Report

NTNU
Norwegian University of Science and Technology
Faculty of Engineering Science
Department of Ocean Operations and Civil Engineering

Prediction modeling of air quality in Geiranger

Dina M. Aspen, Jörg Löffler, Elias Hasle, André L. Ellefsen, Ahsan M. H. A. Kader, Xu Cheng, Børge A. H. Johansen

Report

Prediction modeling of air quality in Geiranger

VERSION

Version number 1

DATE

17.10.2019

AUTHORS

Dina M. Aspen, Jörg Löffler, Elias Hasle, André L. Ellefsen, Ahsan M. H. A. Kader, Xu Cheng, Børge A. H. Johansen

PROJECT NUMBER

296343

NUMBER OF PAGES

30

CLIENT

Stranda Hamnevesen KF

CLIENTS REF.

Rita Berstad Maraak

Norwegian University of Science and Technology

NTNU
Norwegian University of Science and Technology

ABSTRACT

Prediction modeling of air quality in Geiranger

Nature based tourist destinations in the Møre region experience high transport loads over short time periods in the tourist season, with high pollution levels as a consequence. Geiranger and other Norwegian tourist destinations may experience high levels of particle matter (PM), nitrogen oxides (NO_x), sulfur oxides (SO_x) as well as other pollutants during peak season, which may not only cause adverse environmental effects but also be harmful to human health.

The primary objective of the pilot project "Big Data analysis of Air Quality in Geiranger" has been to build air quality prediction models to improve the knowledge about air quality and its variability linked to transportation and meteorological conditions in Geiranger. Prediction models may later be used to evaluate the effect of transportation measures on air quality. In the project, time series data for air quality, transport and meteorological conditions in Geiranger over a four year period has been analyzed and visualized.

This report summarizes findings from initial analyses of air quality, traffic and meteorological data for the Geiranger area in the period 2015-2018. A combination of conventional statistics and machine learning methods have been applied to better explore relationships between these conditions in Geiranger.

The pilot project is funded by Regionale Forskningsfond (RFF) Midt-Norge and has been conducted during spring and summer 2019. The project has been managed by NTNU in Ålesund, with Stranda Hamnevesen as problem owner and project partner, and Bonn University as project partner.

Contents

Sammendrag

Naturbaserte turistdestinasjoner i Møreregionen opplever ofte stor transportbelastning over korte tidsrom, med høye forurensningsnivåer som konsekvens. Geirangerfjorden og andre norske turistdestinasjoner kan i høysesong oppleve forhøyede nivåer av blant annet svevestøv og nitrogendioksid med potensielt helseskadelige effekter.

Pilotprosjektet «Big data analyse av miljøkvalitet i Geiranger» har som målsetting å utvikle ny kunnskap om sammenhenger mellom luftkvalitet, meteorologiske forhold og transportaktivitet i Geiranger gjennom å bygge prediksjonsmodeller for luftkvalitet. Slike modeller vil kunne bidra til å analysere effekten av ulike logistiske og teknologiske tiltak i transportsystemet på luftkvalitet i området. I prosjektet har tidsseriedata for luftkvalitet, transport og meteorologiske forhold over en fireårsperiode blitt analysert og visualisert.

Denne rapporten oppsummerer funn fra analyser av data for luftkvalitet, meteorologiske forhold og transportaktivitet i Geiranger i perioden 2015-2018. En kombinasjon av konvensjonell statistikk og maskinlæring har blitt brukt for å analysere data.

Fra disse preliminære prediksjonsmodellene kan vi konkludere følgende:

1. Prediksjonsmodellene følger *trendene* i luftforurensingsnivåer men klarer ikke å predikere *topper* for konsentrasjonsnivåer av luftforurensning
2. NO_2 predikeres mer nøyaktig enn PM1.
3. Data om meteorologiske forhold øker nøyaktigheten i prediksjonsmodellene
4. Det er ikke mulig å skille bidrag til luftforurensningsnivåer fra land- eller sjøtransportaktivitet med prediksjonsmodellene

Videre forskning burde se nærmere på hvordan bidrag fra sjø- og landtransport kan skilles i disse modellene, hvilken rolle meteorologiske forhold har på luftkvalitet, samt utvikle en bedre forståelse for akkumulering av PM over tid. Bedre oppløsning på data vil muliggjøre slike analyser. Bedre prediksjonsmodeller er også nødvendig for å bidra til beslutningsstøtte for å vurdere ulike forbedringstiltak i ulike scenarier. En annen nyttig utvikling vil også være å koble prediksjonsmodeller til visualiseringsverktøy for å skape brukervennlige grensesnitt for aktører involvert i transportplanlegging.

Summary

Nature based tourist destinations in the Møre region experience high transport loads over short time periods in the tourist season, with high pollution levels as a consequence. Geiranger and other Norwegian tourist destinations may experience high levels of particle matter (PM), nitrogen oxides (NO_x), sulfur oxides (SO_x) as well as other pollutants during peak season, which may not only cause adverse environmental effects but also be harmful to human health.

The pilot project “Big Data analysis of Air Quality in Geiranger” aims to improve the knowledge about air quality and its variability linked to transportation and meteorological conditions in Geiranger through building air quality prediction models. Such models may be used to analyze the effect om air quality from implementing technological and logistical measures in the transportation system. In the project, time series data for air quality, transport and meteorological conditions in Geiranger over a four year period has been analyzed and visualized.

This report summarizes findings from initial analyses of air quality, traffic and meteorological data for the Geiranger area in the period 2015-2018. A combination of conventional statistics and machine learning methods have been applied to better explore relationships between these conditions in Geiranger.

From the preliminary prediction models, we may conclude that:

1. Prediction models follow the *trends* of air pollution levels but do not predict the magnitude of concentrations *peaks*.
2. NO_2 is more accurately predicted than PM1.
3. Data on meteorological conditions increase accuracy in prediction models.
4. It is not possible to attribute air quality levels to land- or sea traffic activity with the current prediction models

Future research efforts in this direction should look more closely into clarifying the contribution from sea- and land traffic, the mediating role of meteorological conditions and the accumulation effects of PM over time. Improved resolution of data is a means to this end. Prediction model development is also necessary to better facilitate decision support for relevant measure evaluations across future scenarios. Another useful extension could also be to integrate models in visualization tools to provide a user-friendly interface and facilitate decision making among transportation planners.

1. Introduction

Nature based tourist destinations in the Møre region experience high transport loads over short time periods in the tourist season, with high pollution levels as a consequence. Geiranger and other Norwegian tourist destinations may experience high levels of particle matter (PM), nitrogen oxides (NO_x), sulfur oxides (SO_x) as well as other pollutants during peak season [1-3], which may not only cause adverse environmental effects but also be harmful to human health [4]. While the transport load and air pollution levels are high and growing, destination image and visitor experiences are also at risk of degradation. From the perspective of planners, efficient measures need to be identified and implemented [5]. This requires knowledge about the dynamic relationship between air quality, transportation and meteorological conditions.

The pilot project "Big Data analysis of Air Quality in Geiranger" aims to improve the knowledge about air quality and its variability linked to transportation and meteorological conditions in Geiranger through building air quality prediction models. In the project, time series data for air quality, transport and meteorological conditions in Geiranger over a four year period has been analyzed and visualized.

The pilot project is funded by Regionale Forskningsfond (RFF) Midt-Norge and has been conducted during spring and summer 2019. The project has been managed by NTNU in Ålesund, with Stranda Hamnevesen as problem owner and project partner, and Bonn University as project partner.

2. Project structure and data

2.1. Project work process

The overall work process of the pilot project is divided into four main steps as shown in Figure 1. Initially, data was collected (step 1) and processed to combine in a project database (step 2). The data was then analyzed (step 3) using Pearson's correlation coefficient and further used along with the deep learning technique Long short-term memory (LSTM) to build prediction models. The data analysis methods in step 3 are described in appendix I. In addition to project reporting (step 4), the experiences from the pilot is used to develop new projects in the future utilizing data and machine learning techniques for decision support.

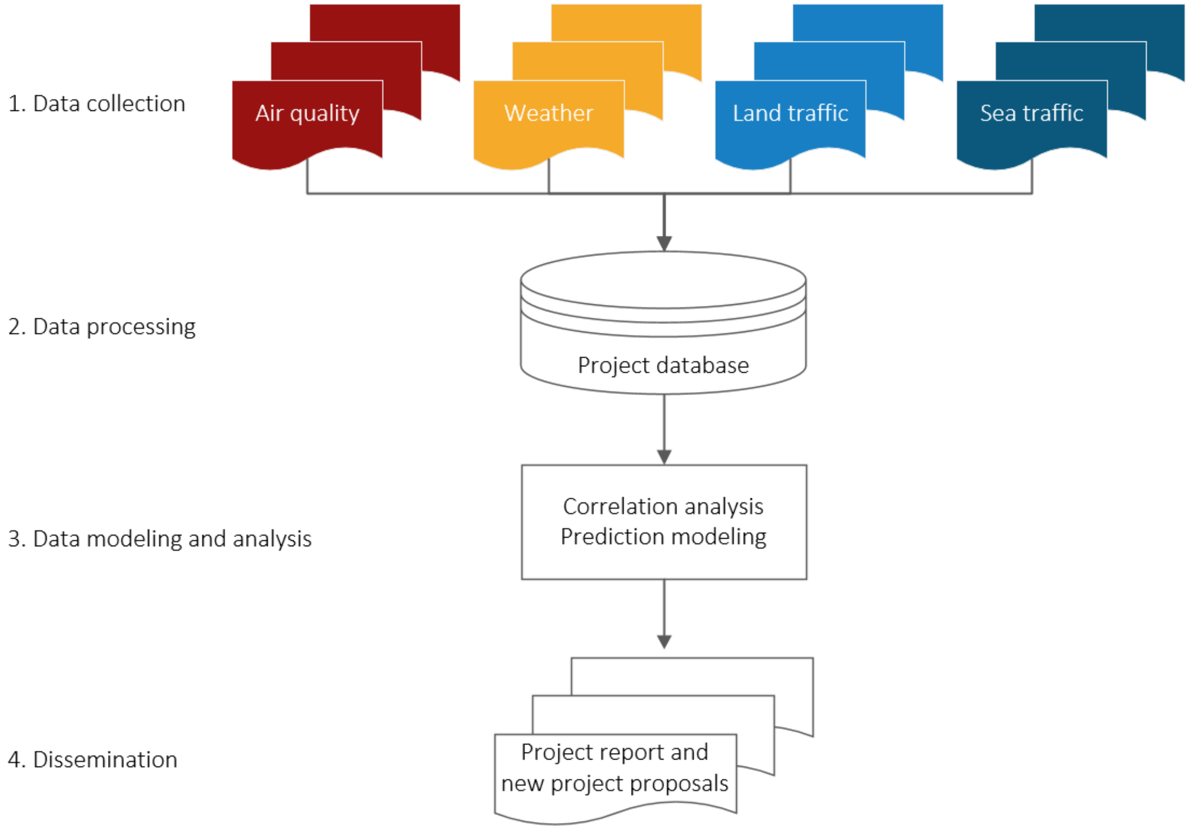


Figure 1: Overall work process of the pilot project

2.2. Project data

All the data has time series format with the resolution of one hour. Data observations are either the aggregate (count) or mean value of selected variables. The analysis window for this pilot follows the time period for the accessed air quality dataset, which starts on 04.06.2015 at 17.00 and ends on 14.09.2018 at 10.00 AM. This totals to 28 746 hours of observations.

Table 1 lists the datasets used in the project.

Table 1: List of datasets used in the data analysis project

	Air quality	**Meteorological**	**Land traffic**	**Sea traffic**
Origin	Bonn University	Bonn University	NPRA	STPA, Sea-web
Measurement points	3	1	3	1
Variables	PM NO_x	Air pressure Precipitation Wind direction Wind speed Relative air humidity Air temperature Global radiation	Vehicle size Vehicle direction	GT total (size) PAX total (capacity) Number of vessels Installed KW for tier 0-3 engines

Air quality and meteorological data was obtained from Bonn University. This data has initially been collected in the Geiranger Air quality monitoring program and covers air concentrations of particulate matter (PM), nitrogen oxides (NO_x), sulfur oxides (SO_x) and most recently, noise. This data pilot focus on PM and NO_x pollution levels in Geiranger.

The air quality data has been collected from three different measurement stations in the Geiranger area, as shown in Figure 2. The meteorological data has only been collected at the station in Geiranger centre (the leftmost purple point on the figure).

Land traffic data has been collected from the Norwegian Public Road Authorities (NPRA) database for traffic registration points in the area. The yellow circles on Figure 2 shows the location of traffic registration stations. This data is a record of cars passing the station separate directions over the course of an hour. The data extracted from the database follows the analysis period (2015-2018).

Sea traffic data was obtained from Stranda Port Authorities (STPA) records for ships calling to port in Geiranger. The data was supplemented with database information from the Sea-web Ships database to provide more details and technical specifications about the vessels. A time series dataset was then manually created for the analysis period (2015-2018).

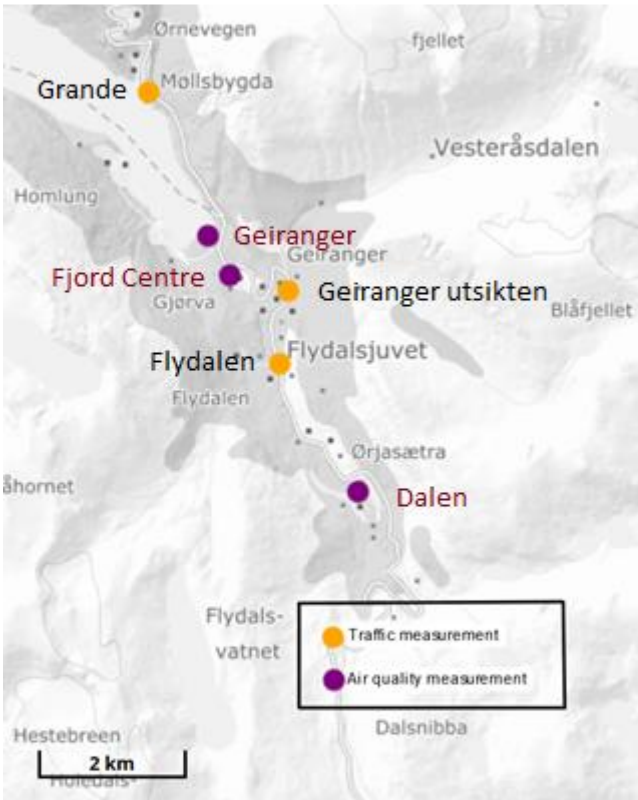


Figure 2: Location and name of measurement stations for land traffic and air quality data

3. Results from Geiranger air quality data analysis

3.1. Historical air quality in Geiranger

University of Bonn and Stiftinga Geirangerfjorden Verdsarv has since 2015 collected data on air quality and meteorological conditions within the Geiranger Air Quality Monitoring Program. Results from these measurements have been published in annual reports [see 1, 3] that comment on the state of air quality in Geiranger. The program monitors particles, gases and since 2017 also sound levels in the World Heritage Area. In the following sections, we analyze air quality in Geiranger for the entire period 2015-2018, focusing on particulate matter (PM) and NO_x gases.

Particulate Matter (PM)

PM is a mixture of solid and liquid particles suspended in the air. Constituents of PM may be sulfates, nitrates, ammonium and other inorganic ions. Allergens and microbial compounds are also found in PM [6]. PM is separated into different fractions based on the particle diameter. Particles with a diameter of less than 2.5 micrometers (µm) is classified as PM2.5 (also called fine PM), while particles with a diameter less than 10 µm is classified as PM10 and so on.

PM may be emitted directly into the air (primary PM) or form in the atmosphere from gases (secondary PM). The origins of PM may be combustion engines, solid fuel combustion for household warming as well as other industrial activities. PM may also originate from road traffic through pavement erosion and abrasion of brakes and tires as well as road salt and sand [6, 7]. Exhaust emissions from combustion engines mostly contributes to fine PM (PM2.5) while other particles from road traffic contributes to the coarser part of PM10 (PM2.5-PM10) [7]. Formation of secondary PM happens through chemical reactions of gaseous pollutants such as nitrogen oxides and sulfur dioxide.

PM may result in both short- and long-term adverse health effects. Both PM10 and PM2.5 contain particles that may be inhaled and penetrate the respiratory system. Effects include aggravation of asthma and other respiratory symptoms as well as lung cancer [6].

PM pollution is regulated in the Norwegian Pollution Regulation where maximum annual concentrations are defined for PM2.5 and PM10, as shown in Table 2. For PM10, a maximum daily concentration is also defined. These levels are legally binding.

In addition to regulation, the Norwegian Institute of Public Health and the Norwegian Environment Agency publish Air Quality Criteria for several pollutants. These are stricter than levels from the Pollution Regulation, but are not legally binding.

Table 2: Air quality threshold levels for PM2.5 and PM10

	Averaging period	Pollution Regulation [8]	Air Quality Criteria [9]
$PM_{2.5}$ [µg/m³]	Day		15
	Year	15	8
PM_{10} [µg/m³]	Day	50, 30 times per year	30
	Year	25	20

Figure 3 shows hourly average for all PM fractions measured at the Geiranger station over the four-year period. As the plot shows, a primary contributor to PM in Geiranger is PM1, i.e. particles with a diameter of less than 1 µm, with some peaks registered for PM in the range 4-10 µm and 10-100 µm.

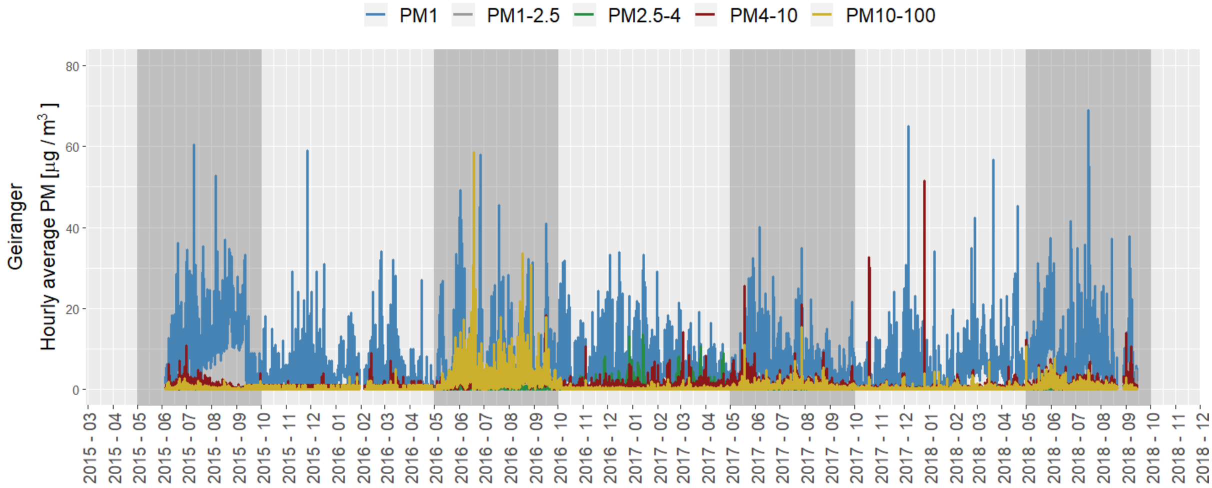


Figure 3: Hourly average concentrations of PM in Geiranger 2015-2018

Furthermore, daily and annual averages for the monitoring period shows that Pollution Regulation air quality levels have not been exceeded. However, the Air Quality Criteria have been exceeded multiple times. For PM2.5, the daily average of 15 µg/m³ have been exceeded mainly between May to September, i.e. the tourist season, as shown in Figure 5. Higher values were also recorded during spring 2018. Annual average concentrations of PM2.5 has not exceeded the Air Quality Criteria at 8 µg/m³ for the years 2016-2017, as shown in Figure 4.

For PM10, Figure 6 shows that the daily average limit of 30 µg/m³ was exceeded mainly during the tourist season in 2016 and spring of 2018. As seen in Figure 4, no exceedance of the annual average limit at 20 µg/m³ PM10 has been recorded during the monitoring period.

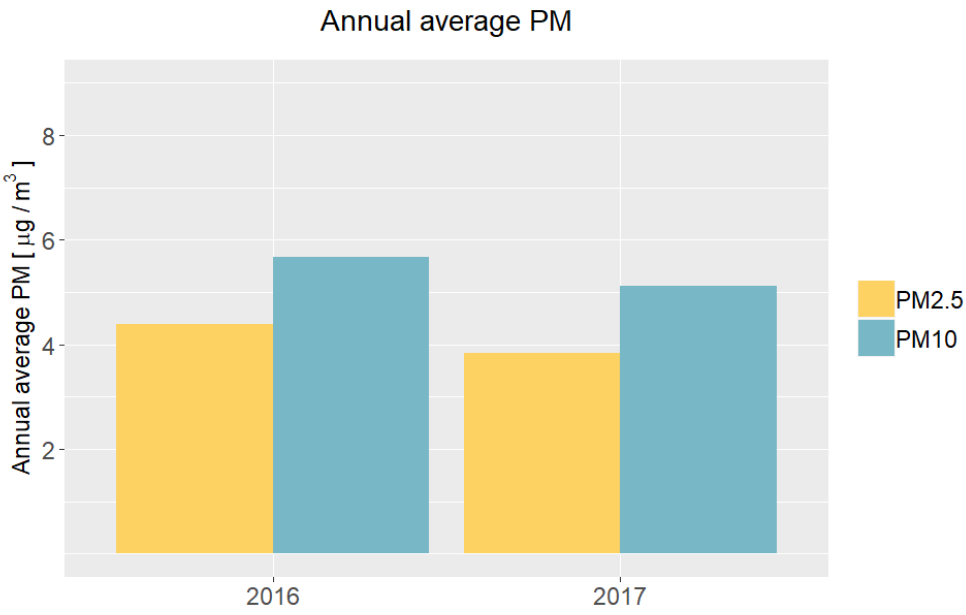


Figure 4: Annual average concentrations for PM2.5 (left) and PM10 (right) at the Geiranger measurement station.

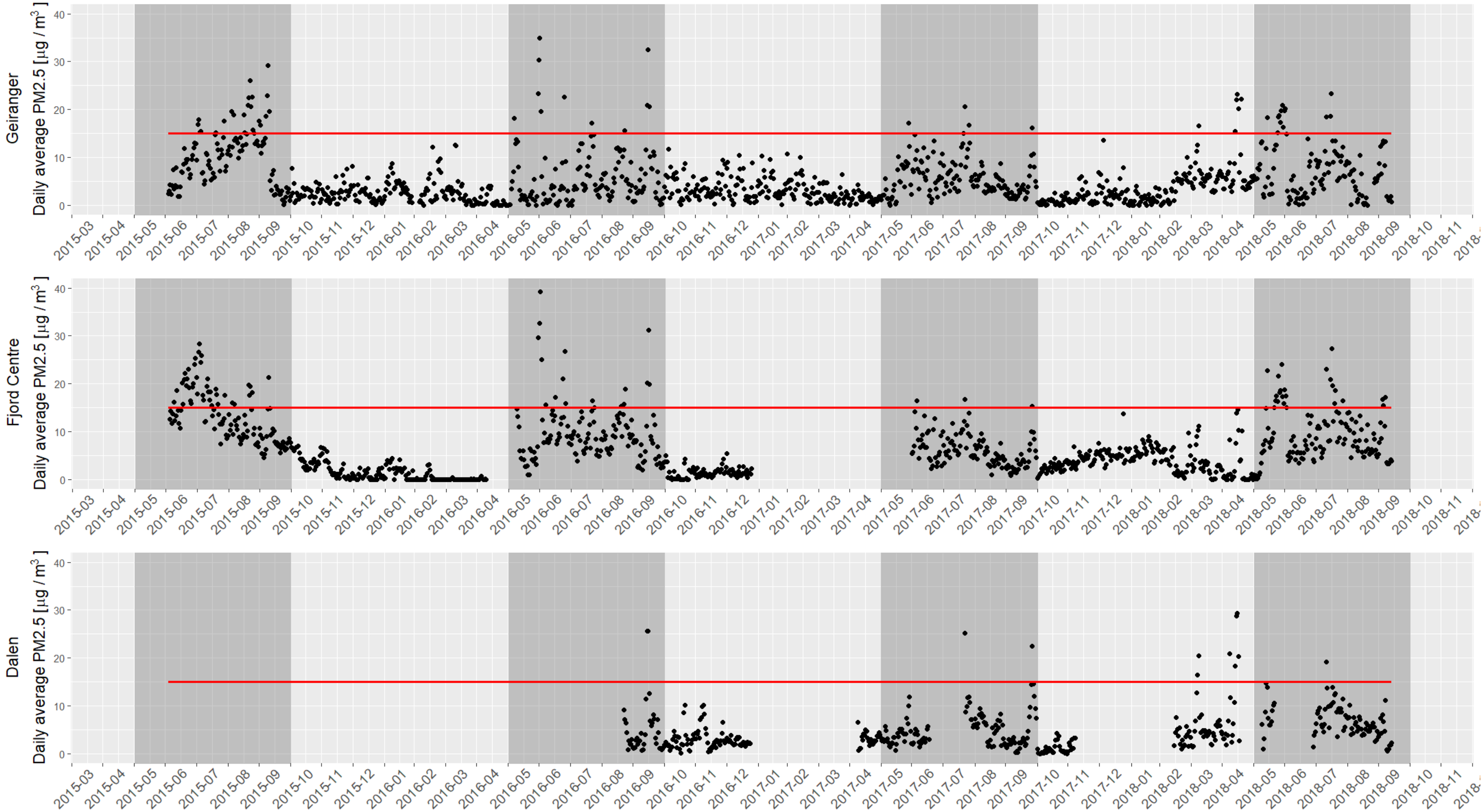


Figure 5: Daily average concentrations of PM2.5 at three measurement stations in Geiranger 2015-2018. The red line shows Air Quality Criteria levels.

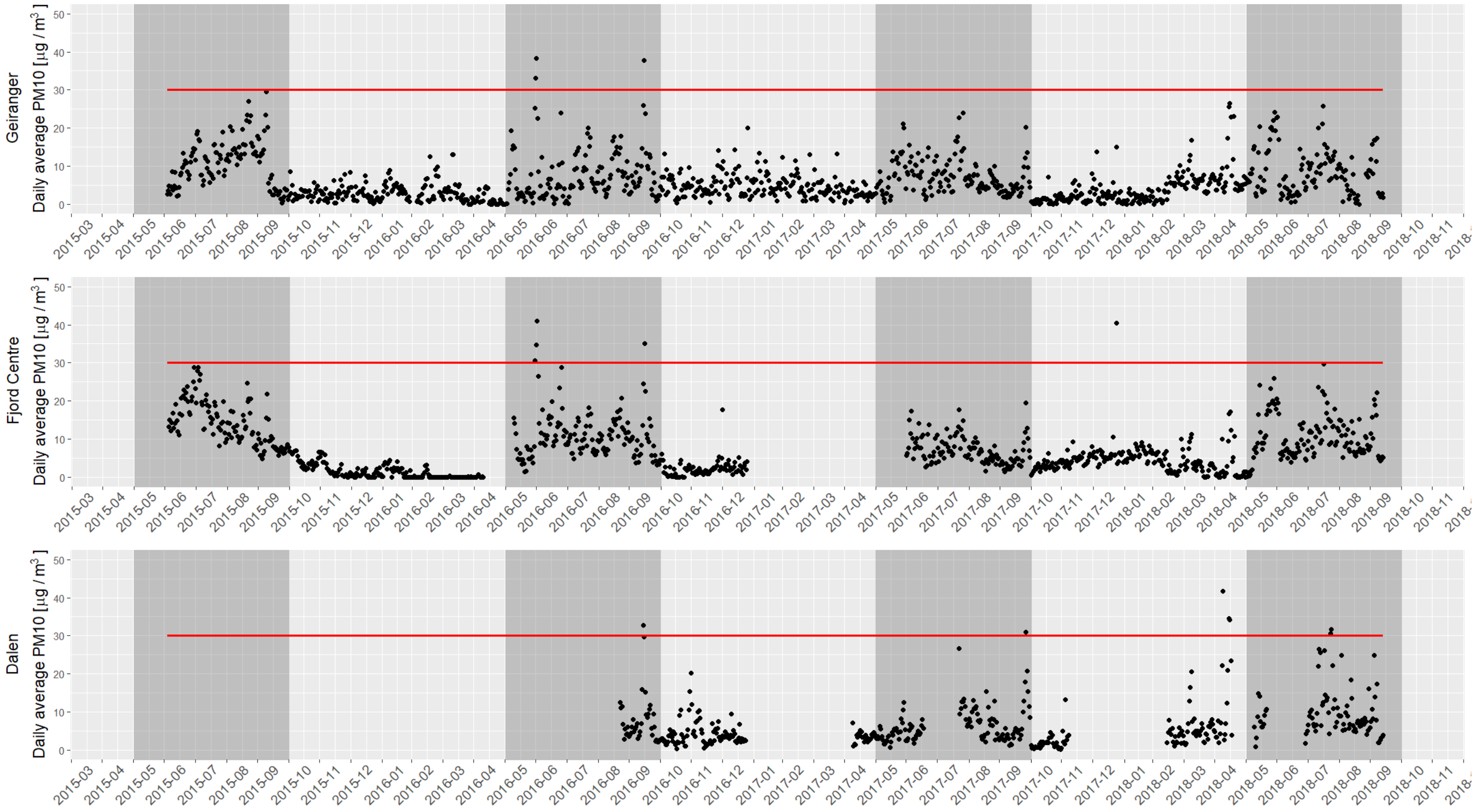


Figure 6: Daily average concentrations of PM10 at three measurement stations in Geiranger 2015-2018. The red line shows Air Quality Criteria levels.

Nitrogen Oxides (NO_x)

NO_x gases consist of nitric oxide (NO) and nitrogen dioxide (NO_2). These gases are formed in combustion processes under high temperatures. An important source of NO_2 is diesel engine combustion, while other origins include industrial activities and waste incineration.

Adverse health effects from NO_2 exposure include respiratory symptoms, development of asthma and increased susceptibility of respiratory infections. Children, elderly people and people with asthma are at greater risk for the health effects of NO_2.

Both Pollution Regulation levels and Air Quality Criteria for NO_2 are defined, as shown in Table 3. For the studied years, the annual threshold of 40 [µg/m³] has not been exceeded. As plots in Figure 5 shows, hourly average NO_2 exceeded the threshold of 100 µg/m³ during July 2018. This took place for four hours around noon on July 17th 2018.

Table 3: Air quality levels for NO_x

	Averaging period	Pollution Regulation [8]	Air quality criteria [9]
NO_2 [µg/m³]	15 min		300
	Hour	200, 18 times per year	100
	Year	40	40

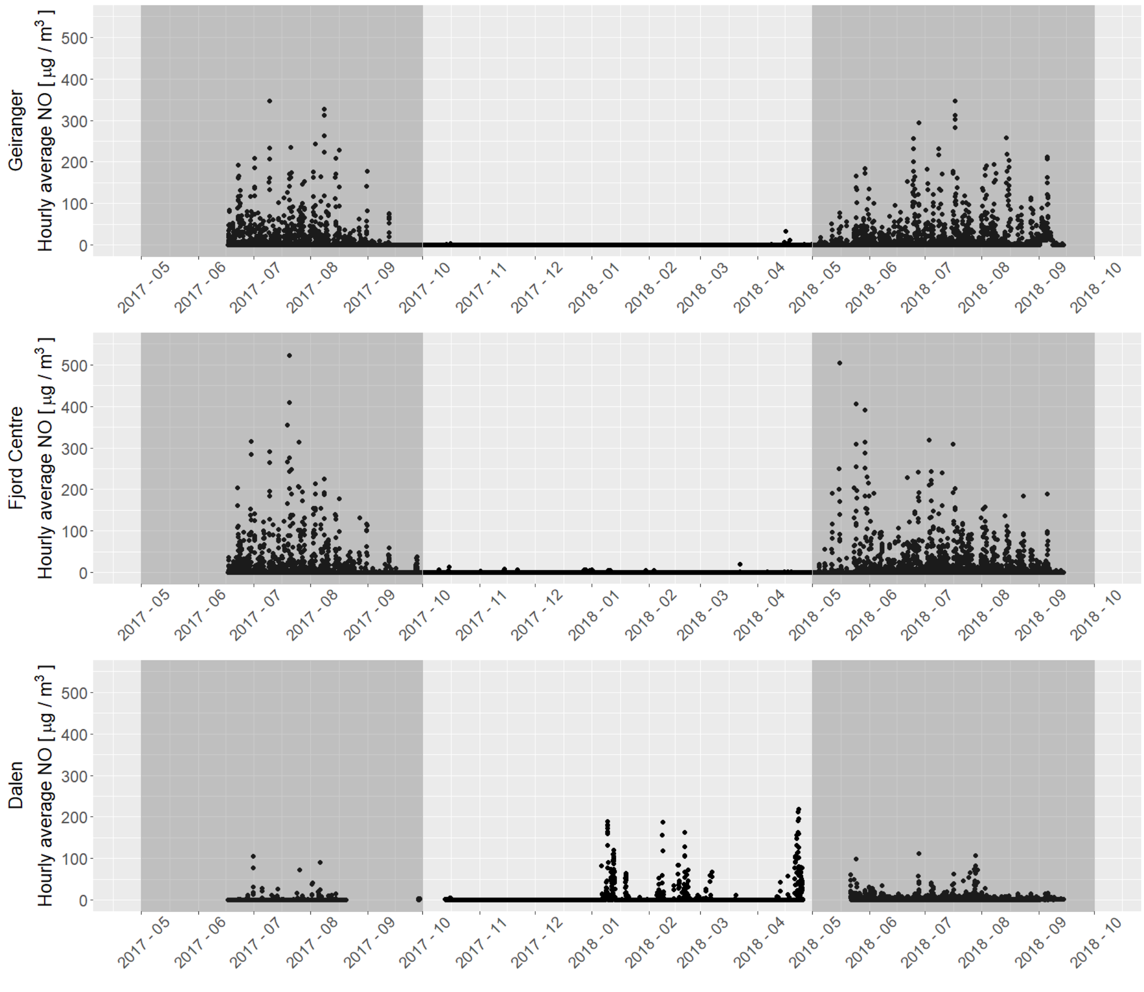


Figure 7: Daily average concentrations of NO at three measurement stations in Geiranger 2017-2018.

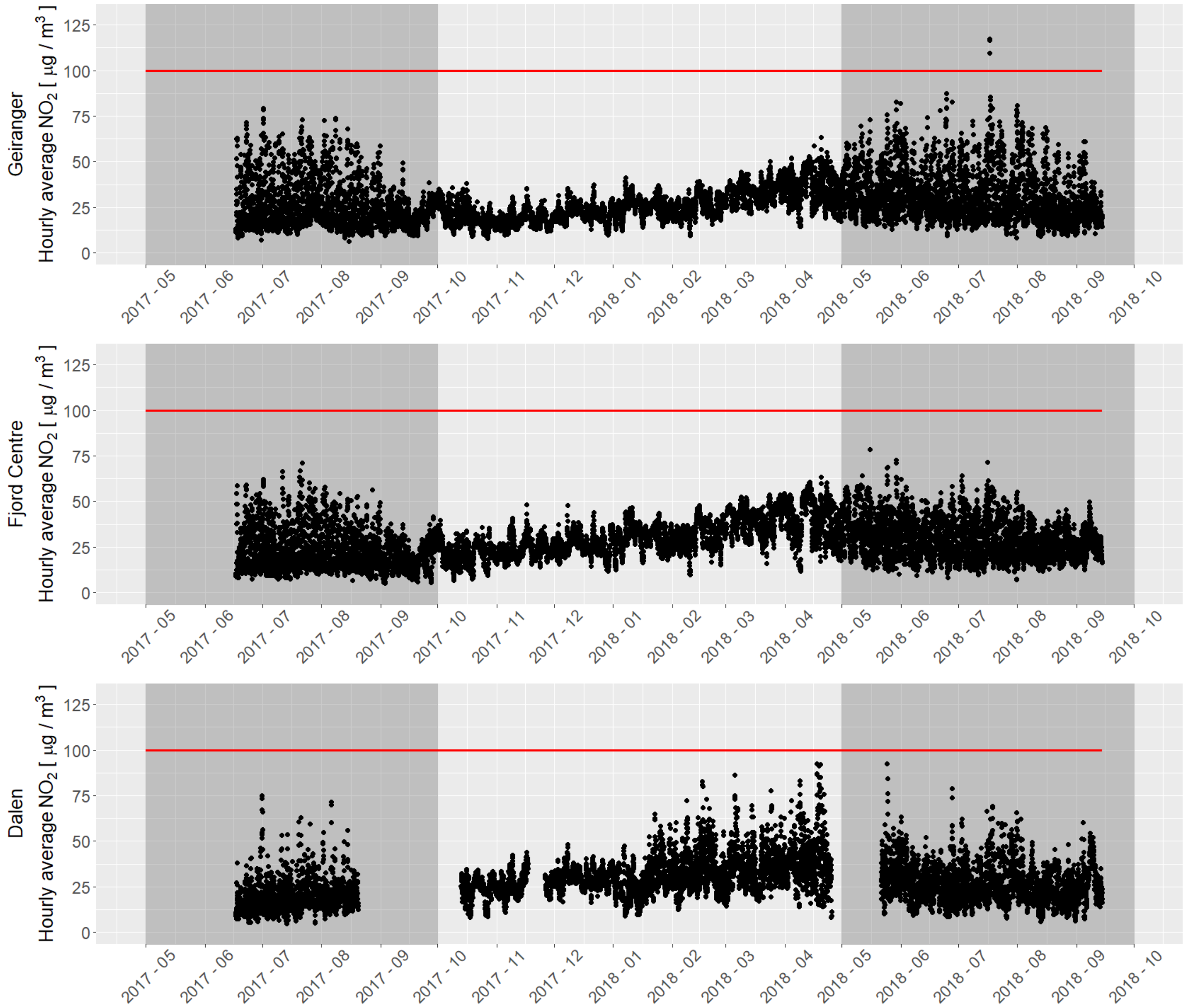


Figure 8: Daily average concentrations of NO_2 at three measurement stations in Geiranger 2017-2018. The red line shows Air Quality Criteria levels.

Further analysis of air quality

From the initial inspection of PM and NOx pollution in Geiranger, some general observations are made:

Firstly, for PM the primary contribution comes from PM1, as shown in Figure 3 and elaborated in [1]. In subsequent analysis and prediction of PM, we therefore focus on PM1.

Secondly, the Geiranger and Fjord Centre measurement stations that have the highest concentrations of PM and NO_x as well as the most frequent exceedances of the Air Quality Criteria for PM. The Geiranger station is furthermore the station with the most continuous data recordings as the other two stations have suffered from power outages and equipment failure for long time segments. In subsequent data analyses and prediction modeling, we will therefore focus on the Geiranger measurement station.

3.2. Temporal variability in data

As seen in section 3.1., air pollution peaks are observed mainly during the tourist season (May – September). In order to further explore data variability over time, mean values for air quality and traffic data for each hour of the day in the tourist (May – September) and non-tourist season (October – April) are compared. Figure 9 shows concentration of particulate matter (PM) and nitrogen oxides (NO_x) for the Geiranger measurement station. Note that time is defined in UTC (Coordinated Universal Time). For the tourist season, local time in Geiranger is UTC+2 hours.

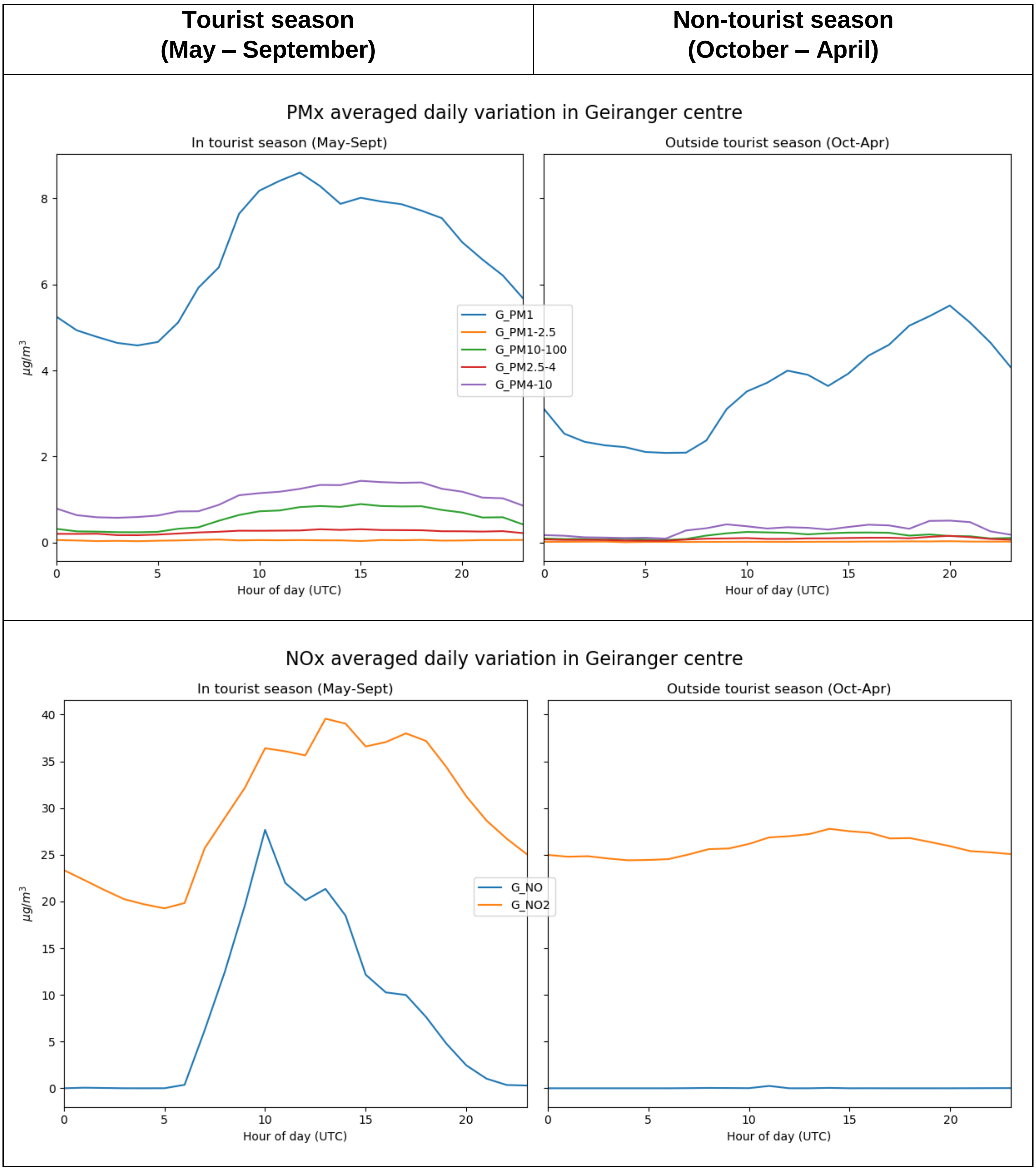


Figure 9: PMx and NOx pollution for tourist and non-tourist seasons. Mean values for time of day at UTC

As can be seen from figure 9, PM and NO_x levels are far higher in the tourist season than outside the tourist season. In the tourist season, mean hourly PM1 levels begin to rise around 07.00 local time and peaks around 14.00. A similar pattern is seen for NOx, where the mean hourly air concentration rises from 07.00 and peaks at 12.00 for NO and around 16.00 for NO_2.

Another observation is that average hourly PM1 levels never reaches zero or even near-zero levels in the tourist season or during the rest of the year.

The land- and sea traffic in the area also has a significantly higher volume during the tourist season compared to outside the tourist season as shown in Figure 10. The land traffic is measured as the number of cars passing the Grande measurement station in both directions, averaged on an hourly basis for the two seasons. For ship traffic, the total KW at different tier levels present in the area is summed and averaged on an hourly basis for the two seasons. The land- and sea traffic follow an almost identical pattern with a steady growth in volume measured as mean hourly volume from 07.00 with a peak at around 13.00.

The plots in Figure 9 and Figure 10 show that the trends for air quality and traffic volumes align. In the subsequent sections, we explore these relations in a more statistically rigorous manner.

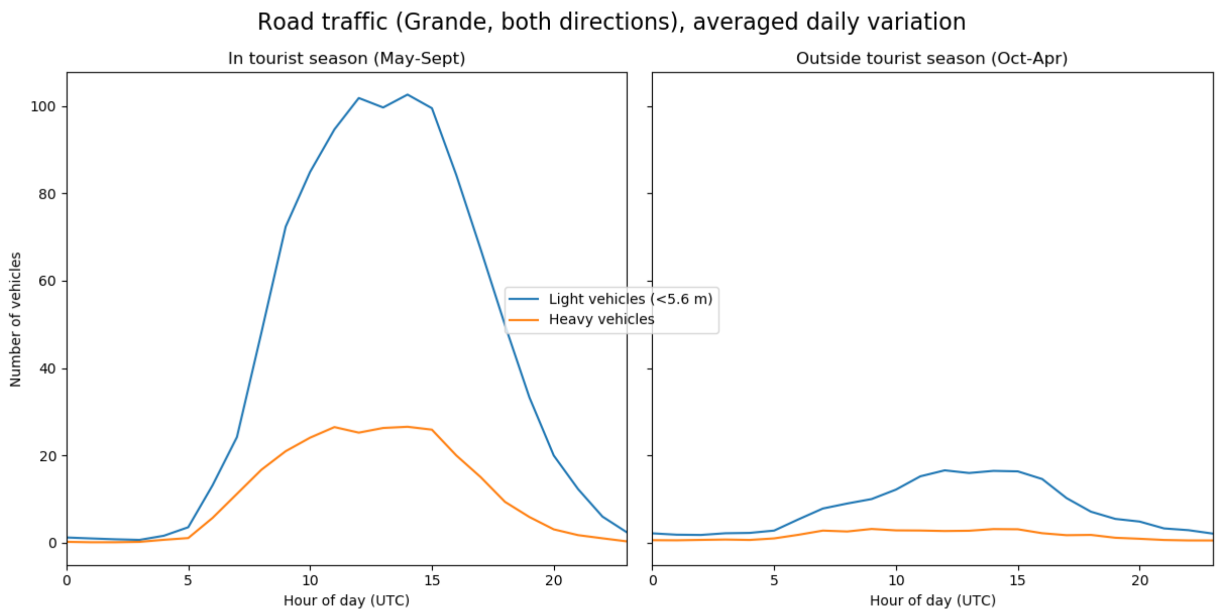
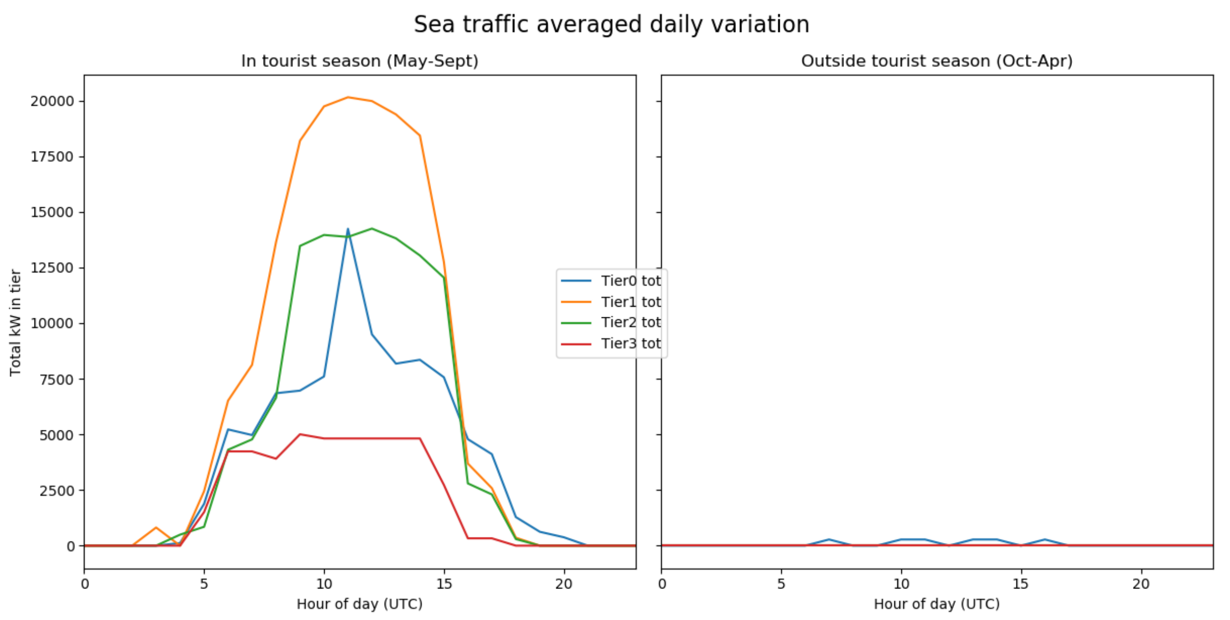
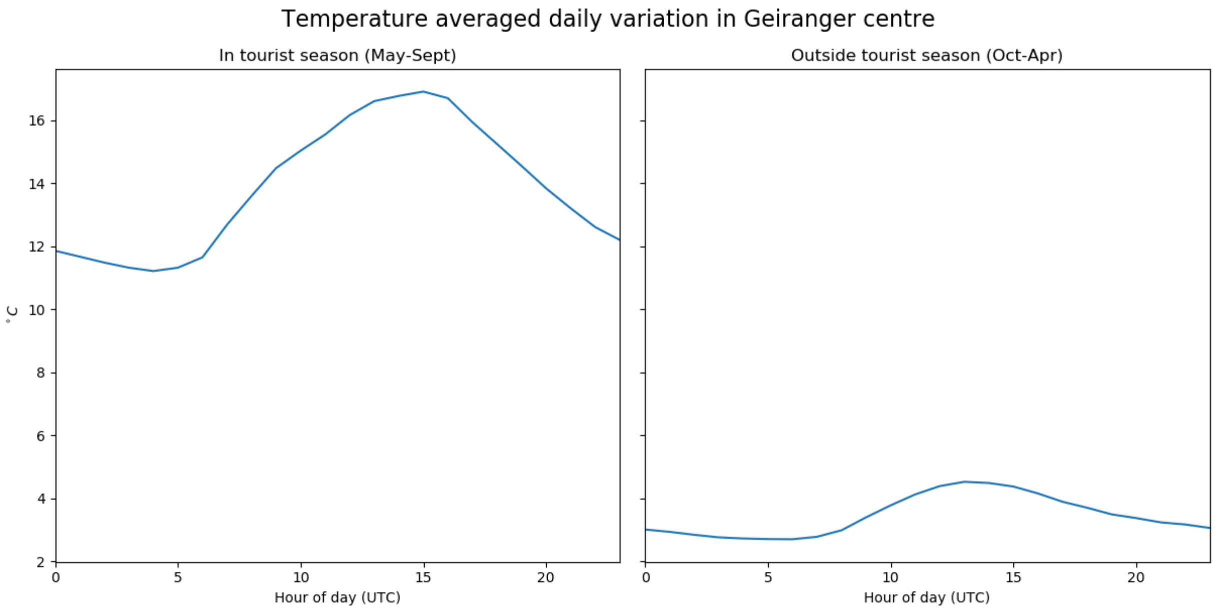
Tourist season (May – September)	**Non-tourist season (October – April)**
	
	

Figure 10: Land and sea traffic volumes for tourist and non-tourist seasons. Mean values for time of day.

Tourist season (May – September)	Non-tourist season (October – April)



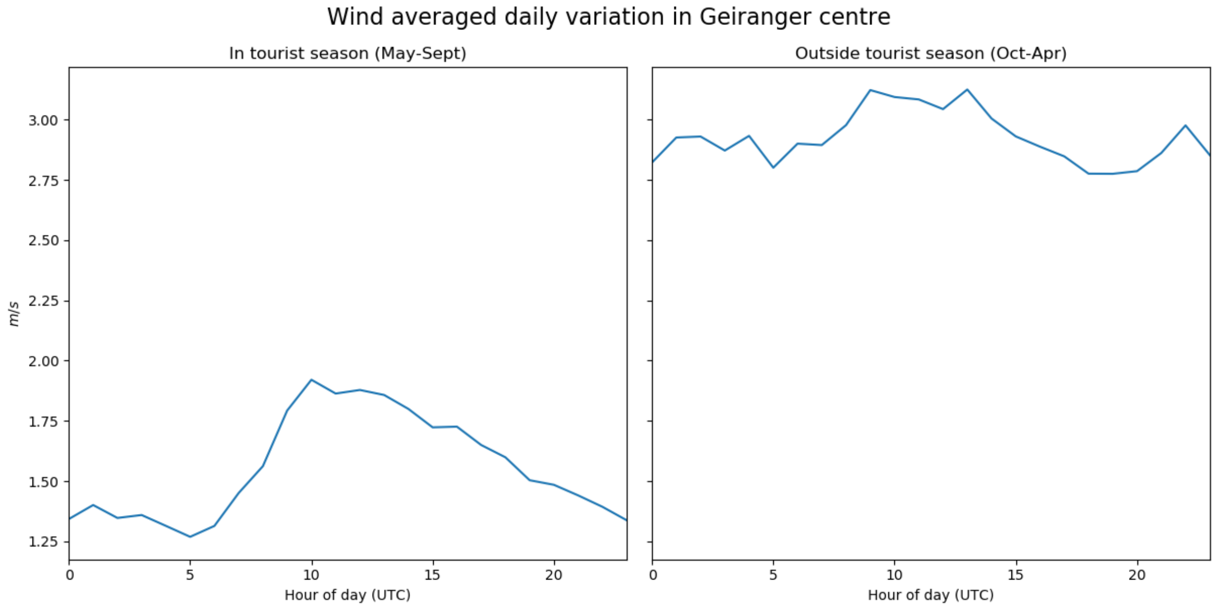


Figure 11: Temperature and wind speed for tourist and non-tourist seasons. Mean values for time of day.

3.3. Correlation analysis

To obtain a more precise measure of the relationship between air quality, traffic and meteorological conditions, we use the *Pearson correlation coefficient (Pearson's r)*. This measure allows to determine the linear relationship between two variables. A perfect linear positive relationship gives a coefficient of 1, while perfect negative relationship gives a coefficient of -1. A correlation coefficient of 0 means that there is no linear relationship between the two variables, i.e. they are uncorrelated. Appendix I elaborates on the computation of Pearson correlation coefficient.

Figure 12 shows correlation for hourly measurements of pollution variables to traffic and meteorological variables during the tourist season data (May – September). The plot has three sections where sea traffic variables are grouped to the left, followed by road traffic variables in the middle and meteorological variables to the right. As is clear from the plot, correlations to NO and NO_2 are overall stronger for all variables than correlations to PM1. All bars show correlation coefficients significant at 0.01 level.

For sea traffic variables, total gross tonnage (GT tot) and total passenger capacity for all ships at berth in the port of Geiranger (PAX tot) as well as number of cruise ships (Cruise) shows a moderate correlation to NO and NO_2. Road traffic also moderately correlates to NO and NO_2. The tier level data for cruise, ferry and hurtigruten is recorded as the total installed main and auxiliary engine capacity for these vessels when at berth in the port of Geiranger. Correlations for these variables are weak compared to the abovementioned sea traffic variables. For road traffic, results show that data from the Flydalen and Grande traffic registration stations correlate more strongly than data from Utsikten registration station. Both sea- and land traffic variables compare in strength of correlation to NO and NO_2, with positive coefficients in the interval 0.4 to 0.5.

For meteorological variables, the first observation is that variables correlate most strongly to NO_2 overall. Temperature and radiation are positively correlated to NO_2 while relative humidity has a negative correlation to NO_2. The same can be said for NO, with the exception that NO has no significant correlation to relative humidity. Correlation between PM1 and meteorological conditions are weaker than for NO and NO_2.

Pairwise correlations do not allow to distinguish causality, i.e. that one variable causes the other. This makes it difficult to determine what the unique contribution of meteorological and transport variables to air quality levels. As shown in Figure 11, both temperature and wind cycles align well with both traffic, as shown in Figure 10, and air quality levels, as shown in Figure 9. They are still useful to identify variables of interest for further analysis and setting up prediction experiments, as described in chapter 3.4.

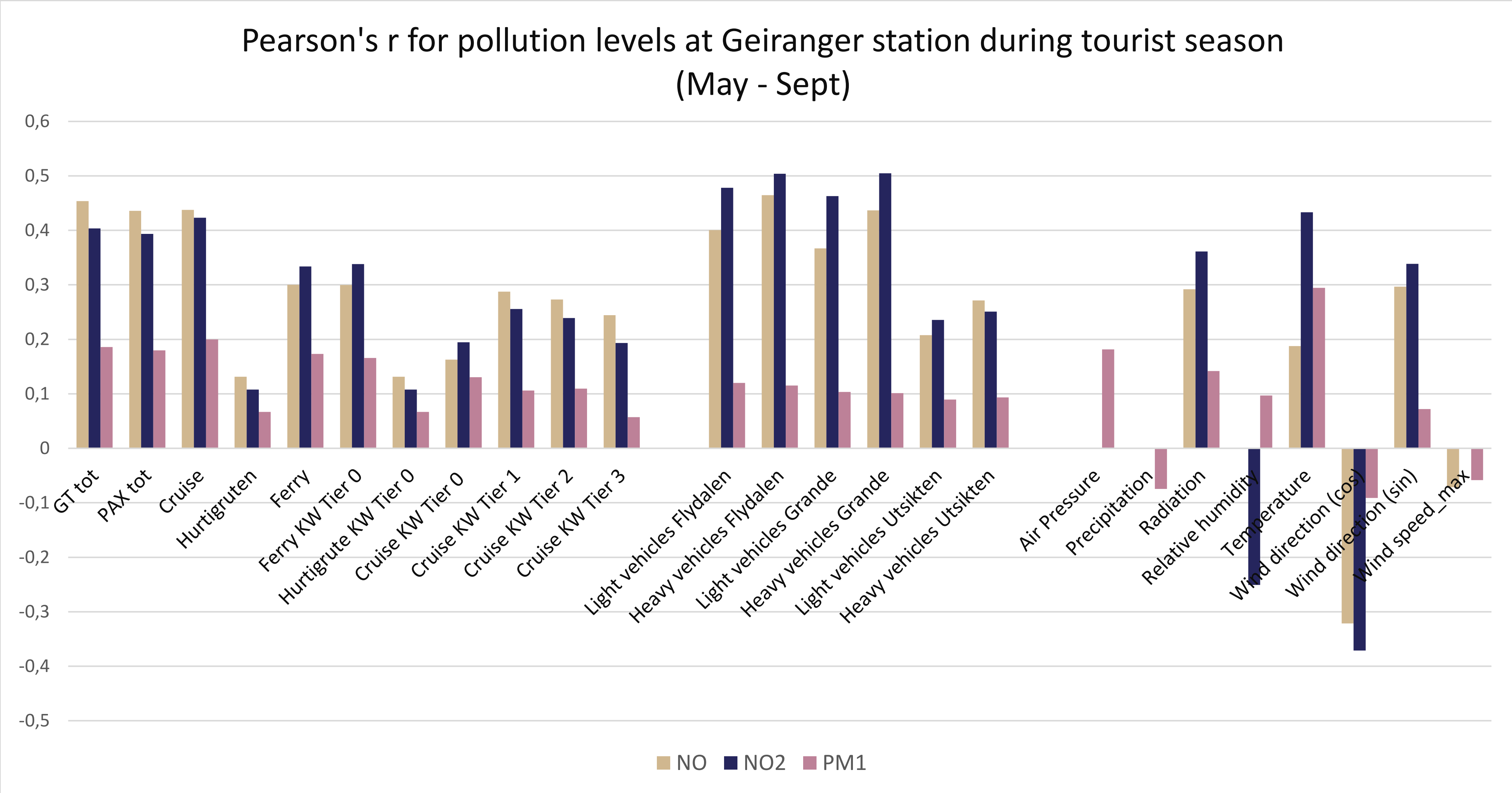


Figure 12: Correlation values for NO, NO2 and PM to traffic and meteorological conditions during the tourist season, significant at 0.01-level

3.4. Predictions of air quality in Geiranger

The primary objective of the pilot study is to explore the possibility of predicting air quality levels in Geiranger using information about traffic and meteorological conditions. Prediction models may be used to evaluate air quality given future scenarios and implementation of improvement measures if predictions reach a high level of accuracy. An important feature of prediction models is to evaluate whether predefined air quality levels are exceeded, such as e.g. Pollution Regulation or Air Quality Criteria levels. In this chapter, we present the results from prediction modeling experiments using machine learning tools. More information about methods and set up of experiments is described in Appendix I.

In our prediction model experiments, a machine learning algorithm is applied to build a mathematical air quality prediction model. The procedure consists of a *training* and a *testing* step. In the *training* step, the algorithm is provided with information about measured air quality levels as well as traffic and meteorological variables for a subset of all the data to build the prediction model. The algorithm is not given information about date, time and weekday in order to avoid predictions based on these temporal variables at the account of traffic and meteorological conditions. In the *testing* step, the developed prediction model is applied to the remaining data, where air quality levels are predicted based on information about traffic and meteorological conditions. The predicted value is compared to the measured value to evaluate the performance of the prediction model.

Experiment 1: Predicting air quality based on traffic and meteorological conditions

Figure 13 shows the best results for predicting PM1 levels at the Geiranger measurement station. The blue line shows measured values while the orange line shows predicted values. As the plot shows, the trend for PM1 levels during tourist months is captured by the model, but it does not perform well in predicting peak PM1 levels. Another observation is that PM1 levels for the non-tourist season is not predicted at any level of precision. This is expected as PM1 levels outside the tourist season is significant but may not be attributed to traffic.

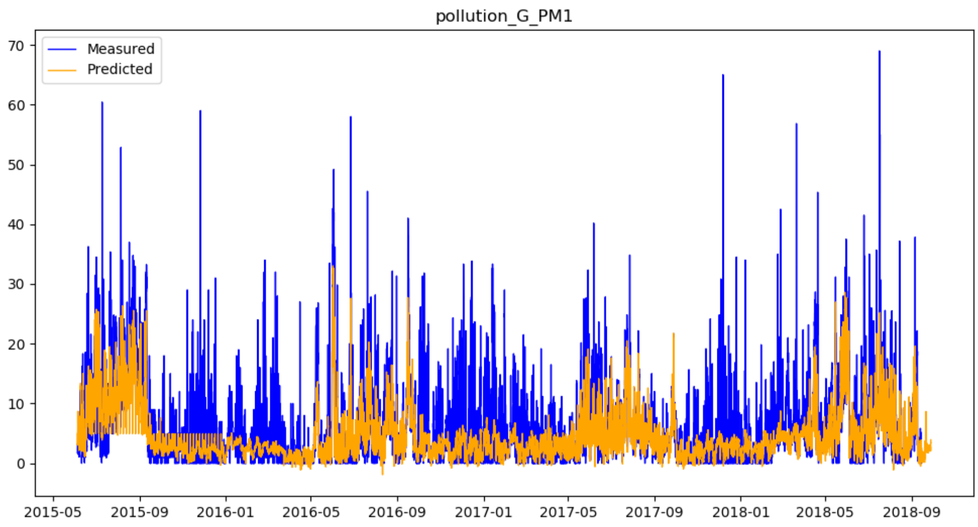


Figure 13: Measured and predicted values of PM1 at the Geiranger station 2015-2018

Figure 14 shows the best results for predicting NO levels at the Geiranger measurement station. As this data was collected from June 2017 onwards, there is no measured data for the previous years. The prediction model is still applied to this data as well. The plot shows that the prediction model only captures the trend and does not capture the peaks of NO concentrations.

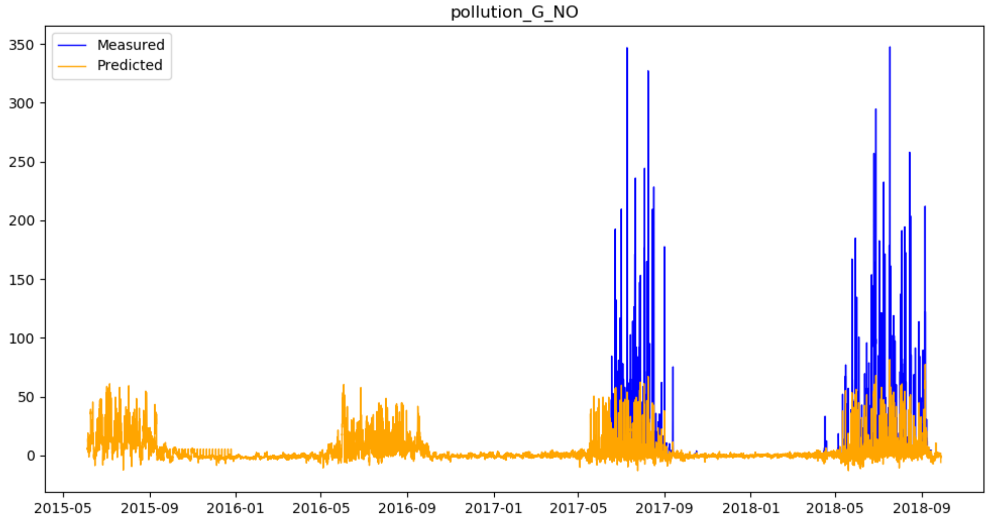


Figure 14: Measured and predicted values of NO at the Geiranger station 2017-2018

Figure 15 shows the best results for predicting NO_2 levels at the Geiranger measurement station. As is the case with NO, data is only measured from June 2017 but the model is still applied to data back to 2015. In this case, the model captures the trend and is closer to predicting the peaks than for PM1 and NO.

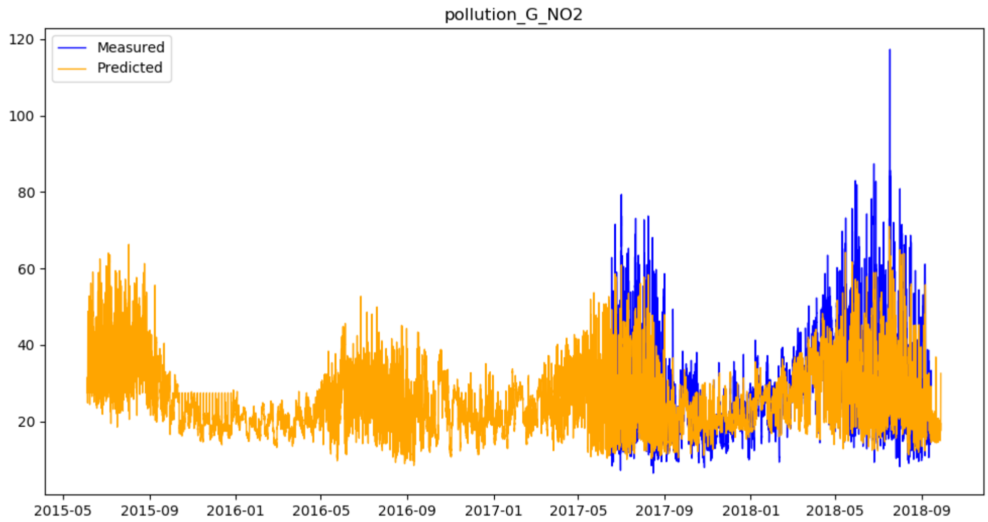


Figure 15: Measured and predicted values of NO_2 at the Geiranger station 2017-2018

Experiment 2: Predicting air quality with masked variables

In this experiment, the training and testing step of prediction modeling is performed without information about either meteorological conditions or traffic conditions. This helps to see how the original model compare in accuracy and help determine the importance of these conditions in making predictions. Prediction plots from all experiments are provided in Appendix II.

In order to evaluate model accuracy, we use the mean squared error (MSE). The lower MSE, the better fit the prediction has to the measured data. It is important to note that the MSE evaluation is only indicative as it varies between runs, i.e. the same experiment may yield slightly different MSE when run several times. Figure 16 shows MSE when predicting various pollutants and masking subsets of variables. As the plot shows, the lowest MSE for prediction of all pollutants is obtained when all traffic and meteorological variables are included (Full model). When meteorological variables are excluded from the analysis (Weather masked), MSE increases significantly irrespective of which pollutant is predicted. This means that prediction accuracy for pollutants decreases the most when meteorological variables are excluded.

For exclusion of both land- and sea traffic variables (All traffic masked) simultaneously, the MSE also increases compared to a full model. This increase is most evident for NO predictions, while it is considered small for PM1 predictions. It is worth noting that for NO_2 and PM1, excluding meteorological variables (Weather masked) gives a lower prediction accuracy (higher MSE) than excluding all information about traffic conditions (All traffic masked).

Finally, excluding either road traffic (Road traffic masked) or sea traffic (Sea traffic masked) does not seem to strongly affect MSE significantly when compared to a full model. In fact, a model with only road traffic and meteorological data might suffice in predicting the dataset used for predictions. As seen in section 3.2., road and sea traffic shows a temporal alignment during the tourist season, which could both be due of daily visitor cycles and that sea traffic generates some road traffic. This might be problematic if the prediction model is to evaluate a future scenario where for instance sea or road traffic is tuned, e.g. removed or altered in a significant manner.

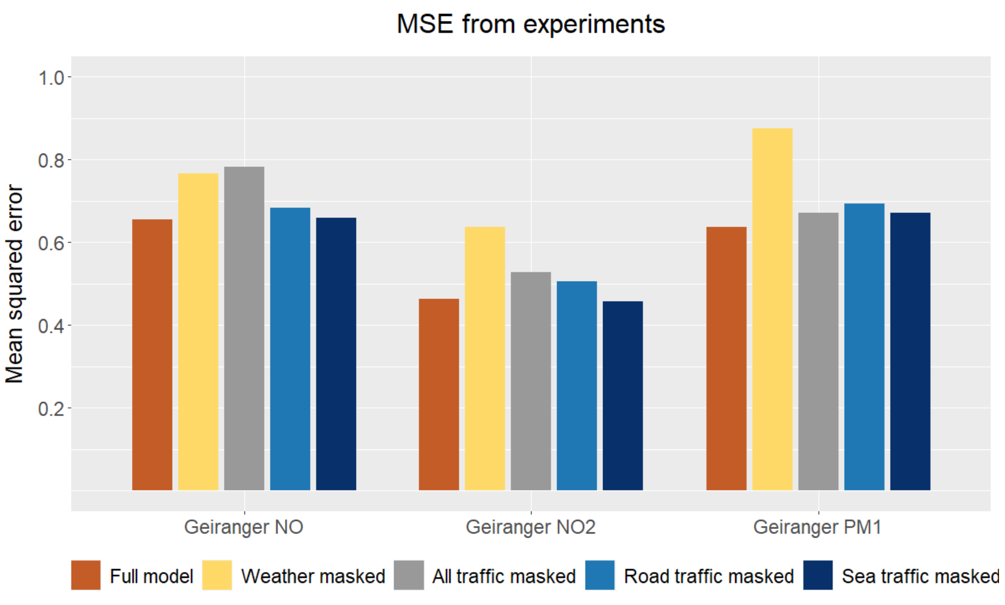


Figure 16: Mean squared error from prediction experiments for air quality at Geiranger measurement station

4. Conclusions and further work

This report summarizes findings from initial analyses of air quality, traffic and meteorological data for the Geiranger area in the period 2015-2018. A combination of conventional statistics and machine learning has been applied to better explore relationships between these conditions in Geiranger.

From initial observations of data, we see that air quality, traffic and meteorological conditions in Geiranger follows aligned temporal cycles over the course of a day during the tourist season (May – September). Outside the tourist season, these relationships are weaker, particularly for PM1. The covariance during tourist season makes it difficult to separate the different contributions to air pollution from land- and sea traffic as well as determine the mediating role of weather.

From the prediction models developed in the project, we may conclude that:

1. Prediction models follow the trends of air pollution levels but do not predict the magnitude of concentrations peaks.

All models underestimated the peak concentration levels for all pollutants studied. The model seemed to be far better at predicting NO_2 peaks than NO peaks. For PM1, the model identified elevated levels, but did not peak close to the measured data it attempted to predict.

2. NO_2 is more accurately predicted than PM1.

NO_2 is predicted at the highest level of accuracy among the pollutants included in the study followed by NO and PM1. PM1 concentration levels vary on a much finer scale and less instantly when traffic loads change compared to NO and NO_2, which could explain the low prediction accuracy. PM1 also seems to accumulate over time in Geiranger, which means that measurements do not only include recent emissions but also emissions from earlier transport activity. Outside the tourist season, PM1 levels sustain while traffic activity is significantly reduced. This could both be because of other emission sources as well as accumulation.

3. Meteorological conditions increase accuracy in prediction models.

We have not determined whether (or to what extent) this is in fact due to an important mediating role, or because weather contains information about season and time of day, or both. Among the meteorological parameters, radiation and temperature have a positive correlation to PM1, NO and NO_2 during the tourist season, i.e. higher temperatures and more radiation are observed along with higher pollution levels. Wind speed has a negative correlation to PM1 and NO. Precipitation does not seem to strongly correlate with pollution levels in Geiranger.

4. It is not possible to attribute air quality levels to land- or sea traffic activity with the current prediction models

This could be because there is a strong covariance of road and sea traffic in our dataset. Sea traffic also generates some road traffic, so road traffic contains information about sea traffic activity levels. This reduces the usability of our prediction models if they are to be used for future scenario explorations.

The objective of this pilot was to generate initial familiarity with the relationship between air quality, traffic and meteorological conditions in Geiranger and to predict air quality using machine learning tools. Although our results show that predictions are possible, a more elaborate development and analysis of data would help determine these relationships in a more precise manner.

In order to increase accuracies in prediction models to better capture peak concentration of pollutants, more data could be included in the analysis and models should be further improved.

The Geiranger Air Quality Monitoring Program is currently piloting new technology to conduct measurements at a higher resolution. This could potentially help understand the pollution dynamics of the finer fractions of PM far better than our current data permits. As PM does not strongly correlate to our traffic and meteorological variables, nor is predicted accurately, it is useful to see its spatial distribution in the area in a more precise manner.

Information about both road- and sea-traffic could also be characterized more elaborately, especially road traffic data should be explored at a higher resolution than in this pilot project. A potential issue to be addressed is that daily variations of mean road- and sea-traffic align quite well, which makes it difficult to attribute their separate activity levels to air quality. Future efforts towards this end would improve the usability of prediction models as it would enable to explore scenarios that are fundamentally different from observations in our dataset, e.g. scaling volume for road- and/or sea traffic, air abatement technology installations in ships and vehicles etc.

Additional extensions of this pilot also involve including other variables of interest. Both SOx and noise data could be included in future modeling and analysis as they may have adverse effects on environmental quality, human health and the general well-being of visitors and residents in the area. This data is already collected and should easily be added to the existing project database.

Additional data that could be subject to analysis in conjunction with the existing database are socio-economic information about visitors to the area. The SUSTRANS project conducted visitor surveys during July 2018, mapping visitor behavior, preferences, spending and satisfaction. This information, along with demographic variables, visitor volume etc., could also be interesting to explore using machine learning tools to see, if and how they are linked to traffic, weather and air quality conditions. Data from additional sources could also be interesting in this regard, e.g. by video footage and telephone data to determine crowding levels.

Air quality monitoring was initiated in Flåm 2019, a cruise destination in Aurlandsfjorden with several properties similar to Geiranger. Pooling this data with the Geiranger Air Quality Monitoring Program could be useful to learn more about pollution dynamics in these types of areas.

Finally, the translation of analytical results to operable action is necessary. The developed prediction models may be used to support decision making in Stranda Hamnevesen and other transportation authorities to see potential consequences of alternative courses of action. As part of this pilot, a visualization for the Geiranger area based on the project database is underway. The visualization tool could be further developed in a more interactive manner to allow to alter conditions and simulate implementation of measures in transportation planning.

Although this pilot has investigated relationships between air quality, traffic and meteorological conditions in Geiranger, several extensions are possible both in the short- and long-term. A main project following the results of this pilot should be developed to further explore how Geiranger can further develop as a smart, safe and sustainable fjord.

References

[1] J. Löffler, "Long-Term Air Quality Monitoring Program. UNESCO World Natural Heritage "Geiranger Fjord", Norway," University of Bonn 2016.

[2] Hanne Weggeberg *et al.*, "Utslipp til luft og sjø fra skipsfart i fjordområder med stor cruisetrafikk - kartlegging av forslag til tiltak," Rambøll2017.

[3] J. Löffler, "Long-Term Air Quality Monitoring Program. UNESCO World Natural Heritage "Geiranger Fjord", Norway," University of Bonn 2018.

[4] Folkehelseinstituttet, "Svar på henvendelse om luftforurensning fra cruiseskip i Geiranger," 2017.

[5] (2016). *Budskapsnotat til samferdselsminister Ketil Solvik-Olsen*.

[6] World Health Organization, "Health effects of particulate matter. Policy implications for countries in eastern Europe, Caucasus and central Asia," 2013, Available: http://www.euro.who.int/__data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf.

[7] Miljødirektoratet, "Grenseverdier og nasjonale mål. Forslag til langsiktige helsebaserte nasjonale mål og reviderte grenseverdier for lokal luftkvalitet," 2014, Available: https://www.miljodirektoratet.no/globalassets/publikasjoner/m129/m129.pdf.

[8] Klima- og miljødepartementet, "Forskrift om begrensning av forurensning (forurensningsforskriften)," 2004, Available: https://lovdata.no/dokument/SF/forskrift/2004-06-01-931.

[9] Nasjonalt Folkehelseinstitutt, "Luftkvalitetskriterier. Virkninger av luftforurensning på helse," Oslo 2013, Available: https://www.fhi.no/globalassets/dokumenterfiler/rapporter/2013/luftkvalitetskriterier---virkninger-av-luftforurensning-pa-helse-pdf.pdf.

APPENDIX I: Method descriptions

Data processing

Merging datasets

All data sources were joined into a single multivariate time series. The timestamps of the datasets are used to align them into one complete dataset. For the air quality and meteorological data, the timestamp is consistently UTC+1, while all other datasets were initially stored with UTC+1 (winter time) and UTC+2 (summer time) adjustments. All the data was converted to consistently follow a UTC+0 timestamp.

Handling missing data

For the air quality and meteorological data sets, some data is missing due to power outage, equipment failure and equipment upgrades during the sampling period. Some variables have long segments of missing data, and even though some of the variables are on and off in a synchronized manner, valuable data would have been lost if one chose to throw away all entries that have data missing. Instead, several source group presence indicators were added to the dataset, to serve as explicit input to the prediction models. Source group presence is 1 if the associated group of variables are present and 0 if they are not.

No data “cleansing” was done.

Standardization

In Machine Learning, it is a recommended practice to have inputs of similar scale. Also, since the experiments involve predicting multiple variables, the outputs need to be of similar scale, so that the errors of some variables do not dominate those of others. To achieve this, the dataset was standardized, meaning the entries in each field was shifted against the mean and divided by the standard deviation. Missing data were then replaced by zeros, which correspond only to relatively harmless average values in case the source group presence variables are not used by the active model.

Correlation analysis

The purposes of the correlation analysis is to understand potential statistical relations between influential factors, like e.g. weather conditions, to the air quality data. For this purpose, the so-called *Pearson correlation coefficient* was analyzed for pairs of variables in the data sets. The coefficient value is in the range -1 to +1 and indicates how strongly two variables are linearly related to each other. A value of 0 indicates that there is no linear relation between the two variables, while a value of +1 or -1 indicates that the two variables are perfectly linear or inversely linear respectively. The Pearson’s correlation coefficient for variables X and Y is obtained from

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

where n is the sample size, X_i and Y_i are sample points, and $\bar{X}$ and $\bar{Y}$ are sample means for the respective variables.

In order to evaluate whether correlations are statistically significant, i.e. not due to random variations, we use α=0.01 adjusted using a Bonferroni correction. The correction compensates for the increase in likelihood of rejecting a null-hypothesis when testing multiple hypotheses. As we correlate 567 pairs of variables in our correlation analysis, the adjusted significance level becomes

$$\alpha = \frac{0.01}{567} = 1.76 * 10^{-5}$$

Prediction modeling

A valid definition of artificial intelligence (AI) is the effort to automate intellectual tasks normally performed by humans. This has been the main goal for researchers and data-analysts within the AI field all the way back since the 1950s. In the past, AI developments were limited by the lack of computational power. However, this has changed drastically in recent years.

Today, machine learning (ML) is an extremely active sub-field of AI. The core idea of ML is that humans input data and expected answers to an algorithm, which the algorithm will train on for several iterations. After the training procedure, the algorithm is able to produce original answers in a real-life system, for example, predictions of PM values in Geiranger.

The data collected from Geiranger is time series data. In other words, the sensor measurements change during time. The Long-short term memory (LSTM) is considered the state-of-the-art ML technique for time series data since it is able to remember relevant time information from the past before making new predictions. This procedure is performed by utilizing a memory cell which is controlled by three gating units; the forget gate, the input gate, and the output gate.

- The forget gate (f_t) decides which past information is no longer relevant for the next prediction.
- The input gate (i_t) decides which new information the memory cell will input and store.
- The output gate (o_t) decides which information the LSTM will output to make the next prediction.

Through these gates, the LSTM has the power to remove or add information in order to both learn past information and current information. This makes it extremely suitable to make predictions on time series data. The LSTM is illustrated in the Figure below. A simple explanation is that several LSTM-blocks will roll over each sensor measurement in the data set in the time direction. X_{t-1} is the previous time step, X_t is the current time step, and X_{t+1} is the next time step.

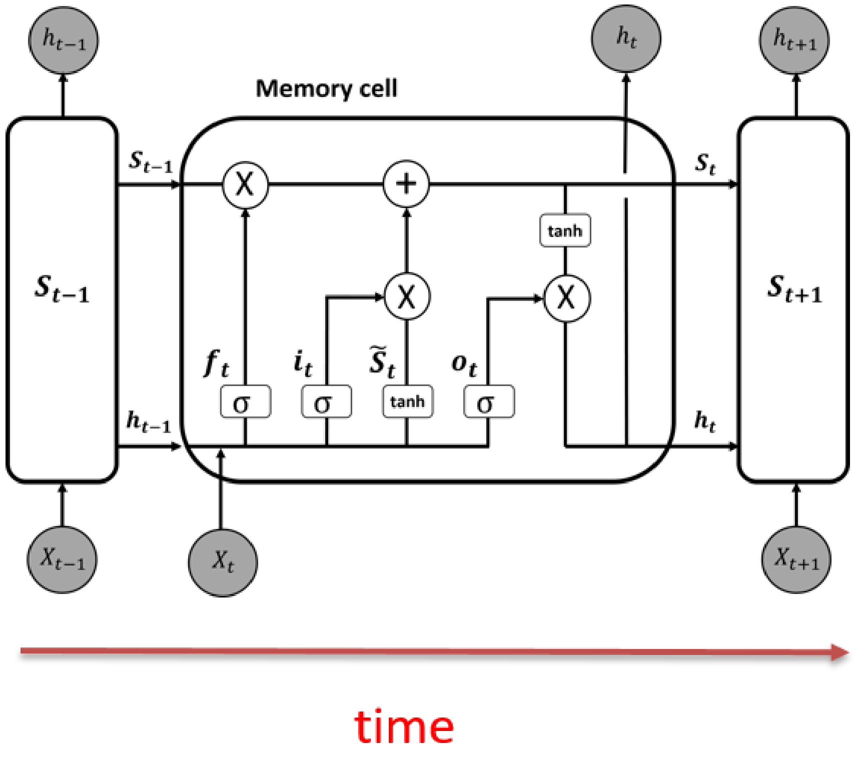


Figure 1 - the long-short term memory

Set-up for data prediction models in the pilot project

To allow structured comparison of different models, a custom object-oriented model hierarchy and training setup was created for the project. This framework is configurable by simple modification of parameters.

The model output can be of two types:

1. Direct predictions of the target variables. (Used in the pilot study.)
2. Parameters for a conditional probability distribution that target variables are assumed to be sampled from. (Experiments not completed yet.)

The two types of models are both trained and evaluated differently. To reduce sensibility to extreme data values, the direct prediction models are trained on a robust pseudo-Huber loss. They are however *evaluated* on the mean squared error of standardized values. The distribution models are optimized to maximize the likelihood of the parameters with respect to the data, such that a reasonable distribution of the target variables follows from a given sequence of input variables. This means the distribution models incorporate a sense of confidence in their own prediction, in form of the probability density at the predicted values. This also provides a way to handle extreme data values. The evaluation of distribution models during cross-validation is based on the same likelihood objective, but for comparisons of predictive performance again the mean squared error is used.

The model types that were considered for direct predictions were: Multilinear Regression, FNN, Recurrent Linear Model, LSTM

The model types that were considered for distribution output were: FNN and LSTM

Short descriptions of models

Multilinear regression is a baseline that contains and expands upon the power of linear correlation analysis, to predict the values of variables from linear relationships.

Feedforward Neural Networks, or FNN, are the basic standard type of neural network. They work by alternating between affine transformations (like Multilinear regression) and nonlinear activation functions. FNNs are proven to be universal approximators of bounded functions in real numbers, capable of representing much more complex relations than linear regression models. The implementation used in experiments is a Self-normalizing Neural Network with alpha dropout.

When using sequential input, it may be possible to exploit that the relationship between subsequent observations are similar. This motivates the usage of Recurrent Neural Networks, RNNs, for sequential analysis. RNNs have feedback connections that propagate information from earlier observations, called a "state", and use the same parameters on each step in the time window. That means the parameters encode the same relation at each step.

A very basic variant of RNN is a Discrete-Time Linear State-Space Model, wherein all transformations are linear (thus it may not qualify as a neural network, but it has the same flow as an RNN). I call the model DTLTI for Discrete-Time Linear Time-Invariant. The calculation of a DTLTI on an initial state and an input window of limited length in principle evaluates to a multilinear expression. However, the weight sharing between time steps allows it to generalize better on small datasets.

A more advanced variant is the non-linear LSTM. It is a very successful type of RNN, which is capable of general computation on sequential inputAgain, when applied to a limited input the calculation could have been represented by a deep FNN (inputs on multiple layers and multiple activation functions in each layer), but during training, an FNN does not profit from a structure that generalizes over sequential transitions. The LSTM in these experiments is configured with peephole connections, skip connections, layer normalization and recurrent dropout.

The distribution models try to fit a conditional Gaussian Mixture Model (reference needed) with parameters and mixture weights generated by the neural networks.

Training

All the models are trained using a variant of Gradient Descent, which means that the total loss function on a batch of input-output pairs is differentiated with respect to the (many) parameters, and a small parameter step is made in a direction that appears to make the loss smaller. This is repeated very many times. Eventually the loss will stop improving.

A common problem is overfitting, which means the model learns to remember the training data rather than representing the underlying patterns. This is typically a problem when the dataset is modest in size. In a Machine Learning context, that can be said about our dataset. The most complex models tried in this study have millions of adjustable parameters in total, compared to less than three million numbers in the dataset, whereof only about half a million are pollution measurements, whereof only about 100 000 are in the set of highly prioritized variables.

A way to avoid overfitting is to keep models small, that is to have few adjustable parameters in them. This is one of the main advantages of the linear/affine models over the big LSTM and big FNN models. If a simple model can make useful predictions, it may in some cases be preferred over a complex one that may be biased toward the training data.

Another measure against overfitting is to train with dropout, which means some layer outputs are purposefully "lost" during training, forcing the network to incorporate redundancy, and reducing its ability to exactly memorize the training data.

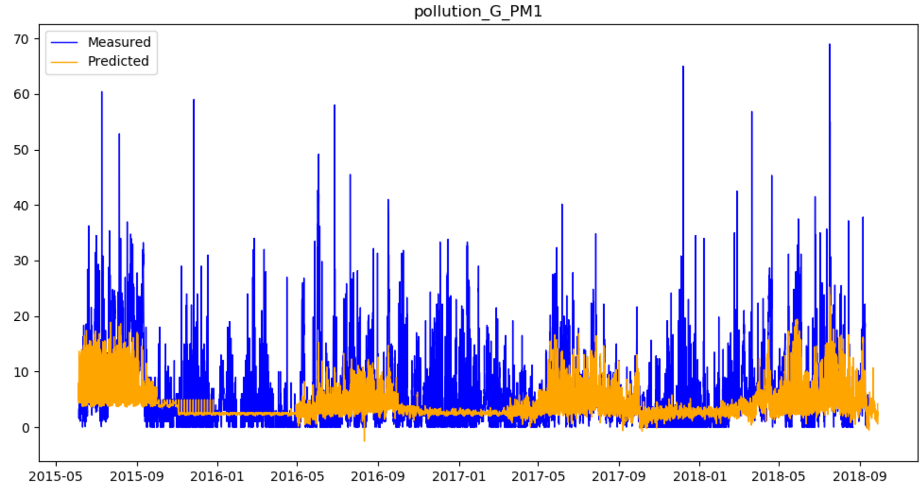
Overfitting can be kept under control by splitting the dataset into one part for training and one part for validation. Keeping track of the loss on the validation set during training facilitates the simple technique of *early stopping*, which is stopping the training when the validation loss starts to increase again, ideally when at the global minimum, the deepest valley of the loss landscape. A variant of this method was used in the experiments.

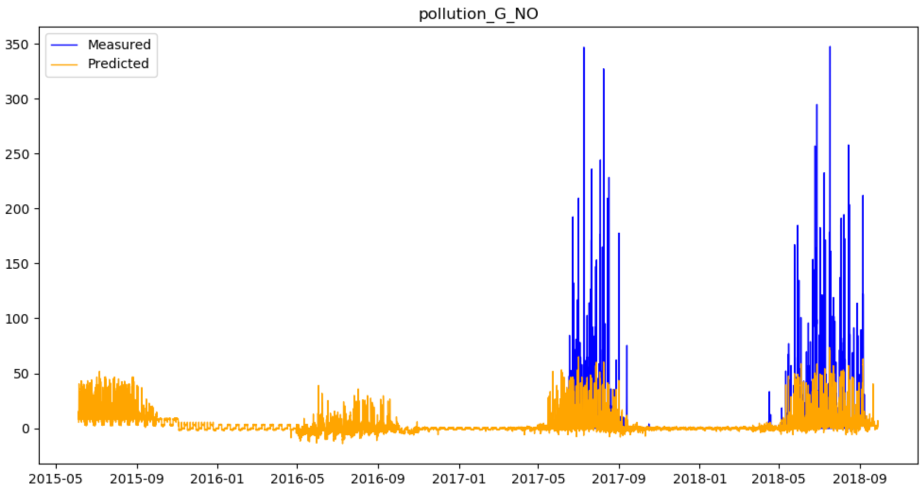
The problem with splitting the dataset, is that it throws away scarce and valuable data. Therefore, for this project, K-fold Cross-Validation was employed. It consists of making a number K complementary training-validation splits of the data, and judging the models based on the average validation loss. Moreover, in this project the CV was implemented as an ensemble model that combines the outputs of the constituent models, and the splitting was implemented as a masking of errors or outputs of each submodel, based on timestamps. This enabled the models to be trained together and validated together.

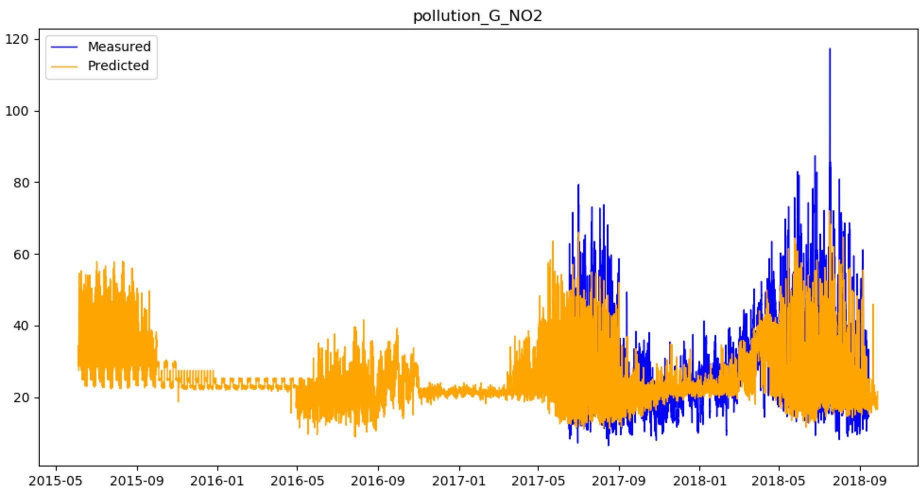
A difficulty with splitting the dataset is that the source presence varies so much yet tend to be the same over hundreds of neighboring samples. A simple splitting where the submodels got ordered 1/K slices of the dataset would thus make the models focus on very different variables. Instead a method of sharding was used, in which the samples were binned into time windows (shards) of 104 hours, and shards were alternately assigned as validation data to each submodel. To avoid information leakage from training to validation, a quarantine equal to the input window length (9 hours) was held out of each shard.

The model that has generated predictions for the figures is a small (width=16, depth=2, parameters=13653) LSTM network trained to predict *all* 21 pollution variables. The rationale for using all pollution variables as targets, rather than only the prioritized ones, was two-fold: 1. This gives the model more information about the underlying system. 2. It further helps combatting overfitting to a selected subset.

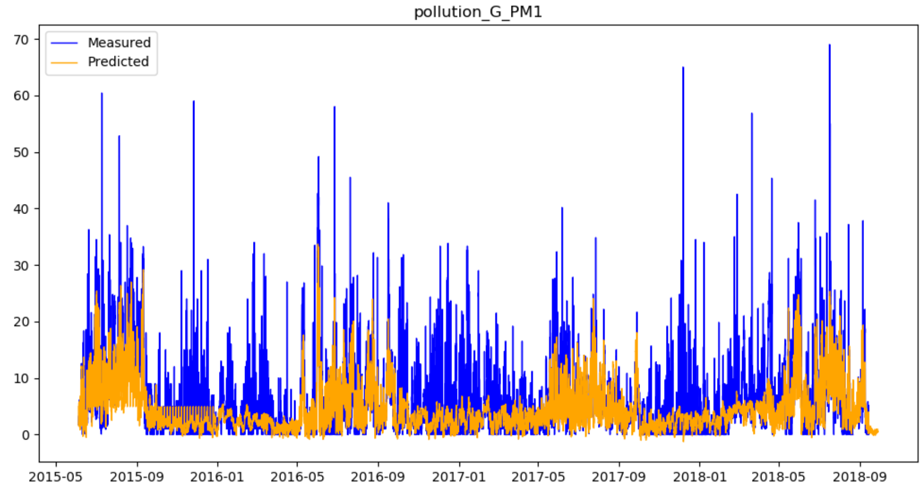
APPENDIX II: Plots from prediction experiments
A: Masking meteorological variables

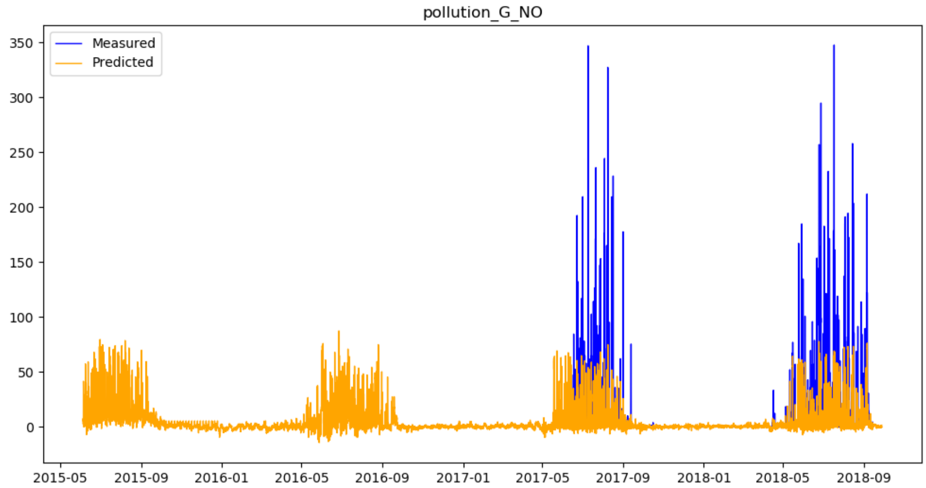


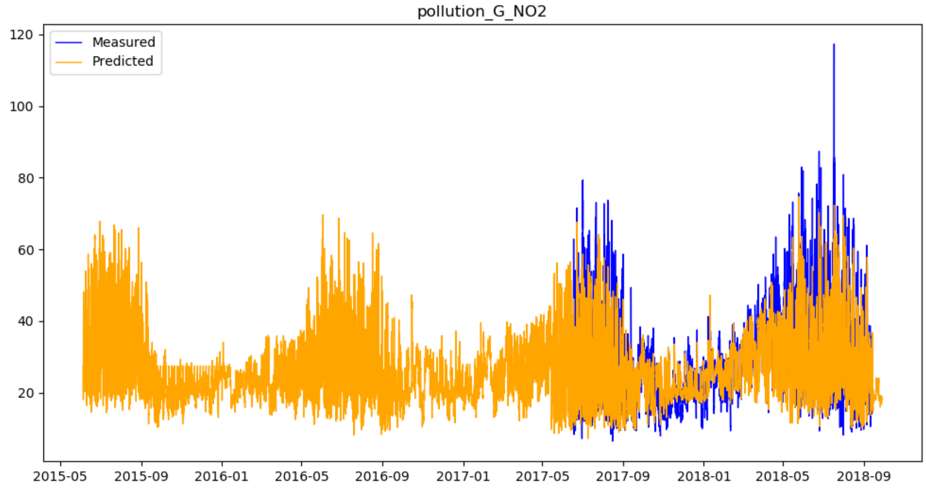




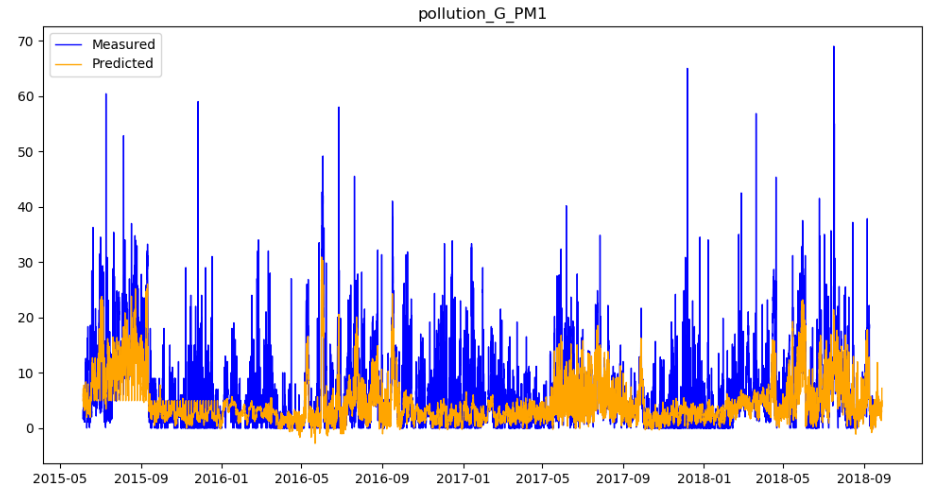
B: Masking road traffic variables

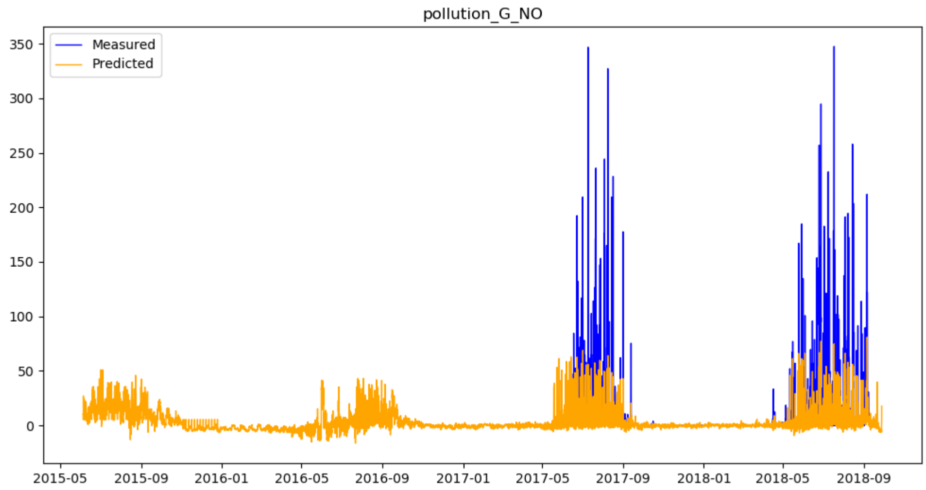


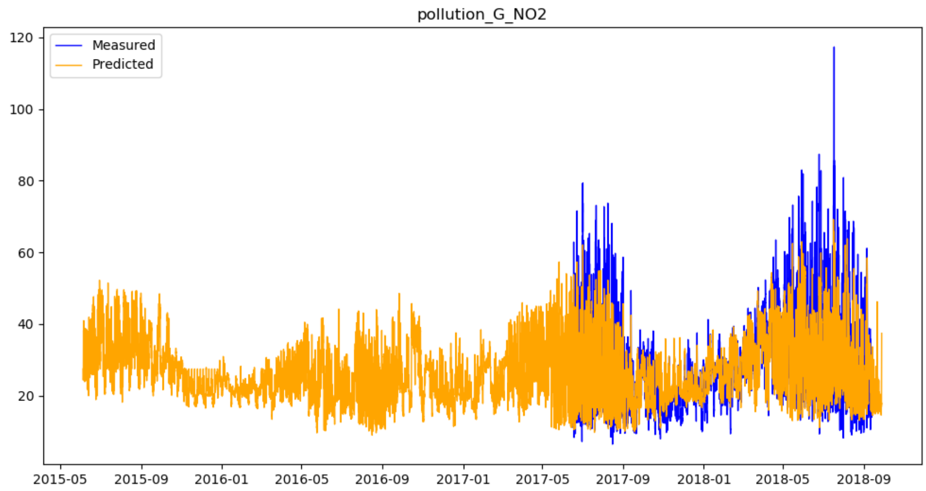




C: Masking sea traffic variables







NTNU

Norwegian University of
Science and Technology