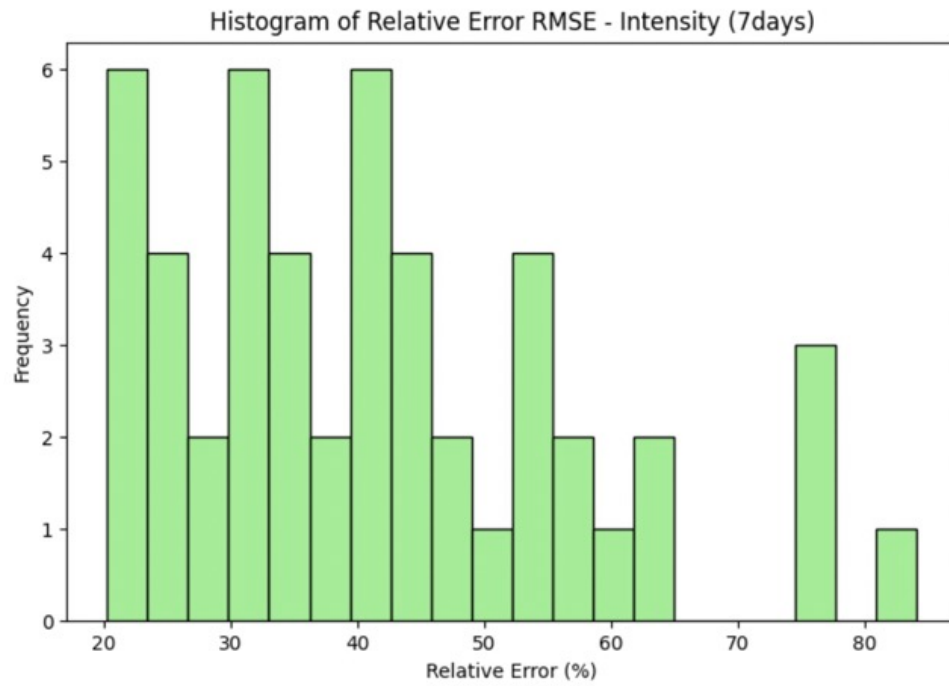


Figure G.3: Histogram of RRMSE for Intensity - 7 Days Horizon

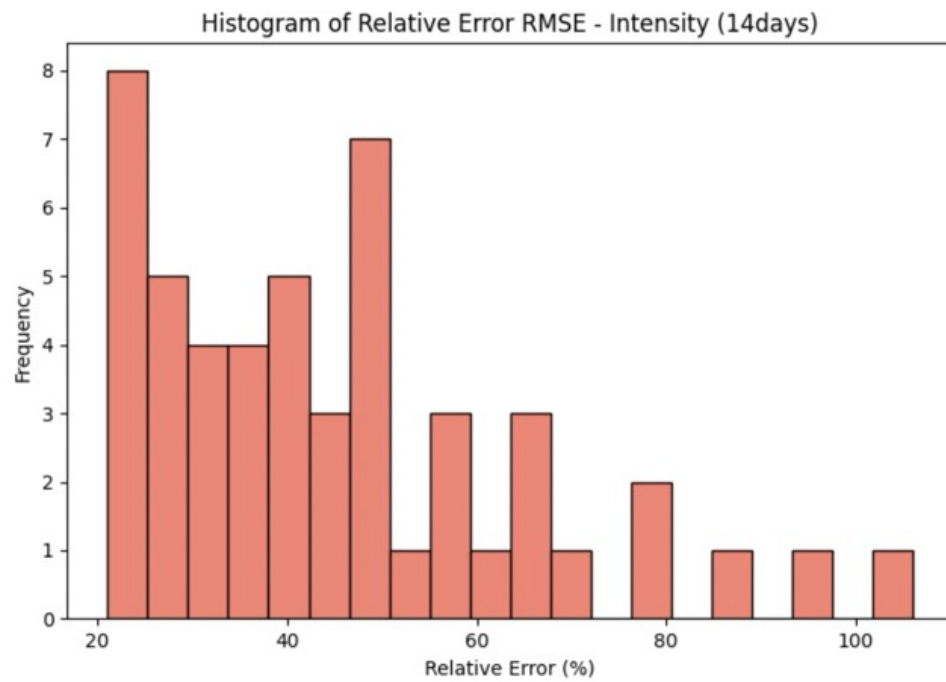


Figure G.4: Histogram of RRMSE for Intensity - 14 Days Horizon

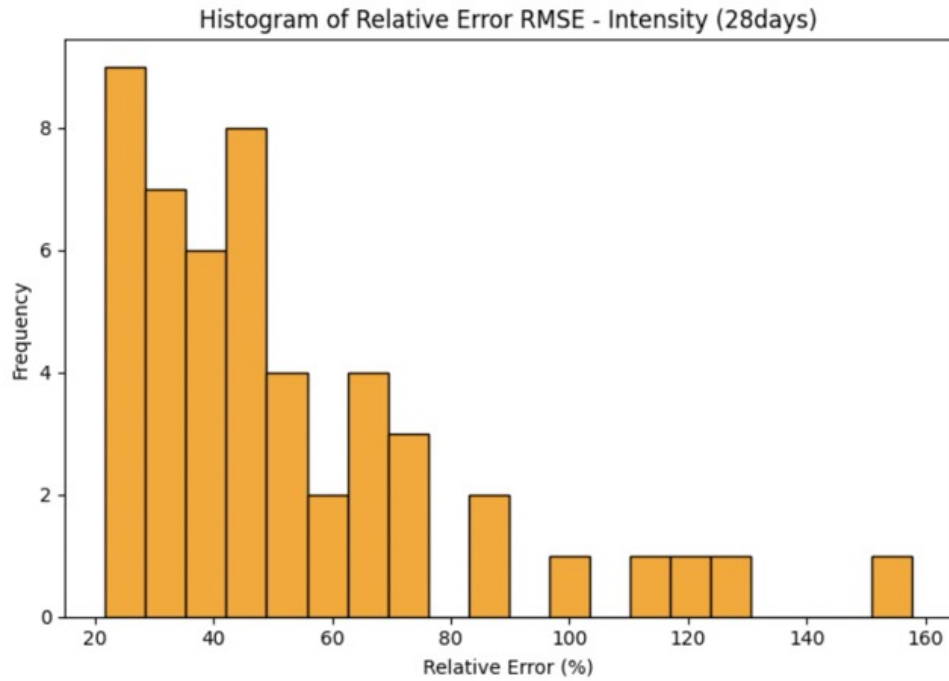


Figure G.5: Histogram of RRMSE for Intensity - 28 Days Horizon

The histograms for the relative error of MAE at each horizon are presented below:

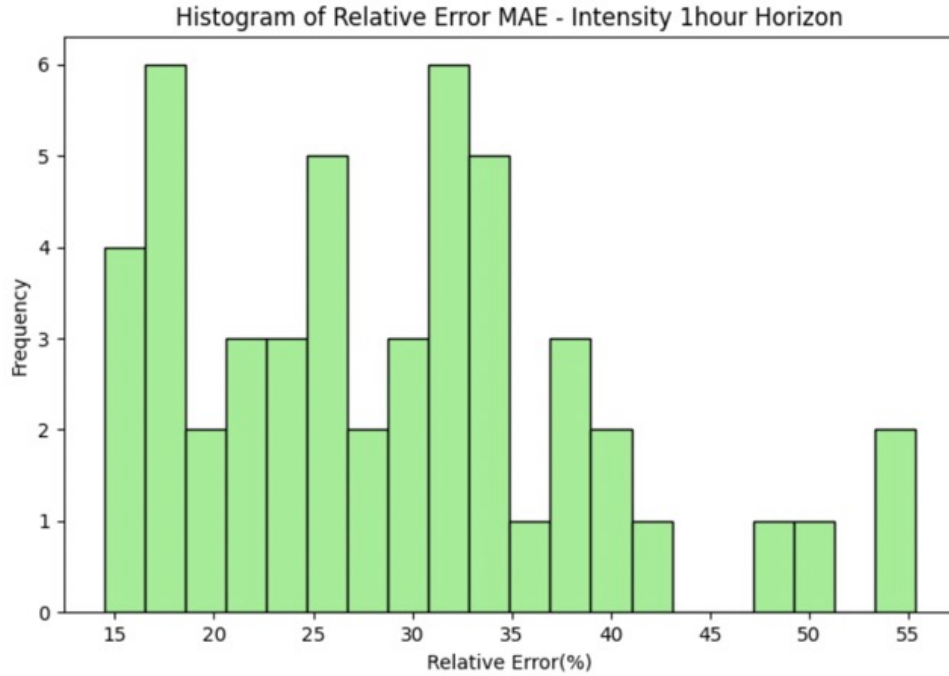


Figure G.6: Histogram of RMAE for Intensity - 1 Hour Horizon

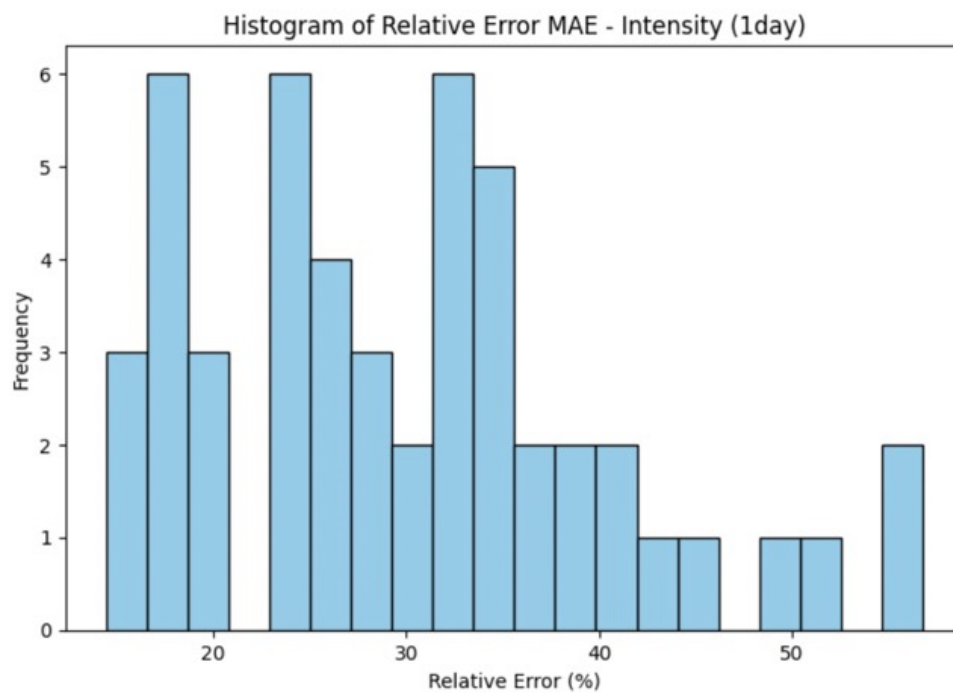


Figure G.7: Histogram of RMAE for Intensity - 1 Day Horizon

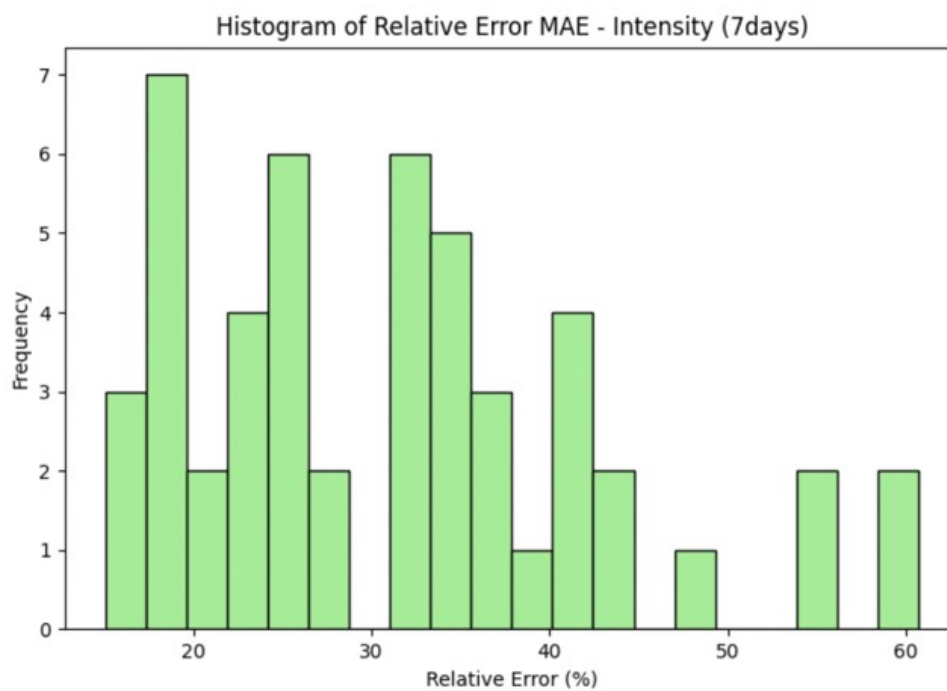


Figure G.8: Histogram of RMAE for Intensity - 7 Days Horizon

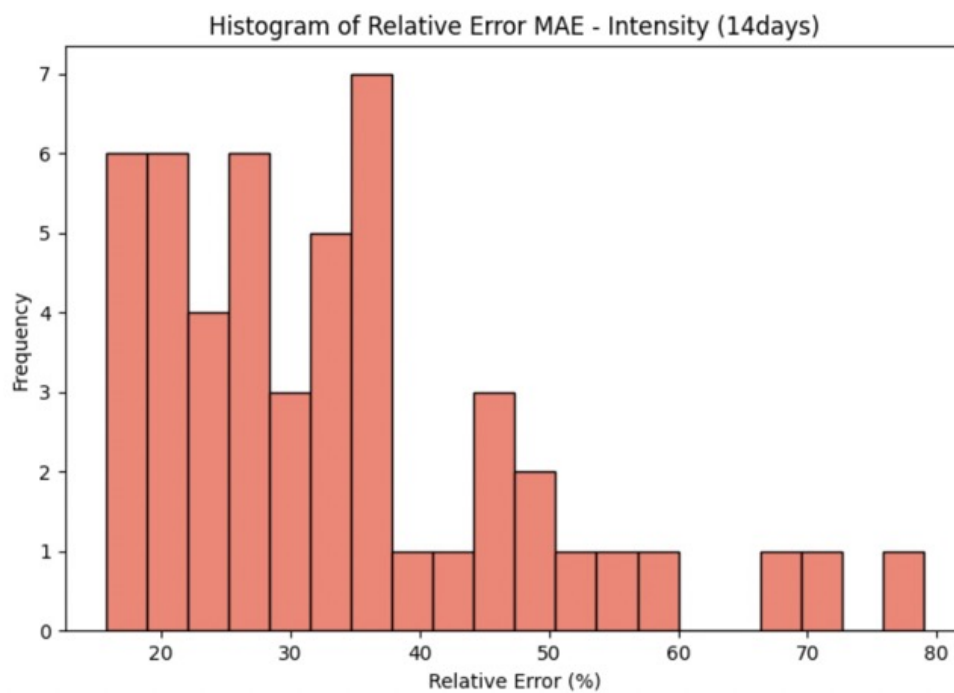


Figure G.9: Histogram of RMAE for Intensity - 14 Days Horizon

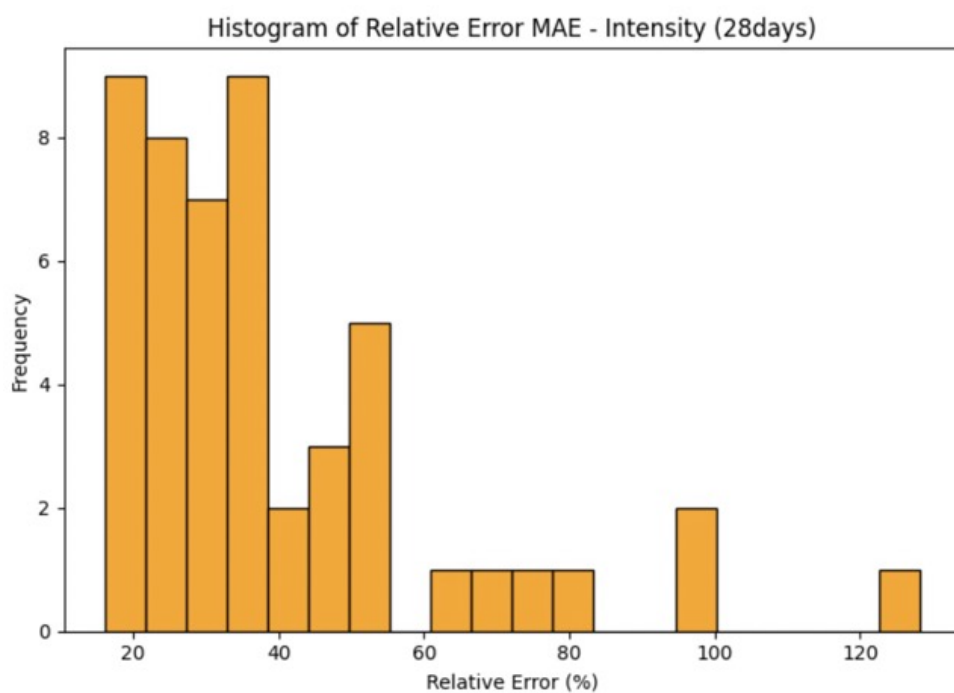


Figure G.10: Histogram of RMAE for Intensity - 28 Days Horizon

Appendix H

Histograms of Relative Errors of RMSE and MAE for Occupancy for 50 Detectors

This section presents histograms illustrating the distribution of relative errors for Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for occupancy forecasts at various forecast horizons. The following horizons were considered: 1 hour, 1 day, 7 days, 14 days, and 28 days. The histograms provide a visual representation of the error distribution, helping to identify the variability and performance of the forecasting model at each horizon.

- **1-hour horizon:** The forecast horizon of 1 hour captures short-term traffic patterns in occupancy, which tend to show lower error variability compared to longer horizons.
- **1-day horizon:** At this horizon, the system captures daily variations in occupancy, providing insight into the model's performance over a full day.
- **7-day horizon:** Weekly trends in occupancy are captured at this horizon, with errors likely increasing as the forecast period extends.
- **14-day horizon:** The two-week horizon tests the model's ability to forecast over medium-term periods, where occupancy patterns may be influenced by external factors such as holidays or events.
- **28-day horizon:** Long-term forecasts, such as the 28-day horizon, present the highest level of uncertainty, as traffic conditions may change significantly over a month.

The histograms for the relative error of RRMSE at each horizon are presented below:

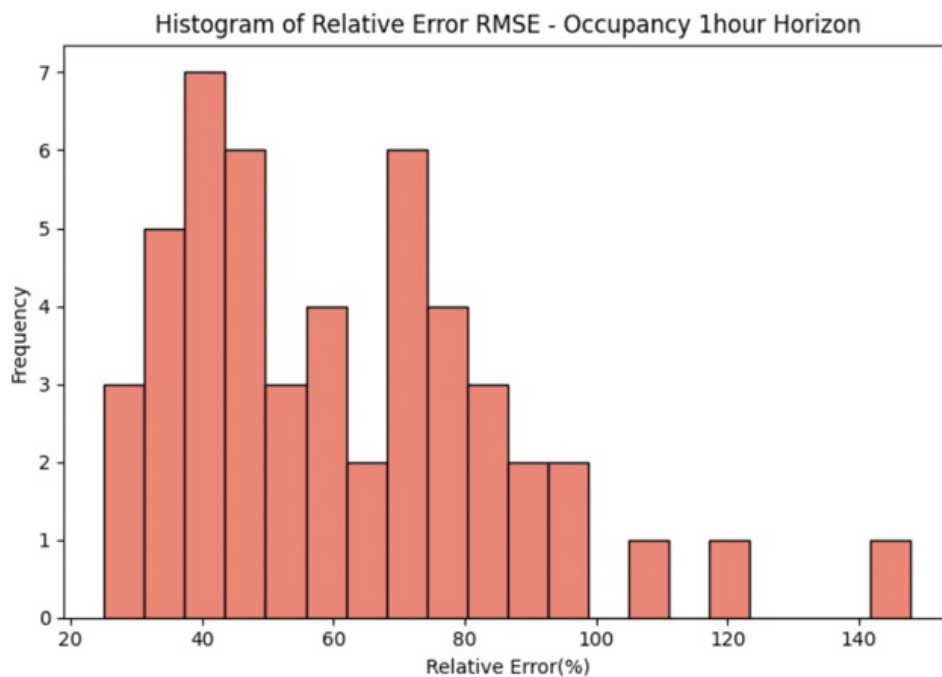


Figure H.1: Histogram of RRMSE for Occupancy - 1 Hour Horizon

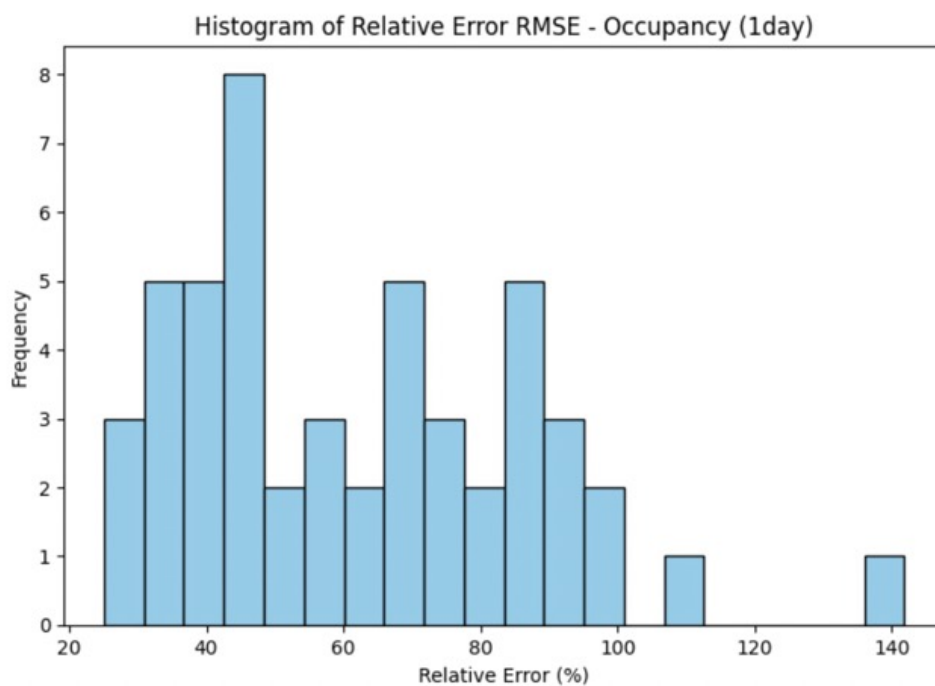


Figure H.2: Histogram of RRMSE for Occupancy - 1 Day Horizon

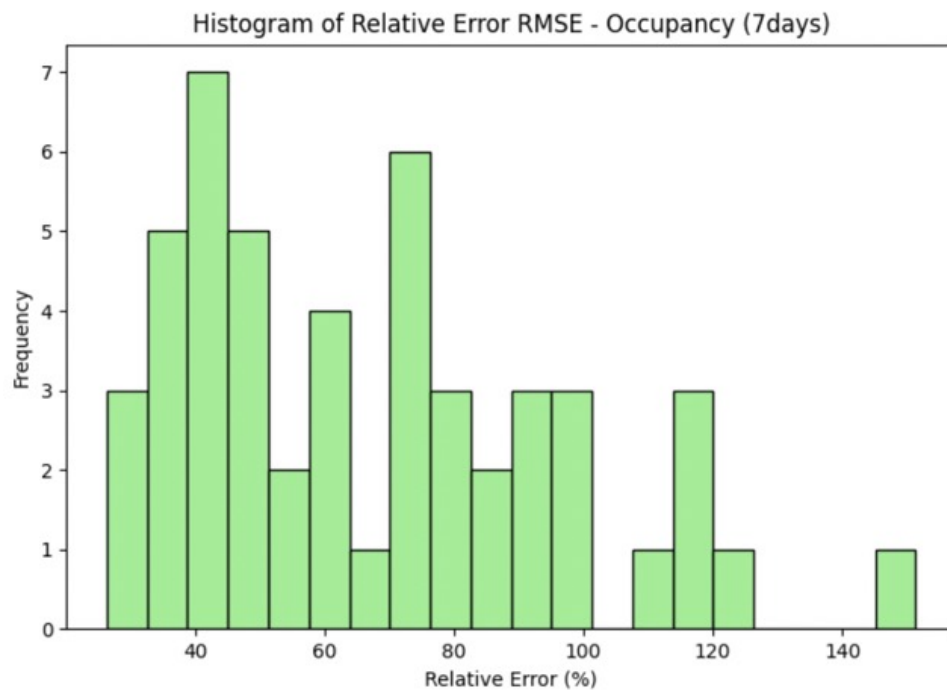


Figure H.3: Histogram of RRMSE for Occupancy - 7 Days Horizon

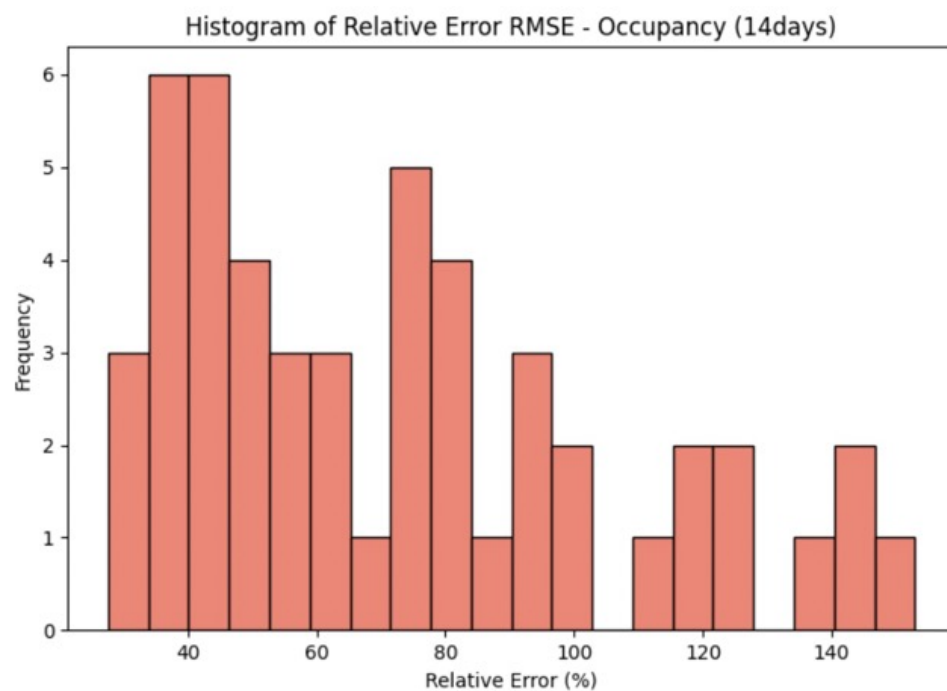


Figure H.4: Histogram of RRMSE for Occupancy - 14 Days Horizon

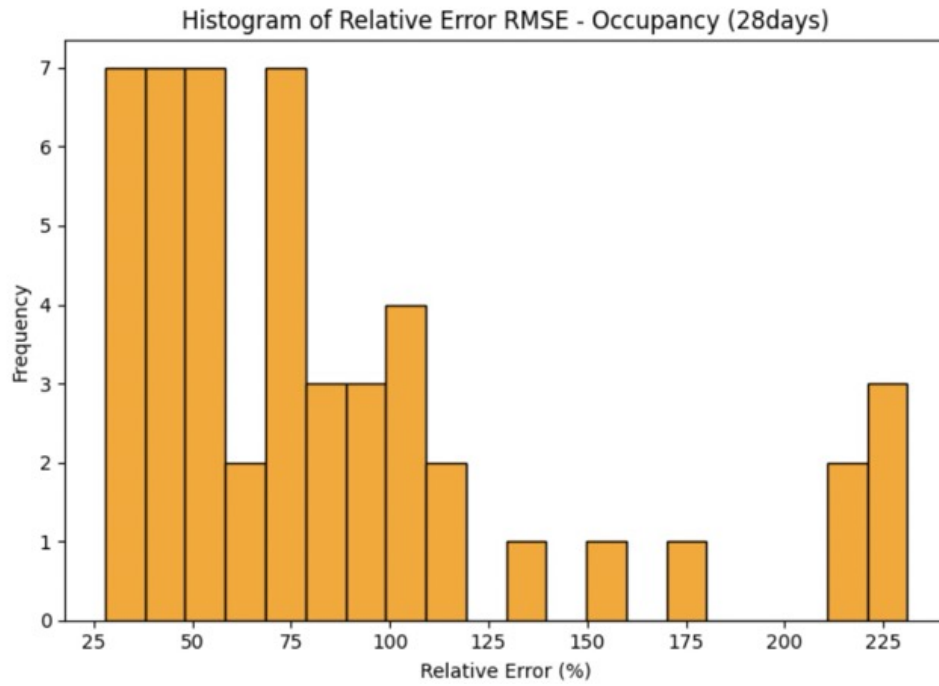


Figure H.5: Histogram of RRMSE for Occupancy - 28 Days Horizon

The histograms for the relative error of RMAE at each horizon are presented below:

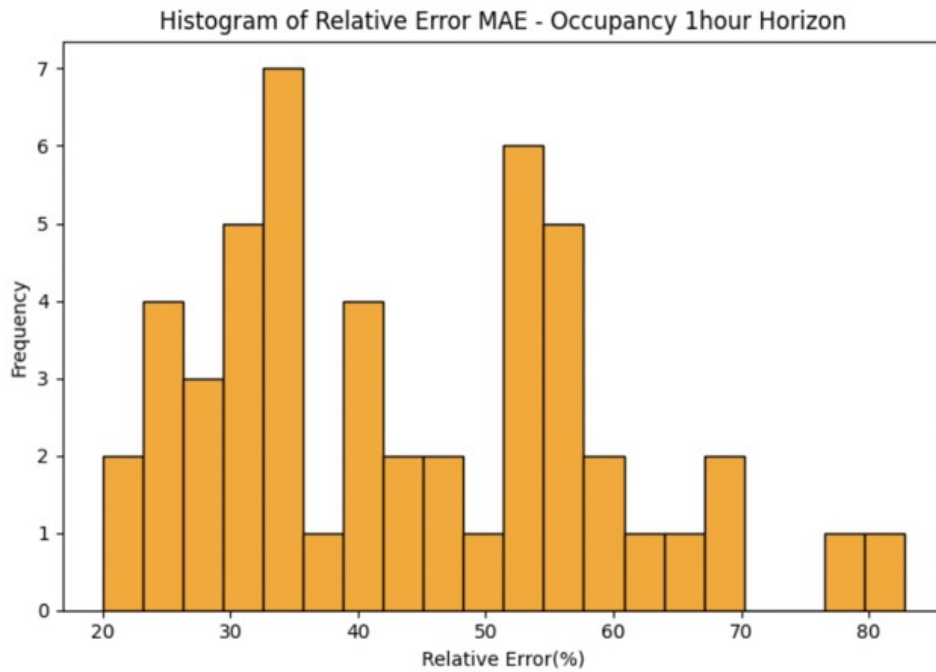


Figure H.6: Histogram of RMAE for Occupancy - 1 Hour Horizon

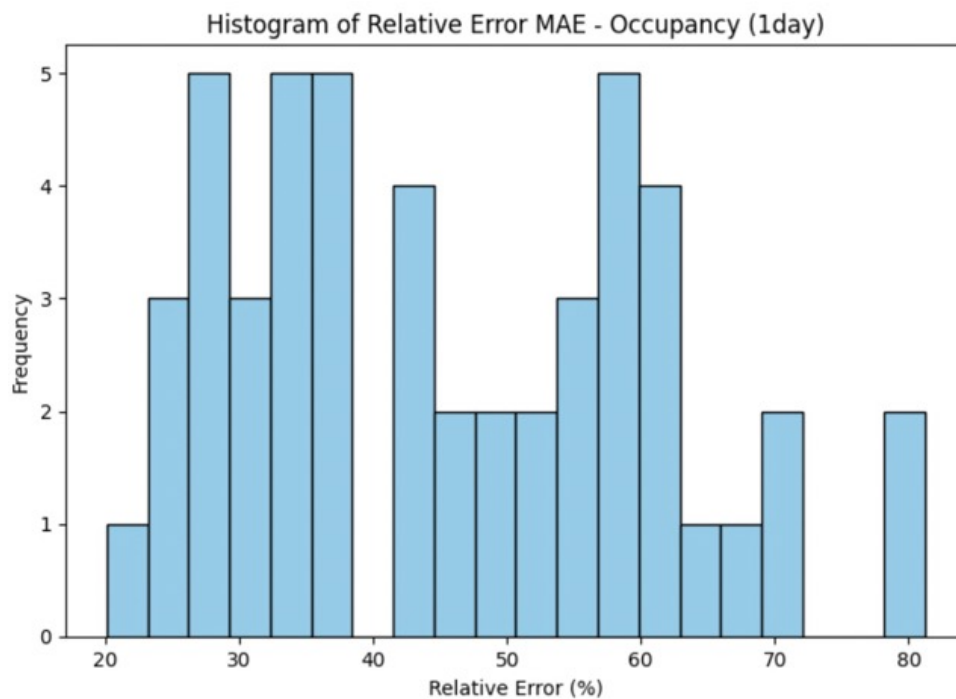


Figure H.7: Histogram of RMAE for Occupancy - 1 Day Horizon

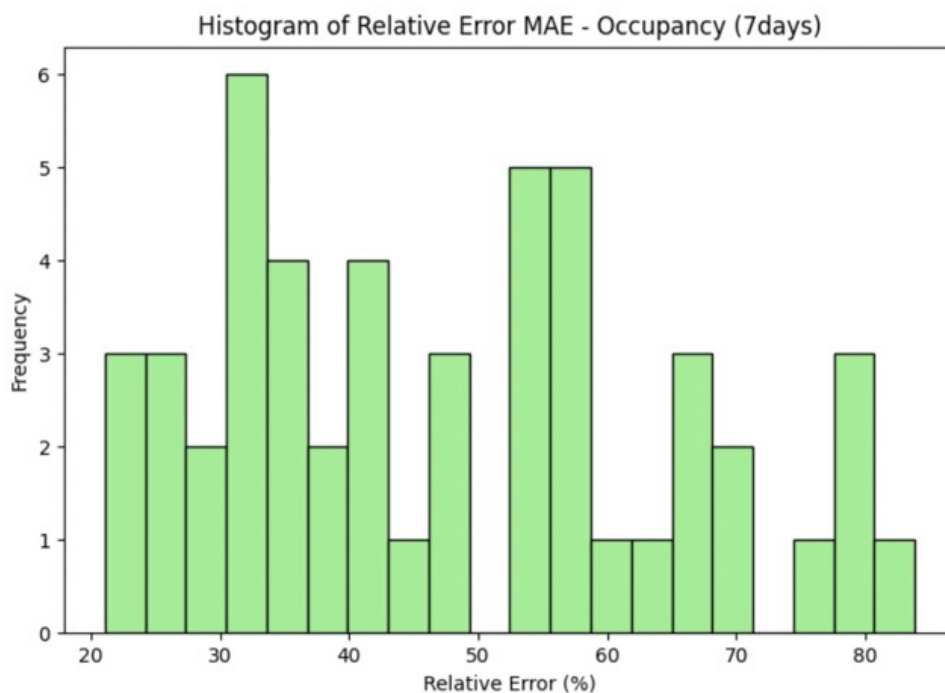


Figure H.8: Histogram of RMAE for Occupancy - 7 Days Horizon

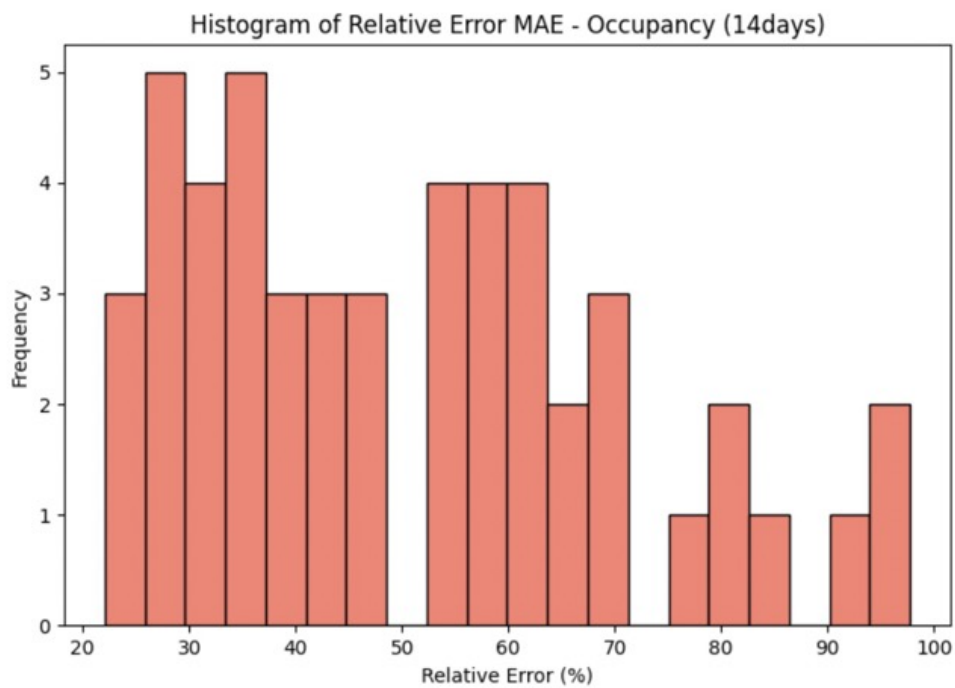


Figure H.9: Histogram of RMAE for Occupancy - 14 Days Horizon

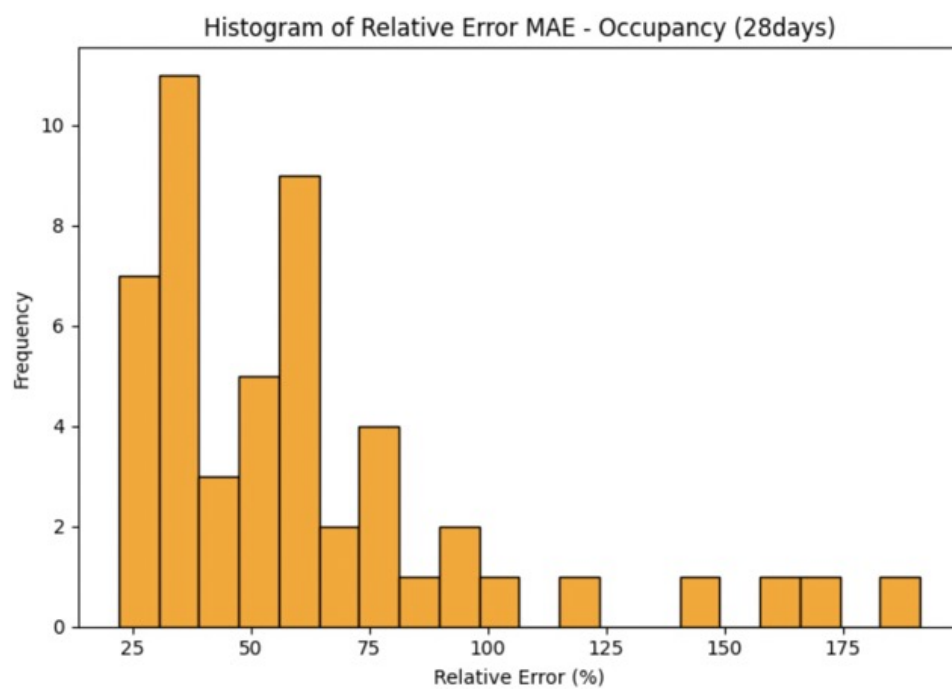


Figure H.10: Histogram of RMAE for Occupancy - 28 Days Horizon