Modelling air quality around highways in the Netherlands

Tamar Vooijs
4596951

Delft University of Technology

TU Delft

Modelling air quality around highways in the Netherlands

by

Tamar Vooijs

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday August 25, 2022 at 11:00 AM.

Student number:	4596951	
Project duration:	February 7, 2022 – August 25, 2022	
Thesis committee:	Dr. ir. R. A. Verzijlbergh,	TU Delft
	Dr. ir. S. van Cranenburgh,	TU Delft
	Dr. T. Verma,	TU Delft

TU Delft

Modelling air quality around highways in the Netherlands

Abstract

More and more research shows the substantial health repercussions of air pollution. Therefore, improving air quality is high on political agendas in modern societies. In the Netherlands, particularly around major roads, NO_2 standards set by the government are often exceeded. Air quality models are used to monitor air quality values and design policies to reduce air pollution. Currently, authorities in the Netherlands use a Gaussian Plume Model for decision-making, but this model paints a rather skewed view of reality due to its underlying assumptions. This research contributes to academic knowledge about air quality modelling by evaluating two innovative model types, a physics-based LES model and a data-driven regression model, for their usage in decision-making to improve air quality. This is done by comparing the performance of both models with the performance of a Gaussian Plume Model for predicting NO_2 levels around a large highway in the Netherlands. Also, two combinations of the LES model and the regression model are examined. It is concluded that both the LES model and the regression model show potential for accurately predicting air quality around highways in the Netherlands. The LES model is particularly suitable for predicting high NO_2 levels, and the regression model is considered suitable for predicting the average NO_2 levels over a longer timeframe. A model in which the LES results were combined with a regression model outperformed the original models and is therefore considered to hold the most potential for usage within air quality policy.

TU Delft

Acknowledgements

The thesis in front of you concludes my master's degree in Engineering, Technology and Policy. Within my master's, I acquired knowledge about grand challenges that find themselves at the intersection of technology and society. I found it very rewarding to learn how we can responsibly use technology to create a more sustainable and livable world. Therefore, I was pleased to get the opportunity to work on this master thesis project. Since I became acquainted with the impact of air quality on human health, it did not leave my mind anymore. Working on a project with such societal relevance and being able to put in my passion for modelling and data science was a great joy to me.

Firstly, I would like to thank my supervisors at TU Delft for their guidance and knowledge during this master thesis. I want to thank Remco Verzijlbergh for his involvement and enthusiasm; our conversations and weekly meetings always left me with an extra boost of motivation and curiosity about the topic. I truly appreciated the trust and the freedom he granted me to conduct this study in a way that suited me. I would also like to thank Sander van Cranenburgh for his help and insightful input, which pushed me in the right direction and greatly improved this thesis. Furthermore, I want to thank Bert Enserink for ensuring that the societal aspect of the study was not forgotten and for adding a different perspective to the thesis. Also, I would like to thank Trivik Verma for stepping in and providing help in unusual circumstances.

Furthermore, I was happy to be able to spend a lot of my days during my thesis at Whiffle. I really enjoyed the open and inspiring atmosphere and the interesting talks during the coffee breaks, lunches and 'borrels'. Special thanks to my fellow graduates, Annette and Tim, for their nice company and patience in answering all my elementary physics questions.

Lastly, I would like to thank my friends and family who have supported and encouraged me during my study. My friends from Laga and my TB friends, who made my time in Delft so enjoyable. My roommates Evi and Angela, who were always willing to listen to my struggles and provided me with welcome distractions. My dear friend Sonia, my parents and my sister Myrthe for being there for me and believing in me. And of course, my boyfriend Laurence, who, in recent months, has also taken on the roles of editor-in-chief and motivational coach, for all of his support and for always making me feel better, even during stressful moments.

The past half year has been filled with hard work, but also with a lot of fun. Many thanks to all of you!

Tamar Vooijs
Delft, August 2022

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
AQF	Air quality forecasting
GPM	Gaussian Plume Model
GPU	Graphics Processing Unit
CPU	Central Processing Unit
GRASP	GPU-Resident Atmospheric Simulation Platform
LES	Large Eddy Simulation
WHO	World Health Organization
RIVM	National Institute for Public Health and the Environment
I&W	The Ministry of Infrastructure and Water Management
EMA	Environmental Management Act
OLS	Ordinary Least Squares
NO_2	Nitrogen dioxide
PM	Particulate Matter
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error

1 Introduction

The Court of Appeal in Brussels ruled last year that the Flemish government had not acted adequately to protect citizens against air pollution (Hof van Beroep Brussel, 2021). As a result of this ruling, high penalty payments have to be paid by the Flemish government, and they quickly have to draw up an ambitious air quality plan that describes how they will eliminate violations of air quality standards. A similar charge lingers over the Dutch state. In 2017, the interest group Milieudefensie started a lawsuit in which they demanded that the Dutch government undertakes steps to comply with European air quality regulations (Rechtbank Den Haag, 2017; Milieudefensie, 2021). In both cases, the plaintiff invokes the right to clean air, as enshrined in Directive 2008/50/EC on ambient air quality and cleaner air for Europe (*Directive 2008/50/EC*, 2008).

Air quality is a recurring problem in the Netherlands. Postma (2005) states that the availability of clean air is not self-evident in a densely populated and traffic-intensive country like the Netherlands. A couple of years ago, the Netherlands belonged to the countries with the highest air pollution in Europe (Guerreiro et al., 2016). In recent years, air quality levels in the Netherlands have improved, but the values are still far below the World Health Organization's (WHO) recommended air quality guidelines (Atlas leefomgeving, n.d.). This means that many people in the Netherlands are exposed to unhealthy levels of air pollution.

In the last years, there has become more attention for the consequences of air pollution. Due to recent studies that showed that health consequences of prolonged exposure to air pollutants were more severe than known before (WHO, 2021b), improving air quality has risen on the political agenda in the Netherlands. In 2021, the National Institute for Public Health and the Environment quantified the health impacts of air pollution in the Netherlands and found that more than 10000 people suffer from chronic lung disorders, 4000 underweight infants are born each year, and 1200 people develop lung cancer (De Ingenieur, 2021). In general, air pollution leads to an average lifespan reduction of thirteen months in the Netherlands (www.vzinfo.nl, n.d.). In terms of disease burden, the effect of air pollution is equivalent to being overweight and larger than excessive alcohol consumption (RIVM, n.d.).

In 2020, almost 40 percent of air pollution emissions in the Netherlands were attributable to traffic or transport (RIVM, n.d.). In addition, exceedances of air quality standards as defined by the European Commission often occur around main traffic arteries (Compendium voor de leefomgeving, n.d.). Monitoring air pollution levels around highways is very important to maintain healthy levels of air quality. It has been shown that nitrogen dioxide (NO_2) is a good indicator of traffic-related air pollution since concentrations higher than the background concentration can be measured till at least 1000 meter distance from the road (Fischer et al., 2000). However, measurement devices are expensive and do not provide insight into future concentrations. As a result, air quality models have been developed that are capable of interpolating and predicting pollutant concentrations. The Dutch government uses variations of the so-called Gaussian Plume Model (GPM) for modelling air quality (Infomil, n.d.-c).

The model, however, paints a rather skewed view of reality because of its underlying assumptions. This raises the question of whether the model is the best fit to support air quality policy. Outcomes of Gaussian Plume Models generally have a large insecurity range, and the model interpretations deviate among users (De Ingenieur, 2021). An example of a misconception in the interpretation of model results was the speed limit increase at the A10 highway in Amsterdam from 80 km/h to 100 km/h. Prior to the intervention, the executive ministry expected a rise in NO_2 of 0.1 $\mu g/m^3$ based on predictions of the model. Afterward, the concentrations turned out to be 3.8 ug/m3 higher on average than for the original speed limit. The model used for the calculations did not consider the impact of the surrounding buildings on the dispersion of the air pollutants (Infomil, n.d.-c). The higher concentrations due to the increase in speed limits lead to discontent and protests among residents. This example shows how wrong model results led to inadequate policy implementation, illustrating the importance of accurate air quality models.

Current advancements in sensor technology, computer science, and local weather forecasting can enable new ways of air quality forecasting that could lead to more accurate predictions. An innovative model with high potential for air quality forecasting is the atmospheric model created by Whiffle, in which the dispersion of air pollutants is modelled as a result of air movement and turbulence. This type of model is called a Large Eddy Simulation (LES). Other methods for modelling air quality that show potential are machine learning models. Research by Kang et al. (2018) shows that using big data and machine learning approaches such as neural networks and random forests yield promising results for predicting air quality.

This research compares three different models for predicting air quality around highways in the Netherlands. These include the physics-based LES-model created by Whiffle and a data-driven regression model. To put results into context, a GPM is used as a benchmark model. This research aims to contribute to decision-making aimed at improving air quality policy in the Netherlands by exploring innovative types of air quality models. This study intends to answer whether the innovative models are more suitable for supporting decision-making about air quality around major roads in the Netherlands than the existing GPM.

1.1. Related work

This section summarizes the existing literature on the role of air quality models in policy-making, followed by various model types created so far. Afterward, the knowledge gap is identified, and the research question is addressed.

1.1.1. The role of air quality modelling in society

Poor air quality causes adverse long-term health effects (Kang et al., 2018), leading to reduced quality of life for individuals and considerable societal costs (Pervin et al., 2008). Consequently, officials have introduced laws to regulate those concentrations in many countries (Delavar et al., 2019). Air pollution models can play an essential role in monitoring air pollution. Research by Kang et al. (2018) reveals that air quality evaluation and forecasting models can support decision-makers while developing policies to enhance air quality. They are used, amongst others, to determine the location of new pollution sources, establish source emission limits and design air-pollution control strategies (Weil et al., 1992). Also, Baklanov & Zhang (2020) emphasizes the importance of air quality predictions for societal purposes. They found that many countries use air quality modelling for designing preventive measures to minimize exposure to unhealthy levels of air pollution. The usage of air quality models goes far back in time. Miller (1982) stated that air quality models already played a significant role during the promulgation of the Clean Air Act Amendments in the US in 1970. Due to its economic and societal relevance, air quality forecasting evolved into a distinct discipline.

1.1.2. Types of air quality models

Various approaches, developments, and applications of air quality models have been created over recent years (El-Harbawi, 2013). Air quality models can be divided roughly into two categories; models in which chemical and atmospheric conditions are simulated based on physics formulas and models in which patterns are derived from historical data (Seaman, 2000). In this research, the first category is referred to as physics-based models, and the latter category of models is called data-driven models.

The GPM is a physics-based model that uses atmospheric conditions and relations to model pollutant dispersion (Miller & Hively, 1987). For a prolonged period, the GPM model was the most commonly employed in physics-based models, according to Daly & Zannetti (2007). Daly & Zannetti (2007) stated that this model type was particularly suitable for understanding the diffusion properties of plumes emitted from large industrial stacks. However, due to its underlying assumptions, the model presents a strongly idealized and simplified image of reality (Tomas et al., 2015). One of the assumptions is that atmospheric turbulence is stationary and homogeneous (Abdel-Rahman, 2008). This assumption does not hold in reality and has significant consequences for the model's outcomes since the dispersion of pollutants in the lowest part of the atmosphere is determined by the presence and structure of turbulence (Weil et al., 1992). Furthermore, the GPM can not easily cope with obstacles in its scenery, as it does not account for the effect of a change in turbulence structures due to the buildings (Reed et al., 2005). This also applies to the GPM that the government uses to map air pollution. In the validation of this model, it is stated the model is not suitable for calculating concentrations of air pollutants in areas where buildings are situated around the roads ("Technische beschrijving van standaardrekenmethode 2 (SRM-2) voor luchtkwaliteitsberekeningen", 2014). In general, the GPM is suitable for providing generalized solutions for air pollutant dispersion but leads to incorrect and misleading results in situations with more complex weather- or terrain conditions (Bieringer et al., 2021).

A type of simulation model that can simulate atmospheric processes in more detail is the LES. In this model, large turbulent flows in the planetary boundary layer, the lowest part of the atmosphere, are solved explicitly (Grylls et al., 2019). The LES modelling approach has been successfully used to predict airborne dispersion in many atmospheric scenarios and terrain conditions (Bieringer et al., 2021). LES simulations used to be very computationally expensive, but developments in computational capabilities are generating new possibilities for its usage (Grylls et al., 2019). An implementation of an LES model is the atmospheric model created by Whiffle. This model runs almost entirely on Graphics Processing Units (GPUs) and is therefore much faster than the standard Central Processing Unit (CPU) (Schalkwijk et al., 2015). This model is named GRASP, which stands for GPU-Resident Atmospheric Simulation Platform (*Whiffle*, n.d.). In the literature, also other attempts are found in which LES is used for the prediction of air pollution (Tseng et al., 2006; Grylls et al., 2019).

Another approach for modelling air quality is the usage of data-driven models. These models search patterns in historical data to predict air quality, instead of simulating a physical relationship between emissions and ambient concentrations (Daly & Zannetti, 2007). There has recently been much attention to machine learning techniques in this field. Sinnott & Guan (2018) state that the emergence of machine learning models brought new opportunities for air quality forecasts. In addition, Baklanov & Zhang (2020) mentions that machine learning models can generate accurate air quality predictions. A relatively simple machine learning approach that is accurate in many situations is Ordinary Least Squares (OLS) regression (Bonaccorso, 2017). Other research has already shown that OLS regression can achieve an adequate accuracy when predicting air quality (Mahanta et al., 2019).

Also, hybrid models that use both data-driven techniques and atmospheric relations to predict air quality show high potential Baklanov & Zhang (2020). Ryan (2016) demonstrates the power of hybrid models, as they state that an physics-based model, combined with statistical post-processing, holds the most theoretical promise for air quality modelling.

1.2. Research approach

In this research, an LES-based model and a regression model are researched and compared to each other and to a benchmark GPM. Afterward, combination of the models is examined to improve the air quality predictions even more. The research adds to the academic literature by comparing the performance of these model types with each other for the specific case of predicting air quality around highways in the Netherlands. To the best of our knowledge, no earlier attempt has been made to compare the performance of those three modeling approaches for predicting air quality levels around highways. Based on this research gap, the following research question is defined:

How could air quality predictions around highways in the Netherlands be improved in order to support decision-making aimed at enhancing air quality levels?

The research approach that is employed to answer the research question is the modelling approach. Three different models are compared with each other. The first model is the GRASP model, which is until now, mainly used for ultra-high resolution weather forecasts. In this research, the performance of the model for predicting NO_2 values around highways is assessed. The second model in this research is a regression model. The third model is a GPM model that serves as a benchmark. The thesis is exploratory in nature, as the main goal is to identify the potential of innovative models for improving air quality predictions. This is done by applying the models to a specific case and comparing their performances on different criteria.

1.2.1. Research sub-questions

This section defines five sub-questions that individually cover a part of the aforementioned research question.

1. How does air quality modelling support decision-making in the Netherlands?

The first sub-question provides insight into the usage of air quality modelling to support decision-making in the Netherlands. For defining the requirements of the air quality models, it is essential to know in which context the models add generative value to the policy-making process. An overview of current air quality policies and regulations is given, and important actors are identified. Afterward, the role of air quality models in the decision-making process is discussed.

2. What are the requirements of air quality models to adequately support decision-making?

Based on the identified purposes of air quality models in the Netherlands in the first sub-question, the requirements to which air quality models should comply are distinguished. Furthermore, performance metrics are chosen to evaluate how well the models meet these requirements. Those performance metrics are focused on how similar the predictions are to the NO_2 measurement data.

3. What is the performance of a LES-model, a regression model and a GPM for predicting NO_2 values around highways in the Netherlands?

Firstly, the GRASP and GPM simulations are prepared by implementing a linesource to simulate the highway and by defining the settings such as run-time, domain, and resolution of the runs. The regression model is trained on the training set of the data. After the models have run for the chosen timeframe the performance of the GRASP model, the regression model and the GPM model are assessed by using the performance metrics. The performances of the models are compared to each other, and the strengths and weaknesses of the models are discussed.

4. Could the combination of a regression model and the GRASP predictions enhance the accuracy of the predictions?

Based on the results of sub-question 3, it is examined whether a combination of both models can capture the strength of both models and lead to more accurate results. It is checked whether the combination of the model leads to higher performance on the performance metrics than the original models.

5. Which model is considered to be most suitable for usage in decision-making concerning air quality?

The last sub-question assesses the suitability of the examined models for supporting decision-making aimed at improving air quality around the Netherlands' highways. This is done by evaluating the models against the requirements of air quality to support decision-making.

1.2.2. Research outline

Figure 1.1 shows the research flow diagram of the paper and fits the research questions to the the corresponding research methodology that is used.

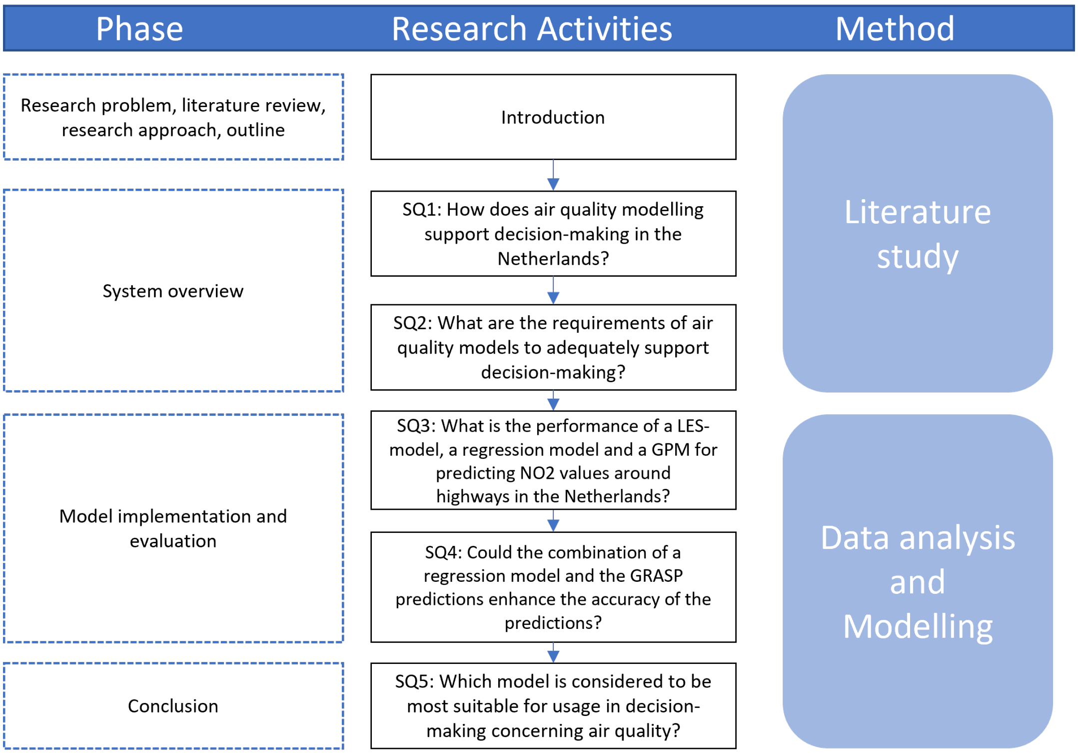


Figure 1.1: Research Flow Diagram

To answer the first two sub-question, literature is conducted. This includes articles found through searching via search engines such as Google Scholar and Scopus, but also governmental websites and publications from research institutes in the Netherlands. The third and fourth sub-question are answered in chapter 3 and 4, which consists of the model implementation and the model results. The fifth sub-question is answered in chapter 5, in which a conclusion is drawn and the answer to the main question is formulated explicitly.

2 System overview

In this chapter, firstly an overview of the main actors that are involved in the air quality policy is given. Secondly, air quality regulation and the role of air quality models in this process are discussed. In the first two sections an answer to the first sub-question is given; How does air quality modelling support decision-making in the Netherlands? Afterward, the requirements with which air quality models should comply to be suitable in the decision-making process are identified, and this answers the second sub-question.

2.1. Actors

The landscape of actors involved in the air quality policy in the Netherlands is complex, as it concerns people from many different sectors and aggregation levels. Air pollution crosses international borders but simultaneously can have very local impacts. Furthermore, air quality regulation often inhibits economic activity (Henderson, 1995). Therefore, governmental actors have conflicting interests as they are responsible for stimulating economic activity and monitoring air quality. This paragraph discusses the most critical actors interested in air quality around highways in the Netherlands.

2.1.1. Governmental actors

In the European Union, the European Environmental Agency (EEA) and the European Commission (EC) warrant that the air quality standards outlined in directive 2008/50/EG are adhered to by member-states (Europadecentraal, n.d.). The EC monitors how well member states implement the EU directives based on environmental information provided by the EEA.

The Ministry of Infrastructure and Water Management (I&W) is accountable for air pollution originating from traffic on a national level in the Netherlands (www.rijksoverheid.nl, n.d.). The Ministry of Health, Welfare and Sport monitors air quality values nationally to check whether it meets the standards. This is typically done by the the National Institute for Public Health and the Environment, also called the RIVM. The RIVM possesses measurement stations and developed the current air quality models. Furthermore, they are responsible for alarming and informing the population whenever air pollution exceeds certain values. The models developed at the RIVM are made publicly available. They are used by the Ministry of I&W to assess the impact of new infrastructural projects on air quality. One of the executive agencies, Rijkswaterstaat, operates the large highways in the Netherlands and has to ensure that exceedances of the European directive are prevented.

However, the Ministry of I&W is also responsible for the accessibility and traffic flow in the Netherlands. Measurements that benefit the latter goal often contradict their objective to improve air quality. These conflicting interests make the position of the Ministry of I&W complex when it comes to air quality around highways. An example that illustrates this complex position was the increase of the maximum speed at the A10 in 2016, as described in 1. On the one hand, this measure improved traffic flow, but on the other hand, it harmed air quality (Ligterink & Smokers, 2016). Accurate air quality models are very important to this actor to address these difficulties. The air quality models provide the possibility to optimize traffic flow while simultaneously monitoring air quality values so that they do not exceed the limits.

2.1.2. Interest groups

Multiple interest groups represent the interest of air quality in the Netherlands. Examples of large interest groups are Milieudefensie and Longfonds. Their goal is to represent citizens by defending the right to breathe clean air. As mentioned in chapter 1, Milieudefensie has launched a lawsuit against the Dutch government, claiming that the government should do more to ensure clean air. Other activities of interest groups include informing citizens about the consequences of air pollution, and the current air quality status in the Netherlands. Next to the large interest groups that focus on air quality on a national level, there are also local interest groups committed to a more specific issue. An example is Kerngroep Ring Utrecht, who demonstrated against the road expansion of two large highways in Utrecht in 2021.

2.1.3. Citizens science initiatives

Next to the interest groups, there are citizen science communities established that monitor air quality. Citizens can join this initiative by purchasing their own measuring equipment and adding measurements to a database. A well-known example of such a citizens science initiative that stores air quality measurements collected by citizens is Sensor.community (formerly Luftdaten). They publish large-scale measurements of air quality indicators. Those citizen science initiatives aim to provide insight into air quality at different locations.

2.1.4. Other actors

Engineering firms and research institutes are also actors with a significant impact or an interest in air quality. The government often employs engineering firms to calculate the effect of, for example, infrastructural interventions on air quality. To do this, they use models created by the RIVM. Furthermore, research institutes like TNO provide new insights regarding air quality based on their research. Furthermore, Dutch citizens are involved; in particular, vulnerable people who suffer from lung disease or who live close to highways are important actors, as they are the most affected by policies concerning air quality around roads.

2.2. Air quality regulation

In the Netherlands, an essential element of air quality regulation is monitoring air quality levels. Multiple guidelines and threshold values exist that attempt to quantify upper limit values that can ensure healthy air quality. The RIVM tracks values of air quality based on measurements and interpolations. This paragraph gives an overview of the air quality standards applicable to the Netherlands. Afterward, measures to improve air quality used by the government are discussed.

2.2.1. WHO Global air quality guidelines

The Global Air Quality Guidelines of the WHO provide global recommendations on thresholds for key air pollutants that pose health risks. The guidelines are applicable worldwide and are scientifically substantiated (WHO, 2021a). Last September, the limits for common air pollutants were sharpened after recent studies proved the health consequences of air pollution to be more severe than known before (WHO, 2021b). The WHO Global air quality guidelines are recommendations and not legally binding (Infomil, n.d.-c). Table 2.1 shows the guidelines for the different air pollutants set by the WHO.

2.2.2. European standards

In EU directive 2008/50/EG, limit values and target values concerning air pollutants are defined. Limit values should not be exceeded by the member states. Sometimes target values are set next to the limit values. In this case, the member state must reach the target values within a certain period. The member state should be able to prove that they are committed to reach these values and that they have undertaken steps that demonstrate their dedication (*Europese wetgeving luchtverontreiniging*, n.d.). The limits specified in the directive are shown in Table 2.1.

2.2.3. Dutch Environmental Management Act

In Dutch legislation, thresholds are included in the Environmental Management Act (EMA). Many limit values arise from the EU directive (Infomil, n.d.-d). Next to the limit values, an alarm threshold and an information threshold are specified at which the government is obliged to inform the entire population or vulnerable people in the population respectively (Infomil, n.d.-d). The table below shows the limit values specified in the EMA. For NO_2, there are two standards set. The first one is the yearly average, which may not exceed 40 μ g/m^3, and the second is an hourly average of 200 μ g/m^3. In practice, it appears that, in particular, the yearly average is exceeded from time to time at different locations in the Netherlands (Compendium voor de leefomgeving, n.d.). Next to NO_2 limits, thresholds for PM_{10} and $PM_2 \cdot 5$ are set. PM is short for particulate matter; $PM_2 \cdot 5$ and PM_{10} are, respectively, the particles in the air with a diameter smaller than 2.5 and 10 μ m. Also, for benzene a threshold is set, but this one is seldom exceeded.

Instance	**Pollutant**	**Type**	**Concentration**	**Note**
World Health Organization	NO_2	Yearly average	10 $\mu g/m3$	
	PM_{10}	Yearly average	15 $\mu g/m3$	
	$PM_2 \cdot 5$	Yearly average	5 $\mu g/m3$	
European directive & Environmental Management Act	No2	Yearly average	40 $\mu g/m3$	
	No_2	Hourly average	200 $\mu g/m3$	Maximum of 18 exceedances
	PM_{10}	Yearly average	40 $\mu g/m3$	
	PM_{10}	Daily average	50 $\mu g/m3$	Maximum of 35 exceedances
	$PM_2 \cdot 5$	Yearly average	20 $\mu g/m3$	
	Benzene	Yearly average	5 $\mu g/m3$	
European directive	$PM_2 \cdot 5$	Target value	14.4 $\mu g/m3$	

Table 2.1: Air quality standards applicable to the Netherlands

Whenever the air quality thresholds exceed the accepted concentration, the government is required to take extra measures. Those measures may be taken at three levels: national, regional, and locally (Infomil, n.d.-a), and can be distinguished into two distinct categories: emission controlling and immission controlling measures (Infomil, n.d.-b). The first category aims to reduce air pollution by decreasing the sources of air pollutants. For traffic-related air pollution, this can be done through, for example, stimulating public transport, reducing the maximum speed, or attempting to alter the traffic composition. The second category of measures is aimed at reducing air pollution exposure to humans. Examples of immission reducing measures are building walls around highways that impact the dispersion pattern of the pollutants, or spatial planning, such that people do not live or work around places with high levels of air pollution.

2.3. Usage of air quality models

Models play an important role in different stages of the decision-process concerning air quality. In the Netherlands, the Monitoringstool created by the Ministry of I&W is the model used for this purpose (*Aerius Air calculation tool*, n.d.). In this tool, two different calculation models are included for modelling air quality values around highways, called SRM1 and SRM2. Both models are based on the Gaussian Dispersion Formula and can be classified as GPMs (Wesseling & Van Velze, 2015). SRM1 is applied for inner-city traffic and SRM2 for suburban traffic. The SRM1 and SRM2 models are deployed for multiple purposes. First of all, the models are used by RIVM for monitoring air quality to check whether the values are below the limit values as described in Table 2.1 (*Aerius Air calculation tool*, n.d.). If the values exceed limits, additional measures are required. Furthermore, the SRM1 and SRM2 method is also used to calculate the impact of infrastructural changes on the air quality in the surrounding area (*Nieuw Normen- en Handhaving-stelsel Schiphol*, 2016). Before an infrastructural project is approved, an environmental impact assessment needs to be written (Rijkswaterstaat, n.d.). In this provision, a paragraph about the impact of the project on air quality has to be inserted. To compose this section, government institutions or engineering companies must use the SRM1 or SRM2 method in the Monitoringstool that is made publicly available.

2.4. Requirements of air quality models

In general, air quality models serve a societal purpose and modellers in this area should pursue transparency, involvement of multiple views, replication and analysis of sensitivity and uncertainty within the modelling process (Saltelli et al., 2020). Transparency includes clear communication about assumptions and uncertainties in the modelling phase. This requirement is independent from the type of model that is used, but in general models with a higher complexity are more prone to become a black box than more simple models. Therefore, complex models need more explanation and communication about the processes within the model to clearly address limitations and assumptions within the model (Saltelli et al., 2020). This should be taken into account while assessing different air quality models.

As stated in section 2.3, air quality models in the Netherlands are used for assessing current air quality values, but also for environmental impact studies and issuing permits of new emission sources. In general, the air quality models used for those regulatory purposes have to provide spatial distribution of high episodic concentrations and of long-term averaged concentrations for comparison with air quality guidelines (De Leeuw et al., 1997). In the Netherlands, the binding guidelines are set in Directive 2008/50/EC and the EMA. Therefore, air quality models for policy making in the Netherlands should be able to monitor whether the air quality values do not exceed those limits. In the case of NO_2, this means that a model can be used to calculate yearly values and hourly values since, for both periods, a limit value is set. For modelling averages per year, the model should be able to capture the trend over a longer timeframe. For modelling hourly values, it is crucial that high No_2 values are modelled, as the government has the task of minimizing exceedances of the hourly threshold.

In the following chapters, the GRASP model, the GPM, and a regression model are implemented to predict values around highways and be validated accordingly. Based on this analysis, a conclusion can be made about the suitability of those models based on the requirements mentioned in this paragraph. In this research, only the performance of the models is included to assess the suitability of the model. Computational time and maintenance of the models are excluded from the scope of the research.

3
Methodology

In this chapter, the methods that are used to assess the performance of the three models for predicting air quality around major roads are discussed. To achieve this, firstly the performance metrics that are used for comparing the performance of the models are described. Secondly, the process of gathering the data required for the simulations is described. Afterward, the set-up of the respectively GRASP, the GPM, and the regression models is defined. This chapter defines the methodology that needs to be followed to answer the third and the fourth sub-questions, as defined in chapter 1.

3.1. Performance metrics

To compare the performance of the GRASP model, the GPM and the regression model, metrics are defined that measure the similarity between the models' predictions and the measurement data. The performance metrics used in this research are given in equation 3.3. These include the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE), and the Pearson Correlation Coefficient. The MAE measures the deviation between the actual and the predicted values. The MAE is scale-dependent and has the same unit as the variable. Furthermore, the RMSE is a common metric and is considered a good error metric for numerical predictions (Christie & Neill, 2021). It is interesting to examine the difference between the RMSE and the MAE, as a large difference between those values means a large variation between the errors of the model. The RMSE and the MAE focus on the error between the models' predictions and the measurement data. However, even with a high error, a model prediction can still provide valuable insights when there is a high correlation between the predictions and the measured value. Therefore, next to the MAE and the RMSE, the Pearson correlation (r) is used to check the relation between the predicted and measured values. A higher r-value means a higher correlation between the measurement data and the model predictions.

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n} |y(i) - \hat{y}(i)|^2}}{N} \tag{3.1}$$

$$MAE = \frac{\sum_{i=1}^{n} |y(i) - \hat{y}(i)|}{N} \tag{3.2}$$

$$r = \frac{n(\sum \hat{y} * y - (\sum \hat{y})(\sum y)}{\sqrt{(n \sum \hat{y}^2 - (\sum \hat{y})^2)(n \sum y^2 - (\sum y^2)}} \tag{3.3}$$

where;
$\hat{y}$ = Predicted value
y = Measured value
N = Number of samples

The scores of the models on the performance metrics are compared with each other and used as criteria to assess the suitability of the models for predicting air quality around major roads in the Netherlands. Next to the performance metrics, the results are interpreted based on the behavior of the models that is visualized in the figures.

3.2. Data collection

This section goes through the datasets required for the simulations. This is done by addressing the availability and quality of the data, as well as its limitations.

3.2.1. NO_2 measurements

To assess the performance of the models, the predicted values of the models are compared with NO_2 measurement data. The RIVM monitors NO_2 values to assess whether the values meet the air quality standards and publishes them at Luchtmeetnet.nl. The measurements are in compliance with the European regulation that specifies requirements for testing legal standards (*Vaststellen van luchtkwaliteit*, n.d.). As a result, the measurements can be considered high-quality data. The quality of the measurements is important since the data is measured per hour and an error in the data has a large impact on the performance of the models. The RIVM has a total of 44 measurement stations around the country. Of all those measurement stations, only one station is directly adjacent to the highway. This station is located at the A2 close to Breukelen. The measurement station and its geographical location are showed in Figure 3.1a and Figure 3.1b. Since this is the only station located directly next to the highway, the measurements of the station show the most accurate impact of traffic emissions from highways. Therefore, the data originating from this measurement station is used to compare with the model predictions. The hourly measurements can be easily extracted from the site of Luchtmeetnet for the years 2012 till 2021.

(a) Position of measurement station in landscape

(b) Geographic location of measurement station

Figure 3.1: Measurement station RIVM directly next to the A2 Breukelen

3.2.2. Traffic data

The emission levels of the simulated highway in the GRASP model and GPM are quantified over time using historical traffic data. Also the regression model uses the traffic data as input data. Traffic characteristics that have an impact on air pollution are found to be traffic speed, intensity, and composition (Zalakeviciute et al., 2018; Energy News Network, n.d.; Johansson et al., 2009). Data has been searched that contains these data-points for highways in the Netherlands. For most large roads in the Netherlands, traffic data is gathered by the National Road Traffic Data Portal (NDW). Their data contains historical information about traffic intensity, composition, and corresponding average speed per mile marker at a specific location. However, while examining the data for the location defined in paragraph 3.2.1, vehicle composition is missing in most situations, and only the average vehicle speed is available in the dataset. Research states that average speed is not a good indicator of traffic emission, as instantaneous speed fluctuation has a more significant impact on the strength of the emissions than average speed (Panis et al., 2006). They found that the instantaneous speed and acceleration profiles might be considerably different for the same average speed, resulting in different fuel consumption and emission levels. Therefore, it might not be representative to consider the average speed when implementing emission levels from traffic data. Due to those availability issues, in this research, only traffic intensity is used for computing the emission strength. The traffic intensity is measured in vehicles per hour. At the NDW, the traffic intensities are stored for almost every hectometer sign. Therefore, the hectometer signs at the A2 closest to the measurement station for both directions are used for conducting the traffic intensities. These are hectometer signs 46.0 Li and 46.0 Re. The quality of the traffic data is high for most periods, as there are not much missing values in the data. However, for the year of 2018, the data contains of many invalid values. A modelling period is chosen for the models for which the quality of the traffic data is high.

3.2.3. ERA5 Weather data

The models use ERA5 data to incorporate large- and mesoscale weather effects. ERA5 data is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). The ERA5 dataset includes reanalysis data of multiple atmospheric, land, and oceanic climate variables (*Era5*, n.d.). Reanalysis data provides a thorough overview of past weather and climate by combining observations and models (Hersbach et al., 2020). The data is globally complete and consistent in time and is therefore sometimes referred to as 'maps without gaps' (Hersbach et al., 2020). The data span the planet on a grid of 30 km, and they use 137 levels to resolve the atmosphere from the ground up to an altitude of 80 km (*Era5*, n.d.). In GRASP, ERA5 data is used to define large-scale boundary conditions in order to obtain realistic weather conditions (Wiegant & Verzijlbergh, 2019). Furthermore, ERA5 data is used as input for the GPM and regression models. From the ERA5 dataset, four variables are selected. These include wind direction, wind speed, solar insolation, and cloud coverage.

3.2.4. Data limitations

As mentioned above, the measurement data for NO_2 concentrations are collected entirely from one measurement station, since this is the only measurement station that publishes NO_2 concentrations next to a highway. The performance of the models is therefore based on the performance for this specific location. For traffic patterns, this is not a problem as these are expected to be similar at different places along the highways in the Netherlands. However, the models might perform differently at other locations next to Dutch highways, as the landscape might differ. In this research, a site was chosen with few buildings in between the highway and the measurement station. Performance of the models can be different if buildings are present between the highway and the measurement station. Another limitation is the lack of data on speed patterns of traffic and vehicle composition on the highway over time. In the current implementation of the models, only traffic intensity is used as an indicator for traffic emissions. Still, the composition of the cars, their speed, and the presence of traffic jams all affect the level of the emissions. This lack of data applies to all three models; therefore, the comparison of the model performance is still considered accurate. However, it is expected that the model performances would improve whenever this is added to the models since it would increase the similarity of the simulated location and the location of the measurement station. Also the data of the traffic intensity is sometimes lacking, as for some years the data consists of many invalid values. This has implications for the period that has been used to train the regression model on.

3.3. Timeframe

To compute the final results, the models were run for a year. As all seasons occur in one year, this period is expected to comprise most of the weather conditions that prevail in the Netherlands. Due to the lockdowns in 2020 and 2021, patterns in traffic intensity were different and not representative during the corona crisis. Therefore, a timeframe before the corona pandemic is chosen. The year that has been selected is 2019. The regression model is split in a training and a testset, and only the results of the regression model on the test set are representative for the performance of the model. Ideally, the test set would therefore comprise a period of a year as well to allow for comparison between the models. However, due to availability issues in the traffic data, the year 2018 cannot be used as training data. Therefore, the regression model consists of a training set of 8 months and a testset of 4 months.

3.4. GRASP

In this section, first the background of the GRASP model is discussed. Afterward, the implementation of the GRASP model is described. This is done by explaining the implementation of the roadway, going through the terrain conditions, and presenting the model set-up.

3.4.1. Background of GRASP

As mentioned in chapter 1, GRASP is an application of an LES model. LES models have been developed to simulate turbulent flows. This is done by simulating the turbulence caused by the largest eddies in the planetary boundary layer and using parametrizations to simulate the smaller eddies. The larger eddies are resolved using the Navier-Stokes equations, and the smaller eddies are treated by a sub-grid model to reduce computational costs. The larger eddies are filtered based on the grid size. This process is shown in Figure 3.2.

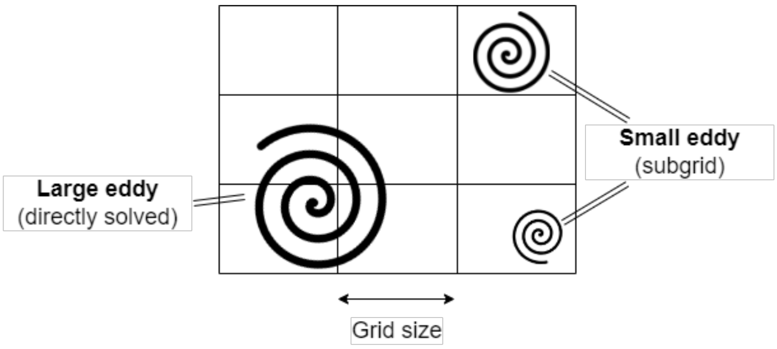
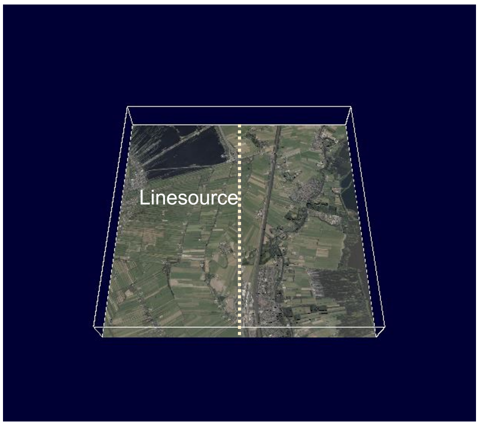


Figure 3.2: Filtering of eddies in Large Eddy Simulation

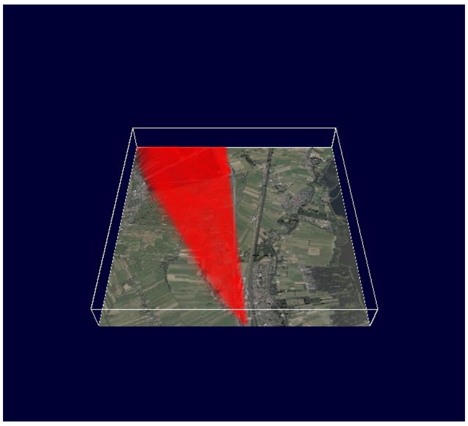
Dispersion of air pollutants in the atmospheric boundary layer is dominated by turbulence since pollutant particles are transported by the turbulent flow (Tomas et al., 2015). As turbulence is resolved within the GRASP model, the model is suitable for predicting the air pollutant concentrations coming from an emission source. In this specific case, the emission source is the highway. Although the computational effort required for the LES has dramatically been reduced since it is possible to run them on a GPU instead of a CPU (Bieringer et al., 2021), the GRASP model still requires considerably more computational power than the other solutions in this research.

3.4.2. Implementation GRASP simulations

In GRASP, the pollutant emissions are assumed to behave as a passive tracer. This implies that the concentration of the pollutant is so low that it does not impact the motion of the fluid. A line source is implemented to simulate the highway as an emission source. This line source spans the domain from north to south with an angle corresponding to the overall direction of the road. In GRASP, a line source is a straight line and cannot simulate the curves of the highway. An emission source is called a tracer in GRASP. A visualization of the implemented line source in GRASP can be seen in Figure3.3. Figure 3.3a shows the GRASP simulation at time t=0. Figure 3.3b shows the GRASP model during the run. The red color represents how the air pollutants from the line source disperse over the domain. In Figure3.3b, the wind is coming from a south-eastern direction, which can be seen in the way that the pollutants spread over the area.



(a) Snapshot of GRASP simulation at timestep t=0

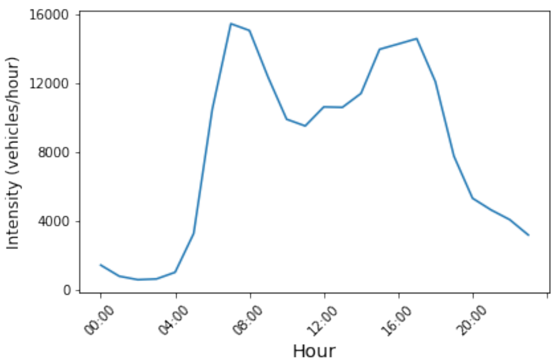


(b) Snapshot GRASP simulation during the run

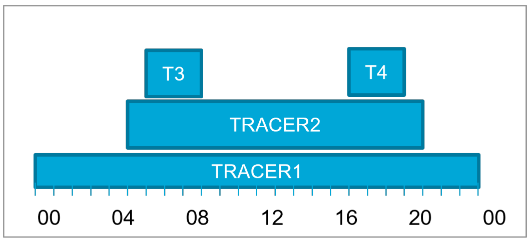
Figure 3.3: Implementation of the line source in GRASP

Simulation of traffic intensities

Furthermore, the traffic intensities described in paragraph 3.2.2 are used as input data for the model to define the emission level of the line sources. In GRASP, the emission strength of the tracer cannot be made variable over time. However, it is possible to implement multiple tracers in the model, which can be turned of/on during the run. Looking at the traffic data in Figure 3.4, a clear daily pattern can be distinguished in which two peaks are visible. Those peaks represent the rush hours that occur twice a day, see 3.4a. Since the GRASP tracer can not be variable over time, a workaround is chosen in which multiple tracers together recreate the daily traffic profile. This workaround is visualized in Figure 3.4b. The different blocks each represent a tracer turned on and off during a model run. The emission strengths of the tracers are simulated such that at rush hour, the sum of the active tracers equals one. The values belonging to the different tracers in the GRASP model are included in appendix A.



(a) Average daily traffic intensity at Breukelen A2 in 2019



(b) Implementation of traffic profile in GRASP as tracers

Figure 3.4: Simulating traffic intensity in GRASP

Terrain conditions

The simulation is set up so that the conditions at the A2 near Breukelen are as accurately replicated as possible to legitimize the comparison between GRASP and the measurement data. In GRASP, obstacles in the domain are taken into account through the usage of a height map in which altitudes from a Dutch elevation map are included. The heights of the obstacles in the domain are depicted in Figure3.5. The red star in the picture represents the measuring station.

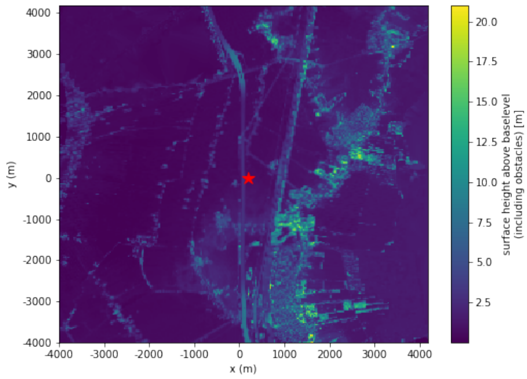


Figure 3.5: Height map of simulation domain GRASP (the red star represents the measuring station)

Set-up GRASP simulations

The model parameters are defined in Table 3.1. The values of the parameters were established by choosing an initial value and calibrating these values through model runs. The domain size should be large enough to represent the situation at the selected location accurately but should not be larger to reduce run-time. The domain size consists of two elements, the grid size and the number of grid points. The grid size is the resolution of the model and defines the distance of the boxes over which the atmospheric variables are averaged. Variable N stands for the total number of grid points. The grid points multiplied by the grid size together define the domain size. The center of the grid was chosen such that the highway is positioned in the middle of the domain. In GRASP, a metmast can be used to record statistics of a range of variables at a specified location in the domain. Different spots were tested to decide upon the position of the metmast. The metmast must be close to the highway to simulate the measurement station, as this one is also placed right next to the highway. However, the line source and the metmast have to be placed in different grid cells to simulate the effect of turbulence between the line source and the metmast. Therefore, two positions of the metmast are examined in different cells next to the line source. The first position was selected one grid box away from the line source. This represents approximately 40 meters since the grid size is 32 by 32 meters. The second location is two grid boxes away from the metmast, which is about 70 meters away from the line source. It was seen that the trend in the model was better maintained at the location closest to the line source. The graphs belonging to this analysis are placed in appendix B. Based on this analysis, the location of the metmast at one grid cell away from the line source was selected. The values of the variables are computed every minute ($t_{sampling}$), but stored only every hour to save computational power (t_{write}).

Parameter	Variable	Value
Domain size	L_x, L_y, L_z	8192 x 8192 x 1024
Grid size	d_x, d_y, d_z	32 x 32 x 8
Grid points	N_x, N_y, N_z	256 x 256 x 128
Grid center	Lat, Lon	52.20126, 4.98660
Metmast position	Lat, Lon	52.20142, 4.98820
Sampling interval	$\Delta t_{sampling}$	60
Writing interval	Δt_{write}	3600

Table 3.1: Model set-up GRASP

3.5. Gaussian Plume Model

As mentioned in chapter 2, the current model used for defining air quality policy in the Netherlands is a GPM. Therefore, the GPM is included as a benchmark in this research, as the GRASP model and the GPM should, at least to some extent, perform better than the GPM to add value for decision-making about air quality.

3.5.1. Background of Gaussian Plume Model

According to Reed et al. (2005) Gaussian models are the most common mathematical models used for air dispersion. They are built on the assumption that the contaminant will diffuse according to a normal distribution (Reed et al., 2005). The GPM is most often used to model point-source emissions such as chimneys. The Gaussian dispersion formula that belongs to these models, as described by Jain (2015), is shown below.

$$X = \frac{Q}{2\pi * u_s * \sigma_y * \sigma_x}[\exp(-0.5 * \frac{y}{\sigma_y}^2)][\exp(-0.5 * \frac{H}{\sigma_z}^2)] \tag{3.4}$$

where;
X = hourly concentration at downwind distance x, $\mu g m^-3$,
u_s = mean wind speed at pollutant release height, m/s,
Q = pollutant emission rate, $\mu g/s$,
σ_y = standard deviation of lateral concentration distribution,
σ_z = standard deviation of vertical concentration distribution,
H = pollutant release height (stack height), m,
y = crosswind distance from source to receptor, m,

The stability coefficients σ_y and σ_z are computed using the Pasquill-Gifford-Turner Stability Classifications, as described in Turner (1970). This classification distinguishes six stability classes and defines the relationship between the stability classes and the coefficients. The theory defines the stability coefficients based on solar insolation, cloud coverage, and wind speed. In general, it can be stated that the atmosphere is more stable during nighttime due to low solar insolation. During the daytime, the stability coefficient is based on the wind speed and the strength of the solar insolation. At night time, the stability is classified based on the cloud coverage and the wind speed. The Pasquill-Gifford-Turner Stability classifications and their corresponding stability coefficients are included in appendix E.

Given that it is an analytical solution computing the NO_2 concentrations using the GPM equation requires little computational effort (Tomas et al., 2015). However, due to its underlying assumptions, the GPM presents a highly simplified picture of reality. Those assumptions include constant meteorological conditions, flat terrain, a negligible deposition, and a conical plume shape (Reed et al., 2005). Furthermore, a line source can not be easily implemented in the model since the integration over a finite line with the Gaussian dispersion formula is only accurate whenever the wind is perpendicular to the line source (Briant et al., 2013).

3.5.2. Implementation GPM

For the implementation of the GPM, an open-source code created by P. Connoly (2018) is modified and used. The Gaussian dispersion formula, as presented in section 3.5.1 forms the base of this code. ERA5 data is used for the weather variables to make an accurate comparison between GRASP and the GPM. This includes wind speed, wind direction, solar insolation, and cloud coverage. Wind speed, solar insolation, and cloud coverage are used for computing the atmospheric stability coefficients.

Simulating traffic intensities

For simulating the traffic intensities, the same approach is used as in the implementation of the GRASP model as described in section 3.4.2.

Terrain conditions

Since the GPM assumes a flat terrain, no obstacles are included in the simulation set-up. Therefore, the only similarity between the domain and the A2 near Breukelen's landscape is the relative location of the roadway and the metmast.

Set-up GPM

The domain size and resolution are comparable with the GRASP domain size as seen in Table 3.2. However, due to implementation choices in the models, the grid size and grid points are slightly different for the GPM and GRASP model. The position of the metmast is chosen such that the metmast is one grid box away from the line source. As the Gaussian dispersion formula is dependent on its previous values, the sampling interval equals the writing interval, which is one hour.

Parameter	Variable	Value
Domain size	L_x, L_y, L_z	8000 x 8000 x 500
Grid size	d_x, d_y, d_z	50 x 50 x 10
Grid points	N_x, N_y, N_z	160 x 160 x 50
Grid center	Lat, Lon	52.20126, 4.98660
Metmast position	Lat, Lon	52.20126, 4.99028
Sampling interval	$\Delta t_{sampling}$	3600
Writing interval	Δt_{write}	3600

Table 3.2: Model set-up GPM

As mentioned in paragraph 3.5.1, a line source can not easily be implemented in the GPM. Therefore, instead of a line source, a sequence of point sources is used. In Figure3.6, a cross-section of the GPM in the z-direction is shown. The figure shows the sequences of points and the emission dispersion over the xy-plane. The red star represents the location of the metmast in the model. For the implementation of the traffic intensity, the same method is used as for GRASP, in which the emission source follows the values of the average traffic profile in the same proportions. The values used for the emission source are included in appendix A.

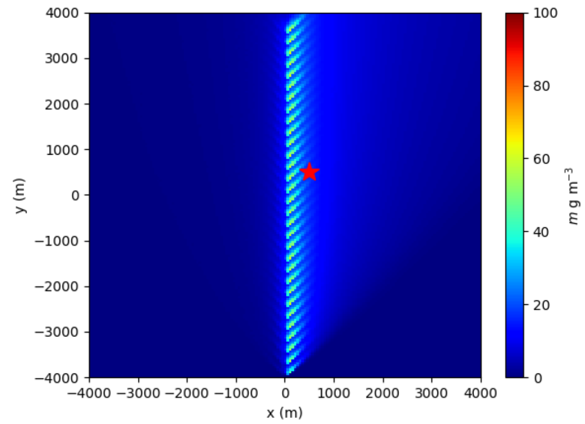


Figure 3.6: Implementation of the line source in GPM (the red star represents the measuring station)

3.6. Regression models

The prior models, GPM and GRASP, are based on physical formulas that simulate atmospheric conditions. Besides these approaches, also a data-driven approach is used to predict NO_2 values near the A2 at Breukelen. The computed model is a polynomial regression that fits within the OLS method. As mentioned in section 1.1, OLS regression is a relatively simple machine learning technique that is shown to be accurate in many situations and has already been proven to be suitable for predicting air quality in Mahanta et al. (2019).

3.6.1. Formalisation of the regression model

Equations of model

In polynomial regression, the relationship between the independent and dependent variables is represented by an nth degree polynomial. The Polynomial regression model that is created is fit with the OLS method, in which the variance of the coefficients is minimized under the Gauss Markov Theorem. The maximum order or degree of the polynomial coefficients can be optimized for the model. This is done by computing the RMSE on the predictions of the regression models with different degrees. The degree that leads to the lowest RMSE is chosen. The formula for a second-order multiple regression with three regressors (x_1, x_2, x_3) is the following:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + +\beta_{23} x_2 x_3 \tag{3.5}$$

where;
$\hat{y}$ = Predicted value for dependent variable
x_i = Regressor value
β_i = Effect parameters

Model input

The dependent variable in the regression model is the NO_2 concentration. The regression model uses three different features for the prediction of the dependent variable, namely the ERA5 windspeed, ERA5 wind direction and the traffic intensity. These are called the regressors. As the ERA5 wind direction is a circular variable, it has to be transformed before it can be used as input data for the regression model. This is done through a method described by Jammalamadaka & Lund (2006). In this method, the sample (non-circular) direction is computed by first taking the sine and cosine of the wind direction, as shown in equation 3.6.

$$\hat{a} = \begin{cases} \arctan \sin a_i / \cos a_i, & \text{if } \cos a_i \geq 0 \\ \arctan \sin a_i / \cos a_i + \pi, & \text{if } \cos a_i < 0 \\ \text{Undefined}, & \text{if} \sqrt{\sin {a_i}^2 + \cos {a_i}^2} = 0 \end{cases} \tag{3.6}$$

where;
$\hat{a}$ = Mean direction of sample,
a_i = Circular wind direction,

The sample mean direction $\hat{a}$ is no longer a circular variable, and can be used as input for the regression model.

Model implementation
The regression model is implemented using the scikit-learn library in python (Pedregosa et al., 2011). The data used for creating and testing the regression model consists of one year and 2 months. The data is split into a training- and a test-set. The training set includes 70% of the data and is used to create the model. The other 30% of the data is allocated for the test-set and is used to test the model's prediction capacity. Furthermore, the maximum degree of the polynomials is specified. To find the degree for which the model yields the best results, the RMSE on the test set is calculated for different degrees of the model. Choosing the right degree for the model is important; when the degree of the model is chosen too low, the model cannot capture the relation between the input and output data. However, a degree that is too high leads to overfitting of the model to the training set. In Figure3.7, the degree of the model is plotted against the RMSE. The plot shows that at a degree of 3, the RMSE is the lowest. Therefore, the chosen degree for the model is 3.

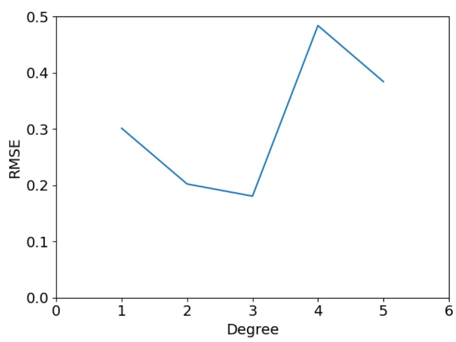


Figure 3.7: Analysis of polynomial degrees for regression model

3.7. Combination of models

Next to comparing the models, it is examined whether a combination of the GRASP model and the regression model can increase the performance of the models. Two regression models are created for this purpose. In the first model, the GRASP pollutant predictions are added as a regressor, next to the ERA5 windspeed, ERA5 wind direction, and the traffic intensities. This regression model is created to see whether the outcomes of GRASP contain additional information that improves the regression model.
The second model uses the wind speed and direction predicted in GRASP for the location of the measurement station instead of the ERA5 data. In this model, the regressors are the GRASP wind speed, GRASP wind direction, and the traffic intensities. For the construction of both models, the data is again divided between the training- and the test-set with the ratio of 0.7/0.3. The degrees of the models are computed using the same approach as in 3.6. The degree chosen for both regres-

sion models with GRASP input is three. In appendix C, the results of the different degrees for the regression models with GRASP data are shown.

4 Results and discussion

In this chapter, the results of the models are presented. First, the post-processing process of the models' predictions is discussed. Subsequently, the results of the GRASP model, the GPM and the regression model are presented and discussed. Afterward, the results generated by a combination of the regression model and the GRASP results are shown and compared with the original models. This chapter concludes upon the third and fourth sub-question, as defined in chapter 1.

4.1. Post-processing of outcomes

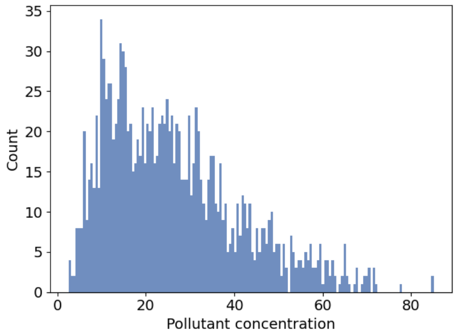
The results of the models are post-processed to make the output data of the models suitable for comparison. In this section, the steps taken in this process are discussed. The analyses in this section are based on model runs for two months.

4.1.1. Normalization of outcomes

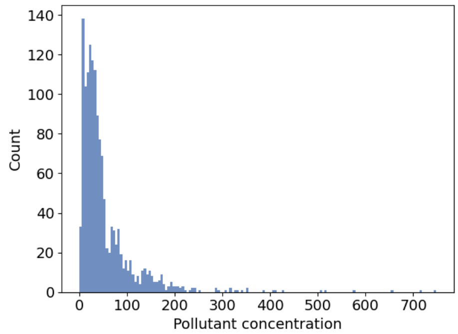
The emission strength of the line sources, based on the traffic intensities that are implemented in the GRASP model and the GPM, follow the same pattern that is based on Figure 3.4 (see appendix A for exact values). However, due to implementation choices in the GPM and GRASP model, the absolute pollutant emissions cannot be compared with each other. It is chosen to re-scale the outcomes to allow for the comparison between the output values of the different models. The re-scaling method that is used is normalization. Frequently used techniques for normalization are Z-score normalization and min-max scaling. The results data characteristics are examined to decide upon the normalization method. Almost all outcome datasets follow a positively skewed distribution, as seen in the distributions in Figure 4.1. Since Z-score normalization requires normally distributed data (Ali et al., 2014), this is not appropriate. In mix-max normalization, there are no assumptions about the distribution of the data. For this reason, min-max scaling is chosen. The formula that is used for performing normalization is the following:

$$X_{new} = (X - X_{min})/(X_{max} - X_{min}) \tag{4.1}$$

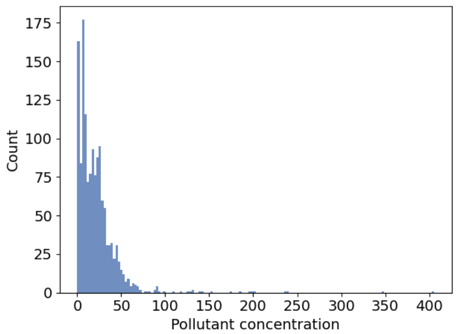
where;
X_{new} = normalized model prediction,
X = model prediction,
X_{min} = minimum value in dataset,
X_{max} = maximum value in dataset



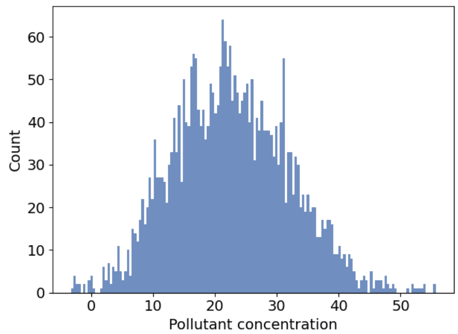
(a) Measurements



(b) GRASP



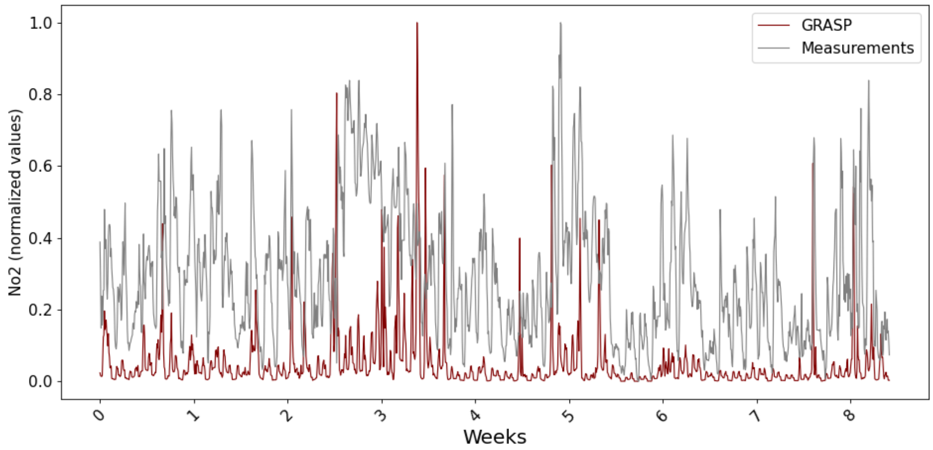
(c) GPM



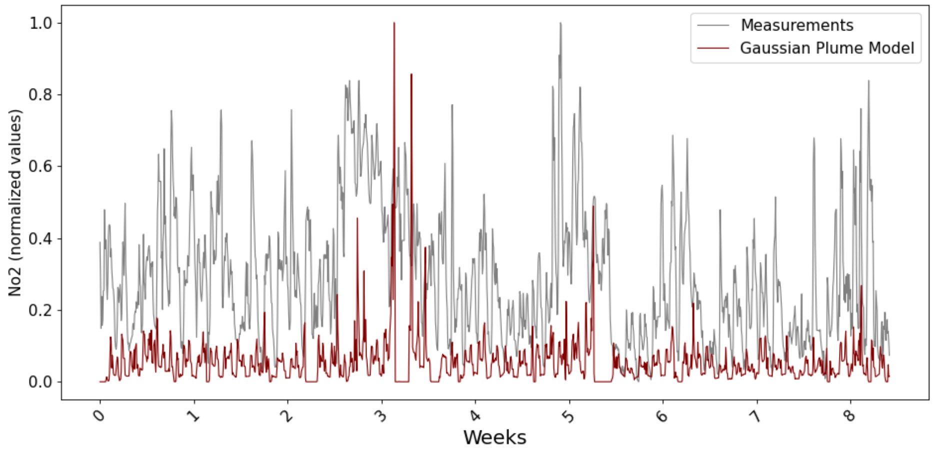
(d) Regression model

Figure 4.1: Distributions of GRASP, GPM, regression model results before postprocessing and distribution of NO_2 measurements

When looking at the distribution in Figure 4.1 closely, it can be concluded that both the GRASP and the GPM outcomes contain outliers since some of the pollution concentrations, shown on the x-axis, become very high compared to the average value. However, min-max scaling requires a dataset without outliers. The impact of min-max scaling on the GRASP and GPM results with outliers can be seen in Figure 4.2. Due to the min-max scaling, the outliers cause a plot in which most values are close to zero.



(a) GRASP predictions



(b) GPM predictions

Figure 4.2: Timeseries of normalized model predictions with outliers in the data

4.1.2. Outliers

In the previous section, it was shown that both GPM and GRASP results contain outliers. This section examines the characteristics of the outliers in the GPM and GRASP results, and a procedure to deal with those outliers is defined.

Outliers in GRASP

To determine which values in the GRASP results are outliers, a thresholding method is used based on the interquartile range (IQR). In Yang et al. (2019), it is stated that defining outliers based on the interquartile range is a frequently used approach for defining an outlier threshold. In general, the threshold in the IQR thresholding method is defined as follows:

$$T_{min} = Q1 - c * IQR \tag{4.2}$$

$$T_{max} = Q3 + c * IQR \tag{4.3}$$

$$IQR = Q3 - Q1 \tag{4.4}$$

where;
T_{min} = Minimum threshold,
$Q1$ = First quartile,

$Q3$ = Third quartile,
c = Constant value defining conservativeness,
IQR = Interquartile range,
H = pollutant release height (stack height)

In the GRASP and GPM results, only outliers above the third quartile are present; therefore, only formula 4.3 is used. For c, a value of 3 is chosen, as Taylor (2020) defines this as the value for which strong outliers are removed. It is vital that high NO_2 predictions are not falsely identified as outliers, as high NO_2 values are critical to be predicted. Therefore, a conservative, relatively high value is used for c. In appendix D, it is visualized which values are considered outliers in the GRASP results.

Whenever the outlier values in GRASP are examined more closely, it can be seen that most of these values share the same wind characteristics. In Figure 4.3, the GRASP results are plotted against the wind speed and the wind direction.

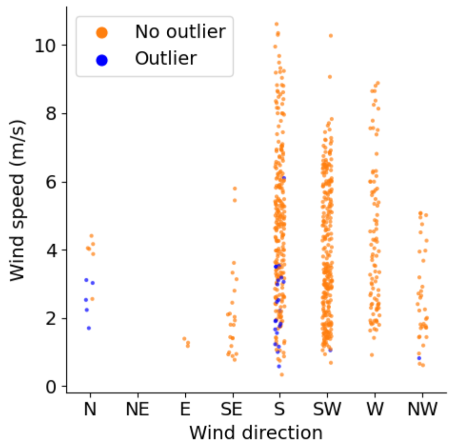
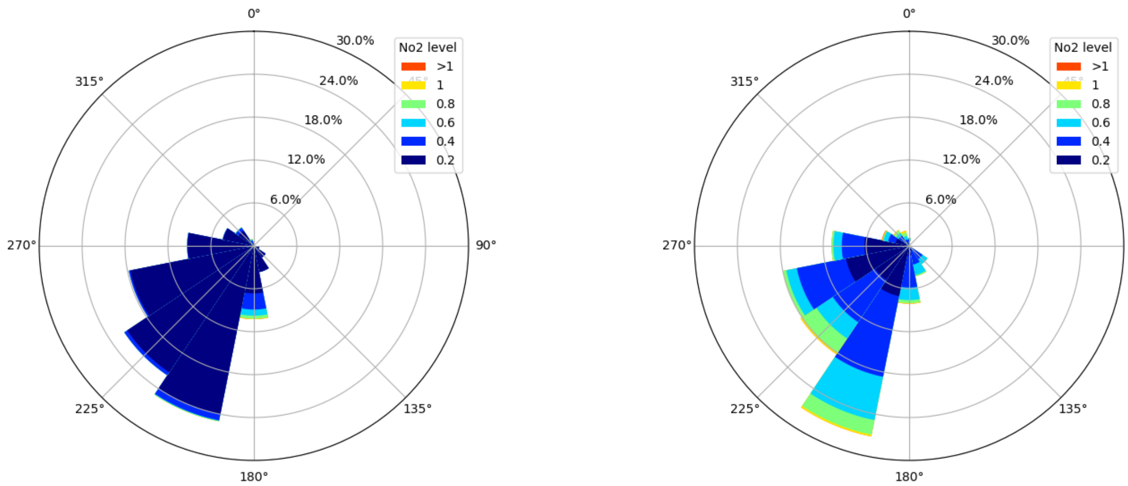


Figure 4.3: Wind speed and wind direction of outlier values in GRASP results

Low wind speed generally results in a smaller dispersion area and, therefore, to higher pollutant concentrations. If the wind comes from the north or south, the wind direction is parallel to the highway. These wind directions result in an accumulation of air pollutants over the length of the highway, as air pollutants move over the direction of the highway and not away from the highway. In other words, the high values of the outliers in GRASP can be partly explained by their weather characteristics. However, it seems like the GRASP model overestimates the effect of those weather characteristics, as the concentrations reach unrealistically high values much higher than the measurements, as can be seen in the distribution graphs in 4.1, when looking at the values on the x-axis that are reached by the GRASP model and the GPM.

In the Figures 4.4, the pollution roses of both the GRASP outcomes and ERA5 data with the NO_2 measurements are visualized. A pollution rose is a means to illustrate the frequency distribution of wind direction temporally correlated with a pollutant. The wind speed, wind direction, and the normalized pollutant predictions of GRASP are used to create the pollution rose in Figure 4.4a. The pollution rose in Figure 4.4b uses the wind direction and wind speed from ERA5. The pollution roses show that between 24% and 30% of the wind speed came from a south-western direction in the considered time period. This can be seen in the plot as the plume towards 225 degrees of the circle is within the two most outer rings, representing 24% and 30%. The wind directions of GRASP and ERA5 are slightly different since GRASP predicts the wind direction more locally. The colors represent the NO_2 values. In the pollution rose for GRASP, it can be seen that the pollution levels are at their maximum when the wind is coming from the south. This is a wind direction parallel to the A2, where the air pollutants are stacking up over the length of the highway instead of dispersing over a larger area. This effect is not to the same extent visible in the pollution rose of the ERA5 data and the measurements. From the pollution rose of GRASP, it becomes clear that the outliers in GRASP come from a wind direction parallel to the highway. When comparing both pollution roses in figure 4.4, it is seen that the effect of the normalization of the GRASP data with outliers results in a majority of low values, seen from the dark blue color. For the pollution rose of the ERA5 data and the measurement, the normalization results in a wider variety of normalized NO_2 levels, seen from the wider variety of colors in Figure 4.4b.



(a) GRASP

(b) Measurements and ERA5 data

Figure 4.4: Pollution rose of GRASP after normalization with outliers and pollution rose of NO_2 measurements and ERA5 data

Based on this analysis, the hypothesis is that the combination of a low wind speed and a wind direction parallel to the highway leads to high NO_2 values in GRASP. The reason that GRASP predicts such high values for those conditions has most probable to do with the fact that GRASP does not account for turbulence on a very small scale (yet). Thus, the slightly larger eddies may or may not reach the metmast, resulting in an all-or-nothing effect. This effect is reinforced by the fact that the turbulence created by the cars on the highway is not included in the GRASP model, as a simple line source is used to simulate the highway. Therefore, the hypothesis is that the combination of the latter explanation and the unfavorable weather conditions cause large outliers in the GRASP results.

Outliers in GPM

The same approach for determining the outlier threshold as in GRASP is used for the GPM. The value chosen as outlier threshold is shown in context to the results in appendix D. The GPM results contain fewer outliers, so the outlier threshold is lower than for GRASP based on equation 4.3. In Figure 4.5, the outliers are visualized in a scatter plot that shows the atmospheric stability on the x-axis and the wind speed on the y-axis. Atmospheric stability is categorized in the six categories of the Pasquill-Gifford-Turner Stability Classifications. Atmospheric stability of 1 represents an extremely unstable atmosphere, and atmospheric stability of 6 is the most stable atmosphere. It can be seen that the outliers in the GPM occur when the atmosphere is stable and the wind speed is low. This corresponds to literature about the GPM, in which it is stated that a stable atmosphere causes high peaks in the outcomes of the GPM (Abdel-Rahman, 2008). Furthermore, just as in the GRASP outcomes, a low wind speed leads to a smaller surface of dispersion and, thus, causes higher pollutant concentrations. This can be seen on the y-axis of 4.5.

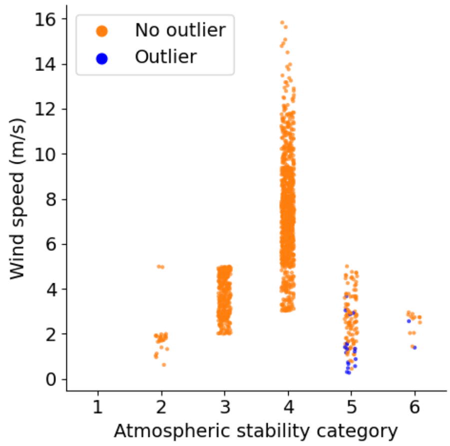


Figure 4.5: Wind speed and atmospheric stability of outlier values in GPM results

Treatment of outliers
The previous sections discussed the conditions under which outliers occur in the GRASP and the GPM model results. It was shown that both models seem to estimate unrealistic high values under certain circumstances. These values highly impact the outcome data and make the normalization unfeasible. Therefore, it is decided to remove the outliers from the outcomes of the models. This is done using forward fill in which the outlier value is replaced by the last non-outlier value.

4.2. Comparison of GRASP, GPM and regression model

This section presents the results of the post-processed output data of the GRASP model, the GPM, and the regression model. The GRASP and the GPM results show a period of a year. For the regression model, only the test set is visualized in the results, containing four months of data. In this section, firstly, the scores of the models on the performance metrics are presented. Secondly, the distributions of the results and the timeseries are visualized. Furthermore, the performances of the models in different circumstances are shown. After the results are presented, they are interpreted and discussed.

4.2.1. Performance metrics

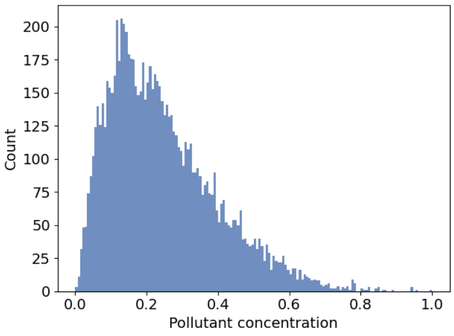
Table 4.1 shows the scores of the models on the performance metrics. Both the GRASP model and the regression model outperform the GPM in terms of RMSE, MAE, and Pearson correlation. The regression model scores best compared to the GRASP and the GPM model.

Original models			
	MAE	RMSE	Pearson correlation
GRASP	0.1633	0.2160	0.4154
GPM	0.2169	0.2598	0.28230
Regression model	0.1471	0.1807	0.6612

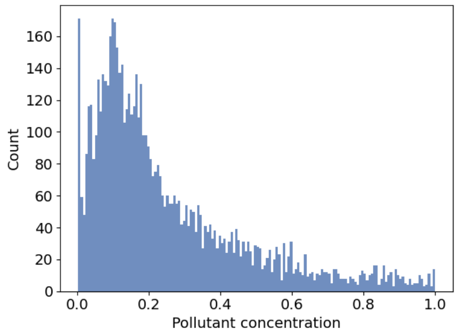
Table 4.1: Performance metrics of GRASP, GPM and regression model

4.2.2. Distributions

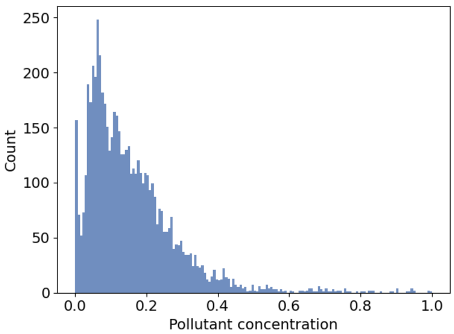
In Figure 4.6 below the distributions of the predicted pollutant concentrations are shown. Both the GRASP results, the GPM results and the measurements follow a positively skewed distribution. The results of the regression model approximates a normal distribution. Based on these plots, the distribution of the GRASP model and the GPM look most similar to the distribution of the measurements.



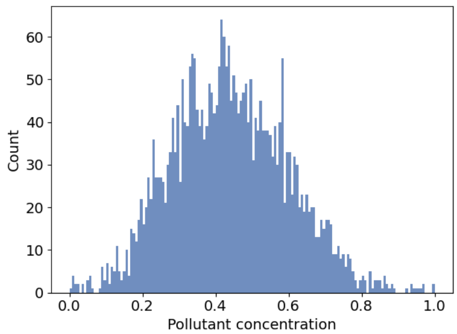
(a) Measurements



(b) GRASP



(c) GPM



(d) Regression model

Figure 4.6: Distributions of GRASP, GPM, regression model results and NO_2 measurements

4.2.3. Performance of models for different circumstances

The performance of the models is also evaluated for the different months in the year and for the different wind directions. This is done by computing the RMSE per month and per wind direction.

Performance per month

In Figure 4.7, a visualization of the RMSE of the models for all months in the year is shown. The results of the regression model consist of four months, as only the test set is taken into consideration. It can be seen that the performance of GRASP is relatively stable, ranging from an RMSE score of approximately 0.12 to 0.25. In general, the RMSE for the GRASP results is slightly higher in the winter period (November till February). The performance of the GPM model has more variations over the months, ranging from 0.12 in April to 0.34 in February. Also, the RMSE of the GPM results is larger in winter. The RMSE values for the regression model are only computed in the winter, with a low variation per month over the test set.

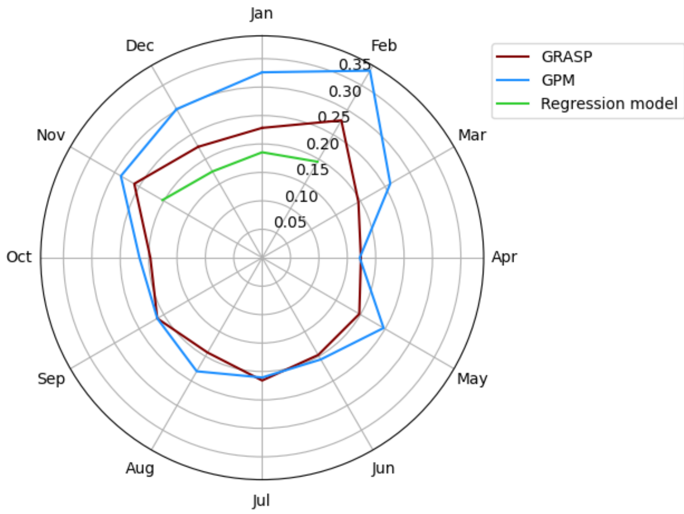


Figure 4.7: RMSE of the GPM, GRASP model and regression model per month over a year

Performance for different wind directions

Figure 4.8 shows the RMSE for the different wind directions. The GRASP model shows the lowest RMSE for a western wind direction, and the highest RMSE for a northern wind direction. The western wind direction is the wind direction for which the pollutants are blown from the highway towards the metmast, which generally leads to higher pollution than the eastern wind direction. The northern direction is also the wind direction that occurred least often, as seen in the pollution roses in appendix F. Hence, a large error already significantly impacts the wind direction's score, so this value is less representative. For the GPM, the southern wind direction leads to the highest errors. The RMSE values of the regression model are relatively stable and do not show large differences in performance for different wind directions.

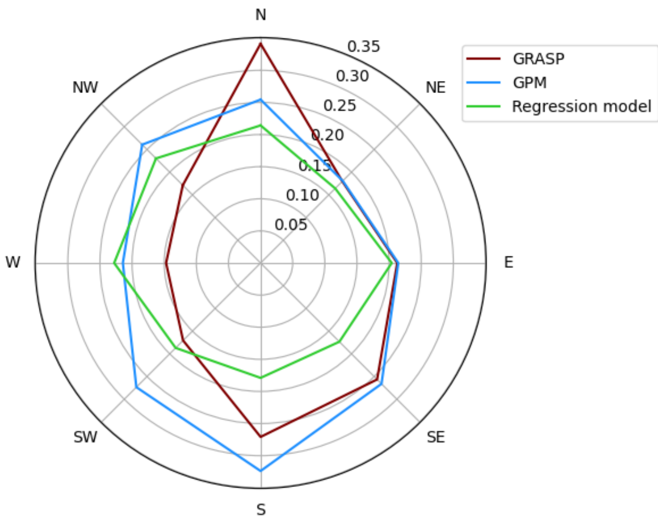
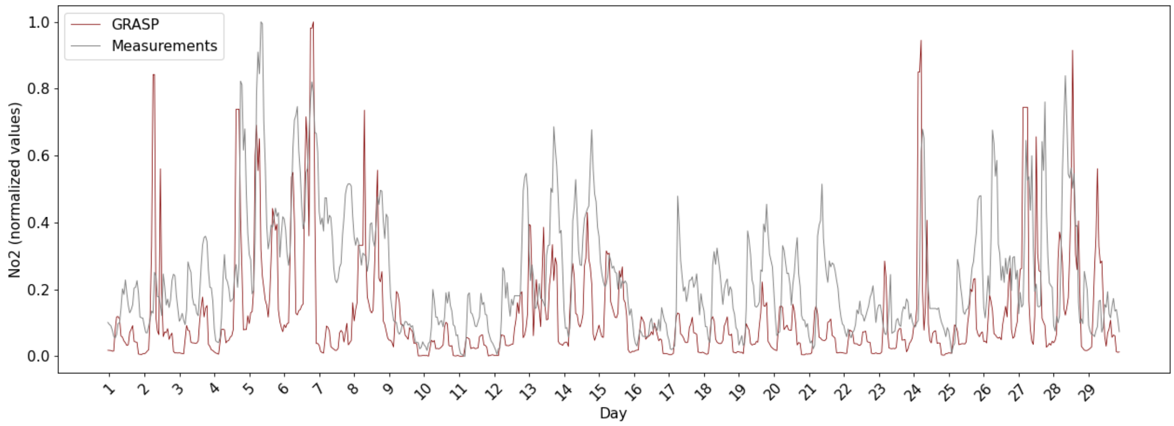


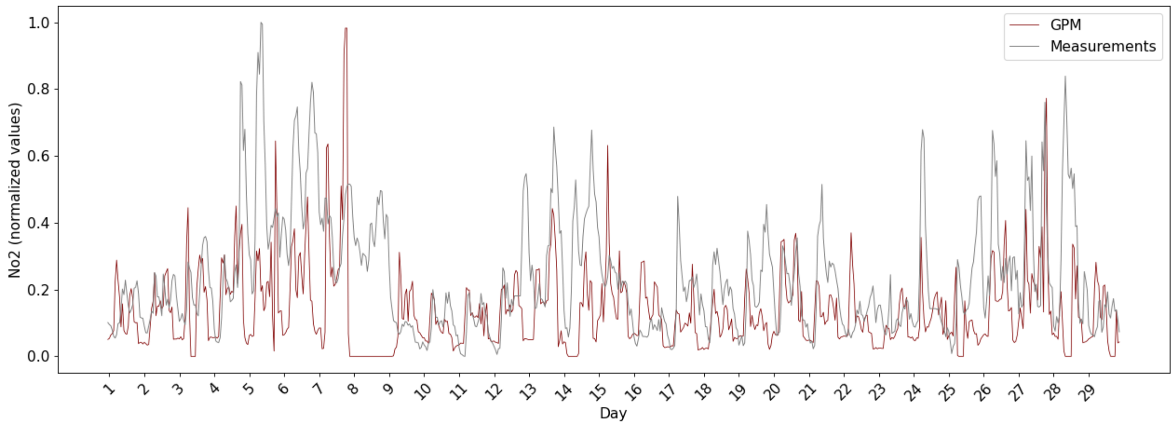
Figure 4.8: RMSE of the GPM, GRASP model and regression model for different wind directions

4.2.4. Time plots

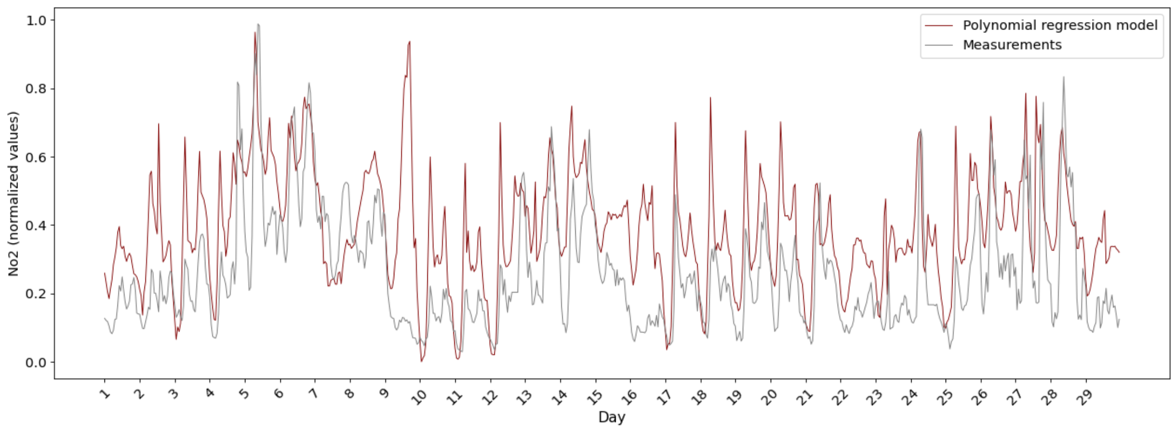
The timeseries 4.9 show the predictions of the models and the measurements for one month, namely February 2020. It is chosen to visualize one month to keep the plots uncluttered. The timeseries are shown to get insight in the behaviour of the models.



(a) Comparison GRASP and measurements over a month



(b) Comparison GPM and measurements over a month



(c) Comparison regression model and measurements over a month in the test set

Figure 4.9: Timeseries showing the behavior of models over a month

Daily and weekly trend

As visible in the timeseries, the measurement data shows daily and weekly patterns. The first pattern consists of two clear peaks daily, caused by the rush hours that occur twice a day. Next to the daily pattern, also a weekly pattern is visible. It can be seen in the data that after approximately five days with high peaks in the rush hours, the peaks are lower for two days. This pattern coincides with the week and weekend days. For example, 22 and 23 February were weekend days, and 24 till 28 were weekdays.

These patterns are, to some extent, visible in the results of the models. The results of the GRASP model and the GPM show commonalities in capturing the weekly and daily trend, as shown in Figure 4.9a and Figure 4.9b. In the predictions of the GRASP model and the GPM, a daily pattern can be distinguished that approaches the daily variations in the measurement data. The weekly trend is more difficult to discern in the GRASP model results and the GPM result. This notion is explained by a limitation in the model set-up of the GPM and the GRASP model. Both models use an average traffic profile for the emission strength of the line source. Because of this, the models cannot predict variations in pollutant concentrations due to differences in traffic intensities over the days, such as the difference between weekend and weekdays. Variations in pollutant concentrations between days are solely caused by differences in meteorological conditions in the models. To address this limitation, the GRASP model and the GPM could be adapted in future research so that the emission strength of the line source in the models is reliant on actual traffic data.

The predictions of the regression model over the test-set in Figure 4.9c show that the regression model captures both the daily and the weekly trend of the pollutant concentrations. This leads to a low MAE and RMSE and a high Pearson correlation. However, the amplitude of the daily trend is generally higher than in the measurement data, and the regression model overestimates the pollutant concentrations during most hours. Despite this, the outcomes of the regression model show that a large part of the trend in the measurement date can be explained by variations in the three regressors; wind direction, wind speed and traffic intensity. It should be noted that the regression model is the only model in the research that does not require the usage of an average traffic profile. This explains at least a part of the better performance of the regression model.

Sensitiveness of GPM and GRASP model for outlier value

Another fact that stands out in the timeseries is that the GRASP model underestimates the pollution concentration in most hours. This can also be seen in the GRASP pollution rose in appendix F. Furthermore, with a few exceptions, the amplitude of the daily peaks is relatively small in general. A reason for this behavior is the sensitiveness of the GRASP model to the threshold that is chosen for which a data point is considered to be an outlier. The values are normalized and therefore considered relative to their maximum value. This phenomenon also applies to the GPM model, as a threshold value for outliers is also used to post-process the GPM results. However, the original GRASP results contain more outliers; therefore, a higher outlier threshold value is chosen, and the effect is more visible in the GRASP results. Selecting the threshold for which a data point is regarded as an outlier must be done carefully. If the threshold value for outliers is too high, the average of the outcomes becomes lower than the average of the measurements. This latter explanation applies to the results of the GRASP model, where values are systematically underestimated. However, data points are wrongly considered outliers if the threshold value is too low. This research uses a technique based on the interquartile range to define the outlier threshold for GRASP and the GPM. However, it would be better to solve the outlier problem within the models instead of removing them from the output data to minimize intervention in model performances. As no outliers had to be removed for the regression model, this problem does not apply to the regression model. This is another explanation for the better performance of the regression model compared to the GRASP model and the GPM.

Predicting high NO_2 values

High NO_2 values in the measurements are predicted relatively well by the GRASP model. The most prominent peaks in measurement data coincide at least to some extent with high values in the GRASP data that show on days 5, 6, 24, and 28. The GPM poorly captures the high values in the measurement data. The high values in the GPM results do not coincide with peaks in the measurement data. The highest peaks in the measurement data on days 5, 6, 24, and 28 are not incorporated in the GPM predictions. The regression model captures the high values in the measurement data well. However, the regression model also shows high values that do not reflect the measurement data, for example, during day 9 or 18. Therefore, the high values are less reliable. Based on this analysis, it is concluded that the GRASP model is the most suitable for predicting high values in the measurement data.

4.3. Combination of GRASP and the regression model

This section examines the performance of a regression model completed with data from the GRASP model. First, the results of the regression model with the GRASP pollutant predictions are shown, and afterward, the results of the regression model with GRASP weather data are presented.

4.3.1. Performance metrics

The score on the performance metrics of the two regression models with GRASP data is shown in 4.2. The Pearson correlation coefficient is lower for the two regression models that include GRASP data than for the original regression model. For the regression model with the GRASP pollution predictions, the Pearson correlation is only slightly lower than for the original regression model, for the regression model with GRASP weather predictions this difference is larger. Both regression models outperform the GRASP, GPM and original regression model in terms of MAE and RMSE. Of the two regression models with GRASP data, the regression model with the GRASP predictions scores the best in terms of RMSE, MAE and Pearson correlation.

Combi models			
	MAE	RMSE	Pearson correlation
Original regression model	0.1471	0.1807	0.6612
Regression model with GRASP pollutant predictions	0.1180	0.1480	0.6266
Regression model with GRASP weather predictions	0.1257	0.1580	0.5388

Table 4.2: Performance metrics of models with added GRASP data

4.3.2. Regression model with GRASP pollutant predictions

The performance of the regression model that includes GRASP pollutant predictions is shown in Figure 4.10. The model predicts the pollutant concentrations with lower errors than the original regression model, which became clear from the MAE score and the RMSE score. This is also visualized in Figure 4.11, which depicts the difference between the original regression model and the measurements, as well as the differences between the regression model with GRASP pollutant predictions with the measurements. It is seen that the original model has larger deviations for the measurement data than the regression model with GRASP pollutant predictions. The trend of the two regression models are similar, but the amplitude of the regression model with GRASP pollutant predictions as the initial regression model overestimated the magnitude of the daily trend. Furthermore, most of the high values in the measurement data are captured by the regression model with GRASP pollutant predictions, for example at day 5, 6 and 26. However, not all high values in the regression model with GRASP pollutant predictions correspond to high values in the measurements, for example at day 9, there is a peak visible in the regression model with GRASP pollutant predictions that does not exist in the measurement data. In general, it can be stated that the regression model with GRASP pollutant predictions performs better than both the original regression model and the GRASP model when it comes to predicting pollutant concentrations with low errors. However, it can be questioned whether a regression model is the best method for combining the GRASP predictions with the other data since a phenomenon that often arises in regression models is multicollinearity. Multicollinearity occurs when the independent variables of a regression model are correlated. In this model, both the GRASP predictions and the other predictors are added to the model, while GRASP predictions also depend on weather variables. Problems caused by multicollinearity are (i) the regression model becomes very sensitive to a slight change in the coefficients, and (ii) the regression model's coefficients do not accurately represent the effect of a change in the dependent variable for each 1 unit change in an independent variable. However, it is still possible to obtain a good fit for a model in which collinearity is present (Neter et al., 1996). As the regression model aims to make predictions for pollutant concentrations and the goal is not to obtain insight in quantifying the relationship between the independent and the dependent variable, it is still considered to be fit for the purpose. Nevertheless, the regression model with GRASP pollutant predictions is more prone to overfitting the training set due to the multicollinearity. Therefore, another data-driven model type might generate even better results on the test set. For example, a neural network since multicollinearity is not a problem for neural networks as parameters are fitted using backward propagation.

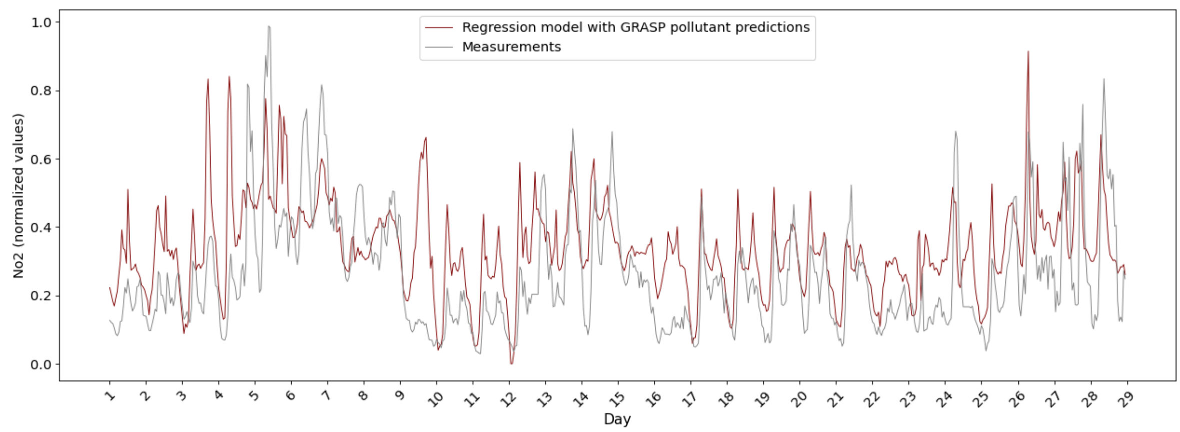


Figure 4.10: Timeseries of regression model with addition of GRASP results as a regressor

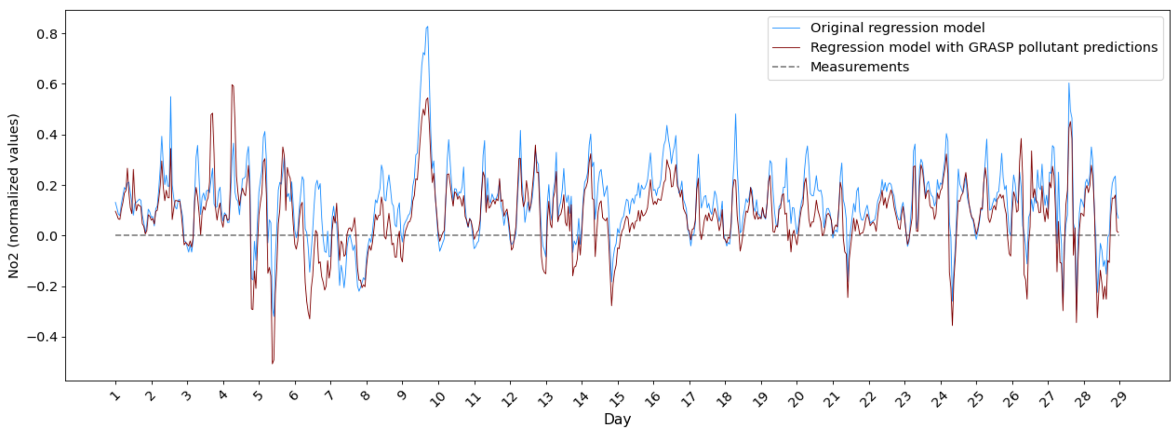


Figure 4.11: Differences between predictions of regression model with GRASP pollutant predictions and measurements compared with differences between original regression model and measurements

4.3.3. Regression model with GRASP weather data

Figure 4.12 shows the regression model when instead of the ERA5 weather data, the GRASP weather data is used for the construction of the model. Also the regression model with GRASP weather predictions predicts the pollutant concentrations with lower errors than the initial regression model, as can be seen in Table 4.2. In particular, the high values in the regression model that were not visible in the measurement data are not present in the regression model with GRASP weather data, which can be seen in Figure 4.13 at for example day 9 and 28. An explanation for the lower errors of the regression model with GRASP weather predictions is that the GRASP weather predictions comprise weather conditions at the exact location of the metmast, where ERA5 data contain more aggregated weather patterns. The differences in wind directions in the pollution roses in appendix F also demonstrate this. This explains the higher accuracy of the model compared with the original regression model. The high values in the measurement data are not captured very well by the regression model with GRASP weather predictions, as the peaks on days 6, 24, and 28 are not captured. Still, the model regression model with GRASP weather predictions could be helpful for computing yearly averages of NO_2 concentrations.

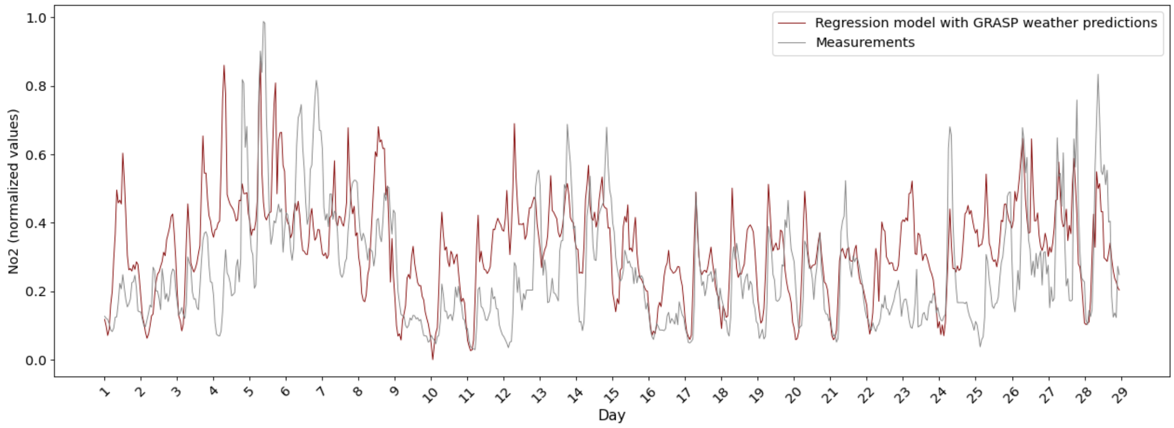


Figure 4.12: Timeseries of regression model with usage of GRASP weather data instead of ERA5 weather data

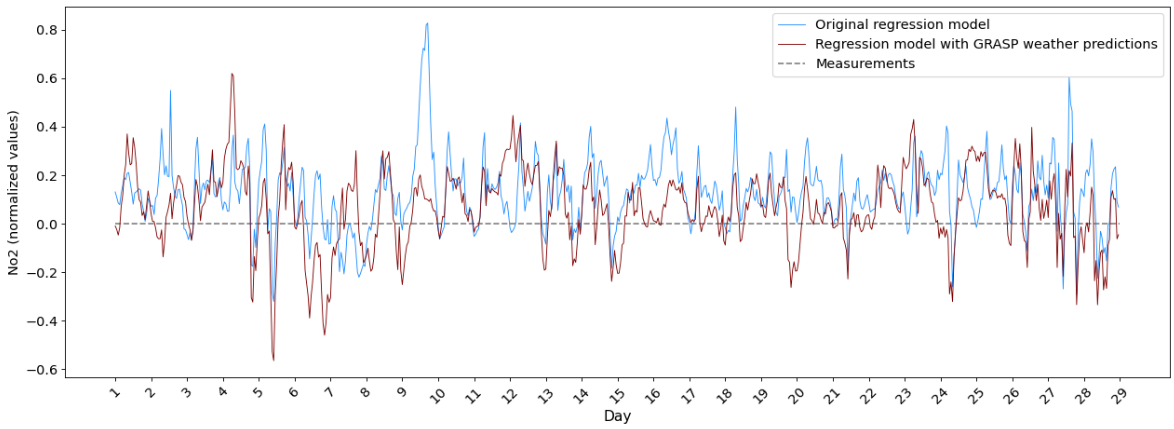


Figure 4.13: Differences between predictions of regression model with GRASP weather predictions and measurements compared with differences between original regression model and measurements

5

Conclusion and further recommendations

Adherence to the limit values defined by the EU in Directive 2008/50/EC is the main priority within air quality policy in the Netherlands. However, the limits set for NO_2 are regularly exceeded around highways. The RIVM is responsible for monitoring the air quality values. To get insight into air quality values, they use a GPM. This research examined whether new, innovative model types such as an LES model and a regression model could result in more accurate air quality predictions. The research question belonging to this study is:

How could air quality predictions around highways in the Netherlands be improved in order to support decision-making aimed at enhancing air quality levels?

To support decision-making in the Netherlands, air quality models should be able to verify if air quality values are below their set thresholds. For NO_2, two thresholds are active: one for the maximum of hourly averages that may only be exceeded multiple times each year and a yearly average that may not be exceeded. Air quality models should be able to monitor whether the air quality values stay below those limits. In the case of NO2, this means that the model can be used to calculate yearly and hourly NO_2 values.

In this research, it is concluded that the current model applied in decision-making in the Netherlands, the GPM, is outperformed by both the GRASP model and the regression models. The GRASP model showed better performance than the GPM, especially in terms of predicting high NO_2 values in the measurements. This could be very useful for identifying whether the hourly limit values are not exceeded.

Although the GRASP model outperforms the current GPM model, there is still room for further improvement in the future. It is seen that the GRASP model is not capable of modeling the variations in traffic intensities between weekdays and weekends. This can be researched in the future by implementing a time-dependent emission profile for the line source. It was also concluded that the GRASP model tends to predict outliers under certain circumstances, which needed to be filtered out in the post-processing step. To reduce the number of predicted outliers, the GRASP model can be extended by including the effect of turbulence on a micro-scale turbulence. With this extension, the performance of the GRASP model becomes less sensitive to the chosen outlier threshold.

The regression models also outperformed the GPM. The regression model led to the lowest error statistics compared to the GPM and GRASP model and is, based on this performance, most suitable for predicting yearly NO_2 averages. The results showed that a large part of the trend in the measurement date could be explained by variations in the three regressors; wind direction, wind speed and traffic intensity. However, the regression model still slightly overestimates most of the NO_2 concentrations and contains high NO_2 values that are not present in the measurement data.

The regression model was extended in this research in two ways; by adding the GRASP pollutant predictions as a regressor and by replacing the ERA5 weather data with GRASP weather predictions. Adding GRASP weather predictions to the regression model instead of ERA5 weather data leads to lower errors in the NO_2 predictions than in the original regression model or in the GRASP model. These smaller errors were mainly due to the reduction of wrongly predicted high values compared to the original regression model. The better performance of the regression model with GRASP weather predictions implies that local weather conditions significantly impact the concentrations measured at the monitoring stations. Adding the NO_2 predictions from the GRASP model to the original regression model resulted in the lowest RMSE and MAE score compared to the other regression models, the GPM and GRASP model. Therefore, combining the GRASP pollutant predictions with a data-driven model seems promising. However, a regression model is not considered the best method to combine the GRASP predictions with a data-driven model due to multicollinearity, which often leads to overfitting of the training data. To improve the predictions even further, a suggested method for examination is a neural network, as multicollinearity is not a problem for neural networks.

The results of the study imply that there is a lot of potential for innovative models such as GRASP and data-driven models for making more accurate air quality predictions. Based on this research, the model that shows the most potential is a combination between the GRASP model and a data-driven model. However, before the implementation of this model within decision-making in the Netherlands, more research is required. The performance of the models is only assessed for one location in this study, which is situated in a rural area with minimal buildings. Examining different locations with other landscapes would be required for drawing final conclusions about the suitability of the models for predicting air quality around major roads. Furthermore, it is interesting to look at other applications of air quality models to see whether the conclusions drawn in this research also apply to those circumstances, for example, the predictions of air quality within urban or industrial areas.

Furthermore, in this research, the suitability of the models in the decision-making process is solely assessed based on the performance of the models. However, other considerations like computational costs and the complexity of maintenance should be considered before deciding upon a model's suitability for societal purposes. Especially within the public domain, criteria such as transparency of the used models are important. The model should not become a black box for the people using it, as this can have major unethical consequences. At this moment, it can be stated that innovative models such as GRASP or machine learning models have a higher complexity than the GPM. Therefore, it is expected that more resources are required by the government to meet the transparency requirements for the GRASP model. However, it is not improbable that those resources will be allocated as more and more research demonstrates the seriousness of the consequences of air pollution and as it has gained greater prominence in the political debate. The implementation and responsible use of new forms of air quality models within policy-making is a pertinent subject of further study.

This research tried to contribute to the domain of air quality modelling. Further research on this topic is highly important, as accurate air quality predictions contribute to more efficient air quality policy. Precise model predictions can prevent misunderstandings about the consequences of changes in spatial planning or can show the effectiveness of measures aiming to improve air quality. It provides the opportunity to know on forehand under what conditions air quality will reach unhealthy values. In those situations, preventive measurements can be taken. For example, on days with unfavorable weather conditions, speed levels on highways can be decreased to maintain healthy air conditions. On other days, when the weather conditions are more favorable, air quality does not have to limit traffic flow. This also applies to other domains, for example, modelling the air quality around industrial areas. In this sense, economic activity and air quality do not obstruct each other but can be optimized to maximize public well-being.

A

GRASP and GPM tracer values

In the GRASP model and the GPM model the total emission strength of the linesource is divided over tracers is four days. The first one represents the background concentration. Based on the averaged traffic profile, Tracer 1 is set on 20% of the total emission strength. Tracer 2 includes the concentrations at daytime. This one is set on 40% of the total emission strength. This means that at daytime, the strength of the linesource is minimal 60% of the maximum emission strength, since tracer 1 and tracer 2 are both working. The third and the fourth tracer represent the rush hours. Those rush hours last for three hours and during rush time, the emission of the linesource is at its max. The tracer values and times are shown in table A.1.

	Duration	Start time	End time	Relative emission strength
Tracer 1	24 hours	00:00	23:59	0.2
Tracer 2	16 hours	05:00	21:00	0.4
Tracer 3	3 hours	06:00	09:00	0.4
Tracer 4	3 hours	15:00	18:00	0.4

Table A.1: Values of tracers in GRASP and GPM based on traffic intensities

B Metmast position GRASP

In Figures B.1 and B.2, the results of the GRASP model for a period of a week is shown. It can be seen that when the metmast is placed further from the linesource, there is less variation in the model results. This is particularly visible between day 5 and day 6, as the model that is located further from the linesource does not pick up differences in No2 concentrations, where the model that is located one grid box from the linesource does. For this reason, the metmast is positioned in the GRASP model at one gridbox away from the line source.

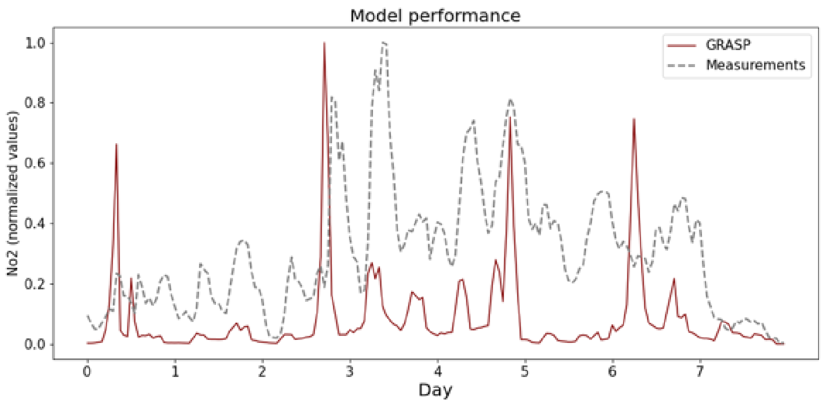


Figure B.1: Location of metmast in GRASP one gridbox from line source

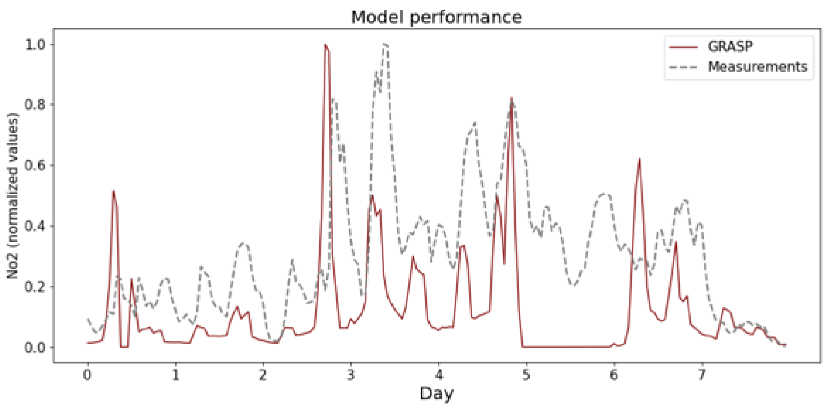


Figure B.2: Location of metmast in GRASP two gridboxes from line source

C

Analysis of degrees for regression models with GRASP input

C.1. Regression model with GRASP pollutant predictions

In Figure C.1 it can be seen that the performance of the model with different degrees is slightly unstable. The model does not clearly converge to one optimum. However, the RMSE is the lowest at a degree of 3, and therefore this value is chosen for the regression model with GRASP pollutant predictions.

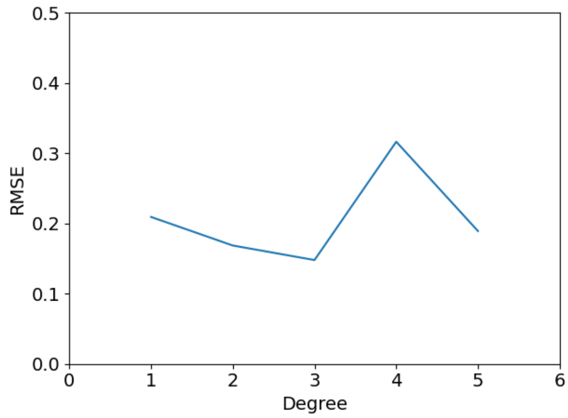


Figure C.1: Analysis of polynomial degrees for regression model with GRASP pollutant predictions

C.2. Regression model with GRASP weather predictions

Figure C.2 shows that the model in which GRASP weather data is added performs the best at a degree of 3. The graph clearly converges to an optimum of 3 and rises again if the degree becomes higher.

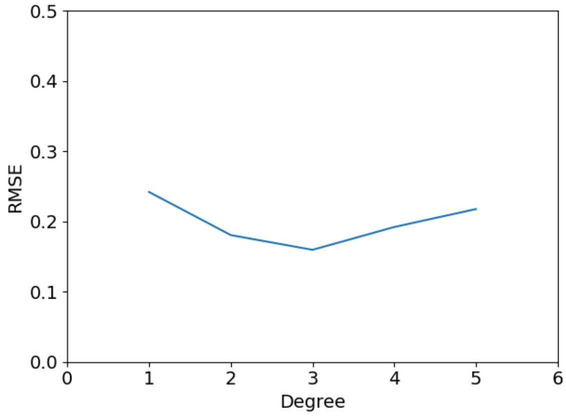
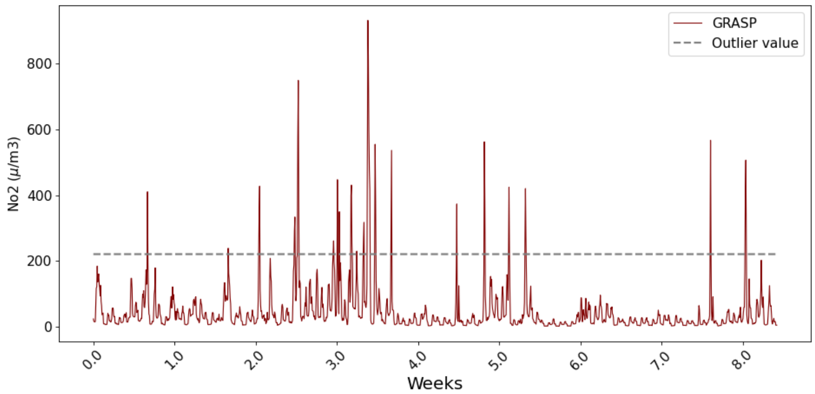


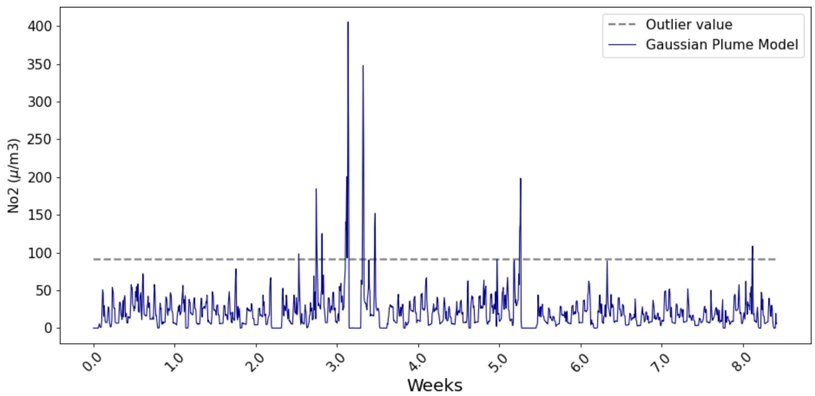
Figure C.2: Analysis of polynomial degrees for regression model with GRASP weather predictions

D
Determining outliers

In the plots below, the outlier threshold value is shown in context to the results. In figure D.1a, it can be seen that the GRASP results contain relatively more outliers than the GPM results, as visualized in D.1b. The method that is used to compute the outlier threshold is based on the interquartile range.



(a) GRASP



(b) GPM

Figure D.1: Outlier values

E

Pasquil-Gifford-Turner Stability Classifications

The figures in this appendix are obtained from Turner (1970).

E.1. Pasquill table

In the table below, the Pasquill-Gifford-Turner Stability categories are defined. These categories are used in the Gaussian Plume Model to represent the atmospheric stability. A stability classification 'A' means that the atmosphere is very unstable and the stability classification 'F' means that the atmosphere is moderately stable.

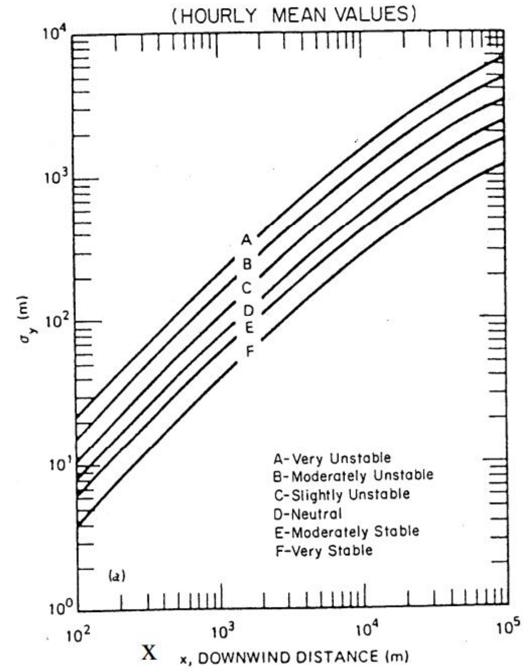
Atmospheric Stability Classifications

Surface wind speed[a] (m/s)	Day solar insolation			Night cloudiness[e]	
	Strong[b]	Moderate[c]	Slight[d]	Cloudy ($\geq 4/8$)	Clear ($\leq 3/8$)
< 2	A	A–B[f]	B	E	F
2–3	A–B	B	C	E	F
3–5	B	B–C	C	D	E
5–6	C	C–D	D	D	D
> 6	C	D	D	D	D

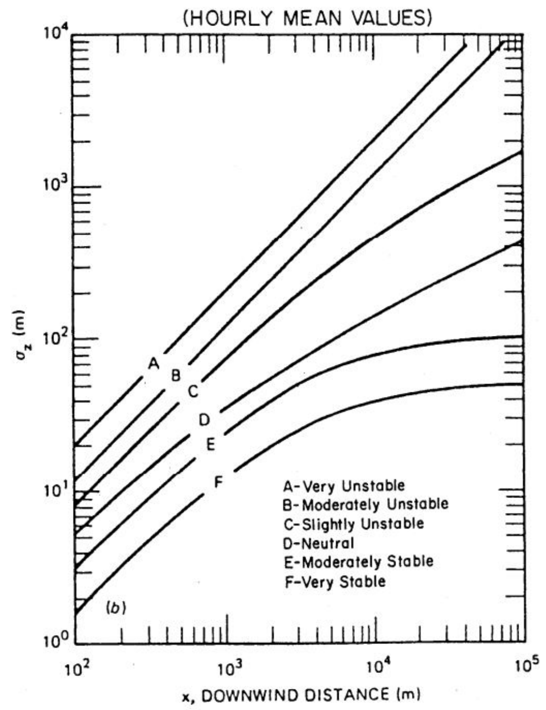
Figure E.1: Atmospheric stability classifications

E.2. Stability classes and dispersion coefficients

For the six stability classes, dispersion coefficients can be read for the different distances from the emission source. The graphs that are created for this purpose can be seen in Figure E.2. These stability coefficients are input variables of the Gaussian Dispersion Formula.



(a) σ_y



(b) σ_z

Figure E.2: Atmospheric stability coefficients

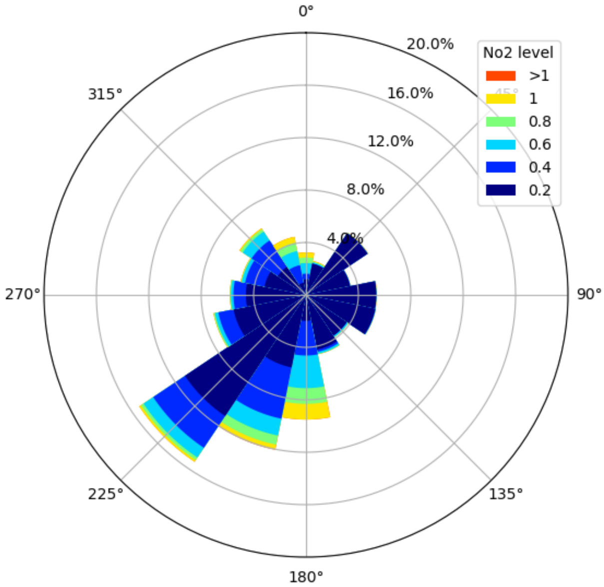
F
Pollution roses

In Figure F.1, two pollution roses are shown. Figure F.1a shows the pollution rose of the predicted wind direction and pollution level by GRASP. Figure F.1b shows the pollution rose of the NO_2 measurement station and the ERA5 wind direction. The wind directions over 2019 predicted by GRASP and analyzed by ERA5 are very similar. It can be seen that the most prevailing wind direction was a south-western wind direction. The wind direction from the north occurred least often in 2019, based on both datasets.

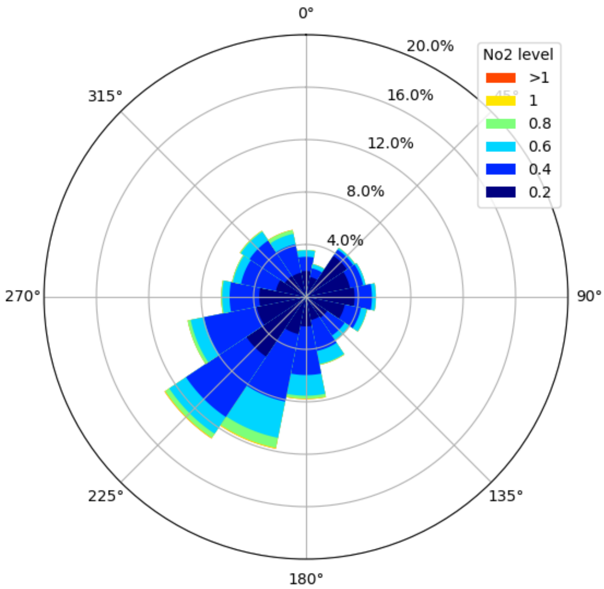
In general, the correlation between the NO_2 concentrations and the wind directions has similar characteristics. Pollution levels are higher whenever the wind comes from the two western quadrants, which can be explained by the fact that the measurement station is located on the east side of the highway. Furthermore, both figures show that the pollution levels are lower when the wind is perpendicular to the highway.

However, there are also differences among the two pollution roses. In the GRASP pollution rose, compared to the other pollution rose, a 'all-or-nothing effect' is visible. There are more values that belong to the highest pollution category, as can be distinguished with the yellow color, but there are also more values that fall within the lowest pollution category, which can be seen with the dark blue color. The pollution rose of the measurements and ERA5 data has more values in the mid-range of the pollution categories. It is expected that the explanation that is given in the main text for the presence of outliers also applies to the difference of the pollution roses. It can be seen that the highest values are present for a southern and a northern wind direction, the wind direction for which the wind is parallel to the highway. GRASP overestimates the pollution levels whenever the wind is coming from these directions.

Although the large outliers are removed from the results visible in these pollution rose, it is expected that the dataset still contains too high predictions for this wind direction, as the outlier threshold is relatively high. Due to the effect of this relatively high values in the normalization process, the other values become low after the normalization. Therefore, also many values are present in the GRASP pollution rose that belong to the lowest pollution category, marked with the dark blue color.



(a) GRASP predictions



(b) Measurements and ERA5 data

Figure F.1: Pollution roses of GRASP and measurements for 2019 at the highway near Breukelen

References

Abdel-Rahman, A. A. (2008). On the atmospheric dispersion and gaussian plume model. In *Proceedings of the 2nd international conference on waste management, water pollution, air pollution, indoor climate, corfu, greece* (Vol. 26).

Aerius air calculation tool. (n.d.). Rijksoverheid. Retrieved from `https://www.nsl-monitoring.nl/informatie/rekentool/`

Air quality. (n.d.). Kenniscentrum Europa Dencentraal. Retrieved from `https://europadecentraal.nl/onderwerp/klimaat/co2-en-luchtkwaliteit/luchtkwaliteit/`

Air quality: measurements. (n.d.-a). Rijkswaterstaat. Retrieved from `https://www.infomil.nl/onderwerpen/lucht-water/luchtkwaliteit/thema'/maatregelen/maatregelen-bron-ro/`

Air quality: measurements. (n.d.-b). Rijkswaterstaat. Retrieved from `https://www.infomil.nl/onderwerpen/lucht-water/luchtkwaliteit/thema'/maatregelen/verkeersmaatregelen/`

Ali, P. J. M., Faraj, R. H., Koya, E., Ali, P. J. M., & Faraj, R. H. (2014). Data normalization and standardization: a technical report. *Mach Learn Tech Rep*, *1*(1), 1–6.

Atlas leefomgeving. (n.d.). *Luchtkwaliteit voorgaande jaren.* Atlas Leefomgeving. Retrieved from `https://www.atlasleefomgeving.nl/kaarten`

Baklanov, A., & Zhang, Y. (2020). Advances in air quality modeling and forecasting. *Global Transitions*, *2*, 261–270.

Bieringer, P. E., Piña, A. J., Lorenzetti, D. M., Jonker, H. J., Sohn, M. D., Annunzio, A. J., & Fry, R. N. (2021). A graphics processing unit (gpu) approach to large eddy simulation (les) for transport and contaminant dispersion. *Atmosphere*, *12*(7), 890.

Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.

Briant, R., Seigneur, C., Gadrat, M., & Bugajny, C. (2013). Evaluation of roadway gaussian plume models with large-scale measurement campaigns. *Geoscientific Model Development*, *6*(2), 445–456.

Christie, D., & Neill, S. (2021). Measuring and observing the ocean renewable energy resource. *Reference Module in Earth Systems and Environmental Sciences; Elsevier: Amsterdam, The Netherlands*.

Daly, A., & Zannetti, P. (2007). Air pollution modeling–an overview. *Ambient air pollution*, 15–28.

Delavar, M. R., Gholami, A., Shiran, G. R., Rashidi, Y., Nakhaeizadeh, G. R., Fedra, K., & Hatefi Afshar, S. (2019). A novel method for improving air pollution prediction based on machine learning approaches: a case study applied to the capital city of tehran. *ISPRS International Journal of Geo-Information*, *8*(2), 99.

De Leeuw, F., Berge, E., Gronskei, K., & Tombrou, M. (1997). *Review on requirements for models and model applications*. European Environment Agency.

Directive 2008/50/ec. (2008). Council of European Union. Retrieved from `https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32008L0050`

Do urban speed limit reductions improve air quality? experts say not so fast. (n.d.). *Energy News Network*. Retrieved from https://energynews.us/2020/03/20/do-urban-speed-limit-reductions-improve-air-quality-experts-say-not-so-fast/

El-Harbawi, M. (2013). Air quality modelling, simulation, and computational methods: a review. *Environmental Reviews*, *21*(3), 149–179.

Era5. (n.d.). ECMWF. Retrieved from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5

Europese wetgeving luchtverontreiniging. (n.d.). RIVM. Retrieved from https://www.rivm.nl/ggd-richtlijn-medische-milieukunde-luchtkwaliteit-en-gezondheid/wet-en-regelgeving-luchtkwaliteit/europese-wetgeving-luchtverontreiniging

Fischer, P., Hoek, G., Van Reeuwijk, H., Briggs, D., Lebret, E., Van Wijnen, J., … Elliott, P. (2000). Traffic-related differences in outdoor and indoor concentrations of particles and volatile organic compounds in amsterdam. *Atmospheric Environment*, *34*(22), 3713–3722.

Gaussisch pluimmodel. (n.d.-c). Rijkswaterstaat. Retrieved from https://www.infomil.nl/onderwerpen/lucht-water/luchtkwaliteit/regelgeving/wet-milieubeheer/beoordelen/koppeling/nieuw-nationaal/handreiking-nieuw/handreiking-nieuw/4_2_gaussisch/

Gezondheidseffecten van luchtverontreiniging in context. (n.d.). RIVM. Retrieved from https://www.rivm.nl/ggd-richtlijn-medische-milieukunde-luchtkwaliteit-en-gezondheid/gezondheidseffecten-luchtverontreiniging/context-gezondheidseffecten-luchtverontreiniging

Grylls, T., Le Cornec, C. M., Salizzoni, P., Soulhac, L., Stettler, M. E., & van Reeuwijk, M. (2019). Evaluation of an operational air quality model using large-eddy simulation. *Atmospheric Environment: X*, *3*, 100041.

Guerreiro, C., Ortiz, A. G., de Leeuw, F., Viana, M., & Horálek, J. (2016). *Air quality in europe-2016 report*. Publications Office of the European Union.

Henderson, J. V. (1995). *Effects of air quality regulation.* National Bureau of Economic Research Cambridge, Mass., USA.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … others (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Hof van Beroep Brussel. (2021). *Arrest 2020/ar/1107*.

Jain, R. (2015). *Environmental impact of mining and mineral processing: management, monitoring, and auditing strategies*. Butterworth-Heinemann.

Jammalamadaka, S. R., & Lund, U. J. (2006). The effect of wind direction on ozone levels: a case study. *Environmental and Ecological Statistics*, *13*(3), 287–298.

Johansson, C., Norman, M., & Burman, L. (2009). Road traffic emission factors for heavy metals. *Atmospheric Environment*, *43*(31), 4681–4688.

Kang, G. K., Gao, J. Z., Chiao, S., Lu, S., & Xie, G. (2018). Air quality prediction: Big data and machine learning approaches. *International Journal of Environmental Science and Development*, *9*(1), 8–16.

Ligterink, N., & Smokers, R. (2016). Uitstoot van autos bij snelheden hoger dan 120 km/u.

Limit values and other air quality standards. (n.d.-d). Rijkswaterstaat. Retrieved from https://www.infomil.nl/onderwerpen/lucht-water/luchtkwaliteit/regelgeving/wet-milieubeheer/beoordelen/grenswaarden/

Luchtkwaliteit. (n.d.). Retrieved from https://www.clo.nl/indicatoren/nl0231-stikstofdioxide

Luchtkwaliteit. (n.d.). Rijksoverheid. Retrieved from https://www.rijksoverheid.nl/onderwerpen/luchtkwaliteit

Mahanta, S., Ramakrishnudu, T., Jha, R. R., & Tailor, N. (2019). Urban air quality prediction using regression analysis. In *Tencon 2019-2019 ieee region 10 conference (tencon)* (pp. 1118–1123).

Miller, C. (1982). Cases in the applications of air-quality models in policy making. *Mathematical Models for Planning and Controlling Air Quality*, 23.

Miller, C., & Hively, L. (1987). A review of validation studies for the gaussian plume atmospheric dispersion model. *Nucl. Saf.;(United States)*, *28*(4).

Negen vragen over de luchtkwaliteit rond snelwegen. (2021). *De Ingenieur*. Retrieved from https://www.deingenieur.nl/artikel/negen-vragen-over-de-luchtkwaliteit-rond-snelwegen

Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W., et al. (1996). Applied linear statistical models.

Nieuw normen- en handhaving-stelsel schiphol. (2016, february 16). Schiphol Group. Retrieved from https://www.rijksoverheid.nl/documenten/rapporten/2021/02/16/bijlage-3-mer-nnhs-2020-deel-1-hoofdrapport

Panis, L. I., Broekx, S., & Liu, R. (2006). Modelling instantaneous traffic emission and the influence of traffic speed limits. *Science of the total environment*, *371*(1-3), 270–285.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pervin, T., Gerdtham, U.-G., & Lyttkens, C. H. (2008). Societal costs of air pollution-related health hazards: A review of methods and results. *Cost Effectiveness and Resource Allocation*, *6*(1), 1–22.

Postma, E. (2005). De nederlandse worsteling met de europese richtlijnen voor luchtkwaliteit. *Nederlands Tijdschrift voor Europees Recht*(6), 134-139.

Rechtbank Den Haag. (2017). *C/09/516573 / ha za 16-975*.

Rechtszaak voor gezonde lucht. (2021, Aug). Milieudefensie. Retrieved from https://milieudefensie.nl/recht-op-gezonde-lucht/rechtszaak-voor-gezonde-lucht

Reed, W., et al. (2005). *Significant dust dispersion models for mining operations* (Vol. 9478). Createspace Independent Pub.

Ryan, W. F. (2016). The air quality forecast rote: Recent changes and future challenges. *Journal of the Air & Waste Management Association*, *66*(6), 576–596.

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., ... others (2020). *Five ways to ensure that models serve society: a manifesto.* Nature Publishing Group.

Schalkwijk, J., Jonker, H. J., Siebesma, A. P., & Van Meijgaard, E. (2015). Weather forecasting using gpu-based large-eddy simulations. *Bulletin of the American Meteorological Society*, *96*(5), 715–723.

Seaman, N. L. (2000). Meteorological modeling for air-quality assessments. *Atmospheric environment*, *34*(12-14), 2231–2259.

Sinnott, R. O., & Guan, Z. (2018). Prediction of air pollution through machine learning approaches on the cloud. In *2018 ieee/acm 5th international conference on big data computing applications and technologies (bdcat)* (pp. 51–60).

Taken en verantwoordelijkheden rijkswaterstaat. (n.d.). Rijkswaterstaat. Retrieved from https://www.rijkswaterstaat.nl/wegen/wegbeheer/natuur-en-milieu/luchtkwaliteit/taken-en-verantwoordelijkheden-rijkswaterstaat

Taylor, C. (2020, August 27). *How do we determine outliers in statistics?* ThoughtCo. Retrieved from https://www.thoughtco.com/what-is-an-outlier

Technische beschrijving van standaardrekenmethode 2 (srm-2) voor luchtkwaliteitsberekeningen. (2014). Retrieved from https://www.rivm.nl/bibliotheek/rapporten/2014-0109.pdf

Tomas, J., Pourquie, M., & Jonker, H. (2015). The influence of an obstacle on flow and pollutant dispersion in neutral and stable boundary layers. *Atmospheric Environment*, *113*, 236–246.

Tseng, Y.-H., Meneveau, C., & Parlange, M. B. (2006). Modeling flow around bluff bodies and predicting urban dispersion using large eddy simulation. *Environmental science & technology*, *40*(8), 2653–2662.

Turner, D. B. (1970). *Workbook of atmospheric dispersion estimates: an introduction to dispersion modeling*. Office of Air Program Pub. No. AP-26.

Vaststellen van luchtkwaliteit. (n.d.). Retrieved from https://www.rivm.nl/ggd-richtlijn-medische-milieukunde-luchtkwaliteit-en-gezondheid/blootstelling-aan-luchtverontreiniging/vaststellen-van-luchtkwaliteit

VZinfo. (n.d.). *Gezondheidsgevolgen.* Retrieved 2020-06-10, from https://www.vzinfo.nl/Fysieke%20omgeving/Luchtverontreiniging/Gezondheidsgevolgen

Weil, J., Sykes, R., & Venkatram, A. (1992). Evaluating air-quality models: review and outlook. *Journal of Applied Meteorology and Climatology*, *31*(10), 1121–1145.

Wesseling, J., & Van Velze, K. (2015). Technische beschrijving van standaardrekenmethode 2 (srm-2) voor luchtkwaliteitsberekeningen.

Whiffle. (n.d.). Whiffle. Retrieved from https://www.weatherfinecasting.com/what-we-do/dispersion/

WHO. (2021a, 22). Improving health by reducing air pollution. Retrieved from https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines

WHO. (2021b). *New who global air quality guidelines aim to save millions of lives from air pollution.* Retrieved Jan 27, 2022, from https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution

Wiegant, E., & Verzijlbergh, R. (2019). *Grasp model description validation report.*

Yang, J., Rahardja, S., & Fränti, P. (2019). Outlier detection: how to threshold outlier scores? In *Proceedings of the international conference on artificial intelligence, information processing and cloud computing* (pp. 1–6).

Zalakeviciute, R., Buenaño, A., Sannino, D., & Rybarczyk, Y. (2018). *Urban air pollution mapping and traffic intensity: Active transport application*. IntechOpen.