



Universiteit  
Leiden  
The Netherlands

# Comparing the Seasonal ARIMA Model to the Gaussian Process Regression Model When Making OV-fiets Bicycle Availability Predictions

Pikutis, Liudvikas

## Citation

Pikutis, L. (2024). *Comparing the Seasonal ARIMA Model to the Gaussian Process Regression Model When Making OV-fiets Bicycle Availability Predictions*.

Version: Not Applicable (or Unknown)

License: [License to inclusion and publication of a Bachelor or Master Thesis, 2023](#)

Downloaded from: <https://hdl.handle.net/1887/3775229>

**Note:** To cite this publication please use the final published version (if applicable).



Universiteit Leiden

Faculteit der Sociale Wetenschappen

# Comparing the Seasonal ARIMA Model to the Gaussian Process Regression Model When Making OV-fiets Bicycle Availability Predictions

Master's Thesis

---

Liudvikas Pikutis

Master's Thesis Psychology,  
Methodology and Statistics Unit, Institute of Psychology  
Faculty of Social and Behavioural Sciences, Leiden University

Date: May 2024

Student number: xxxxxxxx

Supervisor: Dr. Julian D. Karch (internal), MSc. Wouter Hordijk (external),  
MSc. Lucas Fijen (external)

## **Abstract**

This thesis presents a comparison of the seasonal Autoregressive Integrated Moving Average (ARIMA) model and the Gaussian Process Regression (GPR) model when predicting OV-fiets bicycle availability in the following 48 hours. Both the ARIMA and the GPR have been used in past research when predicting traffic flow data. However, a proper comparison of a seasonal ARIMA to the GPR is yet to be examined. OV-fiets bicycle availability data of different rental locations was used to make predictions of future OV-fiets bicycle availability. A rolling-origin cross-validation was implemented to tune the hyperparameters and train the models. Results suggest that the GPR performs better than the seasonal ARIMA, however, upon further visual inspection of the predictions, it was concluded that the hyperparameter selection process had faults, which led to biases in predictions. It was found that the seasonal ARIMA was over-differenced, which gave inaccurate forecasts. Advice is given to future researchers about the limitations of over-differencing data.

# TABLE OF CONTENTS

<b>1. INTRODUCTION</b>	<b>4</b>
<b>2. METHODS</b>	<b>8</b>
2.1. OV-fiets Bicycle Availability Data	8
2.2. Seasonal ARIMA Model	11
2.2.1. Non-Seasonal Orders	11
2.2.2. Seasonal Orders	12
2.2.3. Wold Decomposition	13
2.3. Gaussian Process Regression	13
2.3.1. Bayesian Approach and Introduction of Gaussian Process Regression	13
2.3.2. Kernel Function	15
2.4. Stationarity and Seasonality	16
2.4.1. Inspection of Data	16
2.4.2. Autocorrelation Analyses	17
2.5. Cross-Validation Methods	20
2.5.1. k-fold Cross-Validation	20
2.5.2. Rolling-Origin Cross-Validation	21
2.5.3. Model Evaluation Metrics	23
2.6. Baseline Models	23
<b>3. RESULTS</b>	<b>24</b>
<b>4. DISCUSSION</b>	<b>30</b>
4.1. Interpretation of Results	30
<b>5. CONCLUSION</b>	<b>33</b>
<b>REFERENCES</b>	<b>34</b>

# 1. Introduction

Nederlandse Spoorwegen (NS) is the largest passenger railway operator in the Netherlands, running trains on the busiest railway network in Europe (International union of railways, 2024). Besides running and maintaining trains, NS provides numerous additional services, among which is OV-fiets, a public bike rental service. There are over 300 OV-fiets rental locations, mainly located at train stations, and approximately 22,500 OV-fiets bicycles actively in service (Nederlandse Spoorwegen, n.d). Each rental location provides a limited number of bicycles that users can rent. Once all bicycles are rented out, there is no option to replenish the stock besides bicycles being returned at the end of a rental.

This poses a potential problem for OV-fiets users. If a user plans a journey involving the usage of OV-fiets, how can they know whether a bicycle will be available? As of now, only the current availability of OV-fiets is known to the user via an online application. There could be many OV-fiets bicycles available at the time of checking the application, however, by the time the user arrives at the OV-fiets rental location, the number could have dropped to 0 resulting in the inability to continue the journey as planned.

As reservations of OV-fiets bicycles are not part of the product design, one can either wait until someone returns a rented bicycle or choose another form of transportation like a bus to reach their destination. This causes delays and unpredictability in the journey of daily users of OV-fiets. Data available to NS shows that in some OV-fiets rental locations, once the bicycles are rented out, it can take hours until bicycles are available again. Within these locations, users cannot simply wait until someone returns an OV-fiets bicycle.

Many users rely on the availability of OV-fiets bicycles. A lot of OV-fiets users are travelling for work or study related purposes. It is not possible for users to have delays, as it could lead to being late for meetings, interviews, classes, etc. It is part of the mission of NS to provide a door-to-door journey without trouble or changes in the planning. The best that NS can do is to provide insight into OV-fiets availability so that travellers can adapt their journey to avoid delays.

An approach to mitigate such delays is to predict the availability of OV-fiets in the next 48 hours and communicate these predictions to the users of OV-fiets. These predictions would provide insight to the user into the extent to which they can rely on OV-fiets during their journey. Not only would this save a lot of time, but it also creates less uncertainty about whether an OV-fiets will be available to rent. NS is currently working to provide this information to users through journey planning applications.

A relevant field of research for the availability of OV-fiets bicycles is the study of traffic flow. Traffic flow is the research of travellers and the many different interactions that happen during travel times (Treiber & Kesting, 2013). The goal of studying traffic flow is to better understand the transport network and to implement better techniques to make transportation more efficient for travellers, allow travellers to make better route decisions, reduce carbon emissions and allow for better traffic management (Munoz & Laval, 2006). Within traffic flow theory, the main interest are roadways with vehicles such as cars, buses, bicycles, etc. Detection tools fixed on various locations on a roadway commonly measure flow, speed, and density of traffic (Elefteriadou, 2014). Flow refers to the number of vehicles that pass a specific reference point per unit of time, speed refers to the distance covered per unit of time, and density refers to the number of vehicles per unit of length of a roadway (Gerlough & Huber, 1976). For this thesis, the flow of traffic is measured, or in other words, the amount of people that pass through (and rent an OV-fiets bicycle) a specific reference point which is an OV-fiets rental location.

One way to study traffic flow is using prediction models. Predicting traffic flow can be crucial for the improvement of transport systems and making the commuter experience more positive (Elefteriadou, 2014). Accurately forecasting traffic flow can give way for transportation authorities and companies like NS to better allocate their resources, plan for maintenance or construction, and ultimately reduce overcrowding. Insights into future traffic flow may allow travellers to plan their journey leading to better punctuality and customer satisfaction. Furthermore, anticipating potential disruptions in traffic flow allows companies to implement different strategies to overcome any upcoming troubles and allow for a more flowing transportation network (Munoz & Laval, 2006).

Early research of short-term traffic flow from the 1970s used an Auto-Regressive Integrated Moving Average (ARIMA) model to make predictions (Ahmed & Cook, 1979). From then onwards, a wide variety of models have been used and tested to predict short-term traffic flow by different areas of research such as economics, transportation engineering, machine learning or different branches of statistics (Lv et al., 2014). Despite a lot of different models available, the ARIMA is an extremely sound time-series model with a lot of variations available. Variations such as ARIMA with predictor variables, seasonal ARIMA, ARMA, etc. have all been used to predict traffic flow (Kamarianakis & Prastacos, 2003; Williams, 2001; Williams & Hoel, 2003).

One recent study tested the performance of a seasonal ARIMA for short-term traffic flow predictions. The authors argue that a non-seasonal ARIMA is not sufficient for traffic flow prediction (Kumar & Vanajakshi, 2015). Instead, a seasonal ARIMA should be used to reach a high standard for predictions. The seasonal ARIMA model was tested through 24 hours ahead predictions and compared to a historic average and a naive method. It was found that the seasonal ARIMA outperformed both models when predicting short-term traffic flow.

Another study was conducted to predict short-term traffic flow using a seasonal ARIMA model. The authors present multiple arguments and graphs to show that traffic flow data should be modelled using a seasonal ARIMA instead of a non-seasonal ARIMA (Williams & Hoel, 2003). They compared the seasonal ARIMA model to simpler baseline models such as random walk forecast and historical average forecast. One-step ahead predictions of 15 minutes were made, and the results showed that the seasonal ARIMA model outperformed the baseline models. Both studies show that a seasonal ARIMA performs better than simpler baseline models. However, a question remains about the performance of a seasonal ARIMA in comparison to an 'ideal model' or a more robust and solid model than a simple baseline model.

One model that shows robustness and solidity is the Gaussian Process Regression (GPR). It is a flexible machine learning tool that is used for classification, regression, and data modelling (Schulz et al., 2018). It is based on the idea that functions have a normal, or Gaussian distribution. This means that the values of the function at any collection of points also have a Gaussian distribution (Deringer et al., 2021). An example of this is if a point at  $x$  is taken, the

GPR plots out a Gaussian distribution at point  $x$  along the  $y$ -axis to visualise the different possibilities for  $x$ . This allows GPR to capture complex patterns in data and make probabilistic predictions, providing uncertainty estimates for each point on the  $x$ -axis. GPR models are widely used within different fields such as robotics, solving optimization problems or computer visualisation problems, where the ability to handle complex data is extremely useful (Nguyen-Tuong et al., 2009).

GPR has also been tested within the field of traffic flow. One study by Zhao and Sun (2016) looked at predicting traffic flow using a fourth-order Gaussian Process model, which is a variant of the GPR. The model is compared to other models such as a weighted  $k$ -NN and other machine learning models. The proposed variant of the GPR model outperformed most models when compared. Another study looked at a different variant of the GPR for traffic flow prediction and compared it to different models such as Bayesian network, random walk, and ridge regression (Sun & Xu, 2010). It was found that the variant of the GPR model outperformed the other baseline models.

From this, both the ARIMA and GPR approaches are viable options for traffic flow prediction models. A comparison between the two models would be of high interest to see which model outperforms the other when predicting traffic flow data. One study compared the ARIMA model with the GPR model to see which model performs better when making forecasts of one- and two-steps-ahead of 15 minutes per step (Xie et al., 2010). It was found that the GPR model outperformed the ARIMA model in the one-step-ahead forecast and the gap between the models was even larger for the two-step-ahead forecast.

However, questionable choices were made when tuning the parameters of the ARIMA model. Despite stating that the “data set showed seasonal patterns” (Xie et al., 2010, p. 74), the authors decided to use a non-seasonal ARIMA as a result of the *auto.arima* function in R, which optimises the ARIMA parameters by trying every possible combination of parameters and returning the AIC and BIC values for the ARIMA models. While this is a sound method to choose the best fitting parameters, it is important to note that the type of the seasonality, either daily, weekly, monthly, etc., needs to be chosen by hand, and if the choice of the seasonality

does not fit the data well, then the *auto.arima* function may result in a non-seasonal ARIMA. In the end, this choice could have led to the ARIMA model underperforming in comparison to the GPR model.

Multiple previously mentioned studies compared the performance of the seasonal ARIMA and GPR to other baseline models when predicting traffic flow data. One study compared the non-seasonal ARIMA to the GPR (Xie et al., 2010). However, due to the ARIMA model being non-seasonal, an unanswered question remains about the performance of a seasonal ARIMA compared to the GPR. The aim of this study is to compare the performance of the seasonal ARIMA model to the GPR model when predicting traffic flow, more specifically, OV-fiets bicycle availability. I hypothesise that the GPR will outperform the seasonal ARIMA model due to its flexibility and possibility to model complex data.

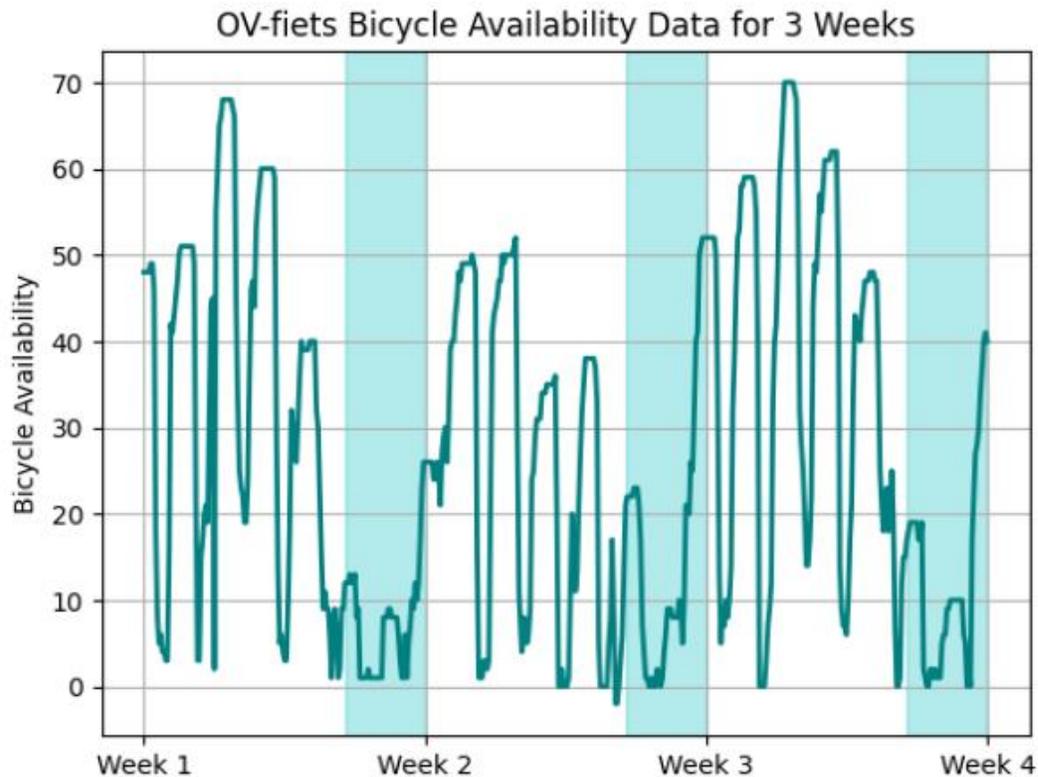
## **2. Methods**

### **2.1. OV-fiets Bicycle Availability Data**

The OV-fiets bicycle availability data consists of information about the amount of OV-fiets bicycles available every minute in all OV-fiets rental locations in the Netherlands between the 24th of January 2023 and the 24th of October 2023. The definition of an available bicycle is that it is present at a rental location and is ready to be rented. An OV-fiets bicycle can be rented for a total of 72 hours (costing 4,55€ for the first 24 hours, afterwards the price increases to 9,55€ per day) before it needs to be brought back to the rental location it was rented from. The data was collected using the OV-fiets Online Service API, which provides the information of the current availability of OV-fiets bicycles in the NS online application used by the public (NS API Management team, n.d.). Every minute, a snapshot of currently available OV-fiets bicycles is taken, processed, and added to the dataset. Data has already been collected and is still being collected every minute for possible future expansions of the data.

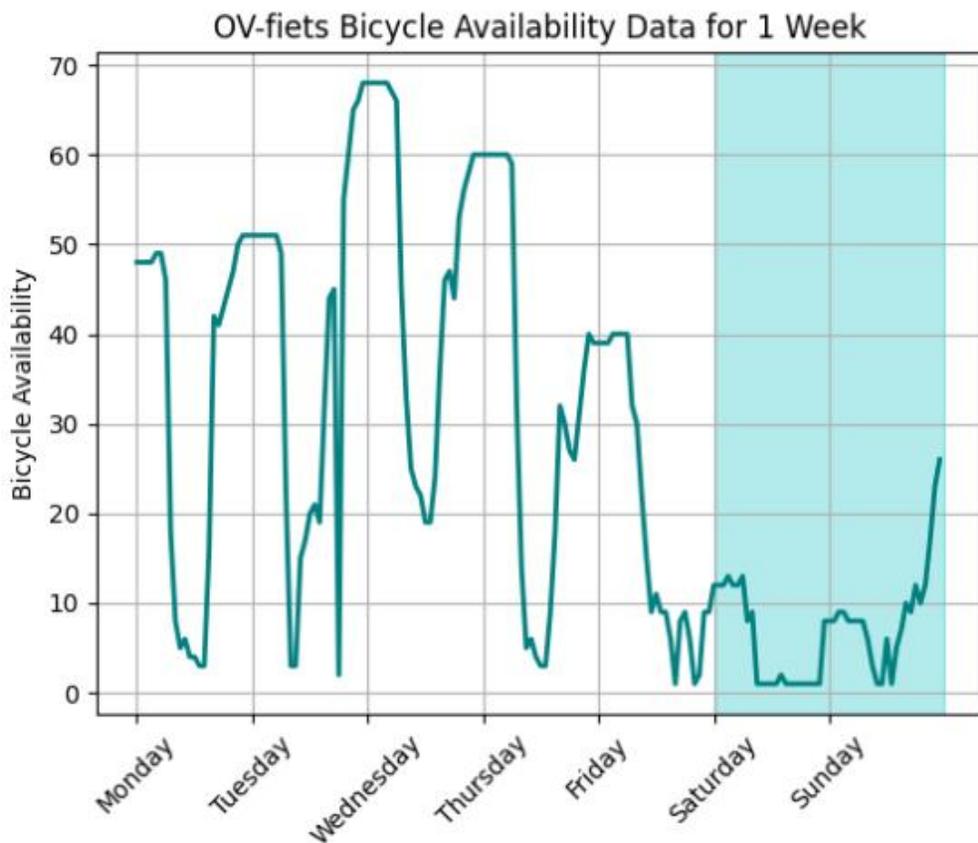
At the time of writing this thesis, there are over 300 OV-fiets rental locations in the Netherlands. Analysing all 300 and more locations would increase the computational load too much, therefore only a select few locations will be analysed. The following stations will be analysed: Amsterdam Amstel, Amsterdam IJzijde, Amsterdam Muiderpoort, Amsterdam Zuidplein, Breda Centrum, Den Haag CS, 's-Hertogenbosch and Utrecht Centraal. These locations are highly visited NS train stations and therefore NS hopes to deploy the models in these stations first.

Furthermore, minute data will be re-coded to hourly data to minimise the time needed to run the analyses. This allows for multiple analyses to be run with fewer computational costs and saves time on computing different methods. However, this is done with a loss of information and possible patterns about bicycle availability between hours, which could lead to better predictions. For example, if most people start renting out bicycles around 7:10AM, this pattern would only be noticed at 8:00AM, when the next data point is inspected by the model. This can be crucial information for the models; however, a greater value was placed on decreased computational load over increased prediction performance of the models.



**Figure 1. OV-fiets Bicycle Availability in Amsterdam Muiderpoort for three weeks. Data ranges from 03/04/2023 to 24/04/2023 starting on a Monday 00:00 and ending on and not including a Monday 00:00. The highlighted areas represent Saturday and Sunday (weekend).**

Figure 1 shows the OV-fiets bicycle availability data in Amsterdam Muiderpoort over a period of three weeks. A steady pattern is observable in the data, where each day the availability goes down to 0 or near 0 between 7:00 AM and 9:00 AM, and then has a large peak back to near full availability between 5:00 PM and 7:00 PM. Figure 2 zooms in on the data by only showing one week. The difference between weekdays and weekends becomes more visible. During weekends (highlighted) the pattern changes slightly with availability being low during the whole day. One reason for this pattern is that people commute to work on workdays, which is why, during peak hours, availability drops to near zero, and then after work hours, the availability increases again when people return their bicycles. During weekends, a slightly different pattern is seen, possibly due to people using OV-fiets bicycles for leisure day trips rather than work-related trips. This leads to bicycles being mostly unavailable during weekends.



**Figure 2. OV-fiets Bicycle Availability in Amsterdam Muiderpoort for one week. Data ranges from 03/04/2023 to 10/04/2023 starting on a Monday 00:00 and ending on and not including a Monday 00:00. The highlighted area represents Saturday and Sunday (weekend).**

## 2.2. Seasonal ARIMA Model

### 2.2.1. Non-Seasonal Orders

The Autoregressive Integrated Moving Average model, or ARIMA for short, is a versatile and easy to implement family of models, often used for time-series data analysis. It looks at past observations and errors to predict the next observation and its error (Shumway et al., 2017). One advantage of an ARIMA model is the ability to adapt it to various complex data. Non-linearity, seasonality or varying volatility are all possible to model with an ARIMA model.

A simple non-seasonal ARIMA model is composed of the Autoregressive model (AR), an Integrated part (I) and the Moving Average model (MA). The AR( $p$ ) model looks at the  $p$  past observations and based on these, predicts the next observation in the time series. It is defined as

$$x_t = \mu + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_p x_{t-p} + \epsilon_t,$$

where  $x_t$  is an observation of a time series at time point  $t$ ,  $\mu$  is the mean of the time series,  $\beta_1, \beta_2, \dots, \beta_p$  are the coefficients,  $\epsilon_t$  is the error at time point  $t$ , and  $p$  is the order of the AR model. If the value for  $p$  is 1, then it will look at one past observation to predict  $x_t$ , if the value for  $p$  is 2, then it will look at two past observations to find  $x_t$ .

The I( $d$ ) part of ARIMA represents the differencing part of order  $d$ . If there is data that is non-stationary around the mean and variance, it is necessary to take the difference of the data to reach stationarity. This is defined as

$$X'_t = X_t - X_{t-1},$$

where  $X'_t$  is a differenced time series. The order  $d$  (not used in the equation) represents the number of times the data should be differenced to reach stationarity. In the cases where a linear trend is present in a time series, one order of differencing makes data mostly stationary around the mean (Weisang & Awazu, 2008). Making data stationary is an extremely important step of ARIMA modelling. Stationary data is more predictable and easier to model because it no longer

has effects of a trend or seasonality. This way, a model can make better and more accurate forecasts (Cheng et al., 2015).

Lastly, the MA( $q$ ) part, which looks at the  $q$  number of past errors and based on these errors, predicts the next observation in a time series. It is defined as

$$x_t = \mu + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t,$$

where  $x_t$  is an observation of a time series at time point  $t$ ,  $\mu$  is the mean of the time series,  $\theta_1, \theta_2, \dots, \theta_p$  are the coefficients,  $\epsilon_t$  is the error at time point  $t$ , and  $q$  is the order of the MA model. If the value for  $q$  is 1, then it will look at one past error to find  $x_t$ , if the value for  $q$  is 2, then it will look at two past errors to find  $x_t$ . The combination of AR( $p$ ), I( $d$ ) and MA( $q$ ) compose an ARIMA model. The full model first differences the data and then applies the AR( $p$ ) and MA( $q$ ) parts, which can be seen as

$$x_t = \mu + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_p x_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t.$$

Simplified, it is just an AR( $p$ ) model added with a MA( $q$ ) model performed on differenced data.

### 2.2.2. Seasonal Orders

In the case that the OV-fiets bicycle availability data contains seasonal effects, there would be an interest in modelling a seasonal ARIMA. Besides the non-seasonal parameters, it also contains an extra four parameters  $P, D, Q, s$ , one of each for seasonal AR, seasonal I, seasonal MA, and the periodicity, respectively. Firstly, a choice of the periodicity/seasonality  $s$  is required. If the data displays yearly seasonality, then a periodicity order of  $s=12$  would be used in the case of a monthly time series data. Continuing with this example, the seasonal AR( $P$ ) and the seasonal MA( $Q$ ) of  $P, Q$  order 1 would look at the observations and errors at time  $t - s - 1$ . In basic terms, it looks at the values of observations and errors one periodicity in the past plus the value of  $P$  (for AR) or  $Q$  (for MA) in the past to predict the next observation.

The seasonal I( $D$ ) part of ARIMA represents the differencing part of order  $D$ . In the case where data has a seasonal effect, it means the data is not stationary around the mean and variance. To make data stationary when there are seasonal effects, the seasonal differencing part of seasonal ARIMA would be implemented. This is defined as

$$X'_t = X_t - X_{t-s},$$

where  $X'_t$  is a differenced time series. The order  $D$  (not used in the equation) represents the number of times the data should be seasonally differenced to reach stationarity. However, it should never be the case where data is seasonally differenced more than once within the seasonal ARIMA (Nau, 2016). This is not the same with non-seasonal  $I(d)$  and it is also possible to have both a non-seasonal differencing and a seasonal differencing. In the end a specification of a seasonal ARIMA  $(p, d, q) (P, D, Q, s)$  model with 7 parameters is required.

### 2.2.3. Wold Decomposition

To minimise the computational power needed to estimate the parameters of an ARIMA model, different theories and analyses can be used to predetermine certain parameters. One theorem that helps determining the  $AR(p)$  and  $MA(q)$  parts of ARIMA is the Wold decomposition. The Wold decomposition theorem states that if a time series  $\{X_t\}$  is stationary around the mean and variance, then

$$X_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} + \eta_t,$$

where  $\varepsilon_{t-j}$  is a white noise process,  $b_j$  is a weight coefficient and  $\eta_t$  is a deterministic time series. In other words, it can be separated into a stochastic series and a deterministic series. The stochastic series can be regarded as a moving average, while the deterministic series can be seen as a sine wave (Williams & Hoel, 2003). The main interest is in the stochastic part because if the time series data is stationary, a moving average model can be applied to make forecasts. If the differencing parameters of the ARIMA is set in a way to make data stationary, a moving average can be modelled. Therefore, the hyperparameter tuning of the (S)AR and (S)I can be ignored, and only tuning the (S)MA hyperparameters is required.

## 2.3. Gaussian Process Regression

### 2.3.1. Bayesian Approach and Introduction of Gaussian Process Regression

Gaussian Process Regression (GPR) is a Bayesian approach to modelling that works based on a mean function and a covariance (kernel) function. One advantage of Gaussian process regression

is the ability to choose and derive your own kernel functions (Wang, 2020). Finding the right kernel function can be quite challenging, however, this freedom gives way to adapt the model to the nuances and peculiarities of the OV-fiets bicycle availability data. A disadvantage of the GPR is that it can take a lot of computational power to run it on a large data set (Corani et al., 2021). This makes the model run very slowly and it can take a significant amount of time to train the model.

GPR is a Bayesian approach to regression that works by inferring a probability distribution over all possible values in the data. Let's assume the data is defined by a linear function:  $y = wx + \epsilon$ . Firstly, the Bayesian approach works by specifying a prior distribution,  $p(w)$ , and then recalculating probabilities based on the observed data. This works through the usage of the Bayes' Theorem, which updates the predicted probabilities of an event by incorporating new information and generating a posterior probability. The same principal works in GPR to compute a posterior distribution and is defined as

$$p(w|y, X) = \frac{p(y|X, w)p(w)}{p(y|X)},$$

$$\text{posterior distribution} = \frac{\text{likelihood} * \text{prior}}{\text{marginal likelihood}}.$$

The posterior distribution  $p(w|y, X)$  is calculated by incorporating both the prior and the observations from a dataset. In the case of predicting new unseen observation,  $x^*$ , within the dataset, a calculation of the predictive distribution is done by finding a function of all possible predictions and multiplying it by its calculated posterior distribution. This is defined as

$$p(f^*|x^*, y, X) = \int_w p(f^*|x^*, w)p(w|y, X)dw,$$

where  $f^*$  is the predicted function and  $x^*$  is the predicted test value. The prior and likelihood are assumed to have a normal distribution (Gaussian). With this assumption, when solving for the predictive distribution, the end product is a Gaussian distribution of all possible values for  $x^*$ .

The mean is the point estimate of  $x^*$  and the variance is the uncertainty measurement.

The Gaussian process regression is a non-parametric model, so instead of calculating the probability of all possible parameters of a specific function, it calculates the probability distribution of all possible functions that fit the data. Specify a prior function, calculate the

posterior distribution using training data and then compute the predictive posterior distribution to make predictions. The Gaussian process prior is defined as

$$f(x) \sim GP(m(x), k(x, x')),$$

where  $m(x)$  is the mean function that is usually set to be equal to 0, and  $k(x, x')$  is the covariance function also known as the kernel function because it tends to consist of a kernel or a combination of multiple kernels.

### 2.3.2. Kernel Function

Choosing an appropriate kernel function comes down to determining an appropriate linear combination of candidate kernel functions and tuning the hyperparameters of said kernel functions to fit the data. Three different kernel functions will be analysed: radial basis function (RBF) kernel, periodic kernel and the linear kernel. One of the most widely used kernel functions in GPR is the RBF also known as the Squared Exponential (SE) (Rasmussen & Williams, 2006). It is defined as

$$k_{RBF}(x, x') = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\ell^2}\right),$$

where  $\ell$  is the length-scale or the length in width of the ‘wiggles’ in a function and  $\sigma^2$  is the average distance of a function away from the mean, it can also be seen as the scale factor.

One observation made about OV-fiets bicycle availability data is the presence of seasonal effects, either daily or weekly. For this purpose, implementing a kernel that measures seasonality is of interest. One such kernel is the periodic kernel (Rasmussen & Williams, 2006) which is defined as

$$k_{per}(x, x') = \sigma^2 \exp\left(-\frac{2\sin^2(\pi|x-x'|/p)}{\ell^2}\right),$$

where  $\ell$  is the length-scale,  $p$  is the periodicity that measures the distance between each period and  $\sigma^2$  is the scale factor. With this kernel a choice of daily or weekly periodicity can be made.

It is also possible to combine kernels either through addition or multiplication, which yields different properties. A kernel that is often combined with other kernels is the linear kernel defined as

$$k_{Lin}(x, x') = \sigma^2 xx',$$

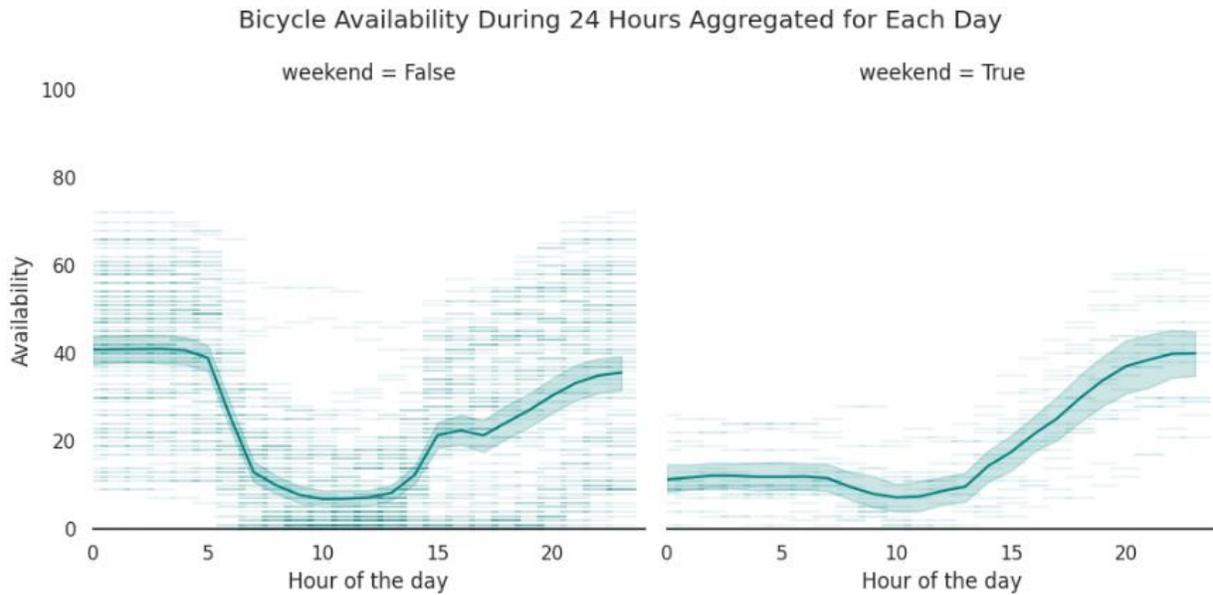
where  $\sigma^2$  is the scale factor which controls the vertical length-scale of the function.

The kernels will be tested in different combinations to see which ones predict OV-fiets bicycle availability best. Further explanations of the kernel selection process will be explained in section 2.5. about cross-validation.

## **2.4. Stationarity and Seasonality**

### **2.4.1. Inspection of Data**

As seen in Section 2.2.3., having data that is stationary around the mean and variance is of key importance for the Wold decomposition. If the data is not stationary, there could be a seasonal pattern present within the data. There are many different types of seasonality, such as daily, weekly, monthly, etc. Some data might even have multiple occurrences of seasonality. To detect whether OV-fiets bicycle availability data has one or more seasonal patterns, a visual inspection of data can be required. Oftentimes, seasonal patterns can be obvious from first glance. Upon inspection of OV-fiets bicycle availability in the rental location Amsterdam Muiderpoort in Figure 3, on every weekday there is a peak in bicycle availability and then a dip close to zero. During weekends, a slightly different pattern is present where availability of bicycles stays close to zero for most of the day. This presents a possibility for a weekly seasonal pattern in the data. It is also likely that there is a daily seasonal effect, however due to the differences between weekdays and weekends, this effect should be weaker than the weekly seasonality. Besides looking at data, correlational analyses can be made to determine whether the data is stationary.

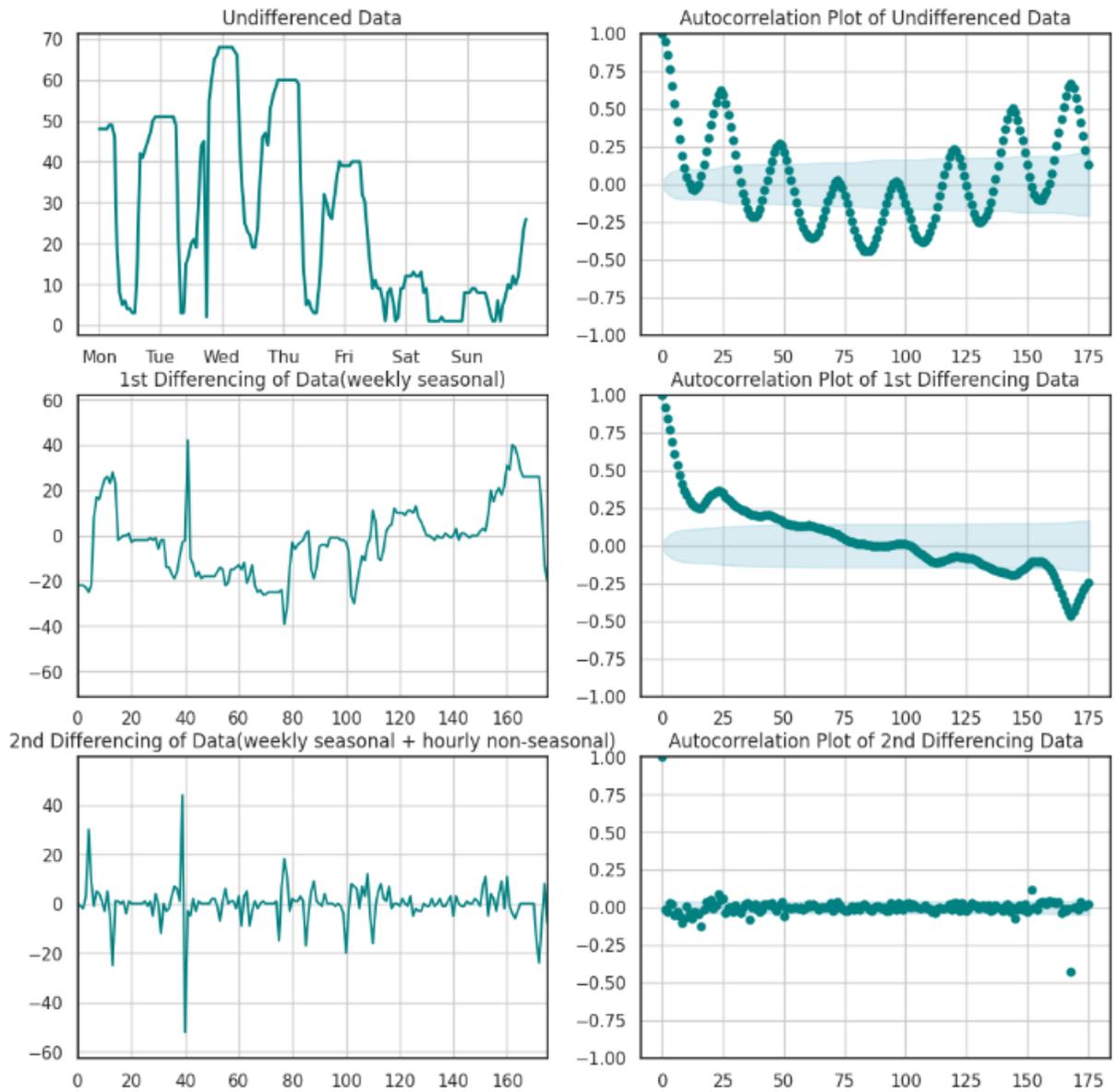


**Figure 3. Bicycle availability aggregated by every hour of the day over the period of 4 months from March to August. Workdays are on the right, and weekends are on the left. Darker colour indicates more data points, lighter colour indicates fewer data points. The straight line represents the mean availability.**

## 2.4.2. Autocorrelation Analyses

Inspections of autocorrelations may give a better insight into the stationarity of data (Perraudin & Vandergheynst, 2017). The autocorrelation plot is computed by finding the correlations between a data point at time  $t$  and the lagged data point at time  $t-1$ , then  $t-2$ ,  $t-3$  until  $t-n$ . These correlations are then plotted on a graph which gives an overview of the randomness within the data. Ideally, the data would spread out randomly around 0, which would indicate that data is stationary (Cressie, 1988; Perraudin & Vandergheynst, 2017). Figure 4 shows six different plots. The left three graphs show OV-fiets bicycle availability data at Amsterdam Muiderpoort plotted undifferenced and then with a 1st order differencing and then a 2nd order differencing. The right three graphs see the autocorrelation plots of said undifferenced/differenced data. Inspecting the autocorrelation plot 4b show that every 24 lags, the correlation reaches a peak and the largest peak is seen after 168 lags, more specifically, after 1 week. This means there is similarity between data points every 24 hours and every 168 hours (1 week). This indicates that there is

both a daily and a weekly seasonality present in the data, however, the weekly seasonality seems stronger than the daily seasonality.



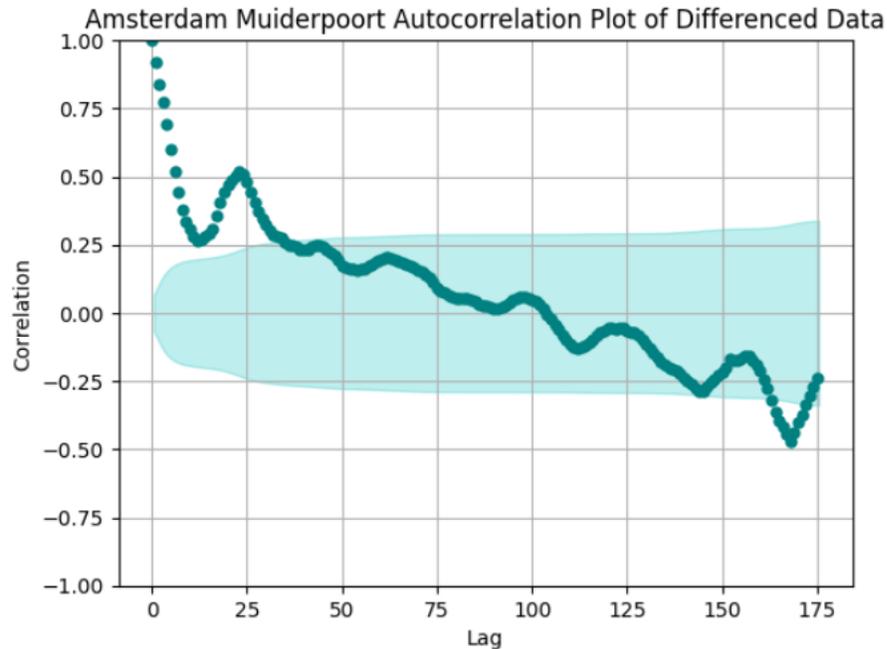
**Figure 4. 4a (top left) Undifferenced data for 1 week. 4b (top right) Autocorrelation plot of undifferenced data. 4c (middle left) 1st weekly seasonal differencing. 4d (middle right) Autocorrelation plot of 1st differencing data. 4e (bottom left) 2nd hourly non-seasonal differencing after seasonal differencing. 4f (bottom right) Autocorrelation plot of 2nd differencing data. X-axis represents the hour.**

Inspection of plot 4b gives a better insight of the seasonal differencing order for the ARIMA that should be chosen to achieve stationarity. With 2 occurrences of seasonality present, a choice was made to go with the seasonal differencing order of 168 lags, which composes the larger, weekly

seasonality. This was chosen over daily seasonality due to plot 4b, which showcased a stronger weekly seasonality than daily seasonality. Plot 4c shows the data after differencing the data shown in 4a by an order of 168 lags, and plot 4d shows the autocorrelation plot of the 1st differenced data. While the autocorrelation plot looks better after differencing, it is not entirely stationary. There is a linear trend present, where the correlation starts off very high and then steadily decreases into the minus with a peak at the 168th lag.

A similar plot to 4d was found by Williams and Hoel (2003) when they observed the autocorrelation plot after weekly differencing traffic flow data. In their case, the data they observed was the week with Christmas and New Year's holidays, which caused unusual patterns in the data. After they observed the autocorrelation plot at a different time, the data was stationary after weekly differencing. In this study, after observing a different time as seen in Figure 5, the data after weekly differencing stayed non-stationary with the same linear trend as seen in plot 4d. This meant that another differencing order is needed to reach stationarity. However, it should be noted that performing 2 or more seasonal orders should never be done as it will not reach stationary data (Nau, 2016). Instead, a non-seasonal differencing order should be implemented along the seasonal differencing order. When a linear trend is present in the autocorrelation graph, typically, a first order differencing is enough to make the data stationary (Weisang & Awazu, 2008).

Plot 4e shows what the data looks like after double differencing. There is a big sharp peak in 4e, this could be due to an outlier which is also seen in 4a. Inspecting the autocorrelation data in 4f, after a 1-lag difference as well as a seasonal differencing order, the data is now mostly stationary. Therefore, a seasonal ARIMA with a differencing order,  $d=1$ , and a seasonal differencing order,  $D=1$ , should make OV-fiets bicycle availability data mostly stationary, and a moving average can be modelled as was seen with the Wold decomposition theorem. Following this reasoning, for the seasonal ARIMA, after seasonal and non-seasonal differencing, only the MA and SMA parts will be modelled without the AR or SAR parts. This is a common model often used for time series analysis (Nau, 2016).



**Figure 5. Autocorrelation plot for 168-hour (1 week) lagged difference of the OV-fiets rental location Amsterdam Muiderpoort. Data ranges from 10/04/2023 to 17/04/2023**

## 2.5. Cross-Validation Methods

Cross-validation is one of the most widely used methods to tune model parameters and estimate prediction error (Berrar, 2019). For the models, cross-validation will be used to tune hyperparameters and determine model performance.

### 2.5.1. k-fold Cross-Validation

Cross-validation works by splitting data into training and testing sets. The model is built using the training set and then tested against the test set to evaluate the performance of the model. Different model parameters are tested to see which parameter combination performs best. One common cross-validation method is k-fold cross-validation. It works by splitting the data into  $k$  ‘folds’ or subsets. The training set is composed of  $k-1$  subsets and the remaining 1 fold acts as a validation set, which is used to test model performance. The model is built with the training set and tested against the validation set. This process is repeated until each fold is used as a

validation set. The final model performance is the average of all model performances achieved across all validation sets.

### **2.5.2. Rolling-Origin Cross-Validation**

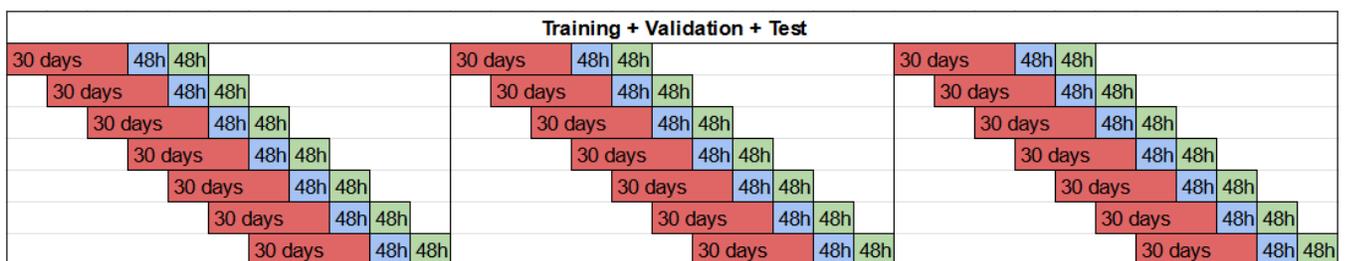
Due to working with time-series data, a typical k-fold cross-validation is not possible. This is because the data is not statistically independent, therefore, values in the past can predict values in the future, or in other words, future data is dependent on past data. Multiple different methods have been proposed to perform cross-validation with time-series data, such as fixed-origin, rolling-origin, rolling-window, etc. (Bergmeir & Benítez, 2012; Tashman, 2000). Typically, researchers would use 10-folds, and subsequently assign each fold to be a test set, while the rest of the data is trained on. With time-series methods, the test set is only the last block and subsequently, new data is added to the end of the previous data to act as the new test set (Bergmeir & Benítez, 2012). Each method for time series cross-validation is a viable option with only slight differences.

For this study, rolling-window cross-validation is used. The key point of this method is to keep the amount of data in the training set constant. With each new fold or loop, as new data is added to act as the test set, the old validation set becomes part of the training set and the beginning of the training data is discarded. This is done multiple times over multiple loops. The training set consists of 30 days, with each day being 24 hours long, a validation set consists of 48 hours, and a test set of 48 hours for fitting the model, hyperparameter tuning and model evaluation, respectively. Afterwards, the next loop will be performed, where the last test set becomes the new validation set, the last validation set becomes part of the training set, the beginning 48 hours of the previous training set are discarded, and new data is added to act as the new test set. This process loops over 7 times, to account for the 7 days of the week.

The prediction periods are intervals of 48 hours. It is natural to assume that predictions furthest away from the training set will be worse than predictions closest to the training set. Furthermore, there can exist differences between different days. This was seen in Figure 3 when comparing availability data during weekdays and weekends. For these reasons, the loop is repeated 7 times,

sliding the test set by 48 hours. This way, the origin of the prediction period will start on every day of the week over the span of 2 weeks. For example, if the origin of the test set is Monday, during the next loop, the origin will be Wednesday, then Friday, etc. until the origin ends with and not including a Monday.

After the 7-day loop, a new sample of the data is taken starting at the end of the last test set and run the 7-day loop once again. In such a way, a total of 3 samples of data will be run through the 7-day loop as can be seen in Figure 6. This is done to consider different timepoints of the year and avoid overfitting the model to a certain time of the year. A similar method of cross-validation was previously done by a group of student interns at NS during the summer of 2023. One key difference they had in their cross-validation method is instead of having a 7-day loop, they performed one loop, then took a new sample of data 48 days ahead to perform the second loop and then repeated this once more for a total of 3 times. This leads to missing out on information about the model’s performance on certain days of the week. If the model performs worse on weekends than on weekdays, this information could be overlooked and will not be observed by the model evaluation metrics. To account for that, the cross-validation with a 7-day loop at different times of the year will take this into consideration and improve on the previously used method for cross-validation. This whole process will be repeated with all 8 OV-fiets rental locations.



**Figure 6. Rolling-Origin Cross-Validation where training set is displayed in red, validation set is displayed in blue and testing set is displayed in green**

One small adjustment will be made for the GPR model. There is no need for hyperparameter tuning via cross-validation within GPR because by default the hyperparameters are tuned via model evidence maximisation using only the training set by the python library *sklearn* which is used to run the GPR model. Instead of hyperparameter tuning, cross-validation will be used to

pick the best kernel combination. Out of all the individual kernel functions, every possible combination of kernel functions will be tested on the validation set. The best performing combination will then be selected and used with the test set.

### **2.5.3. Model Evaluation Metrics**

After obtaining model predictions, computation of prediction errors to evaluate model performance follows. Two model evaluation metrics will be used to compare the predictive performance of the models: root mean square error (RMSE) and mean absolute error (MAE). A larger focus is placed on the MAE, as it is easier to interpret and understand than other error measures, especially from a business point of view like NS (Hewamalage et al., 2023). Furthermore, MAE has already been used before at NS for similar projects using OV-fiets bicycle availability data, which gives possibility for future comparisons between the results found in this thesis and previous work done by NS. For these reasons, MAE will be used as an error measure for hyperparameter tuning as well as model evaluation. Only the final models will have both RMSE and MAE model evaluation metrics, which have been found to be a good combination reflecting robustness, and generalisability of the overall models (Hewamalage et al., 2023).

## **2.6. Baseline Models**

Additionally, 2 different models will be analysed in this study to act as baseline models. One model is a simple ARIMA with  $(0,0,1) (0,1,1,24)$  order. It has a seasonal differencing order  $s=24$ , to model for daily seasonality. The second model predicts the mean for the whole testing period of 48 hours. Both models are trained using the cross-validation method mentioned previously. Using these two models, a baseline is given for the predictions with the simplest of models. This way, when comparing the seasonal ARIMA to the GPR, it is also possible to see whether these models are performing well in comparison to simple baseline models.

### 3. Results

**Table 1.**

*Forecast Performance Comparison for Different Locations*

Model	MAE	RMSE
<i>Amsterdam Amstel</i>		
Seasonal ARIMA	105.41	114.48
GPR	62.97	76.55
Simple ARIMA	<b>41.78</b>	<b>51.12</b>
Mean Baseline	52.5	59.36
<i>Amsterdam IJzijde</i>		
Seasonal ARIMA	169.36	189.96
GPR	118.98	142.55
Simple ARIMA	<b>85.79</b>	<b>105.38</b>
Mean Baseline	101.87	113.41
<i>Amsterdam Muiderpoort</i>		
Seasonal ARIMA	30.62	33.36
GPR	20.23	24.77
Simple ARIMA	<b>14.25</b>	<b>17.54</b>
Mean Baseline	16.28	18.37
<i>Amsterdam Zuid Zuidplein</i>		
Seasonal ARIMA	69.05	75.33
GPR	46.12	56.96
Simple ARIMA	<b>27.45</b>	<b>34.95</b>
Mean Baseline	38.74	43.25
<i>Breda Centrum</i>		
Seasonal ARIMA	93.44	99.97
GPR	49.98	60.01
Simple ARIMA	<b>41.95</b>	50.12
Mean Baseline	42.47	<b>47.91</b>
<i>Den Haag CS</i>		
Seasonal ARIMA	37.52	43.29
GPR	21.91	26.75
Simple ARIMA	26.15	29.89
Mean Baseline	<b>18.65</b>	<b>21.47</b>
<i>'s-Hertogenbosch</i>		
Seasonal ARIMA	80.97	94.31
GPR	75.48	93.03
Simple ARIMA	<b>46.7</b>	<b>55.66</b>
Mean Baseline	61.05	71.39
<i>Utrecht Jaarbeursplein</i>		
Seasonal ARIMA	359.15	385.53
GPR	176.6	211.78
Simple ARIMA	<b>122.43</b>	<b>146.35</b>
Mean Baseline	153.51	172.3

*Note:* ARIMA = Autoregressive Integrated Moving Average. GPR = Gaussian Process Regression.

MAE = Mean Absolute Error. RMSE = Root Mean Squared Error. Numbers in bold indicate the lowest error value within that location.

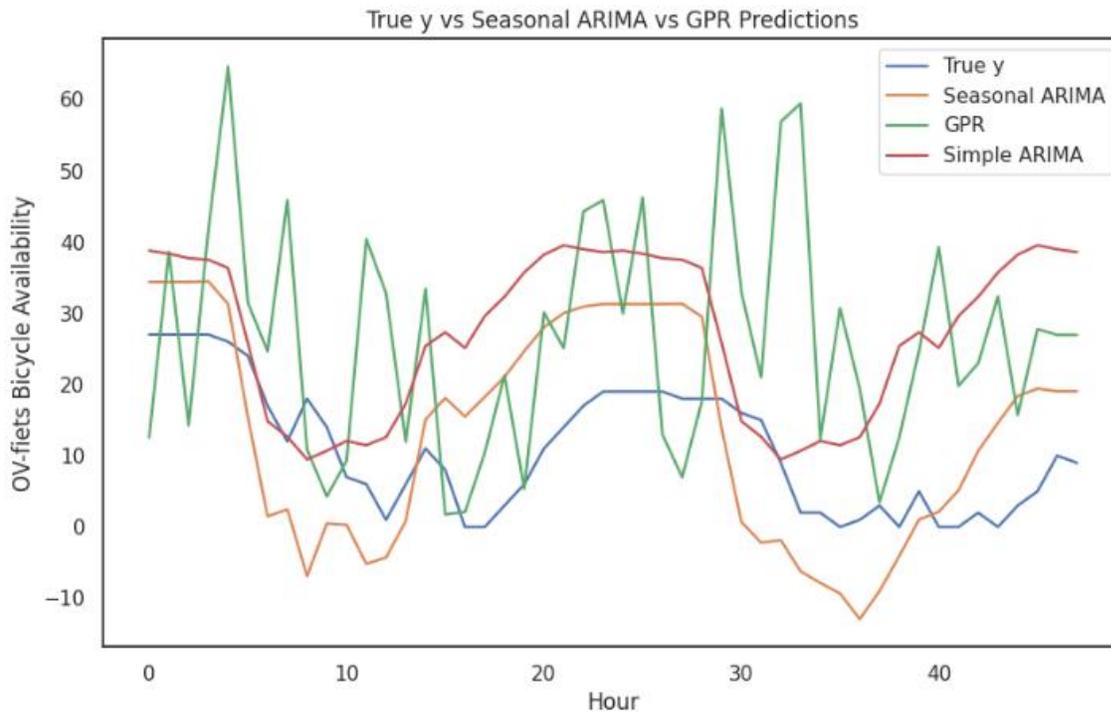
Table 1 shows the results of the seasonal ARIMA, simple ARIMA, GPR and the Mean baseline models on all the OV-fiets bicycle rental locations that were tested. The average MAE and RMSE values were calculated using the values of MAE and RMSE from each test set that was described in the method section about rolling-origin cross-validation (21 in total). Before analysing the results, it is important to note that the values of MAE and RMSE are not comparable between locations. This is because for locations where the total number of bicycles is larger, the MAE and RMSE is going to be larger, and for locations where the total number of bicycles is fewer, the MAE and RMSE is going to be lower. For this reason, a comparison can only be made for MAE and RMSE within locations between models. Note that lower MAE and RMSE values indicate better predictive performance of the models.

Looking at all values within different locations, there is a similar trend. The MAE and RMSE for the seasonal ARIMA model are consistently larger for all locations than the MAE and RMSE of the GPR. This means that across all locations, the GPR has outperformed the seasonal ARIMA model when predicting OV-fiets bicycle availability. Furthermore, calculating the average percent difference between the error estimates of the two models shows that on average, the MAE of the GPR model is 35.35% lower than the MAE of the seasonal ARIMA model. On average, the RMSE of the GPR model is 29.10% lower than the RMSE of the seasonal ARIMA model.

Looking at the baseline simple ARIMA, it was found that across all except one location, the baseline model has outperformed the GPR in terms of MAE and RMSE. The only location where the GPR outperformed the simple ARIMA was in Den Haag CS. Otherwise, the simple ARIMA has shown consistently better results than either model. Calculating the average percent difference between the error estimates of the simple ARIMA and the GPR shows that on average, the MAE of the simple ARIMA is 24.63% lower than the MAE of the GPR model. In terms of RMSE, the error of the simple ARIMA is 25.36% lower than the GPR. Furthermore, on average, the MAE of the simple ARIMA is 52.13% lower than the MAE of the seasonal

ARIMA. Lastly, the RMSE of the simple ARIMA is 48.09% lower than the RMSE of the seasonal ARIMA.

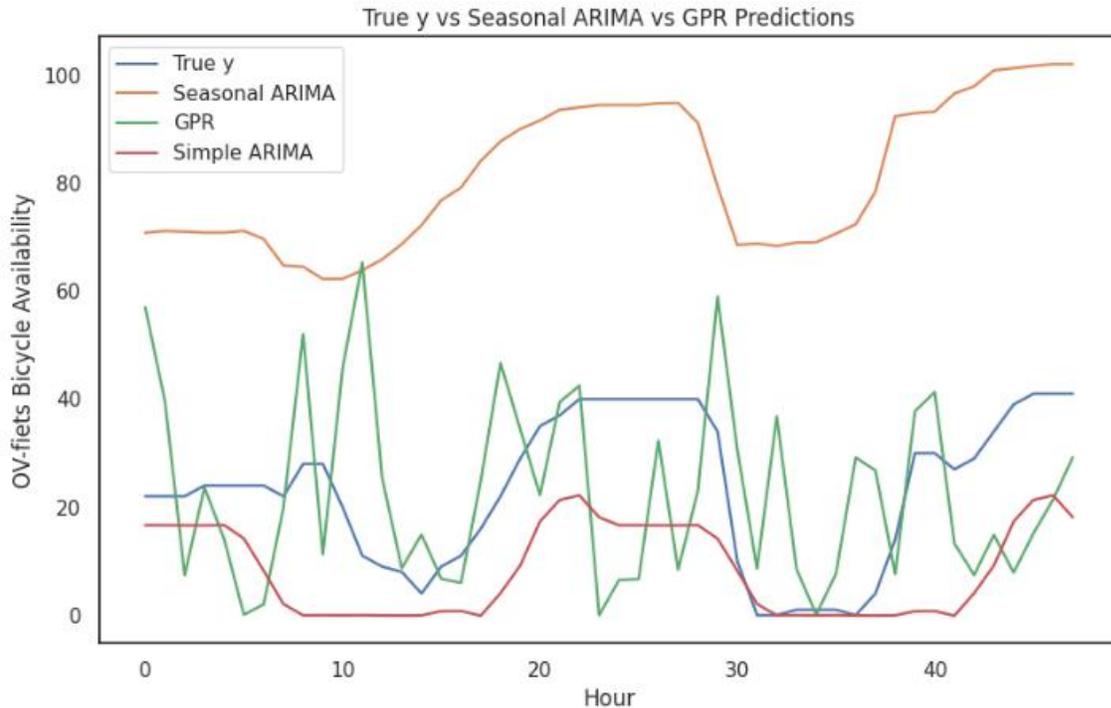
Looking at the mean baseline model, across all locations, it outperforms both the seasonal ARIMA and the GPR in terms of MAE and RMSE. Only in 2 different locations does the mean baseline outperform the simple ARIMA. In Den Haag CS, the mean baseline model outperforms all other models in terms of MAE and RMSE. In Breda Centrum, the mean baseline model outperforms all other models in terms of RMSE only.



**Figure 7. Predictions of OV-fiets bicycle availability for 48 hours. Comparing the seasonal ARIMA, simple ARIMA and GPR to the true availability of OV-fiets bicycles.**

Upon inspection of Figure 7, contrary to the results, the seasonal ARIMA seems to be predicting OV-fiets bicycle availability much better than the GPR. The shape of the seasonal ARIMA predictions matches quite well with the actual availability values. While the shape of the GPR predictions seems to be random and centred around the mean, Figure 8 shows a different type of

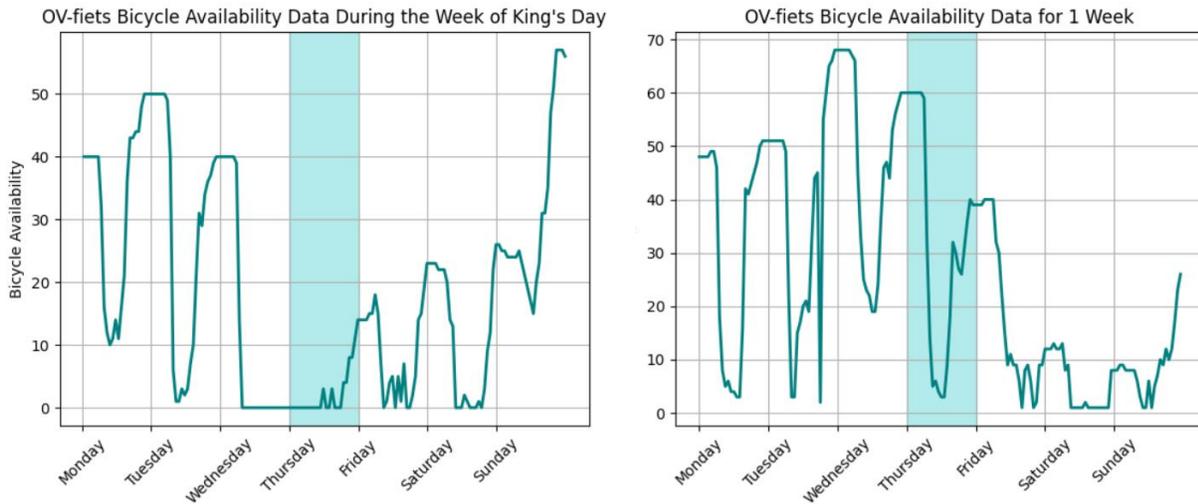
story. The seasonal ARIMA is predicting way off target, while the GPR predictions stay quite random around the mean. Furthermore, the simple ARIMA still predicts quite well, as was seen in the results table.



**Figure 8. Predictions of OV-fiets bicycle availability for 48 hours. Comparing the seasonal ARIMA, simple ARIMA and GPR to the true availability of OV-fiets bicycles. Taken specifically at a time when the seasonal ARIMA predictions are significantly worse than actual OV-fiets availability.**

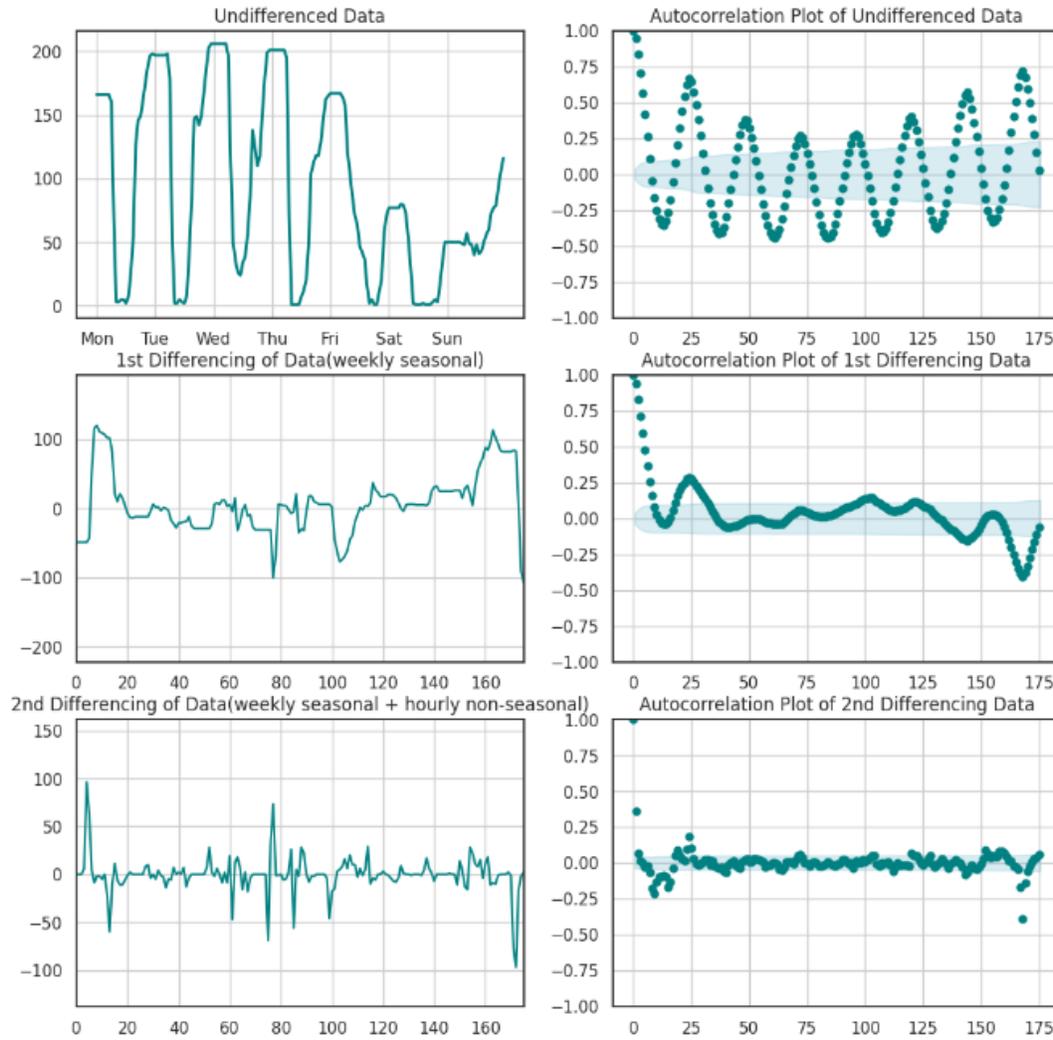
This pattern was present in all locations and at different time points. The seasonal ARIMA would sometimes make very accurate predictions and other times the predictions would be very far from actual data. The GPR also consistently had random predictions around the mean in all locations. The predictions were never as accurate as some of the seasonal ARIMA predictions, however, they were also never as inaccurate as some of the seasonal ARIMA predictions. Thus, while the results in Table 2 indicate that GPR performs better than the seasonal ARIMA, inspecting Figure 7 and Figure 8, makes it less clear which model is better at predicting OV-fiets bicycle availability overall.

After looking at the results, further new analyses were made with the data to help understand why the seasonal ARIMA is underperforming compared to the GPR and the simple ARIMA. The following figures were made with the idea in mind that the seasonal ARIMA model was overfitting to the data, which led to unsatisfactory predictions at different times and locations.



**Figure 9. OV-fiets bicycle availability on the week during King’s Day (left) compared to one week before (right). King’s day (Thursday) is highlighted on the left graph, while a typical Thursday is highlighted on the right graph.**

Figure 9 shows a comparison of OV-fiets bicycle availability during the week of King’s Day to the week before it. A big difference is seen in bicycle availability between the two weeks. On the day before King’s Day and on King’s Day, the availability goes down to 0 bicycles available. In comparison, the week before showed a more stable pattern with bicycle availability decreasing in peak hours and increasing during non-peak hours. These types of fluctuations in data could lead to improper fitting of the models to the data, which could lead to worse predictions than expected.



**Figure 10. Data of Amsterdam Amstel. 4a (top left) Undifferenced data for 1 week. 4b (top right) Autocorrelation plot of undifferenced data. 4c (middle left) 1st weekly seasonal differencing. 4d (middle right) Autocorrelation plot of 1st differencing data. 4e (bottom left) 2nd hourly non-seasonal differencing after seasonal differencing. 4f (bottom right) Autocorrelation plot of 2nd differencing data. X-axis represents the hour.**

Furthermore, it is possible that for different locations, different hyperparameters fit the data better. Looking at Figure 10, by the first differencing, the data is already mostly stationary around the mean. Visually, it looks very similar to the autocorrelation plot after seasonally differencing that was found by Williams and Hoel (2003). For this reason, the authors decided to choose one order of differencing as opposed to the choice of weekly and daily differencing. In this study, the data of Amsterdam Muiderpoort was used to find the optimal differencing

hyperparameters for the seasonal ARIMA. When running the model for other locations, the differencing hyperparameter was not readjusted for the new data. It is highly likely that the model overfits to the sample of data used for model fitting, therefore, it does not transpose well to data on other days or other locations. This could be a big reason why the predictions of the seasonal ARIMA are sometimes far from actual bicycle availability as was seen in Figure 8.

## **4. Discussion**

The following section focuses on the research question and discussion of the results of this thesis. Furthermore, limitations of the thesis will be mentioned and recommendations for future research will be given.

### **4.1. Interpretation of Results**

The purpose of this thesis was to evaluate which of the two models, seasonal ARIMA and GPR, would predict OV-fiets bicycle availability better. It was hypothesised that the GPR would have better predictions due to its flexibility to adapt to different types of data. The results showed that between all locations that were tested, the GPR outperformed the seasonal ARIMA in both model evaluation metrics MAE and RMSE. However, upon further inspection of the visualisations of the predictions, some days, the seasonal ARIMA would predict with a lot of accuracy and on other days it would have extremely poor predictions, while the GPR mainly had random predictions around the mean.

Ideally, the predictions would closely resemble the actual OV-fiets bicycle availability data. The results show that the seasonal ARIMA can predict with high accuracy, but falls short on other days. The reason for this shortcoming of the seasonal ARIMA comes down to the choice of hyperparameters. Despite making autocorrelation plots to find the best fitting hyperparameters, it is very likely that during some days or weeks, the data acts differently from the data that was inspected using the autocorrelation plots. Which could mean that a different set of

hyperparameters is needed to make accurate predictions using the seasonal ARIMA for different times of the year.

One such day when data acts completely different from usual is on April 27th, which is known as King's Day in the Netherlands because it is the day on which the King of the Netherlands was born. It is a national holiday in the Netherlands and people celebrate it by dressing up in orange colours, going out to parties, enjoying different fairs, flea markets, or just being outside and socialising with others (Editors Holland.com, 2023, November 8). When inspecting OV-fiets bicycle availability on King's Day as opposed to the week before it, there is a big shortage of OV-fiets on the day before King's Day and on King's Day. This could be seen as an outlier, nevertheless, the week of King's Day was used for model training. These types of inconsistencies in data could lead to worse predictions than expected.

Another reason which could have led to a bias in the predictions is over-differencing. One study showed that while differencing could be a good way to make data stationary, it could also lead to over-differencing (Hossain et al., 2019). The authors differenced their data twice and found that the forecasting results of an AR model were significantly worse. They conclude that for non-stationary data it is important to not over-difference the data to reach stationarity. As was seen in Figure 10, there is no need to difference the data twice to reach stationarity. It is highly likely that for most locations, only one type of differencing would have proved to be sufficient. However, this was not inspected in advance when building the model on different locations. Over-differencing is the likely reason why the seasonal ARIMA underperformed when forecasting OV-fiets bicycle availability.

Furthermore, when comparing the seasonal ARIMA with 2 types of differencing, weekly and hourly to the simple ARIMA with only 1 type of differencing, daily, a big difference is seen in the results. The simple ARIMA with one differencing outperforms not only the seasonal ARIMA but the GPR and mean baseline models. Furthermore, when inspecting the 48-hour forecasts, only the seasonal ARIMA with 2 types of differencing has a big bias in forecasts, while the simple ARIMA with 1 type of differencing does not have such a bias and stays consistent in its

predictions. This means that an ARIMA that has too many differencing hyperparameters can lead to a large bias in predictions.

The results of the GPR in terms of model evaluation metrics are better, but visually not promising. It seems that the GPR is mainly predicting around the mean with a periodic effect. This is largely due to the choice of the kernel function and tuning of the hyperparameters for the GPR model. There have been attempts made by multiple authors to make the choice of a kernel function easier through automation or other means (Micchelli et al., 2005; Song et al., 2008; Abdesslem et al., 2017). However, these methods prove to be difficult to implement and use with different data sets.

Multiple attempts were made to find a proper kernel combination and fitting hyperparameters for the GPR, however the results proved to be inaccurate. This is the case due to the GPR being an overly complex model, needing a good understanding of the inner works and mathematics behind it to implement a fitting kernel function. It is important that not only a proper selection of kernel functions is made, but also a proper selection of hyperparameters. If either of these is insufficient, it could lead to similar results as seen in Figure 7 or Figure 8.

Comparing these results to previous research is not a fair comparison due to the limitations of this study, however, it is important to look at the results of other authors despite not finding accurate results. Firstly, the research by Xie et al. (2010) found that when predicting traffic flow data, the GPR outperformed the non-seasonal ARIMA model when comparing the values of RMSE. Furthermore, they tested one-step ahead predictions and two-step ahead predictions. The GPR performed significantly better than the ARIMA model when predicting two-steps ahead, while one-step ahead predictions were only slightly better than the ARIMA model (Xie et al., 2010).

Another study looking at short-term wind speed forecasting found that the GPR predicted wind speed better than the ARIMA model in terms of MAE and RMSE (Wang, & Hu, 2015). One-step ahead, two-step ahead and three-step ahead predictions were made and, in all cases, the GPR outperformed the ARIMA model. However, when predicting three-steps ahead, the difference

between prediction errors was very small (0.86 for the GPR, 0.87 for the ARIMA). A study about forecasting the concentration of particulate matter in the atmosphere of Seoul, Korea found that in terms of RMSE, the GPR model outperformed the ARIMA model (Jang et al., 2020). Different kernel functions were tested separately, including the Matern, RBF and Matern + RBF. All three kernel functions outperformed the ARIMA model in terms of RMSE (Jang et al., 2020).

Previous literature shows that overall, the GPR outperforms the ARIMA model and different variations of the ARIMA model. It is highly likely that the complexity of the model allows for better adjustments to the data and therefore, better results. However, as seen from the current thesis, there can be pitfalls to using the GPR and likewise, the ARIMA model. With improper hyperparameter tuning or kernel selection, despite the simplicity or complexity of the model, it is possible to get results that do not meet one's expectations or do not match previous literature. For these reasons it is important to take careful consideration when running different models.

Future research is advised to make sure not to over-difference the data when performing an ARIMA analysis. If any form of biases is present in the results, careful consideration towards the differencing order is advised.

## **5. Conclusion**

In conclusion it was found that the GPR performed better than the seasonal ARIMA model when predicting OV-fiets bicycle availability in the next 48 hours. However, the results prove to be unsatisfactory for both models due to the limitations found when tuning the hyperparameters of both models. The results of the seasonal ARIMA display a bias when looking at the predictions making the prediction error much higher than usual. The GPR model showed better predictions, however, visually, it is seen that further improvements could be made to the kernel function. Further research needs to be conducted with careful hyperparameter tuning considered.

## References

- Abdessalem, A. B., Dervilis, N., Wagg, D. J., & Worden, K. (2017). Automatic kernel selection for gaussian processes regression with approximate bayesian computation and sequential monte carlo. *Frontiers in Built Environment*, 3, 52.
- Ahmed, M. S., & Cook, A. R. (1979). *Analysis of freeway traffic time-series data by using Box-Jenkins techniques* (No. 722).
- Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192-213.
- Berrar, D. (2019). Cross-Validation.
- Cheng, C., Sa-Ngasoongsong, A., Beyca, O., Le, T., Yang, H., Kong, Z., & Bukkapatnam, S. T. (2015). Time series forecasting for nonlinear and non-stationary processes: a review and comparative study. *Iie Transactions*, 47(10), 1053-1071.
- Corani, G., Benavoli, A., & Zaffalon, M. (2021, September). Time series forecasting with gaussian processes needs priors. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 103-117). Cham: Springer International Publishing.
- Cressie, N. (1988). A graphical procedure for determining nonstationarity in time series. *Journal of the American Statistical Association*, 83(404), 1108-1116.
- Deringer, V. L., Bartók, A. P., Bernstein, N., Wilkins, D. M., Ceriotti, M., & Csányi, G. (2021). Gaussian process regression for materials and molecules. *Chemical Reviews*, 121(16), 10073-10141.
- Editors Holland.com. (2023, November 8). *Celebrate King's Day like a local*.  
<https://www.holland.com/global/tourism/getting-around/information/the-royal-family/kings-day-in-holland>
- Elefteriadou, L. (2014). *An introduction to traffic flow theory*. Springer.
- Gerlough, D. L., & Huber, M. J. (1976). *Traffic flow theory* (No. HS-006 783).
- Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, 37(2), 788-832.

- Hossain, Z., Rahman, A., Hossain, M., & Karami, J. H. (2019). Over-differencing and forecasting with non-stationary time series data. *Dhaka University Journal of Science*, 67(1), 21-26.
- International union of railways. (2024, March 20). UIC - International Union of Railways. <https://uic.org/support-activities/statistics/>
- Jang, J., Shin, S., Lee, H., & Moon, I. C. (2020). Forecasting the concentration of particulate matter in the Seoul metropolitan area using a Gaussian process model. *Sensors*, 20(14), 3845.
- Kamarianakis, Y., & Prastacos, P. (2003). Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches. *Transportation Research Record*, 1857(1), 74-84.
- Kumar, S. V., & Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review*, 7, 1-9.
- Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2014). Traffic flow prediction with big data: A deep learning approach. *Ieee transactions on intelligent transportation systems*, 16(2), 865-873.
- Micchelli, C. A., Pontil, M., & Bartlett, P. (2005). Learning the Kernel Function via Regularization. *Journal of machine learning research*, 6(7).
- Munoz, J. C., & Laval, J. A. (2006). System optimum dynamic traffic assignment graphical solution method for a congested freeway and one destination. *Transportation Research Part B: Methodological*, 40(1), 1-15.
- Nau, R. (2016). General seasonal ARIMA models—(0, 1, 1)×(0, 1, 1) etc. URL: <https://people.duke.edu/~rnau/411sdif.htm> (visited on Mar. 12, 2020).
- Nederlandse Spoorwegen. (n.d.). *OV-fiets*. Retrieved March 14, 2024, from <https://www.ns.nl/deur-tot-deur/ov-fiets>
- Nguyen-Tuong, D., Seeger, M., & Peters, J. (2009). Model learning with local gaussian process regression. *Advanced Robotics*, 23(15), 2015-2034.
- NS API Management team. (n.d.). APIs: Details. NS API Developer Portal. <https://apiportal.ns.nl/api-details#api=ovfOnlineServiceApi>
- Perraudin, N., & Vandergheynst, P. (2017). Stationary signal processing on graphs. *IEEE Transactions on Signal Processing*, 65(13), 3462-3477.

- Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian processes for machine learning* (Vol. 1, p. 159). Cambridge, MA: MIT press.
- Saxion. (2022, April 22). *5 things you need to know about Kingsday / Saxion*. Hogeschool Saxion. <https://www.saxion.edu/news/2022/april/kingsday>
- Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85, 1-16.
- Shumway, R. H., Stoffer, D. S., Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. *Time series analysis and its applications: with R examples*, 75-163.
- Song, H., Ding, Z., Guo, C., Li, Z., & Xia, H. (2008). Research on combination kernel function of support vector machine. In *2008 International conference on computer science and software engineering* (Vol. 1, pp. 838-841). IEEE.
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, 16(4), 437-450.
- Treiber, M., & Kesting, A. (2013). Traffic flow dynamics. *Traffic Flow Dynamics: Data, Models and Simulation*, Springer-Verlag Berlin Heidelberg, 983-1000.
- Wang, J., & Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis—Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy*, 93, 41-56.
- Wang, J. (2020). An intuitive tutorial to Gaussian processes regression. *arXiv preprint arXiv:2009.10862*.
- Weisang, G., & Awazu, Y. (2008). Vagaries of the Euro: an Introduction to ARIMA Modeling. *Case Studies In Business, Industry And Government Statistics*, 2(1), 45-55.
- Williams, B. M. (2001). Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling. *Transportation Research Record*, 1776(1), 194-200.
- Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of transportation engineering*, 129(6), 664-672.

Xie, Y., Zhao, K., Sun, Y., & Chen, D. (2010). Gaussian processes for short-term traffic volume forecasting. *Transportation Research Record*, 2165(1), 69-78.

Zhao, J., & Sun, S. (2016). High-order Gaussian process dynamical models for traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 17(7), 2014-2019.