



**Stockholm  
University**

Department of Meteorology

**Evaluation and validation of Copernicus Atmosphere  
Monitoring Service regional ensemble forecast  
of air pollutants and birch pollen in the Stockholm region**

Beatrice Säll

Department of Meteorology  
Master thesis report 30 HP (MO9001)

Autumn 2018

Supervisors:

Christer Johansson, Department of Environmental Science and Analytical Chemistry

Caroline Leck, Department of Meteorology

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	CAMS regional ensemble forecast . . . . .	4
1.2	Pollutants and birch pollen . . . . .	5
1.2.1	Nitrogen dioxide . . . . .	6
1.2.2	Ozone . . . . .	6
1.2.3	Particulate matter . . . . .	7
1.2.4	Birch pollen . . . . .	8
<b>2</b>	<b>Method</b>	<b>8</b>
2.1	Urban and regional observation sites . . . . .	8
2.2	The data and work process . . . . .	10
2.3	Statistical indicators . . . . .	11
2.4	Binary statistics . . . . .	13
2.5	Possible sources of error . . . . .	15
<b>3</b>	<b>Results and discussion</b>	<b>15</b>
3.1	Data set distributions . . . . .	15
3.2	Evaluation of forecast performance . . . . .	16
3.2.1	Nitrogen dioxide . . . . .	16
3.2.2	Ozone . . . . .	19
3.2.3	Particulate matter, PM <sub>2.5</sub> . . . . .	24
3.2.4	Particulate matter, PM <sub>10</sub> -PM <sub>2.5</sub> (coarse mode) . . . . .	28
3.2.5	Birch pollen . . . . .	32
3.3	Statistical indicators . . . . .	34
3.4	Forecast accuracy as a function of forecast time . . . . .	38
3.5	Health risks . . . . .	39
<b>4</b>	<b>Summary and Conclusions</b>	<b>40</b>
<b>5</b>	<b>Appendix I: QQ-plots</b>	<b>45</b>

**6 Appendix II: Pollutant concentration as a function of wind speed****46****Abbreviation list**

APIs – Application Program Interface  
ACES – Department of Environmental Science and Analytical Chemistry at Stockholm University  
CAMS – Copernicus Atmosphere Monitoring Service  
CDO – Climate Data Operators  
CEN – European Committee for Standardization  
CH<sub>4</sub> – Methan  
CO – Carbon monoxide  
CO<sub>2</sub> – Carbon dioxide  
DJF – December, January and February  
ECMWF – European Center for Medium-range Weather Forecasts  
EPA – Environment Protection Agency  
EU – European Union  
EUMETSAT – European Organization for the Exploitation of Meteorological Satellites  
FAIRMODE – Forum for Air quality Modeling  
FAM – Fraction of alarms missing  
FAR – False alarm ratio  
FBI – Frequency bias index  
FGE – Fractional gross error  
FMI – Finnish Meteorological Institute  
H – Atomic hydrogen  
HNO<sub>3</sub> – Nitric acid  
HO<sub>2</sub> – Hydroperoxyl radical  
INERIS – French National Institute for Industrial Environment and Risks  
JJA – June, July and August  
M – Non-reacting molecule  
MET Norway – Norwegian Meteorological Institute  
METEO-FRANCE – French national meteorological service  
MNMB – Modified normalized mean bias  
MQI – Model quality indicator  
MQO – Model quality objective  
NO – Nitrogen oxide  
NO<sub>2</sub> – Nitrogen dioxide  
NO<sub>x</sub> – Nitrogen oxides  
O – Atomic oxygen  
O<sub>2</sub> – Oxygen gas  
O<sub>3</sub> – Ozone  
OH – Hydroxyl radical  
PM – Particulate matter  
PM<sub>2.5</sub> – Particulate matter with diameter up to 2.5 μm  
PM<sub>10</sub> – Particulate matter with diameter up to 10 μm  
POD – Probability of detection  
QQ-plot – Quantile to Quantile plot  
R – Correlation coefficient  
RIUUK – Rhenish Institute for Environmental Research at the University of Cologne  
RMSE – Root mean square error  
RMS<sub>*U*</sub> – Root mean square of the measurement uncertainty  
RV – Reference value  
SMHI – Swedish Meteorological and Hydrological Institute  
SMURBS – Smart Urban Solutions for air quality, disasters and city growth  
UTC – Universal Time Coordinated  
VOC – Volatile organic carbon

**Abstract:** In this study, Copernicus Atmosphere Monitoring Service regional air quality ensemble forecast is evaluated and validated in the Stockholm region for pollutants nitrogen dioxide, ozone, particulate matter with diameter up to 2.5 and 10  $\mu\text{m}$  respectively as well as birch pollen. These forecasts are planned to be used in a public information system of the daily air quality situation in Stockholm. However, it is important to examine how accurate the forecasts are in the region. Evaluation and validation made in this study compares forecasts of each pollutant and birch pollen with corresponding observations at three sites in the Stockholm region. One urban background site and two regional background sites. Both urban and regional sites are included in order to examine the impact of long-range transport air pollutants. Time series comparisons as well as statistical indicators were used in evaluation and validation. This study concludes that Copernicus Atmosphere Monitoring Service regional air quality ensemble forecast performs well in the Stockholm region for pollutants which mostly depend on long-range transport by advected air masses from the European continent. Furthermore, the forecast does not capture variations in concentrations of pollution due to variations in local emissions. This is however not resolved by the forecast model so the result was reasonable. An other conclusion is that forecast time is not a main source of uncertainty in the forecast. Instead, it is again inaccuracy in input of local emissions and the fact that local emissions are not resolved by the forecast which seem to be the largest factor of uncertainty in the forecast of the pollutants which mostly depend on local emissions. Copernicus Atmosphere Monitoring Service regional ensemble forecast is suitable for forecasts of regional background air but not urban air. In the urban region, a forecast with higher horizontal and vertical resolution would be a necessary supplement to account for local emissions. Some limitations in the forecast construction are discussed as well as a few recommendations on what would be important to account for in a forecast used in a public information system of daily air quality.

## 1 Introduction

Ambient air pollution is estimated to cause 7 million premature deaths yearly worldwide (WHO, 2018a). Diseases such as chronic obstructive pulmonary disease, lung cancer, heart disease, stroke can all be worsened or caused by exposure to ambient air pollution. Furthermore, poor air quality can worsen asthma and cause acute respiratory infections (WHO, 2018a). In Europe, Swedish cities have among the lowest concentration of air pollution and levels of several air pollutants have decreased in recent years (Gustafsson *et al.*, 2018). Even so, it is estimated that approximately 7600 premature deaths in Sweden occur yearly as a result of exposure to polluted ambient air (Gustafsson *et al.*, 2018). As a tool to quantify health risks from air pollution, a Health Risk Index has been developed for Stockholm. The Health risk index is based on the relationship between concentration of air pollutants and negative health response, found in an epidemiological study of short-term effects from exposure to various pollutants and pollen in emergency visits regarding asthmatic issues (Johansson *et al.*, 2013; Olstrup *et al.*, 2019). The relationship between concentration and health response provides a percentage of the increase in emergency visits regarding asthmatic issues due to elevated concentrations on pollutants and pollen. According to this health risk index, elevated levels of birch pollen and  $\text{PM}_{10}$  contribute most to the health risk index during their maximum seasons (Johansson *et al.*, 2013).

Air pollution have both natural and antropogenic sources. Antropogenic sources of pollutants are, among other, fossil fuel combustion, heat and power generation, industrial facilities and municipal waste burning (WHO, 2018b). Both local and remote emission of pollutants are important for concentrations of pollution in a certain area, since many pollutants have a residence of longer than a week which is the time necessary for them to remain in an air mass advected over large regions (Wallace & Hobbs, 2006). Long-range transported air pollutions contribute significantly to background concentrations of pollutants in ambient air in the Stockholm region (Jönsson *et al.*, 2013; Targino *et al.*, 2013).

European cities have a responsibility to monitor air quality in accordance with the Air Quality Directive 2008/50/EC and cities in Sweden also in accordance with Luftkvalitetsförordning 2010:477 (Sabelström, 2018). The cities also have a responsibility to inform the general public about current air quality. In order to do so, many cities forecast the air quality situation of the near future. Current air quality forecasts in the Stockholm region are mostly based on local emissions, with some consideration to backward trajectories of air masses in order to get an indication of the contribution from long range transport. This approach does however not provide quantitative information of pollution concentrations from long

range transport. Copernicus atmosphere monitoring service (CAMS) provide free air quality dispersion forecast for entire Europe, including long-range transported pollutions within the entire domain (20°W/30°N/45°E/70°N) (Meteo-France, 2017). These forecasts could therefore result in more comprehensive forecasts of in the Stockholm region. Furthermore, short term air quality forecasts like CAMS regional ensemble forecast could be used when forecasting the Health Risk Index. The department of Environmental Science and Analytical Chemistry at Stockholm university (ACES) is involved in an European Union (EU) project called Smart Urban Solutions for air quality, disasters and city growth (SMURBS), where they will develop a prognostic Health Risk Information system for Stockholm considering health effects of several pollutants and pollen using CAMS regional ensemble forecast (ACES, 2017). However, if CAMS regional ensemble forecast is to be used when implementing such public information systems, its reliability in the region must be evaluated. CAMS does continuous validation of their forecast products for the entire European domain in general. However, air quality conditions in Sweden and nordic countries differ somewhat from the southern European domain. For example use of stubbed tires lead to peaks in the amount of coarse particulate matter during spring. This is not seen in southern Europe (Latvala, 2018b).

## 1.1 CAMS regional ensemble forecast

CAMS is a part of the EU program Copernicus. The Copernicus program is managed by the EU Commission and CAMS is implemented together with EU member states, the European space agency, the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT), the European Center for Medium-range Weather Forecasts (ECMWF), EU agencies and Mercator Océan (Copernicus, 2016). CAMS continuously monitor the atmosphere’s composition on global and regional scales. CAMS regional ensemble forecast is constructed using median output of seven chemical transport models developed in Europe. CAMS regional ensemble forecast is used in this study since previous assessments have found it to be the more accurate than any individual model (Meteo-France, 2017). Input emission data, boundary conditions and meteorological parameter settings are common to all models (Meteo-France, 2017). Surface emission data is from the year 2011 (Meteo-France, 2017). Horizontal and vertical resolution as well as chemistry schemes, land coverage and physical parameterizations, such as turbulence, convection and deposition, vary between the models (Meteo-France, 2017). A brief summary of key features in each model can be found in Table 1 and key features in the ensemble forecast can be found in Table 2.

**Table 1:** Key characteristics, including model name, developing institute, horizontal resolution and vertical resolution, of each individual model Meteo-France (2017).

Model name	Developing institute	Horizontal resolution	Vertical resolution
CHIMERE	French National Institute for Industrial Environment and Risks (INERIS), France	0.15° x 0.1°	Variable, 8 levels from the surface up to 500 hPa
EMEP	Norwegian Meteorological Institute (MET Norway), Norway	0.25° x 0.125°	20 layers up to 100 hPa, with approximately 10 in the boundary layer
EURAD-IM	Rhenish Institute for Environmental Research at the University of Cologne (RIUUK), Germany	15 km	23 layers up to 100 hPa lowest layer thickness about 35 m about 15 layers below 2 km
LOTUS-EUROS	Royal Netherlands Meteorological Institute (KNMI) & Netherlands Organisation for Applied Scientific Research (TNO), the Netherlands	0.25° x 0.125°	4 layers, top at 3.5 km above sea level
MATCH	Swedish Meteorological and Hydrological Institute (SMHI), Sweden	0.2°	52 levels
MOCAGE	French national meteorological service (METEO-FRANCE), France	0.2°	47 layers up to 5 hPa, lowest layer thickness 40 m, 8 layers below 2 km
SILAM	Finnish Meteorological Institute (FMI), Finland	0.1°	69 layers for meteorological pre-processor, 9 layers for chemistry and vertical sub-gridscale

**Table 2:** Key characteristics, including model name, developing institute, horizontal resolution and construction method of the ensemble forecast Meteo-France (2017).

Model name	Developing institute	Horizontal resolution	Construction method
Ensemble	Copernicus Atmosphere Monitoring Service (CAMS), European Union	0.1°	Multi model median

CAMS regional ensemble forecasts are published once a day and each forecast covers 96 hours (4 days). Several pollutants and vertical levels are available as forecast output. CAMS regional ensemble forecast is representative of background air due to the size of the model grid (Marècal *et al.*, 2015). Background air is not sensitive to rapid local variations in air quality, as in for example street canyons (Sabelström, 2017b). Hence, the forecast does not resolve variations in concentration due to local emissions. However, the forecasts have functions simulation rush hour traffic in order to partly account for local emissions (Robertsson, 2018).

Chemistry schemes and aerosol parameterizations included in all individual models are the following: gas phase chemistry, heterogeneous chemistry, aerosol size distribution, inorganic aerosols, aqueous phase chemistry, dry deposition / sedimentation, mineral dust, sea salt, wet deposition. In addition is secondary organic aerosols included in all but the models MOCAGE and LOTOS-EUROS (Meteo-France, 2017). This means that the individual models describe the same atmospheric features, even though schemes are not constructed identically. Each model have the same underlying parameterization of birch pollen emissions but output of birch pollen still differs as a result of the various differences in the model constructions (Sofiev *et al.*, 2015). The pollen module does not include quasi-biannual cycles of birch pollen (Sofiev *et al.*, 2015). The quasi-biannual cycles are important in order to capture pollen concentrations. Sofiev *et al.* (2015) found that the lack of this process contributes to errors in the birch pollen forecast. The forecast pollen module is controlled by temperature and blooming will start after a certain accumulated heat threshold is reached Sofiev *et al.* (2015).

This study focuses on evaluating and validating the CAMS regional ensemble forecast in the Stockholm region to examine the accuracy of the forecast and which factors affect the accuracy. A comparison between urban and regional background air is made in the study in order to get a sense of how the forecast captures local and regional variations of pollutants. The aim of this study is to evaluate the accuracy of CAMS regional ensemble forecast in the the Stockholm region and how the accuracy change with time. The aim is also to try to determine which factors affect the forecast accuracy and if there are any meteorological situations particularly difficult for air quality forecasting. Lastly, forecast accuracy in relation to health risks are considered in order to get a sense of how useful the CAMS regional ensemble forecast would be, for the general public information regarding the air quality situation in the Stockholm region.

## 1.2 Pollutants and birch pollen

Pollutants considered in this study are: ozone ( $O_3$ ), nitrogen dioxide ( $NO_2$ ), particulate matter with diameter smaller than 10 and 2.5  $\mu m$  respectively ( $PM_{10}$  &  $PM_{2.5}$ ). Each pollutant have harmful effects on human health as well as plant matter and animals (Latvala, 2018a; WHO, 2018a). Birch pollen is also included in this study since it is a strong allergen during blooming season in the spring time. It is a cause of increased emergency visits as well as cause of people becoming more sensitive to exposure to other pollutants during the allergy season (Johansson *et al.*, 2013).

Residence time, defined as the average time a certain constituent resides in the atmosphere before being deposited or chemically transformed, varies between the different pollutants (Wallace & Hobbs, 2006). In Sweden, there is a north south gradient in the concentrations of air pollution from long-range transport by advection of air masses, since the concentration decrease with distance to the source depending on the residence time of each pollutant (Sabelström, 2017a). Thus, southern Sweden is more affected by air pollution from the european continent compare to northern Sweden (Sabelström, 2017a). The city of Stockholm is affected both by locally emitted and long-range transported pollutants.

### 1.2.1 Nitrogen dioxide

Antropogenic emissions of nitrogen oxides ( $\text{NO}_x$ ) are mainly through fossil fuel and biomass combustion in power generators and from industries and traffic (WHO, 2018b).  $\text{NO}_x$  is the sum of nitrogen oxide (NO) and nitrogen dioxide ( $\text{NO}_2$ ) in the atmosphere. Emissions of  $\text{NO}_x$  is mainly NO but  $\text{NO}_2$  is rapidly formed in the atmosphere when NO is oxidized by hydroperoxyl radicals ( $\text{HO}_2$ ) (SMHI, 2014a).  $\text{NO}_2$  can worsen bronchitis and asthma, as well as cause respiratory infections and reduced lung function and inhibit growth (WHO, 2018b).  $\text{NO}_2$  is an ozone precursor, thus crucial for tropospheric ozone production (WHO, 2018b).  $\text{NO}_2$  has a residence time of nearly one day in the atmosphere (Wallace & Hobbs, 2006). This means that  $\text{NO}_2$  concentrations in Sweden mostly depend on local emissions since  $\text{NO}_2$  emitted far from Sweden will be chemically consumed before reaching the region. A major sink of  $\text{NO}_2$  is through reacting with the hydroxyl radical (OH),

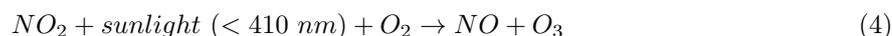


where  $\text{HNO}_3$  is nitric acid (SMHI, 2014a). Nitric acid is removed from the atmosphere through wet deposition.  $\text{NO}_2$  concentrations are highest during winter when fuel consumption is largest. Diurnal cycle in cities peak during rush hour traffic in the morning and afternoon (SMHI, 2014a).

### 1.2.2 Ozone

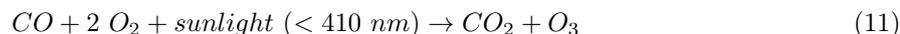
Ozone is a main component in photochemical smog (WHO, 2018b). Negative health effects from ozone include problems with, asthma, reduced lung function and respiratory diseases (WHO, 2018b). In addition to being a pollutant harmful to humans and plants, ozone is also a strong greenhouse gas contributing to global climate change (Pleijel, 1999).

Ozone is a secondary pollutant, meaning it is not directly emitted. Instead, it is chemically produced in the atmosphere when methane ( $\text{CH}_4$ ), carbon monoxide (CO) or other volatile organic compounds (VOCs) are oxidized by photochemical reactions involving  $\text{NO}_x$  (WHO, 2018b). Oxidation of  $\text{NO}_x$  by ozone creates a cycle with no net production of ozone. Key reaction in the oxidation cycle are



where NO is nitrogen oxide, O is atomic oxygen and  $\text{O}_2$  is oxygen gas. The presence of OH and VOCs such as CO will slightly alter this cycle as they result in substances which NO is more inclined to react with than ozone. This will result in a net production of ozone as long as  $\text{NO}_x$  concentration remains below 100 ppt(v) (Wallace & Hobbs, 2006). Key reactions forming ozone from oxidation of OH by CO are





where  $CO_2$  is carbon dioxide, H is a free hydrogen atom and M represents a non-reacting molecule in the troposphere e.g. nitrogen (Wallace & Hobbs, 2006). When  $NO_x$  concentrations exceed 100 ppt(v) reaction in equation (1) will be triggered, terminating the ozone production (Wallace & Hobbs, 2006). In city centers where  $NO_x$  concentrations are generally higher, ozone destruction as a result of equation (1) dominates (SMHI, 2014b). However, sunlight is the dominating factor in determining if ozone is produced. Tropospheric ozone will not be produced if there is no sunlight, even if all ozone precursors are available. The residence time of ozone is approximately one month in the atmosphere and can therefore be mixed in the boundary layer on an intra-hemispheric distance scale (Wallace & Hobbs, 2006). This means that ozone produced over the European continent could be advected in over Sweden.

Ozone levels tend to be highest during summer time, when sunlight is most abundant (Pleijel, 1999). Diurnal cycle of ozone tends to peak some hours after midday. Mainly because photochemical reactions are most effective during midday when the sun is most intense (SMHI, 2014b). Furthermore, ozone concentrations tend to be higher some distance away from city centers since ozone production is not instant and since destruction by high concentrations of  $NO_x$  from traffic is strong in city centers (SMHI, 2014b). Ozone production can become elevated during sunny days with stable stratification and low wind speeds, since the ozone precursors are not diluted in the atmosphere then (Pleijel, 1999).

Ground uptake by plants is a major ozone sink (Pleijel, 1999). The efficiency of the ground uptake depends on the amount of plant matter but also on meteorological conditions such as stability and wind speed. Stable stratification and low wind speeds lead to more effective ground uptake, since dilution effects are decreased during such conditions resulting in ozone accumulation near the surface. This situation often occurs during summer nights. Since ozone is not produced during the night time due to lack of sunlight, concentrations of ozone tend to decrease during nighttime (Pleijel, 1999). Ground uptake can on the other hand be reduced when boundary layer mixing is strong, for example during sunny summer days (Johansson, 1993).

### 1.2.3 Particulate matter

Health issues associated with inhalation of PM depend on the chemical composition of the particle (Latvala, 2018b).  $PM_{2.5}$  can, unlike the coarse particles, penetrate lungs and enter the blood stream, potentially resulting in more negative health effects compared to the coarse particles in which cannot penetrate lungs. However, coarse and fine particles can both cause heart and lung disease as well as increase symptoms of asthma (Latvala, 2018b).

Particulate matter (PM) can consist of a number of substances, such as sulphate, nitrates, ammonia, mineral dust, black carbon and sodium chloride to name a few (WHO, 2018b).  $PM_{10}$  has a diameter of up to 10  $\mu\text{m}$  and  $PM_{2.5}$  up to 2.5  $\mu\text{m}$ . PM have different sources, both anthropogenic such as combustion and tire and road wear and natural such as sand particles or sea spray (Latvala, 2018b). Coarse particles can be separated from  $PM_{10}$  by subtracting with  $PM_{2.5}$  ( $PM_{10} - PM_{2.5}$ ). What remains are particles with diameter between 2.5 and 10  $\mu\text{m}$ . These particles are normally few in number but constitute a large fraction of total mass in  $PM_{10}$  (SMHI, 2014c). Resuspended road dust and tire wear often fall under this category (Latvala, 2018b; Areskoug *et al.*, 2004).

Concentrations of  $PM_{10}$  are highest in cities (Latvala, 2018b). In northern cities, concentrations peak during the spring when studded tires wear on dry roads (Latvala, 2018b).  $PM_{2.5}$  concentrations largely depend on long-range transport from the European continent, thus concentrations of  $PM_{2.5}$  tend to be larger in southern compared to northern Sweden (Latvala, 2018b).  $PM_{2.5}$  concentrations tend to be slightly elevated during winter since fuel combustion in traffic and from residential heating is increased during the cold season. Coarse particles have a residence time of approximately 10 hours (Wallace & Hobbs, 2006). This means that such particles will be mixed locally in the atmosphere.  $PM_{2.5}$  has a

residence time of approximately 4 – 7 days (Wallace & Hobbs, 2006). Thus, such particles can be mixed on an intra-hemispherical distance scale. Wet and dry deposition are the major sinks of PM (Wallace & Hobbs, 2006). Concentrations of PM tend to be lowest during summer months. One reason is the fact that it precipitates most during summer (Stockholm stad, 2018). Other contributing factors are that emissions are lower during the summer and that turbulent mixing is more efficient which leads to higher dilution (Wallace & Hobbs, 2006). Coarse particles emitted in the European continent will generally be deposited before reaching Stockholm. Emissions of PM<sub>2.5</sub>, on the other hand can remain in air masses originating from the European continent.

#### 1.2.4 Birch pollen

Pollen is not an air pollutant as such. However, it is a strong allergen which gives rise to health issues during blooming season. Allergies can be worsened from co-exposure to polluted air (Sofiev *et al.*, 2015). The source of birch pollen is, naturally, blooming birch trees. Meteorological conditions, mostly temperature, control when blooming starts. Birch pollen has a diameter of approximately 20 – 25  $\mu\text{m}$  (Ekeholm, 2018). Birch pollen have a residence time in the order of hours. Thus, local sources are most important for concentrations of pollen. However, remote sources can also contribute if meteorological conditions are permitting. Blooming normally occurs earlier in southern Europe which means that birch pollen events from long-range transport can occur before the local blooming season actually starts (Sofiev *et al.*, 2015). Birch trees have a quasi-biannual blooming cycle with alternating levels of high and low pollen concentrations approximately every other bloom (Sofiev *et al.*, 2017). Pollen is commonly separated into concentration levels low, medium, high and very high (Naturhistoriska riksmuseet, 2016). With limit values depending on the source. For birch the limits are the following: low 1 – 10 grains/m<sup>3</sup>, medium 11 – 100 grains/m<sup>3</sup>, high 101 – 1000 grains/m<sup>3</sup>, very high > 1000 grains/m<sup>3</sup> (Naturhistoriska riksmuseet, 2016).

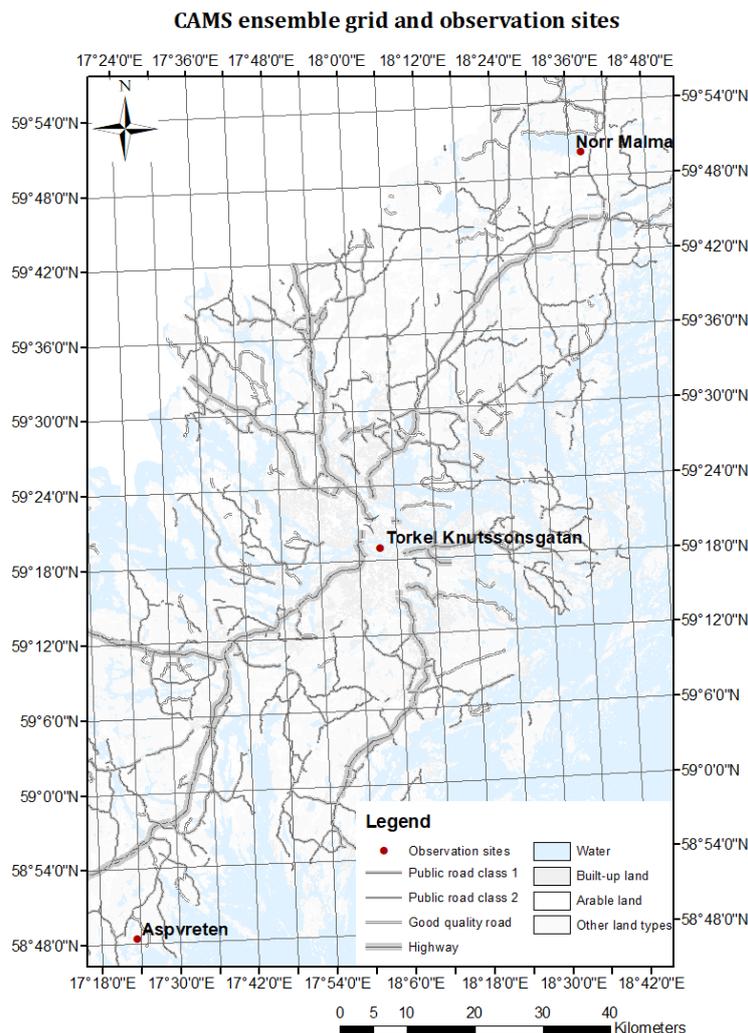
## 2 Method

### 2.1 Urban and regional observation sites

Observations of air pollutants from three sites in the Stockholm region were considered in this study. The three sites are representative of background air as to correspond to the forecasts. Background air is commonly separated into regional and urban, where the regional background air can be found in areas far from emission sources, typically rural areas (Sabelström, 2017b). Urban background air is representative of the concentration of air pollution which inhabitants in a city are generally exposed to. Concentrations in background air is representative of a larger area and does, as oppose to street canyons, not vary rapidly due to local emissions (Sabelström, 2017b). Concentrations of air pollution are generally higher at urban background sites due to proximity of local emission sources. Two regional background observation sites and one urban background observation site was included in the study. Urban background observations came from a roof top station at Torkel Knutssonsgatan (59.316N, 18.057E) and regional background observations from stations Norr Malma (59.832N, 18.631E) and Aspvreten (58.806N, 17.389E). Data of all pollutants were available at the Torkel Knutssonsgatan and at Norr Malma sites. Data of ozone and PM<sub>10</sub> were available at the Aspvreten site. Observations of birch pollen from palynological laboratory at Swedish Natural history museum in Stockholm from an measurement site at roof top level at Stockholm University approximately 6 km north of the site at Torkel Knutssonsgatan.

The site at Torkel Knutssonsgatan at roof top level, is surrounded by an urban environment with dense traffic, street canyons and buildings. The site at Norr Malma is situated approximately 70 km northeast of Stockholm city center, is in a crop field surrounded by forested area and a lake. The site at Aspvreten is no longer active but when it was, it was situated in a forested area approximately 90 km southwest of Stockholm city center. The site was approximately 2 km from the Baltic sea. An overview of the location of the measurement sites and how they are situated in relation to the forecast grid is shown in Figure 1. The urban background site was selected order to try to evaluate how the air quality forecast

perform in a region where local emissions affect the background air. Two regional background sites were included in the study to examine how much land cover affects the quality of the forecast.



**Figure 1:** Schematic map showing the location of all three observation sites in relation to the forecast grid. Each box on the map represents a grid box in the forecast.

The sites at Torkel Knutssongatan and Norr Malma are run by the air quality department at Stockholm municipality (SLB-analys). Stockholm municipality is responsible for monitoring the city's air quality as well as quality control and verification of the observations (Norman & Lövenheim, 2016). Instruments used in at the sites are for  $\text{NO}_2$  an Environment AC32M, for  $\text{O}_3$  an Environment S.A O342M, for  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  a TEOM-1400ab (Tapered Element Oscillating Microbalance) (Norman & Lövenheim, 2016). All instruments are approved by the Swedish Environment Protection Agency (EPA) (Norman & Lövenheim, 2016). The site at Aspvreten was run by ACES. Instruments used were API 400E for ozone and TEOM 1400 - 8500 FDMS for  $\text{PM}_{10}$  (Areskoug, 2018). Pollen were collected by a Burkard Seven Day Volumetric Spore Trap and each sample has been analyzed by a palynologist (Naturhistoriska riksmuseet, 2017).

Environment AC32M measures both  $\text{NO}_2$  and  $\text{NO}_x$  ( $\text{NO} + \text{NO}_2$ ). A constant flow of air is taken in by the instrument which is then mixed with air saturated by ozone. The rapid reaction between  $\text{NO}$  and ozone (equation (5)) utilized by the instrument and a photomultiplier translates intensity of light, sent out by excited  $\text{NO}_2$  molecules created in the reaction, into a voltage. The voltage is directly proportional to the  $\text{NO}$  concentration in the air sample.  $\text{NO}_x$  is measured in an air sample by reducing  $\text{NO}_2$  into  $\text{NO}$  in a heated catalytic converter. The concentration of  $\text{NO}_2$  is determined by the difference in the two samples. Environment S.A O342M determines ozone concentrations by continuously measuring ozone absorption

of ultra violet light. Ozone has an absorption maximum of 245 nm. In a TEOM-1400ab particles are gathered on an oscillating body of glass. The frequency of the glass body changes in proportion to a change in mass on the particle filter. The frequency change is converted into particle concentration. The instrument has two separate air intakes and can thereby measure both  $PM_{10}$  and  $PM_{2.5}$ . (Norman & Lövenheim, 2016). API 400E uses a technique similar to that of Environment S.A O342M and TEOM 1400 – 8500 FDMS a technique similar to that of TEOM-1400ab (Areskoug, 2018). The Burkard Seven Day Volumetric Spore Trap takes in a certain amount of air per time unit. The air then hits a slowly rotating panel covered in tape where particles stick. Rotation speed of the panel is 2 mm/ hour. The pollen trap is emptied every morning during pollen season and the particle content on the tape is analyzed with microscope in order to determine the number of grains. A concentration in grains/ $m^3$  is calculated based on the analysis. This is the standard method of measuring pollen in Europe. (Naturhistoriska riksmuseet, 2017)

## 2.2 The data and work process

The period October 2015 – December 2017 was included in this study. The CAMS data was not available further back in time than October 2015. No model runs or measurements were made in this study. Instead data from ongoing measurements and historical model data was used. 24-hour periods of the forecast was compared to observations. Thus, every 24 hours in the time series of forecast data was a new forecast. Whereas the observations were continuous measurements. The study focused on hours 0 – 23 and hours 73 – 96 of the forecast. The first time period was selected since assumed to be the most accurate in the forecast and thus the most relevant period to evaluate. The second period was included in order to examine how much the accuracy declines over forecasted time. The 48 hours between the first and last day of each forecast were not included in the study. A comparison of only the first and last 24 hours of each forecast was deemed sufficient as to obtain an overview of how forecast accuracy change over time. A more detailed analysis of this could be done with all 96 forecast hours but that was not contained with in the scope of this study. Forecasts of surface concentrations of the pollutants were used. No evaluation of or comparison to any of the individual models was done in this study due to time constraints. Only CAMS regional ensemble forecast was evaluated and validated from comparison with observations.

Ensemble forecast data was retrieved as Application Program Interfaces (APIs) from CAMS website. The APIs contained forecast data for the entire EU domain ( $20^{\circ}W/30^{\circ}N/45^{\circ}E/70^{\circ}N$ ) and in order to only retrieve the data matching the observation points Climate Data Operators (CDOs) were used. Observation data from Torkel Knutssongatan was obtained from Stockholm municipality's data base and data from Aspveten from SMHIs data portal. Both forecasted and observed data consisted of time series of the pollutants with hourly resolution. Observations were in Swedish time zone with daylight savings during winter. However, all data was harmonized into Universal Time Coordinated (UTC) time zone in order to match the forecast data. For each pollutant, only hours with both forecasted and observed concentrations were used. Hours with negative values of  $PM_{10}$  and  $PM_{25}$  observations were excluded. According to the Forum for Air quality Modeling (FAIRMODE), data coverage should be 75 % or higher for a considered time period. This criterion was met for all pollutants and all sites. Data processing was preformed mostly in Python (2018) (version 3.6) but some statistical analysis was done using R (2017) (version 3.5.2) in the editor RStudio (2015) (version 1.1.463) together with module (OpenAir, 2012) (version 2.6.0). An examination of each data set's distribution was done in R (2017) by constructing Quantile to Quantile plots (QQ-plots). A QQ-plot divides the data in quantiles and compares them to quantiles of a theoretical normal distribution. In order to get a sense of diurnal and seasonal variations mean and median values as well as standard deviation and 25th/75th percentiles respectively were calculated for months and hours. Mean and standard deviation was used if the data set had a normal distribution and median and 25th/75th percentiles if the data set was not. Statistical indicators were calculated for each pollutant and site for the whole time period. Parametric statistical indicators, described in section 2.3, were also calculated for the summer quarter, June, July and August (JJA) and the winter quarter December, January and February (DJF).

The ensemble forecast was evaluated at regional and urban background sites for pollutants  $O_3$ ,  $NO_2$ ,  $PM_{2.5}$  and  $PM_{10}$ - $PM_{2.5}$ .  $PM_{10}$ - $PM_{2.5}$  was considered rather than  $PM_{10}$  in order to get a sense of how the forecast accuracy differed comparing fine PM to coarse. The different time periods were considered

in order to examine seasonal and diurnal variations in the data which are not captured in only the full time period is considered. The ensemble forecast's ability to capture general features of the pollutants on annual, monthly and daily time scales was evaluated. Furthermore, each pollutant's dependency on wind speed was examined. Wind speed data was obtained from one meteorological station in Högdalen and from one at the regional background site at Norr Malma. Every data point for each forecast and observation data set was sorted based on wind speed intervals of 1.5 m/s. A mean value of concentrations for each wind interval was plotted against corresponding wind speed.

### 2.3 Statistical indicators

Statistical indicators used for evaluation and validation of the ensemble forecast when examining the air pollutants are either in accordance with those recommended by FAIRMODE or with those used in previous validations of the ensemble forecast produced by CAMS. Thus, these results are comparable with previous and future validations of CAMS forecasts. Birch pollen was examined using binary statistics. The statistical indicators were used to evaluate the first 24 hours as well as to examine how the forecast accuracy varies depending on forecast time by comparing statistical indicator scores for hours 0 – 23 by scored for hours 73 – 96. The statistical indicators are so called parametric indicators. They assume that the distribution of the data set is normal (Wilks, 2011).

Root mean square error (RMSE) is defined as,

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - M_i)^2} \quad (12)$$

where  $N$  is the number of observations,  $M_i$  is each modeled value and  $O_i$  each observed. RMSE has the same unit as the variables. It is a measure of the square root of the average square error, an indication of the spread of individual error of each moment in a data set (Marècal *et al.*, 2015). RMSE is rather sensitive to outliers in the data set, due to the squaring and cannot be used in inter-comparison of different species since RMSE is not dimensionless (Marècal *et al.*, 2015).

Correlation coefficient (R) is defined as,

$$R = \frac{\sum_{i=1}^N (M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (M_i - \bar{M})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}}, \quad (13)$$

where  $\bar{O}$  and  $\bar{M}$  are the observed and the modeled averages respectively, defined as  $\bar{O} = \frac{\sum_{i=1}^N O_i}{N}$  and  $\bar{M} = \frac{\sum_{i=1}^N M_i}{N}$ . R is the mean difference between each forecasted value and the forecast mean multiplied with the mean difference between each observed value and the observation mean normalized by the forecast and observation standard deviation. It is an indicator of the strength of the linear relationship between the forecast and the observations. R is bounded by -1 to 1 and the closer R is to 1 the better the forecast is (Marècal *et al.*, 2015; Janssen *et al.*, 2018).

Modified mean bias (MNMB) is defined as,

$$MNMB = \frac{2}{N} \sum_{i=1}^N \left( \frac{M_i - O_i}{M_i + O_i} \right). \quad (14)$$

MNMB is a measure of the mean bias of forecast and observation normalized by the mean of the forecast and the observations (CAMS, 2017). MNMB is bound by -2 to +2 and the closer MNMB is to zero, the better the forecast is (CAMS, 2017; Marècal *et al.*, 2015).

Fractional gross error (FGE) is defined as,

$$FGE = \frac{2}{N} \sum_{i=1}^N \left| \frac{M_i - O_i}{M_i + O_i} \right|. \quad (15)$$

It is a measure of the absolute value of mean bias of forecast and observation normalized by the mean of forecast and observations. FGE is bound by 0 and 2. The closer FGE is to zero the better the forecast is (Marècal *et al.*, 2015).

MNMB and FGE are dimensionless and relative, since they are normalized, which makes it possible to compare MNMB and FGE scores for different air pollutant to each other (CAMS, 2017; Marècal *et al.*, 2015).

MNMB and FGE were chosen in this study partly because they have been used in previous validations of CAMS forecasts and partly for quality that they can be compared between the different pollutants. Both MNMB and FGE are normalized by the mean of the forecast and the observations. Another possible approach could have been to normalize by the mean of the observations only, but it was not used since it has not been included in other validations of CAMS forecasts.

FAIRMODE has developed an indicator called modeling quality indicator (MQI), which depends on RMSE and the so called measurement uncertainty,  $U(O_i)$ , which is determined using standardized reference values for each pollutant (Table 3). MQI is used to determine if the model quality objective (MQO) set by FAIRMODE is achieved. MQO is defined as the minimum level of quality a model must achieve for policy use (Janssen *et al.*, 2018).  $U(O_i)$  is defined as

$$U(O_i) = k u_c(O_i) \quad (16)$$

where  $k$  is coverage factor, set to  $k = 2$  corresponding to confidence level of approximately 95 % and  $u_c(O_i)$  is the combined uncertainty for all observations (Janssen *et al.*, 2018). The combined uncertainty,  $u_c(O_i)$ , can be decomposed into two components, one proportional to the concentration,  $u_p(O_i)$ , and one non-proportional,  $u_{np}(O_i)$ , related as

$$u_c(O_i)^2 = u_p(O_i)^2 + u_{np}(O_i)^2. \quad (17)$$

As the non-proportional component by definition is independent of concentration, it can be estimated at a concentration level of choice (Janssen *et al.*, 2018). The concentration level is set to a reference value (RV) and  $u_{np}(O_i)$  can be expressed as

$$u_{np}^2(O_i) = \alpha^2 (u_r^{RV} RV)^2 \quad (18)$$

where  $\alpha$  is a fraction of uncertainty of the reference value (Janssen *et al.*, 2018). The proportional component,  $u_p(O_i)$ , can be expressed in a similar way as

$$u_p^2(O_i) = (1 - \alpha^2) (u_r^{RV} O_i)^2 \quad (19)$$

(Janssen *et al.*, 2018). The 95th percentile highest values of all uncertainty values,  $U_{95}(O_i)$ , were selected and previously defined variables get a subscript "95":

$$k u_c = k u_{95,c}^{RV} = U_{95}^{RV}. \quad (20)$$

$U_{95}(O_i)$  is calculated as

$$U_{95}(O_i) = U_{95}^{RV} \sqrt{(1 - \alpha^2) O_i^2 + \alpha^2 \cdot RV^2} \quad (21)$$

(Janssen *et al.*, 2018). Finally an expression for the root mean square of the measurement uncertainty ( $RMS_U$ ) is defined as

$$RMS_U = \sqrt{\frac{\sum_{i=1}^N (U_{95}(O_i))^2}{N}} = U_{95,r}^{RV} \sqrt{(1 - \alpha^2)(\bar{O}^2 + \sigma_o^2) + \alpha^2 \cdot RV^2} \quad (22)$$

**Table 3:** Constants used when calculating measurement uncertainty for each pollutant (Janssen *et al.*, 2018).

	$U_{95,r}^{RV}$	RV	$\alpha$
NO <sub>2</sub>	0.24	200 $\mu\text{g}/\text{m}^3$	0.20
O <sub>3</sub>	0.18	120 $\mu\text{g}/\text{m}^3$	0.79
PM <sub>10</sub>	0.28	50 $\mu\text{g}/\text{m}^3$	0.13
PM <sub>2.5</sub>	0.36	25 $\mu\text{g}/\text{m}^3$	0.30

where  $\bar{O}$  and  $\sigma_o$  are the observed mean and standard deviation respectively (Janssen *et al.*, 2018). Standard values for the variables for each pollutant  $U_{95,r}^{RV}$ , RV and  $\alpha$  are given in Table 3.

MQI is defined as

$$MQI = \frac{RMSE}{\beta RMS_U} \quad (23)$$

where  $\beta$  is a proportionally constant, arbitrarily set equal to 2 as to only allow a deviation between model and observation to be twice that of the measurement uncertainty (Janssen *et al.*, 2018). MQO is fulfilled when  $MQI \leq 1$  (Janssen *et al.*, 2018).

The statistical indicators were calculated for NO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub> and PM<sub>10</sub>. Coarse mode PM (PM<sub>10</sub>-PM<sub>2.5</sub>) and birch pollen was not evaluated using the statistical indicators, since constants used for measurement uncertainty was not available for them. PM<sub>10</sub> can be viewed as a proxy for coarse mode PM since the coarse particles, as previously mentioned, constitute a large fraction of total mass in PM<sub>10</sub> (SMHI, 2014c). Birch pollen was evaluated using binary statistics, described in section 2.4.

One limitation of the MQI:s is that reference values used in the estimation of measurement uncertainty does not differ between urban and regional background. Another is that the constants are not specific for the Stockholm region. The results could have been more accurate if reference values for the Stockholm region would have been available. However, it is not possible to know anything about the magnitude of this discrepancy without examining with reference values for the Stockholm region. The largest limitation with the MQI:s and the other statistical indicators is the fact that they assume that the data has a normal distribution. The scores of statistical indicators could be misleading if the assumption is not met (Wilks, 2011). However, there is an ongoing discussion within FAIRMODE and the European Committee for Standardization (CEN) on how to address this issue (Alpfjord Wylde, 2018). It is necessary to be aware of the limitations of statistical indicators. When statistical indicators which assume a normal distribution are used on data sets which are not, the result is that low and high outliers have to high influence on the result. This could perhaps be mitigated by excluding the very high and very low values however that could mean a data loss which would be to high. Instead it is better to use the statistical indicators as just that, indicators, of trends in the data set whilst being aware of the issue with outliers. The previously discussed statistical indicators are used in this study in order to obtain results which can be compared to previous and future model evaluation and validations on similar data sets. The forecast and the observations may not have perfect normal distributions but that have a similar distributions compared to each other. This together with that fact that normalized statistical indicators were used lead to results which are relevant although not flawless.

## 2.4 Binary statistics

The birch pollen data was separated into 5 categories (no pollen, low, medium, high, very high) according to the concentration levels defined in section 3.2.5. The statistical indicators were not applied to the birch pollen data set since it would not have been representative to analyze the 5 five categories in terms of parametric indicators. After the divide into five categories, the data sets affectively have only five possible values and examining this with parametric indicators would not result in anything meaningful. Instead, binary statistics were applied to the categories in order to get a sense of how accurate the birch pollen forecast was.

Binary statistics divide a data set into two categories: "yes", an event occurred and "no", no event

occurred (Wilks, 2011). This method slightly oversimplifies the analysis of the birch pollen data set since it does not distinguish between the different levels of pollen concentration, however it gives an overview of the performance for the pollen forecast in terms of how accurately the forecast captured the pollen blooming. Both forecasted and observed pollen data were divided into the two categories, so that all levels of pollen concentration (low, medium, high and very high) was set to be "yes" an event occurred and, dates when the concentrations was lower than 1 grains/m<sup>3</sup> was set to "no" event occurred and a binary data table was constructed according to Table 4. April – June for 2016 and 2017 are considered for the pollen data. Thus, 182 dates were considered in the binary statistics calculations if nothing else is stated.

**Table 4:** Binary statistics table with explanation of what the input of each box should be.

	Observed "yes" event	Observed "no" event	Total
Forecasted "yes" event	Hit (a)	False alarm (b)	Forecast "yes" (a + b)
Forecasted "no" event	Miss (c)	Correct non-event (d)	Forecast "no" (c + d)
Total	Observation "yes" (a + c)	Observation "no" (b + d)	Total sum (a + b + c + d = n)

Total of hits, misses, false alarms and correct non-events can be used to analyze the data using various indicators. Probability of detection (POD) or hit rate is defined as:

$$POD = \frac{a}{(a + c)}. \quad (24)$$

POD is a ratio between number of hits and total number of forecasted events (Wilks, 2011). Perfect score of POD is 1, it would indicate that all observed "yes" events was captured by the forecast.

False alarm ratio (FAR) is defined as

$$FAR = \frac{b}{(a + b)}. \quad (25)$$

FAR is the ratio between false alarms and total number of forecast "yes" events. FAR indicates how often the forecast predicted yes event when in fact no event occurred (Wilks, 2011). 0 is perfect FAR score, it would indicate that there were no false alarms in the forecast. FAR was calculated for the entire data set in order to get a sense of the total FAR but also for a modified selection when only birch index high and very high was counted as "yes" events, in order to get a sense of how large FAR was for occasions with the most severe pollen levels. These occasions would be most important to forecast accurately since they have the largest health impacts.

Fraction of alarms missing (FAM) was calculated as:

$$FAM = \frac{c}{(a + c)}. \quad (26)$$

FAM is the ratio between events missed by the forecast and total number of observed "yes" events. FAM was only calculated for the case when only high and very high birch index was counted as "yes" events in order to estimate how frequently the forecast missed the most severe pollen events. FAM indicates how often high/very high levels of birch pollen would have been missed by the forecast as a result of under-forecasting. Perfect score of FAM is zero.

Frequency bias index (FBI) is defined as,

$$\frac{(a + b)}{(a + c)}. \quad (27)$$

FBI is a ratio between total number of times an event was forecasted and times it was observed. FBI indicates if an event is under- or over-forecasted (Wilks, 2011). Perfect score of FBI is 1 which would mean that an event was forecasted the same number of times as it was observed (Wilks, 2011). FBI score lower than one indicates that the event was observed more than it was forecasted, hence the event was under-forecasted. Conversely, FBI score higher than one indicates that the event was over-forecasted (Wilks, 2011).

A limitation with the binary statistical indicators is the fact that null hits between model and observation does not count as a hit although it should. This was however manually computed by analyzing graphs. The fraction of correctly forecasted pollen level (including no pollen) as well as the fraction of days where the forecast was within one pollen level of the observation (not including no pollen) was computed.

## 2.5 Possible sources of error

There was a number of possible error sources in this study. A part from what has already been discussed in previous section regarding the MQI and statistical indicators there were also uncertainty related to the construction of the forecast as well as the observations. The uncertainty of the observations was however not evaluated in this study, since the observations are quality controlled by Stockholm municipality. The harmonization of the individual models is one possible source of error connected to the forecast construction. Each institute which have developed one of the seven models has a responsibility to post-process the model output so that specified requirements of the ensemble forecast are met. The purpose of the post-processing is to harmonize resolution and other parameters which differs between each model. This harmonization is possibly a large source of error in the ensemble forecast since accuracy of each individual forecast could decrease when adapting vertical and horizontal resolution. For example, the depth of the vertical levels varies between the models, yet the concentration of the lowest vertical level is reported to as output to base the CAMS regional ensemble forecast on. This means that a concentration supposed to represent the surface can, in fact be representative of a larger vertical layer. Moreover, the depth of the vertical layer is not of equal size in all individual models. This will reduce the accuracy of the ensemble forecast. Another factor reducing forecast accuracy is that a median does not capture extreme values. However, this is also a strength since it becomes insensitive to outliers. The accuracy of the ensemble forecast depends on the spread in the output from the individual models, a large spread results in a more uncertain ensemble forecast and vice versa.

Another potential source of error in the forecast evaluation is the fact that the comparison is made between point measurement and grid box. In this context, each grid box of the forecast covers a relatively large area (approximately 50 km<sup>2</sup>), thus the coverage inside each grid box is normally not the same through one box. The grid box containing the urban background site covers all of Stockholm city center. The grid box containing the regional background site Norr Malma cover forest and a part of a lake and a road as well as the fields at which the observation site is on. The grid box at the regional background site Aspvreten covers forested area and some roads. Almost no part of the grid box is ocean which is beneficial in this case since it means that the grid box largely contains the same land cover as is found by the observation site. There are small residential areas inside the grid boxes containing both regional sites. Difference in land coverage inside a forecast grid box and the surroundings of the observation sites is one possible source of error. For example, a large grid box could lead to more diluted air as the area where mixing can occur is large. Thus, concentrations could be underestimated by the forecast, compared to if the grid would have been smaller. However, the comparison between a grid box and point measurement is still meaningful since observations of background concentrations are, as previously mentioned, representative of a larger area than the point at which they are measured. The vertical grid is also a possible source of error since could lead to higher dilution effects compared to a point measurement. The fact that the input data of local emissions are from the year 2011 could also be a source of error in this study. Outdated emission data could lead to inaccuracy in forecasted concentrations of pollutants and birch pollen.

## 3 Results and discussion

### 3.1 Data set distributions

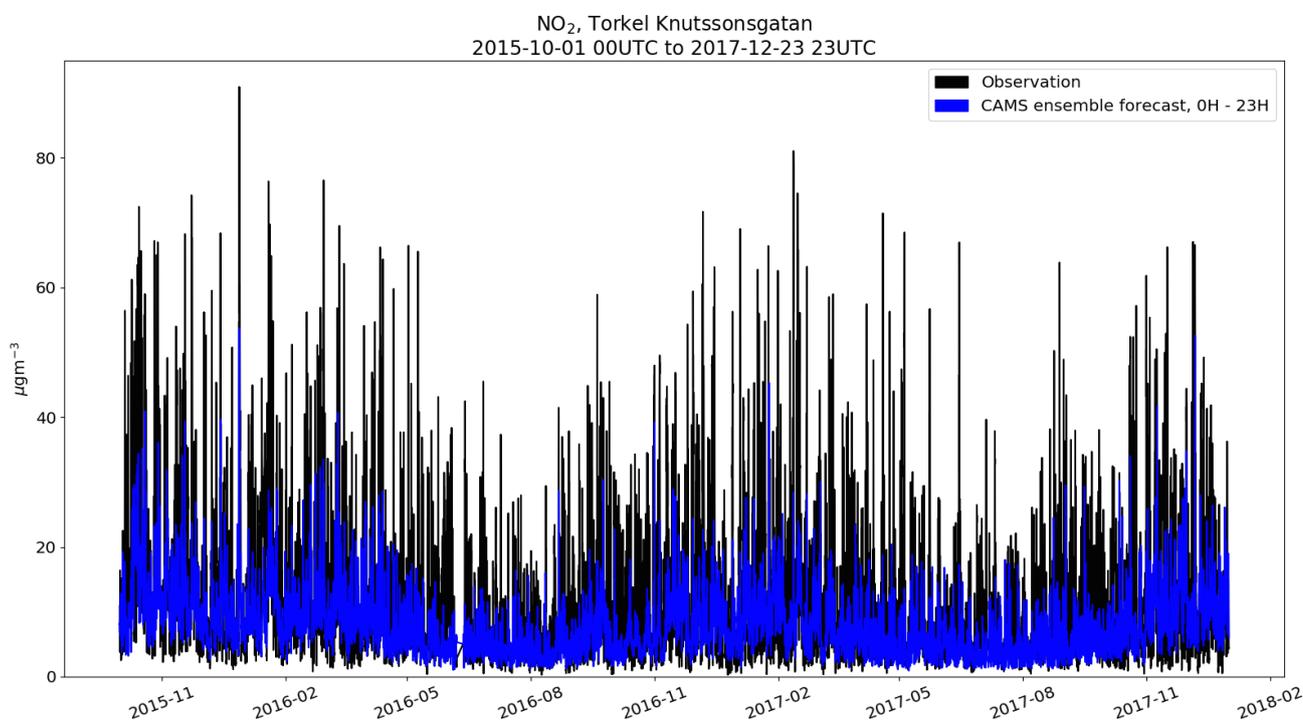
According to QQ-plots the pollutants NO<sub>2</sub>, PM<sub>10</sub> and PM<sub>2.5</sub> were not normally distributed, compared to a theoretical normal distribution. Appendix I shows two examples of the distributions which were determined not to have a normal distribution (Figures 31) and two which were assumed to have a normal

distribution (Figure 32). The deviation from normal distribution was larger for observations compared to the forecast. The distribution of  $\text{NO}_2$  and  $\text{PM}_{10}$  deviate more strongly from a normal distribution at the urban background site compared to the regional background sites, whereas  $\text{PM}_{2.5}$  had approximately the same distribution at both sites. Ozone was approximately normally distributed. There were some outliers at the high and low values but the distribution was assumed to be sufficiently near normal. Deviations were larger for observed ozone concentrations as well and there was not a large difference the deviation from normal distribution between the three sites. The resulting distributions lead to that pollutants  $\text{NO}_2$ ,  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  were examined with calculations of median and 27th/75th percentiles while ozone was examined with calculations of mean and standard deviation. Thus the parametric statistical indicators would be less representative for the behavior of the data sets of  $\text{NO}_2$ ,  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$ . The QQ plots also show that the distribution of each forecast is similar to the corresponding observation. This is a positive result as it would have been problematic if the forecast had predicted a distribution which did not match the distribution of the observations.

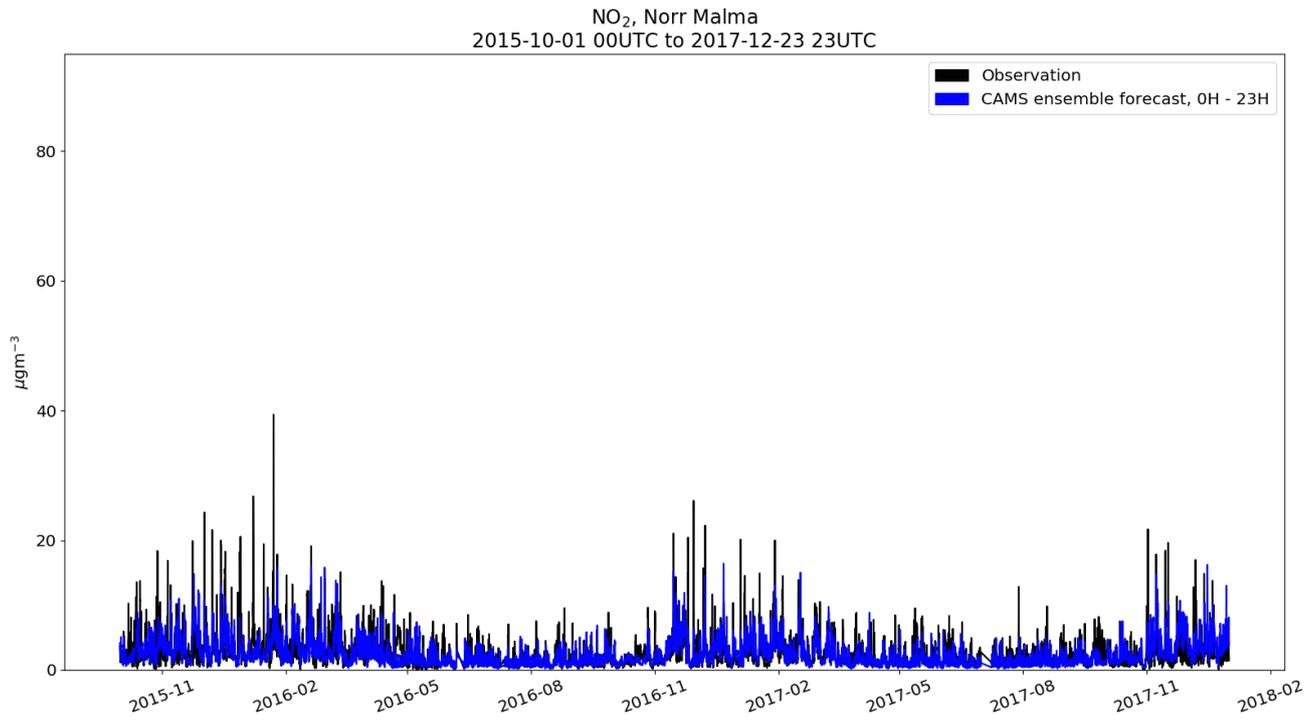
## 3.2 Evaluation of forecast performance

### 3.2.1 Nitrogen dioxide

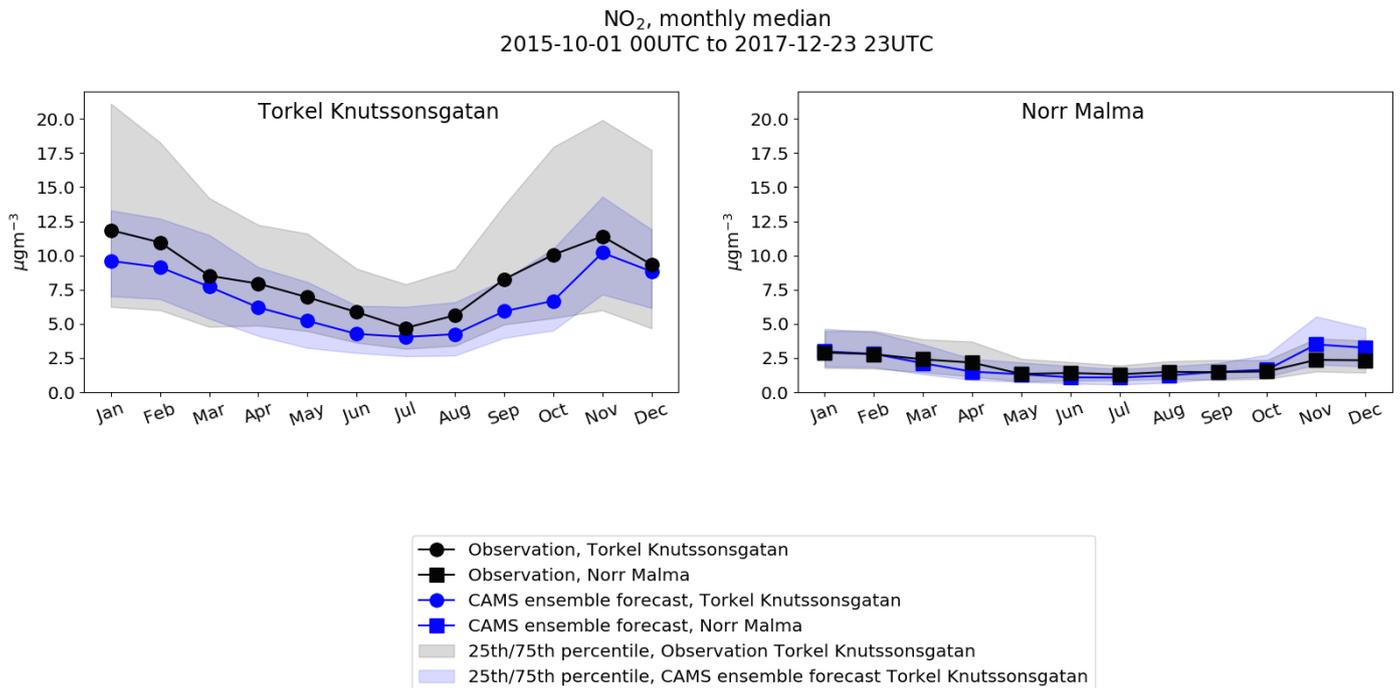
The observations for urban background site Torkel Knutssonsgatan had minimum concentrations during May to September and maximum during October to April (Figure 2). Observations at regional background site Norr Malma displayed similar seasonal variations, however in much lower concentrations (Figure 3). The ensemble forecast displayed similar general annual variations but strongly underestimated peaks of  $\text{NO}_2$  at both sites. High concentrations were particularly underestimated at the urban background site. Hence, there were large differences in the variations of  $\text{NO}_2$  between the urban and regional background sites. Monthly median concentrations of  $\text{NO}_2$  varied a lot with time of year at the urban background site whereas monthly median concentrations at the regional background sites were quite constant during a year (Figure 4).



**Figure 2:** Time series of hourly data of  $\text{NO}_2$  at the urban background site at Torkel Knutssonsgatan October 2015 – December 2017. Y-axis displays concentration of  $\text{NO}_2$  in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively.



**Figure 3:** Time series of hourly data of NO<sub>2</sub> at the regional background site at Norr Malma October 2015 – December 2017. Y-axis displays concentration of NO<sub>2</sub> in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively.

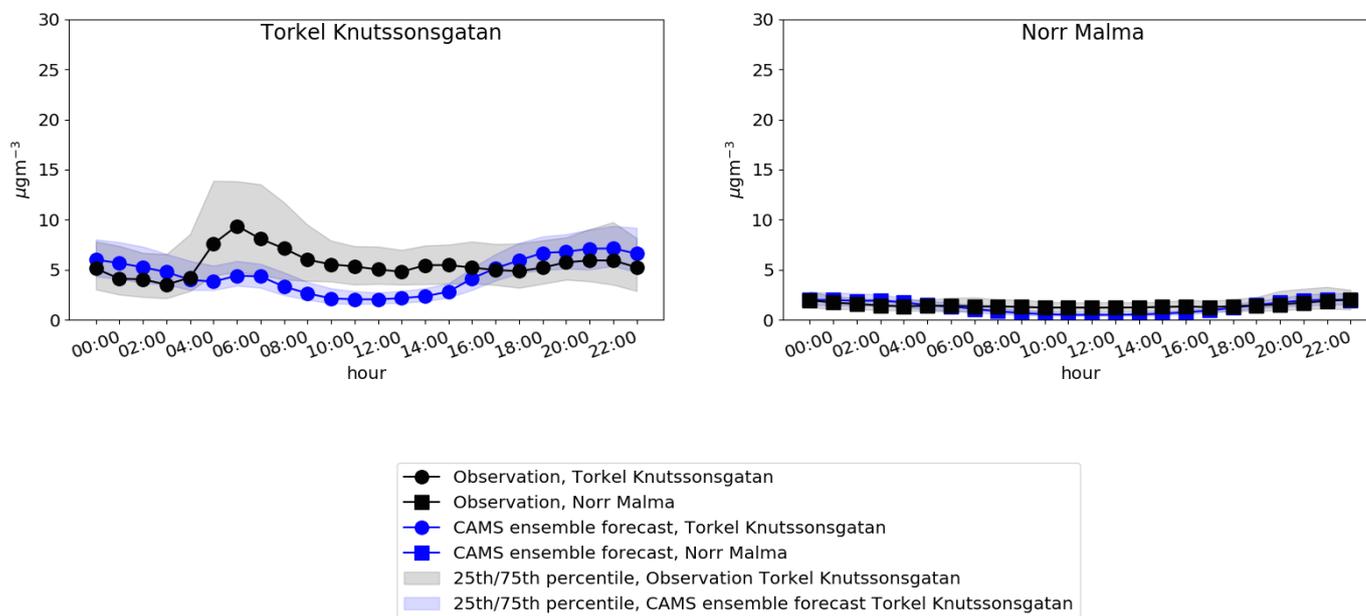


**Figure 4:** Monthly median of NO<sub>2</sub>. Y-axis displays median concentration of NO<sub>2</sub> in  $\mu\text{m}^{-3}$  and x-axis time with monthly resolution. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssongatan and squared curves the regional background site at Norr Malma.

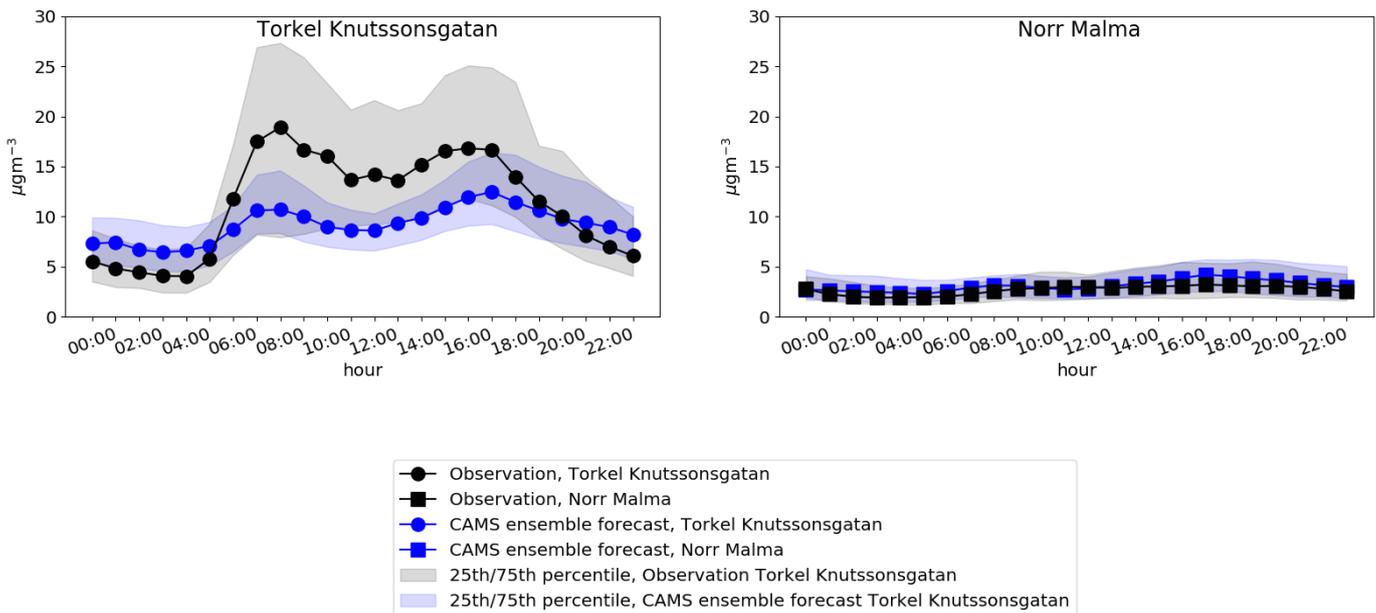
The fact that the forecast was not able to capture high peak periods at the urban background site was expected since the forecast resolution is not high enough to capture variations in concentrations due to local emissions, such as from traffic. Rush hour traffic emissions exists in the forecast models, these do however not seem to be sufficient. The relatively coarse grid together with post-processing of each individual model output and perhaps outdated input emission data lead to underestimation of events of high  $\text{NO}_2$  concentrations. Furthermore, the construction of the ensemble forecast itself lead to underestimation of  $\text{NO}_2$  concentrations. The forecast is a median of seven forecasts, which will by default be less accurate at predicting high and low extreme values.

Diurnal cycle of observations at the regional background site was quite constant during both DJF and JJA (Figures 5 & 6). The ensemble forecast captured the diurnal cycle at the regional background site. At the urban background site, there were diurnal variations in the observed  $\text{NO}_2$  median concentrations during both seasons, particularly large during DJF. During DJF there was a clear diurnal cycle with one peak during morning and one during evening and a decrease in concentrations in between. The peaks correspond to rush hour traffic and the decrease in between was could be due to dilution from turbulent mixing. Day-time median concentrations at the urban background site were underestimated by the forecast by approximately  $10 \mu\text{g m}^{-3}$  during DJF. However, the two peaks with a decrease in-between can be seen in the forecast as well. This is an indication that the forecast was able to partly capture the behavior due to local emissions, even though the magnitude was not the same. The forecast seem to capture the general behavior of diurnal variations more accurately during DJF compared to JJA. This could be due to that local emissions are more important during winter and thus show a clearer signature in the resulting forecast output. Furthermore, boundary layer height, and thereby vertical mixing, could be misrepresented in the forecast. The magnitude of the concentrations were more accurately predicted during JJA. This could be due to the fact that extreme concentrations are lower during summer than winter. Thus, reducing inaccuracy in the forecast, resulting in lower median concentrations.

$\text{NO}_2$ , hourly median JJA  
2016-2017



**Figure 5:** Hourly median of  $\text{NO}_2$ , June, July and August. Y-axis displays median concentration of  $\text{NO}_2$  in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssonsgatan and squared curves the regional background site at Norr Malma. Blue and black shades represent 25th/75th percentile of forecast and observation median.

NO<sub>2</sub>, hourly median DJF  
2015-2017

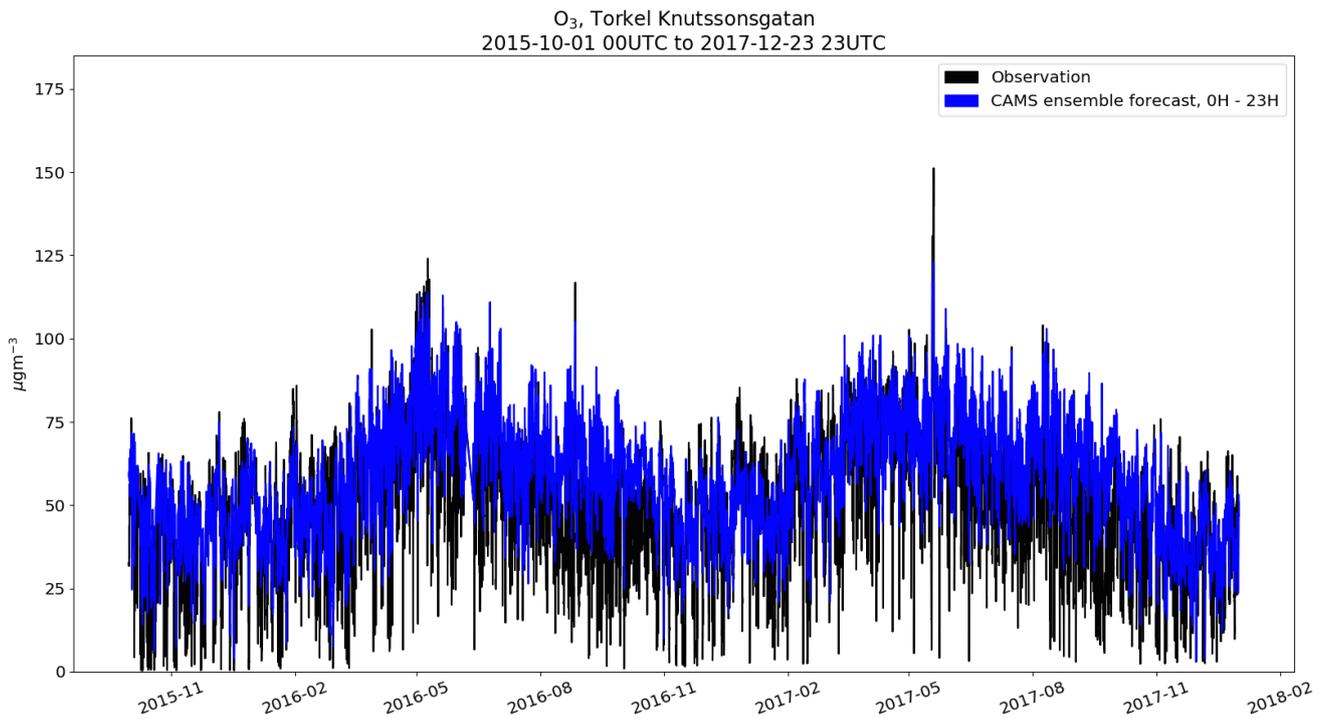
**Figure 6:** Hourly median of NO<sub>2</sub>, December, January and February. Y-axis displays median concentration of NO<sub>2</sub> in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssongatan and squared curves the regional background site at Norr Malma. Blue and black shades represent 25th/75th percentile of forecast and observation median.

The regional background site is situated far from local emission sources and thus largely depend on long-range transport of NO<sub>2</sub>. Residence time of NO<sub>2</sub> is however quite short, approximately 1 day, leading to low concentration at the regional background, since NO<sub>2</sub> from remote sources will be deposited before reaching the site. Winter season peaks at the regional background site are also due to more emissions as a result of increased traffic. However the concentrations were lower compared to at the urban background site since the air is mixed during transport.

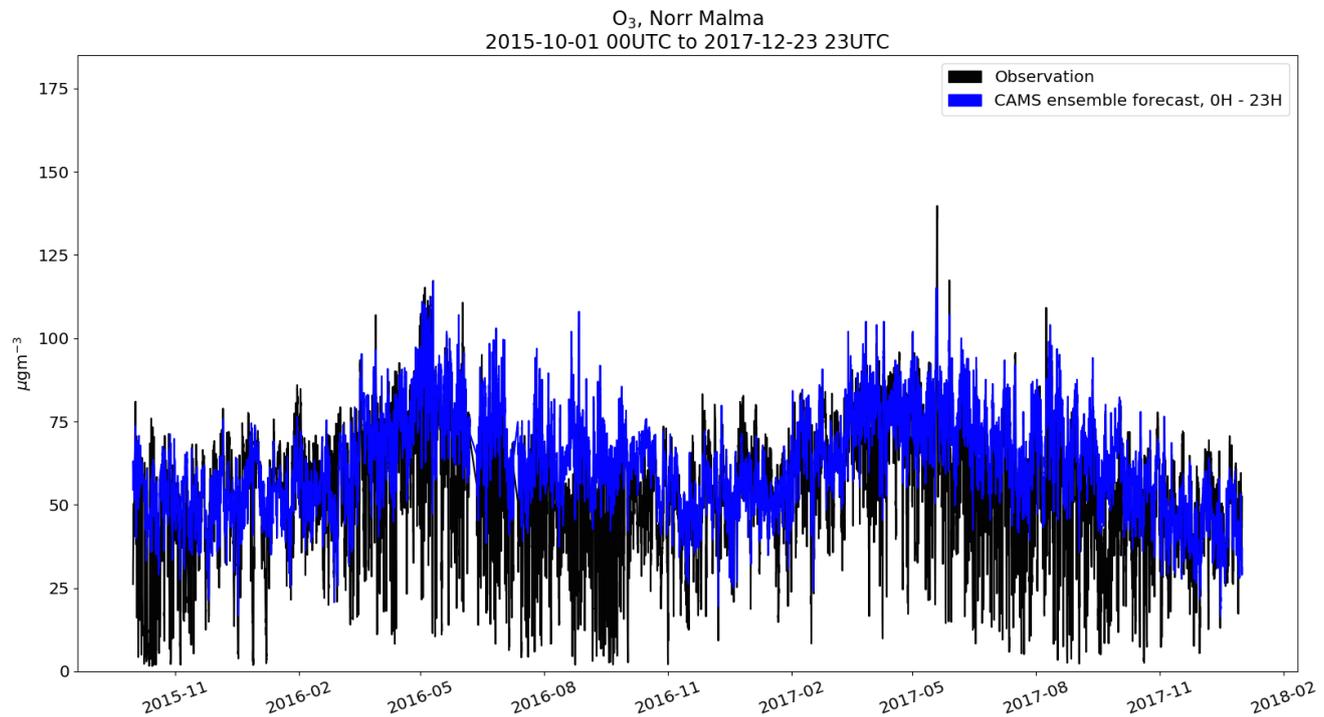
Observed and forecasted mean concentrations of NO<sub>2</sub> decreased with increasing wind speed at the urban background site (Figure 33, Appendix II). This due to that turbulent mixing increases with wind speed. The difference in mean NO<sub>2</sub> concentrations between forecast and observation decreased with increasing wind speed (Figure 33, Appendix II). The points corresponding to the maximum wind speeds were not reliable since they contain very few data points. Mean concentrations of NO<sub>2</sub> at the regional background site decrease slightly with wind speed. The difference between the forecast and the observations was very small. This result agreed with a small impact from local emission of NO<sub>2</sub>.

### 3.2.2 Ozone

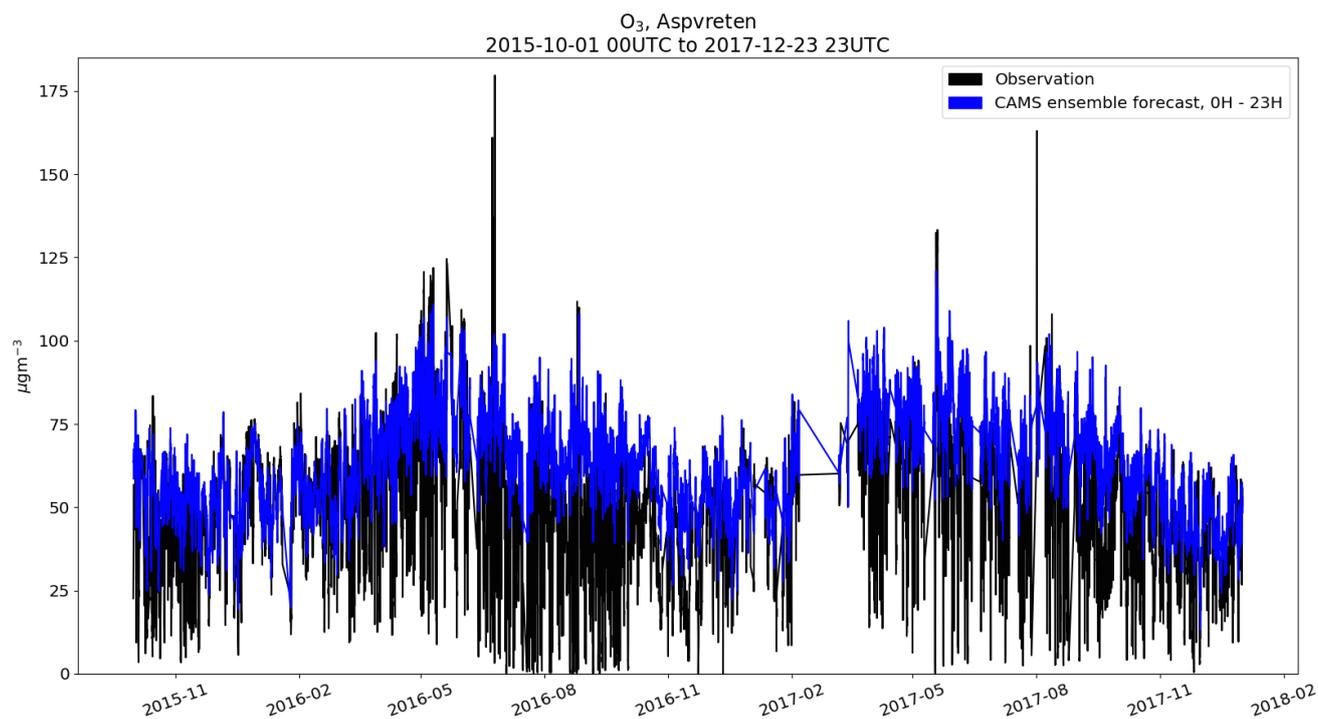
The annual cycle of ozone with maximum during summer months and minimum during winter was seen in the observations at all three sites (Figures 7, 8 and 9). The ensemble forecast captured this general appearance. Variations in observed concentrations of ozone did not differ much between urban and regional background sites, as oppose to variations in concentrations of NO<sub>2</sub> where the median variation was approximately 5 times and 2 times higher at urban background site than at the regional background site during winter and summer respectively. The small variation between urban and regional background sites was generally captured by the forecast of ozone. Observed monthly mean was, out of the three sites, lowest at the urban background site from October to Mars and at the regional background sites from May to August (Figure 10). Monthly mean for the ensemble forecast was lowest at the urban background site all months and the difference between the ensemble forecast and the observations was smallest during the winter months (Figure 10).



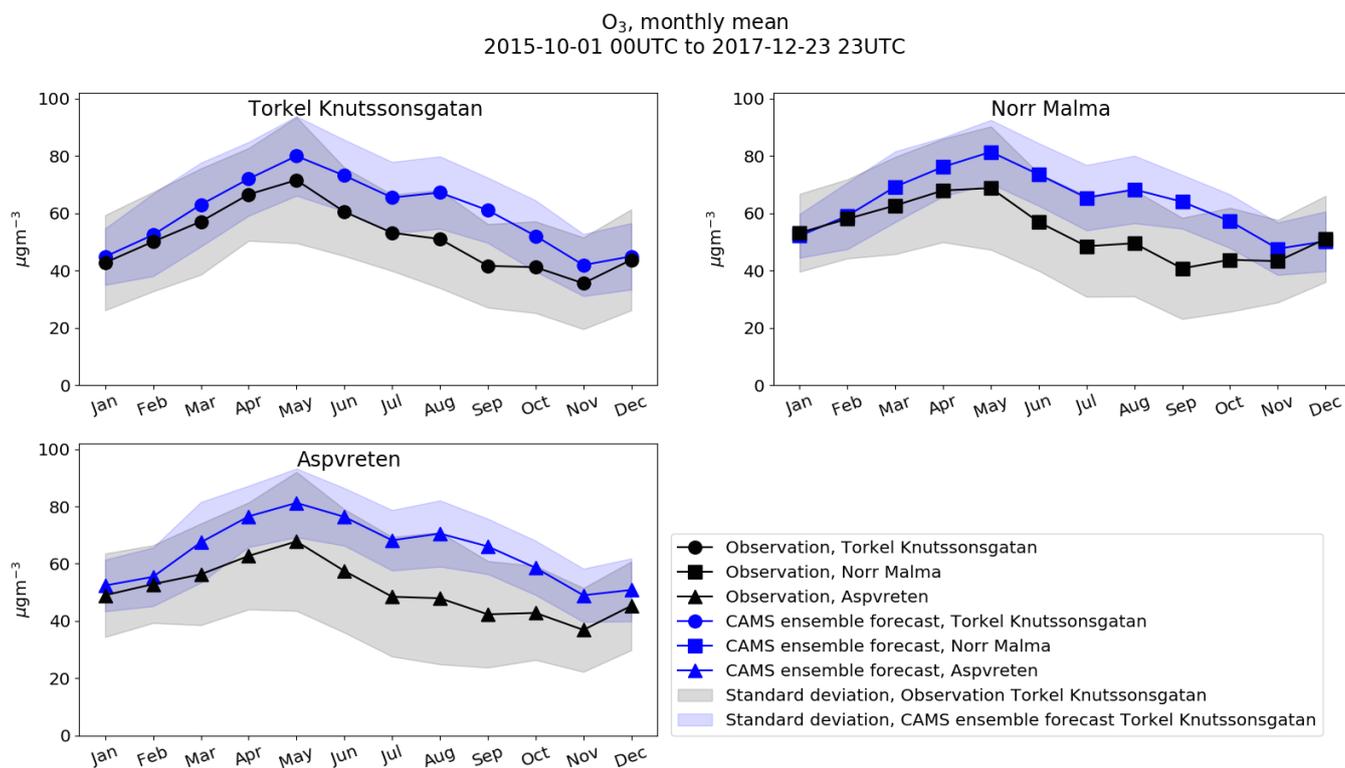
**Figure 7:** Time series of hourly ozone data ozone at the urban background site at Torkel Knutssonsgatan October 2015 – December 2017. Showing O<sub>3</sub> concentration in  $\mu\text{m}^{-3}$  on y-axis and time with hourly resolution on x-axis. Blue and black curves represent forecast and observations respectively.



**Figure 8:** Time series of hourly ozone data at the regional background site at Norr Malma October 2015 – December 2017. Showing O<sub>3</sub> concentration in  $\mu\text{m}^{-3}$  on y-axis and time with hourly resolution on x-axis. Blue and black curves represent forecast and observations respectively.



**Figure 9:** Time series of hourly ozone data at the regional background site at Aspvreten October 2015 – December 2017. Showing O<sub>3</sub> concentration in  $\mu\text{m}^{-3}$  on y-axis and time with hourly resolution on x-axis. Blue and black curves represent forecast and observations respectively.



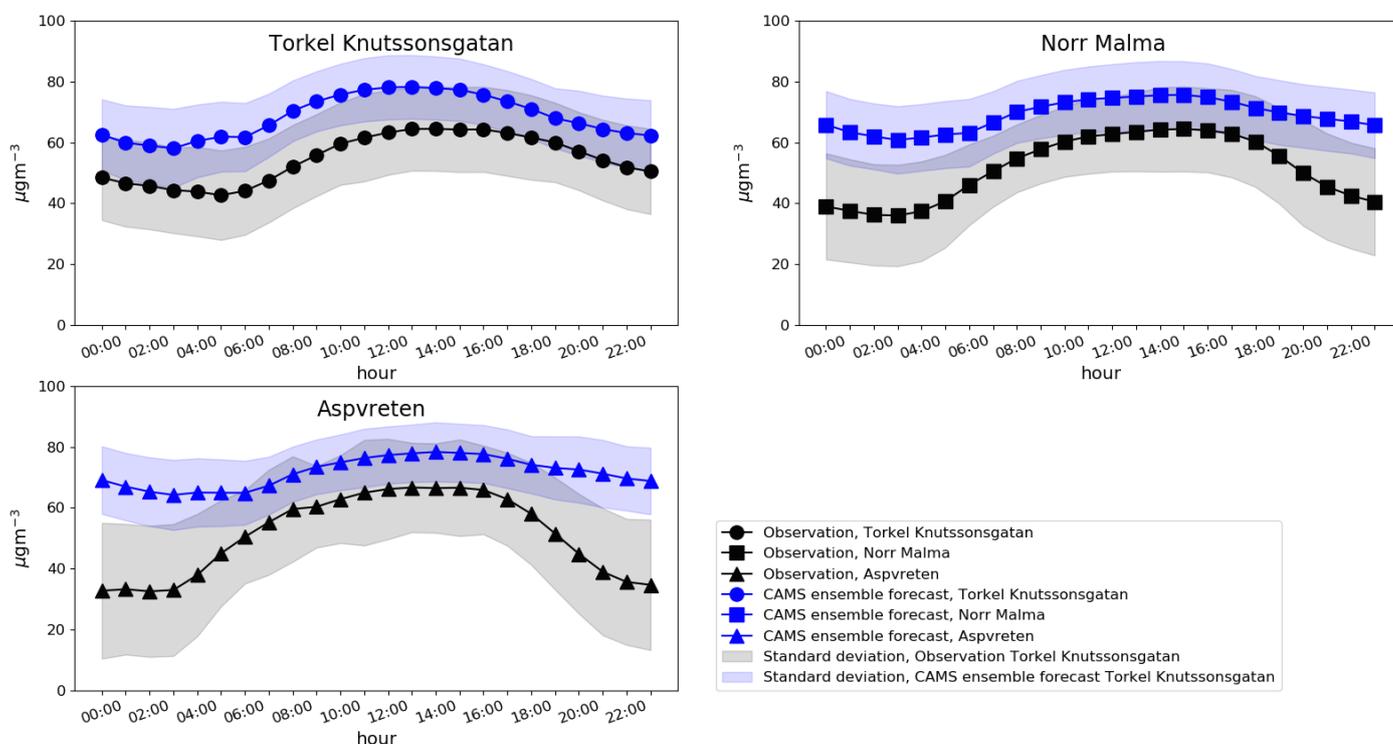
**Figure 10:** Monthly mean of ozone. Y-axis displays mean concentration of O<sub>3</sub> in  $\mu\text{m}^{-3}$  and x-axis shows months. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssongatan, squared curves the regional background site at Norr Malma and curves with triangles the site at Aspvreten.

The ensemble forecast did not capture low concentrations of ozone. The forecast was particularly inaccurate at the regional background sites. High ozone concentrations on the other hand, were captured with

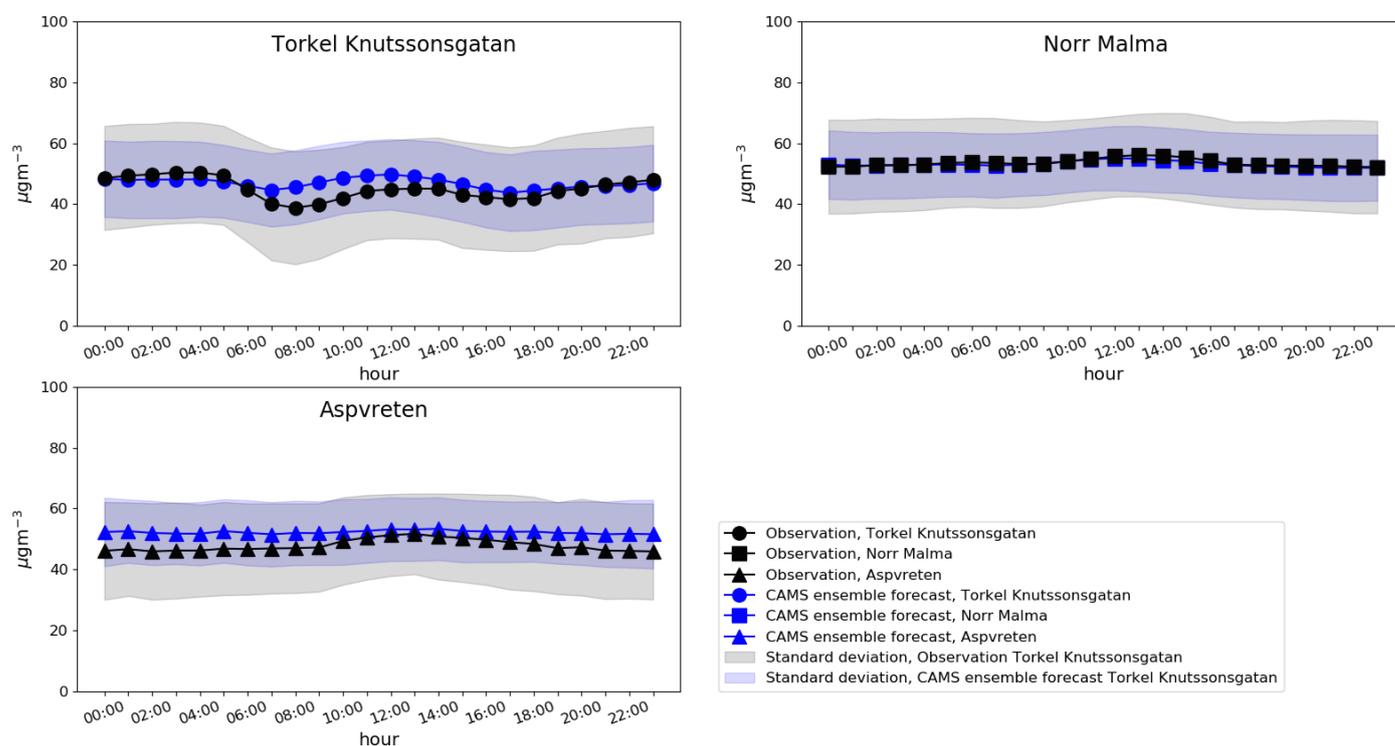
quite high accuracy by the ensemble forecast. The time of extreme peaks were often captured, although the concentration was underestimated by approximately  $25 - 50 \mu\text{gm}^{-3}$ .

Diurnal cycles of  $\text{O}_3$ , with minima during night time and maxima in early afternoon, were visible at all three observation sites during JJA (Figure 11). The difference was largest at the regional background sites. Furthermore, difference between ensemble forecast and observation was largest during the night time. According to the ensemble forecast, night-time mean concentrations of ozone were lowest at the urban background site, the opposite of what was seen in the observations (Figure 11). Diurnal variations were dampened in the ensemble forecast and mean concentrations were underestimated by the forecast all hours at all sites. Diurnal variations of ozone were small during DJF, particularly at the regional background sites where mean concentrations were more or less constant over the course of a day (Figure 12). The ensemble forecast was very accurate at both regional sites, however the accuracy was somewhat lower at Aspvreten where the forecast overestimates the mean ozone concentration slightly. There were two daily minima at the urban background site during hours corresponding to rush hour traffic. This feature was seen in the ensemble forecast but the magnitude of the ozone decrease was underestimated with approximately  $10 \mu\text{gm}^{-3}$ .

$\text{O}_3$ , hourly mean JJA  
2015-2017



**Figure 11:** Hourly mean of ozone, June, July and August. Y-axis displays mean concentration of  $\text{O}_3$  in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssonsgatan, squared curves the regional background site at Norr Malma and curved with triangles the regional background site at Aspvreten. Blue and black shades represent standard deviation of forecast and observation median respectively.

O<sub>3</sub>, hourly mean DJF  
2015-2017

**Figure 12:** Hourly mean of ozone, December, January and February. Y-axis displays mean concentration of O<sub>3</sub> in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssonsgatan, squared curves the regional background site at Norr Malma and curved with triangles the regional background site at Aspvreten. Blue and black shades represent standard deviation of forecast and observation median respectively.

The results indicate that meteorological forecasts as well as emissions fed into the forecast lead to an ensemble forecast which captured large scale annual features of ozone. Limitations in the forecast were however evident on a more detailed scale. During May to August the observed monthly mean was, out the three sites, lowest at the regional background sites. This is likely a result of the combination of high ozone destruction at the regional background sites due to strong dry deposition by plant matter uptake and reduced termination, by equation (1), of ozone production at the urban site due to lower levels of NO<sub>x</sub> during summer months. The monthly ensemble forecast mean was lowest at the urban background site all months and the difference between the ensemble forecast and the observations was smallest during the winter months. Again, indicating that the forecast cannot reproduce the low ozone concentrations at the regional background sites during the summer month.

High mean concentrations of ozone did not differ much between the three sites, during JJA, whereas median NO<sub>2</sub> concentrations varied substitutionally between the regional and the urban sites (Figures 5 – 6 & 11 – 12 ). This suggests that background concentrations of NO<sub>x</sub> was saturated for ozone formation at all three sites. As ozone production is not instant, during time it takes for ozone to form, the air is advected and turbulently mixed. This results in well-mixed daytime maximum background concentrations of ozone. This feature can be seen in observations and was captured by the ensemble forecast, however the mean concentration of ozone was overestimated by the forecast with approximately  $10 \mu\text{gm}^{-3}$ .

Low mean concentrations of ozone differ more between the urban and the regional background sites during JJA than during DJF. The regional background sites display a longer period of low mean concentrations over the course of a day compared to the urban site and the minimum mean concentration was lower at the regional background sites compared to the urban. This is likely due to dry deposition by plant uptake being more effective at the regional background sites as they are richer in vegetation. Whereas the urban site is largely surrounded by buildings and pavement. This sink dominates during summer

when the plants have leaves and is extra strong during stable atmospheric conditions with low wind. Both regional sites are situated in areas ideal for ozone consumption by plant up take, leading to lower night time minimum ozone concentrations at the regional background sites compared to the urban. The minimum mean concentration was slightly lower at the site at Aspvreten than at Norr Malma. This could be due to the fact that the plant matter surrounding the site at Aspvreten (forest) consumes ozone more effectively than the plant matter surrounding the site at Norr Malma (crop field). Low mean concentrations at the regional background sites were strongly overestimated by the ensemble forecast indicating that the forecast representation of dry deposition is not satisfactory. The land surface cover in the individual models contribute to the magnitude of dry deposition in the individual models. There are some roads and residential areas in both grid boxes containing the regional observation sites which would result in a lower ozone consumption compared to the immediate surroundings of the observation sites. It would therefore be possible that the forecast could be improved by altering land cover in order to better represent the actual land surface cover. The vertical resolution is quite coarse this could mean that stratospheric ozone transport was a larger factor than in reality. This could be another reason for the systematic overestimation of ozone concentrations. Another possible reason could be that the chemical process production of ozone is too efficient in the models. The model underestimation of  $\text{NO}_2$  but overestimation of ozone could indicate this.

During DJF lack of sunlight has eliminated the diurnal cycles at the regional background sites and mean concentrations were almost constant all day. The ensemble forecast captured the lack of diurnal variations at the regional background sites quite well during this period compared to JJA. This could be due to the fact that photochemical chemical processes are of less importance during the winter season. Chemical processes are associated with relatively high uncertainty in each individual model. Each individual model uses a different chemistry schemes, harmonizing them in the post-processing lead to increased uncertainties. Hence, a forecast, where these processes are of less importance, would be more accurate than one where concentrations largely depend on chemical processes. The increased ensemble forecast accuracy during winter is likely a consequence of this. During winter the model is more accurate than at summer. The forecast was more accurate at the regional background sites compared to the urban background during DJF. A likely reason for the two daily minima at the urban background site during DJF could be termination of ozone production by equation (1). Winter is associated high levels of  $\text{NO}_x$  from traffic and other fossil fuel combustion, particularly in urban areas since the residence time of  $\text{NO}_x$  is relatively short. This is another example of a local process the model cannot resolve. Examining the accuracy of how the contribution by local emissions of  $\text{NO}_2$  during DJF compares to that of the destruction of ozone by  $\text{NO}_x$  it was evident that the ozone destruction was less accurately captured. The process is a chemical reaction, as opposed to the direct emissions of  $\text{NO}_2$ , and therefore associated with uncertainties in the individual models.

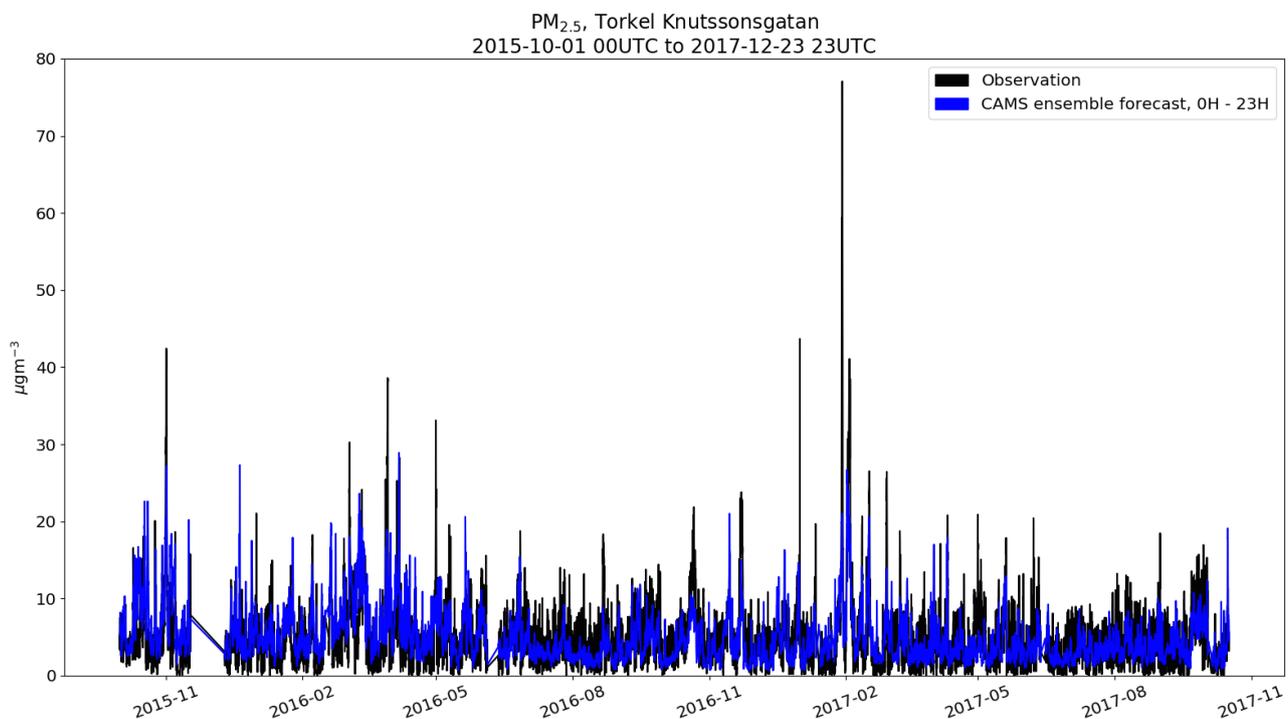
Mean concentrations of ozone increased with wind speed for both forecast and observation (Figure 34, Appendix II). Stronger winds lead to a more well-mixed boundary layer, leading to weaker group uptake by plants, as previously discussed, resulting in higher mean ozone concentrations since the sink is weakened. This behavior was not seen in the mean ozone concentration when the winds increase further than wind speed interval centered around 6 m/s (Figure 34, Appendix II). After a certain wind speed, mixing of the air will lead to ozone being diluted reducing mean concentration despite the weakened sink. Furthermore, only a few samples with high wind speeds was included in the data set. This would increase uncertainty in the result for high wind speeds. The ensemble forecast captured an increase in ozone with wind speed quite well. The forecast became more accurate with increasing wind speed, particularly at the regional background site. This is consistent with other results found in this study indicating that the forecast does not predict low ozone concentrations accurately. Low ozone concentrations occur when the wind speeds are low and conditions are stable.

### 3.2.3 Particulate matter, $\text{PM}_{2.5}$

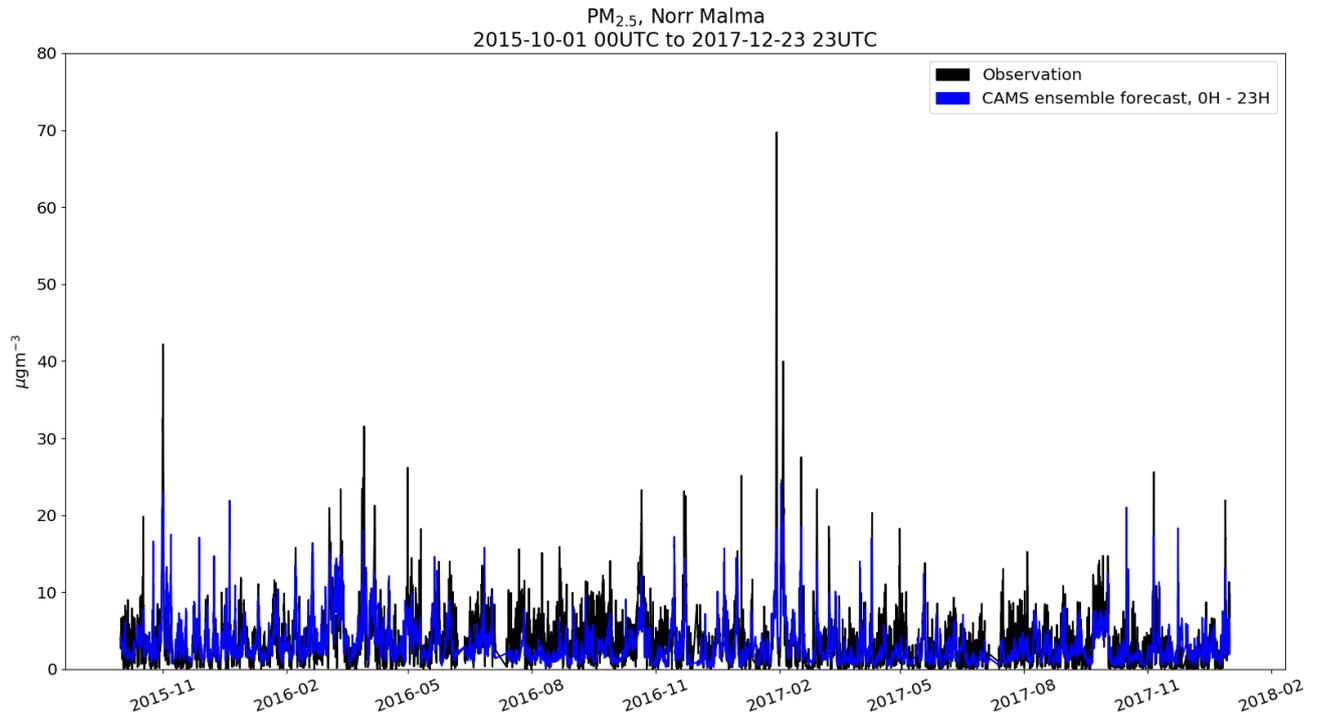
Concentrations of  $\text{PM}_{2.5}$  did not seem to vary much with season at either observation site (Figure 13 & 14). However, there was a slight peak in monthly median concentrations in April – May and September – October at the urban background site (Figure 15). Median concentrations were quite constant at the regional background site with an exception in September where there was a concentration maxima

and in November where there was a concentration minima (Figure 15). Observed monthly median concentrations of  $PM_{2.5}$  were slightly higher at the urban background site compared to the regional background site (Figure 15). The ensemble forecast captured the fact that concentrations were higher at the urban site but it was not able to predict high peaks at either site. Furthermore, the ensemble forecast displayed more seasonal variation at both sites compared to observations, with low levels during summer and elevated levels during the winter/spring (Figure 15). The forecast did not capture the general appearance variations of the monthly median for either observation site. At the urban site, concentrations were slightly overestimated during DJF and underestimated during JJA. Furthermore, the forecast underestimated monthly median concentrations at the regional background site all month of the year. The largest difference between forecasted and observed monthly medians occurred during JJA at the regional background site. A possible explanation of this could be that the processes controlling wet deposition are too effective in the individual models, since that process is strongest during summer. The difference between the two sites was much larger in the forecast than in the observations. This could be due to a too effective dilution of long-range transported pollutants in the forecast.

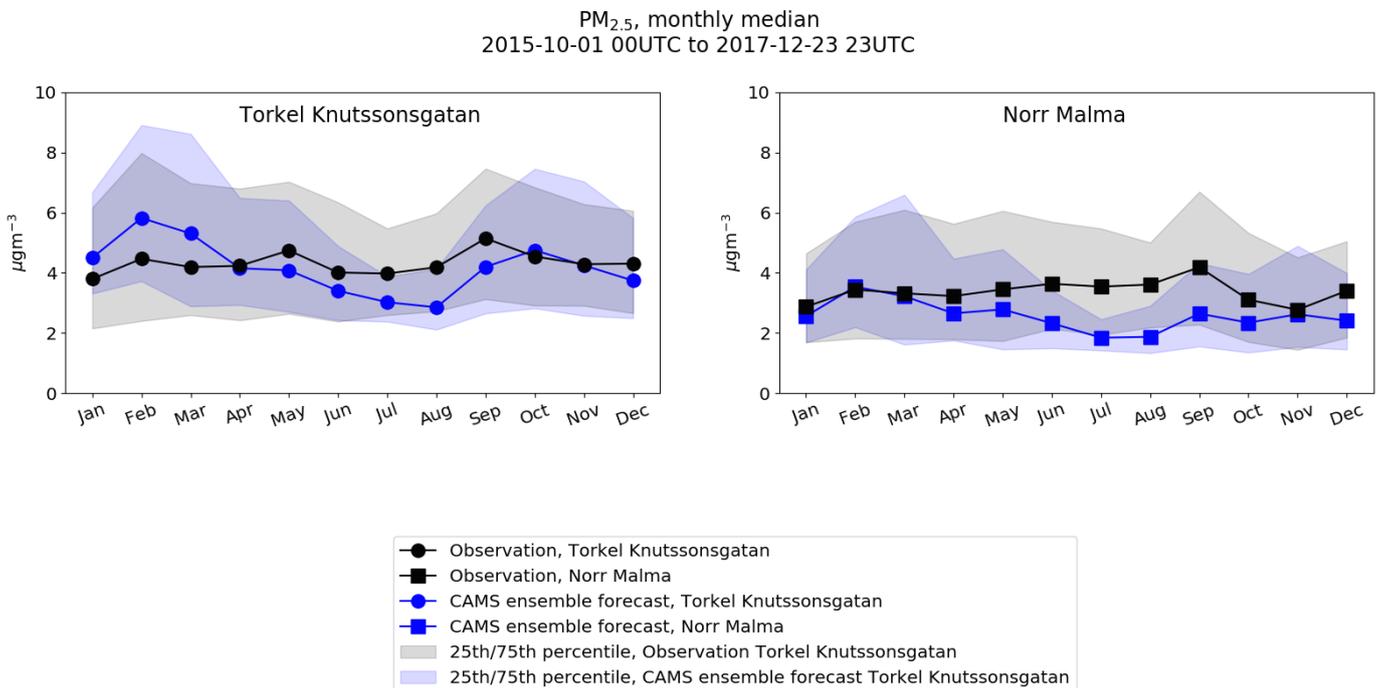
The forecast underestimation of high concentrations could be due to horizontal grid size and the fact that the forecast is an ensemble. For a given grid box, there will be a larger area to spread long-range transported pollutants advected into a region, originating from outside the grid box, compared to a smaller grid box. Thus, levels of advected air pollutants would tend to be underestimated. The vertical resolution of the forecast could have similar impacts.



**Figure 13:** Time series of hourly data of  $PM_{2.5}$  at the urban background site at Torkel Knutssongatan October 2015 – December 2017. Y-axis displays concentration of  $PM_{2.5}$  in  $\mu m^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively.

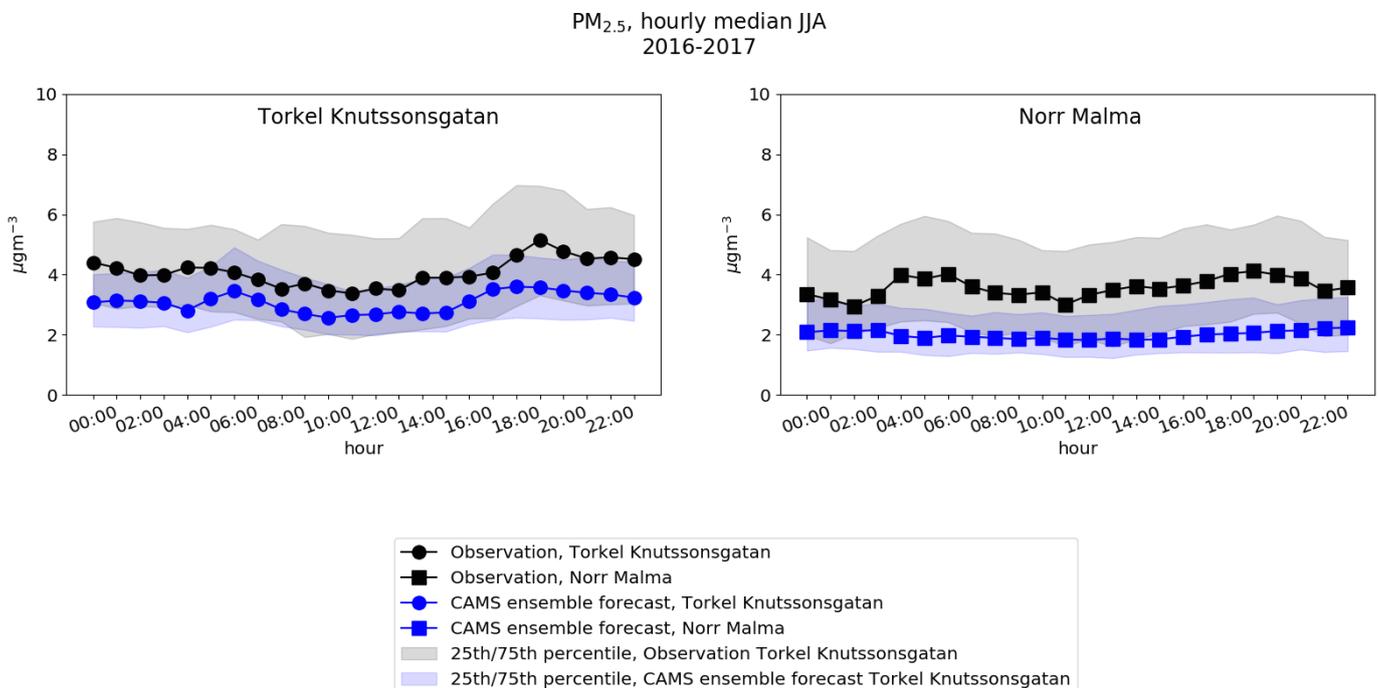


**Figure 14:** Time series of hourly data of  $PM_{2.5}$  at the regional background site at Norr Malma October 2015 – December 2017. Y-axis displays concentration of  $PM_{2.5}$  in  $\mu m^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively.

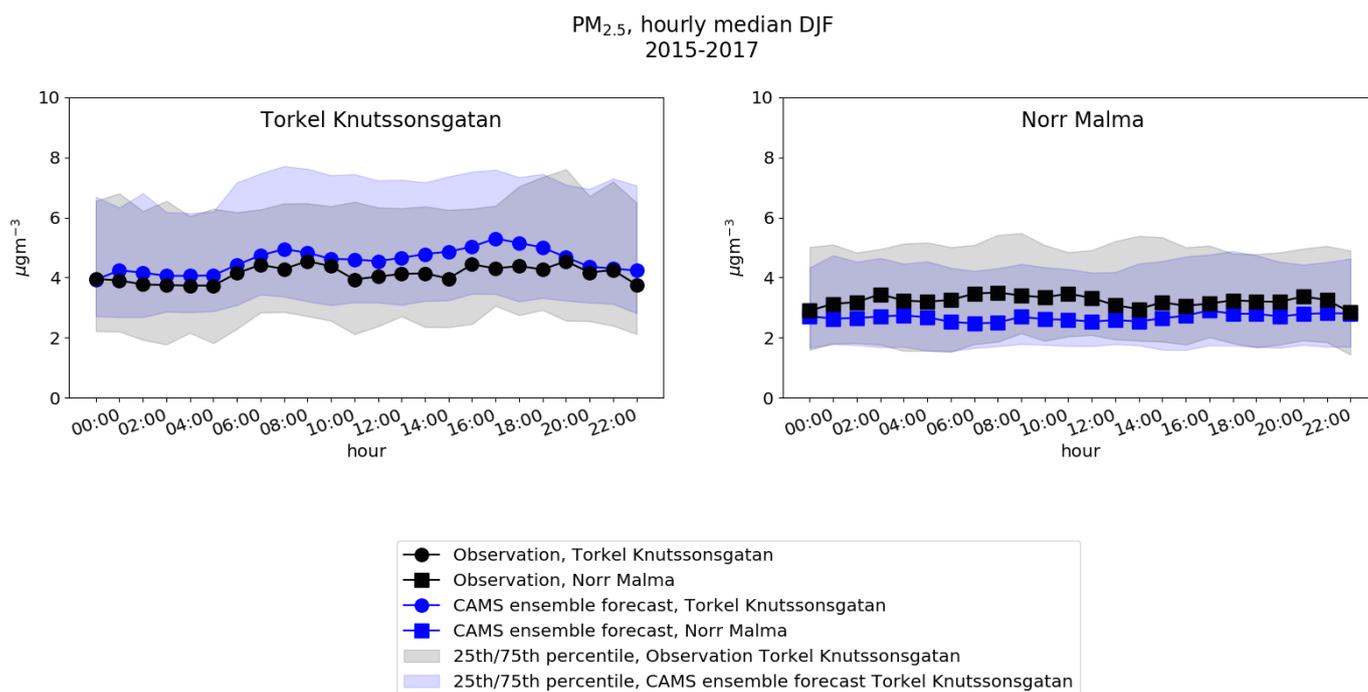


**Figure 15:** Monthly median of  $PM_{2.5}$ . Y-axis displays median concentration of  $PM_{2.5}$  in  $\mu m^{-3}$  and x-axis shows months. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssongatan and squared curves the regional background site at Norr Malma.

During JJA median concentration of  $PM_{2.5}$  only varied slightly with time of day at both of sites. At the urban the small variations were to some extent captured by the ensemble forecast. However, at the regional site, the forecast was constant over the course of the day whereas median concentrations in the observations varied (Figure 16). Median concentrations were underestimated by approximately  $1 - 2 \mu g m^{-3}$  by the forecast all hours at both sites during JJA. During DJF variations in daily median concentrations of  $PM_{2.5}$  were quite small as well (Figure 17). Median concentrations were overestimated by the forecast at the urban background site and underestimated at the regional. Median concentrations of  $PM_{2.5}$  largely depend on long-range transportation by advected air masses. Even so, the general behavior was not captured with particularly high accuracy by the forecast. However, the forecast is more accurate during DJF than during JJA. This suggests that what is causing the discrepancy between the forecast and observations during JJA has to do with chemical processes or the process controlling wet deposition, rather than advection of polluted air long range-transported from the European continent. Chemistry and wet deposition occur strongest in summer whereas long-range transport is active all year. Furthermore, the accuracy was slightly higher at the urban background site both season. This could be a result of the dominating source of  $PM_{2.5}$  at the urban site is direct emissions where secondary aerosol formation by chemical processes is of more relative importance at the regional site which is further from direct emission sources.



**Figure 16:** Hourly median of  $PM_{2.5}$ , June, July and August. Y-axis displays median concentration of  $PM_{2.5}$  in  $\mu m^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssonsgatan and squared curves the regional background site at Norr Malma. Blue and black shades represent 25th/75th percentile of forecast and observation median respectively.

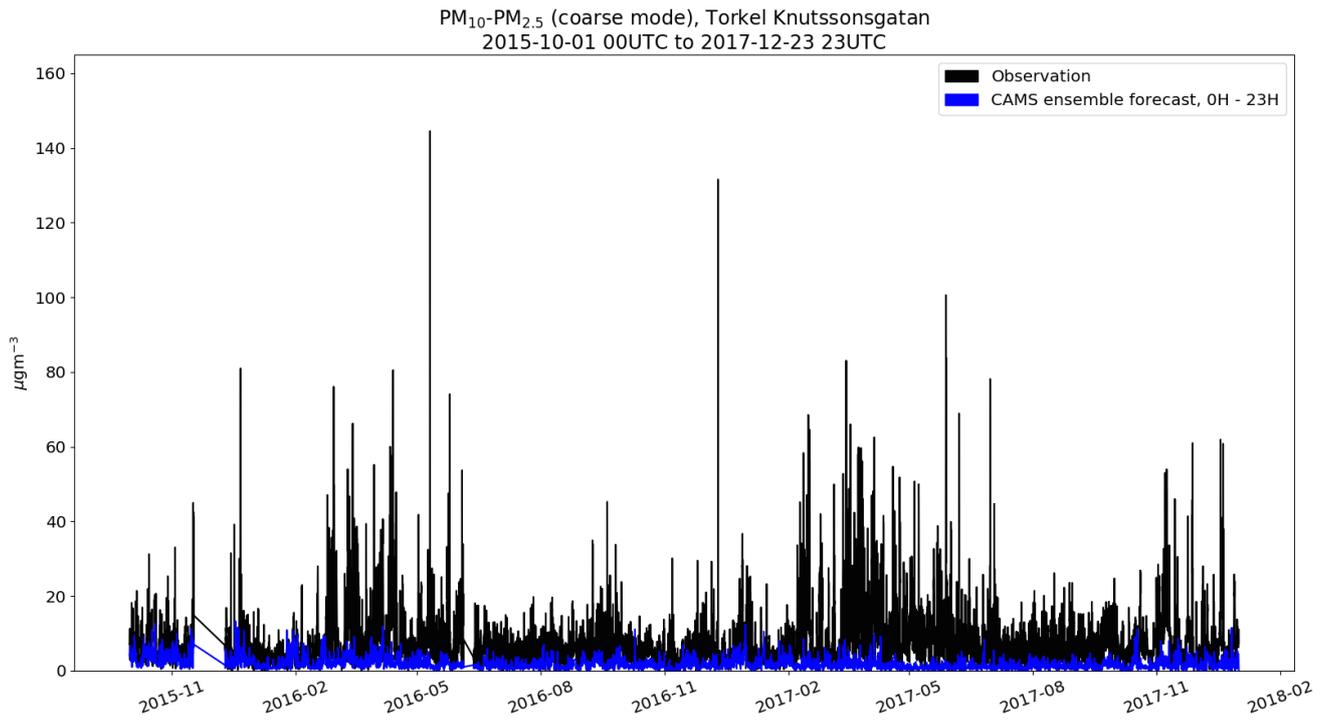


**Figure 17:** Hourly median of PM<sub>2.5</sub>, December, January and February. Blue and black curves represent forecast and observations respectively. Y-axis displays median concentration of PM<sub>2.5</sub> in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Dotted curves represent the urban background site at Torkel Knutssonsgatan and squared curves the regional background site at Norr Malma. Blue and black shades represent 25th/75th percentile of forecast and observation median respectively.

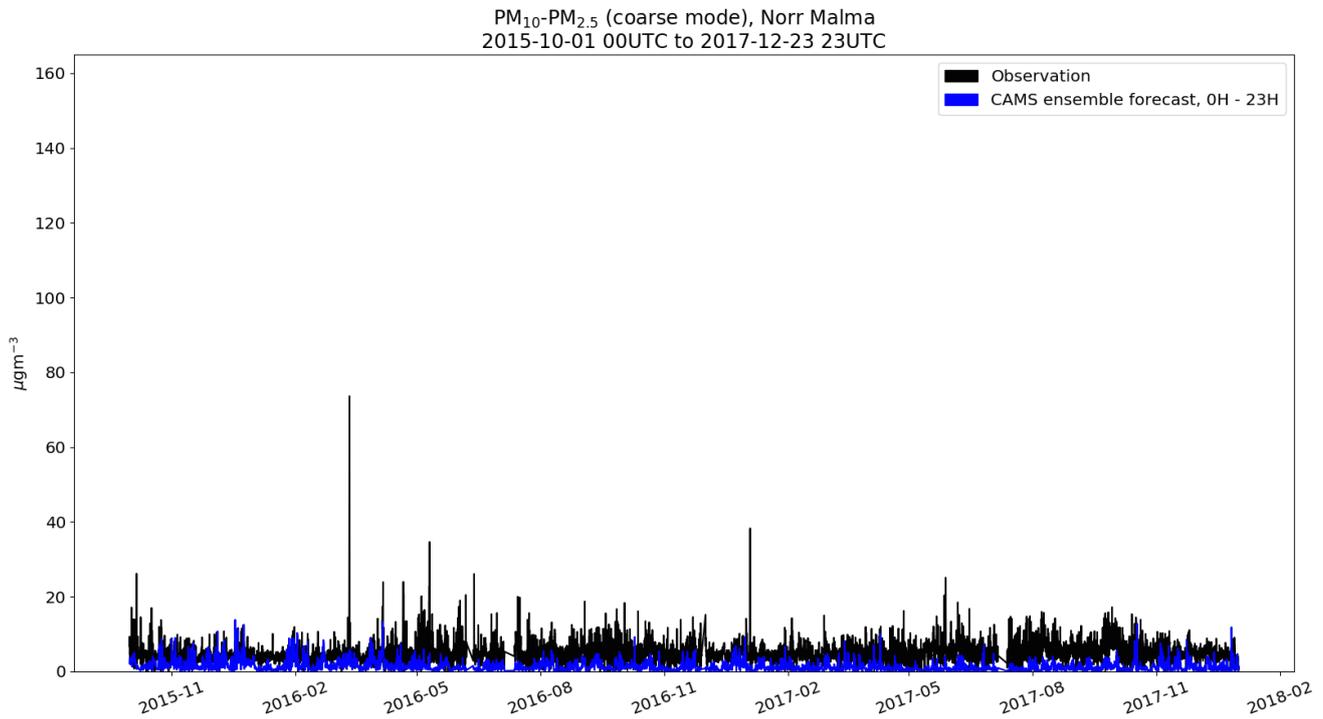
Mean concentrations of PM<sub>2.5</sub> decreased as a function of wind speed for both observed and forecasted data at the regional background site (Figure 35, Appendix II). At the urban background site, observed mean concentrations initially decreased with increasing wind. However, after a certain wind speed, mean concentrations increased. Mean forecasted concentrations of PM<sub>2.5</sub> decreased with increasing wind speed. The decrease seen in the results is likely due to increased turbulent mixing and thereby increased dilution of the particles with increasing wind speed. The observed mean concentrations increase at a certain wind speed at the urban site could be due to local emission sources. When wind speed is high and a local emissions sources are upwind of the observation site, concentrations can increase at the site. The forecast would not resolve this and thus not show the increase. Data at high wind speeds was however not reliable since the data availability was very low.

### 3.2.4 Particulate matter, PM<sub>10</sub>-PM<sub>2.5</sub> (coarse mode)

Observed yearly variations of coarse particulate matter displayed maximum during the period February – May at the urban background site at Torkel Knutssonsgatan (Figure 18). There was very weak yearly variations at the regional background station at Norr Malma (Figure 19). The ensemble forecast underestimated concentrations of coarse particulate matter at both sites, particularly the urban background site. Yearly variations were not captured by the forecast at the urban background site. The forecast did not differ much between concentrations of coarse particulate matter at the urban and regional background sites but there was a substantial difference in observed concentrations between the sites. There was a large peak in the monthly median at the urban background site during March – May, which was not captured by the forecast (Figure 20). Monthly median variations at the regional background site were small. This behavior was captured by the forecast but concentrations were underestimated.

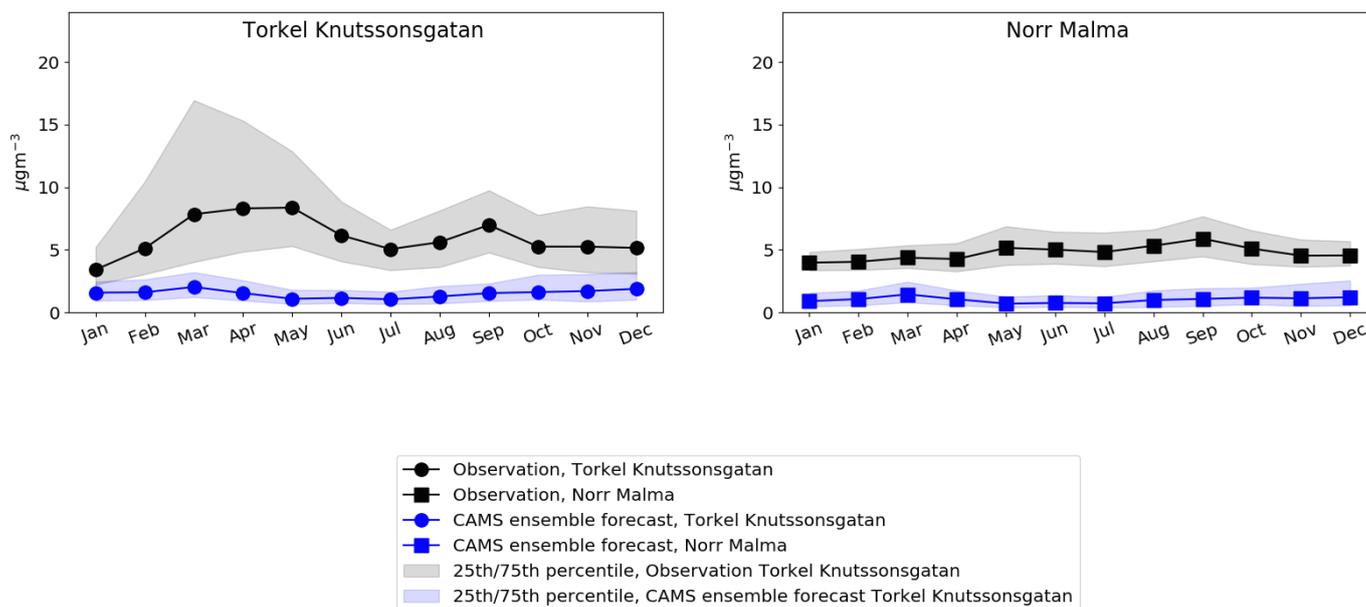


**Figure 18:** Time series of hourly data of PM<sub>10</sub>-PM<sub>2.5</sub> at the urban background site at Torkel Knutssongatan October 2015 – December 2017. Y-axis displays concentration of PM<sub>10</sub>-PM<sub>2.5</sub> in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively.



**Figure 19:** Time series of hourly data of PM<sub>10</sub>-PM<sub>2.5</sub> at the regional background site at Norr Malma October 2015 – December 2017. Y-axis displays concentration of PM<sub>10</sub>-PM<sub>2.5</sub> in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively.

Monthly median PM<sub>10</sub>-PM<sub>2.5</sub> (coarse mode)  
2015-10-01 00UTC to 2017-12-23 23UTC

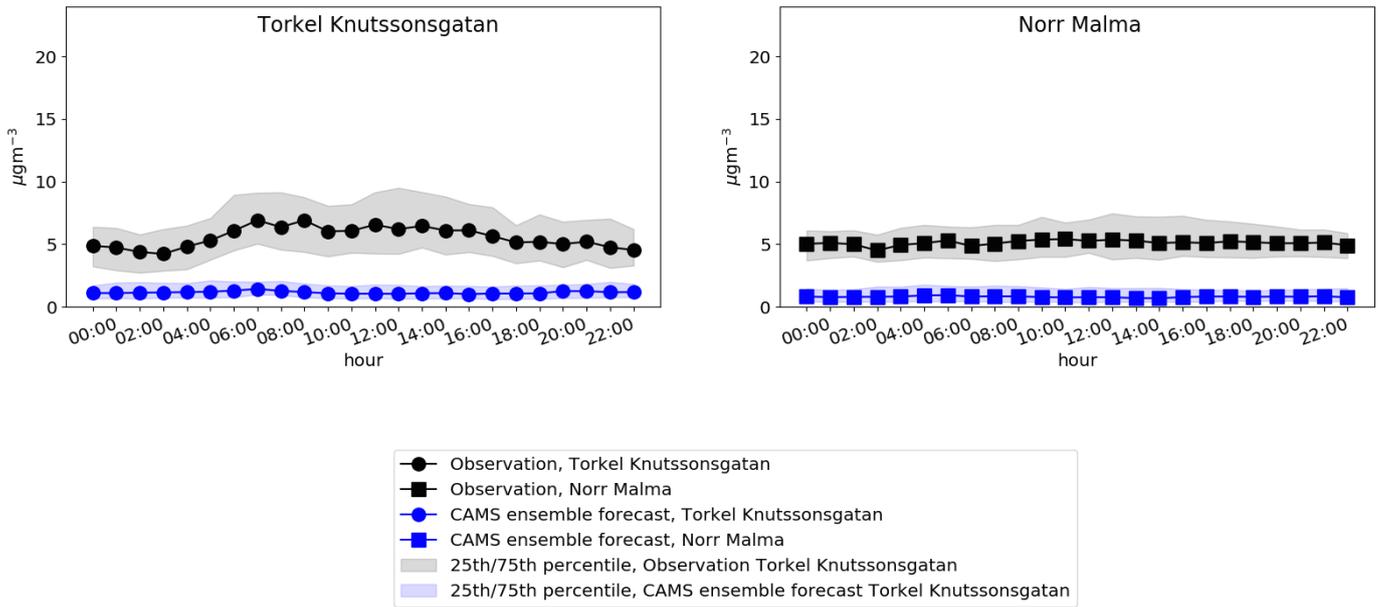


**Figure 20:** Monthly median of PM<sub>10</sub>-PM<sub>2.5</sub>. Y-axis displays median concentration of PM<sub>10</sub>-PM<sub>2.5</sub> in  $\mu\text{gm}^{-3}$  and x-axis time with monthly resolution. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssonsgatan and squared curves the regional background site at Norr Malma.

Median concentration of coarse particulate matter was quite constant over the course of the day at the regional background site both during summer and winter season (Figures 21 & 22). This general behavior was captured by the ensemble forecast, but median concentrations were underestimated by approximately  $5 \mu\text{gm}^{-3}$ . There were diurnal variations in observations at the urban background site both seasons. These were not captured by the forecast.

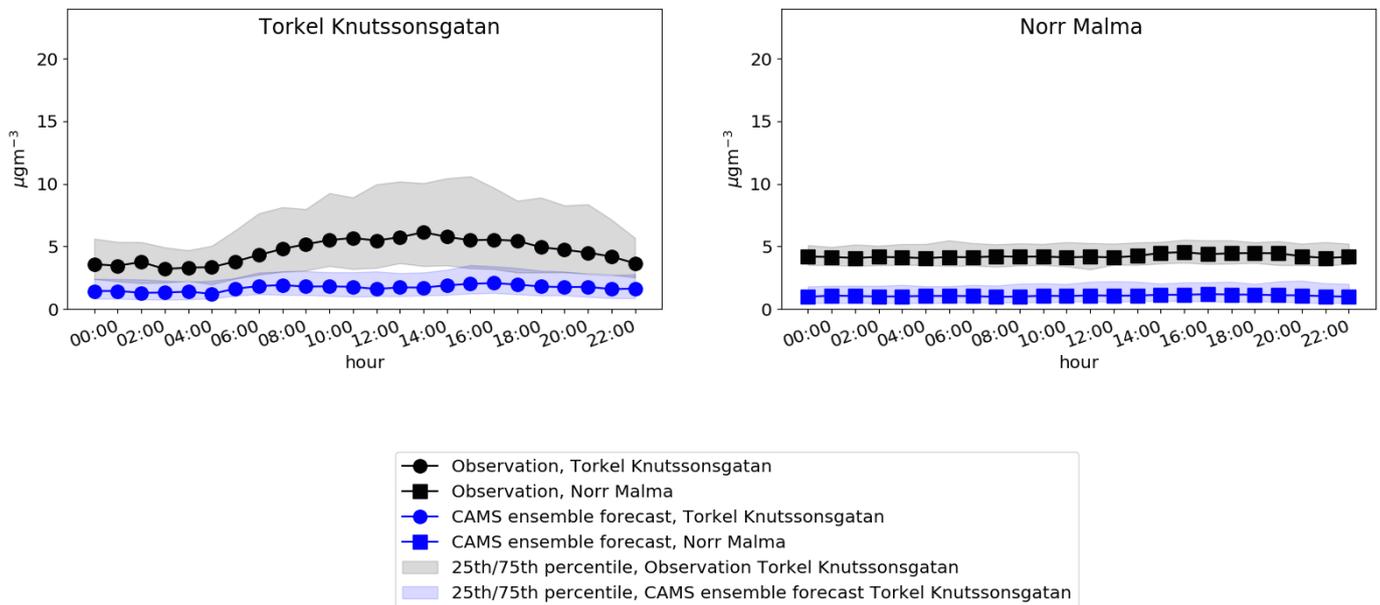
Coarse particulate matter is not advected far with air masses since the particles have a short residence time, approximately 10 hours. Thus, local emissions contribute most to concentrations of coarse particulate matter. The regional background station does not have such sources in its proximity. Hence, the concentrations were lower there than at the urban background site, which is strongly influenced by the coarse particles certain periods of the year. Road and tire wear and resuspension of dust are local sources of coarse particulate matter which contribute to the maximum at the urban background site. The forecast does not resolve these local emission and forecast accuracy is thereby low for coarse particles at the urban site. The underestimation by the forecast at the regional background site could be due to dilution effects being overestimated in the forecast as well as the fact that the grid size is coarse. The systematic underestimation in the forecast could also be due to outdated input emission data. Further examination is however needed in order to determine how large of an effect that would be.

PM<sub>10</sub>-PM<sub>2.5</sub> (coarse mode), hourly median JJA  
2016-2017



**Figure 21:** Hourly median of PM<sub>10</sub>-PM<sub>2.5</sub>, June, July and August. Y-axis displays median concentration of PM<sub>10</sub>-PM<sub>2.5</sub> in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssonsgatan and squared curves the regional background site at Norr Malma. Blue and black shades represent 25th/75th percentile of forecast and observation median respectively.

PM<sub>10</sub>-PM<sub>2.5</sub> (coarse mode), hourly median DJF  
2015-2017

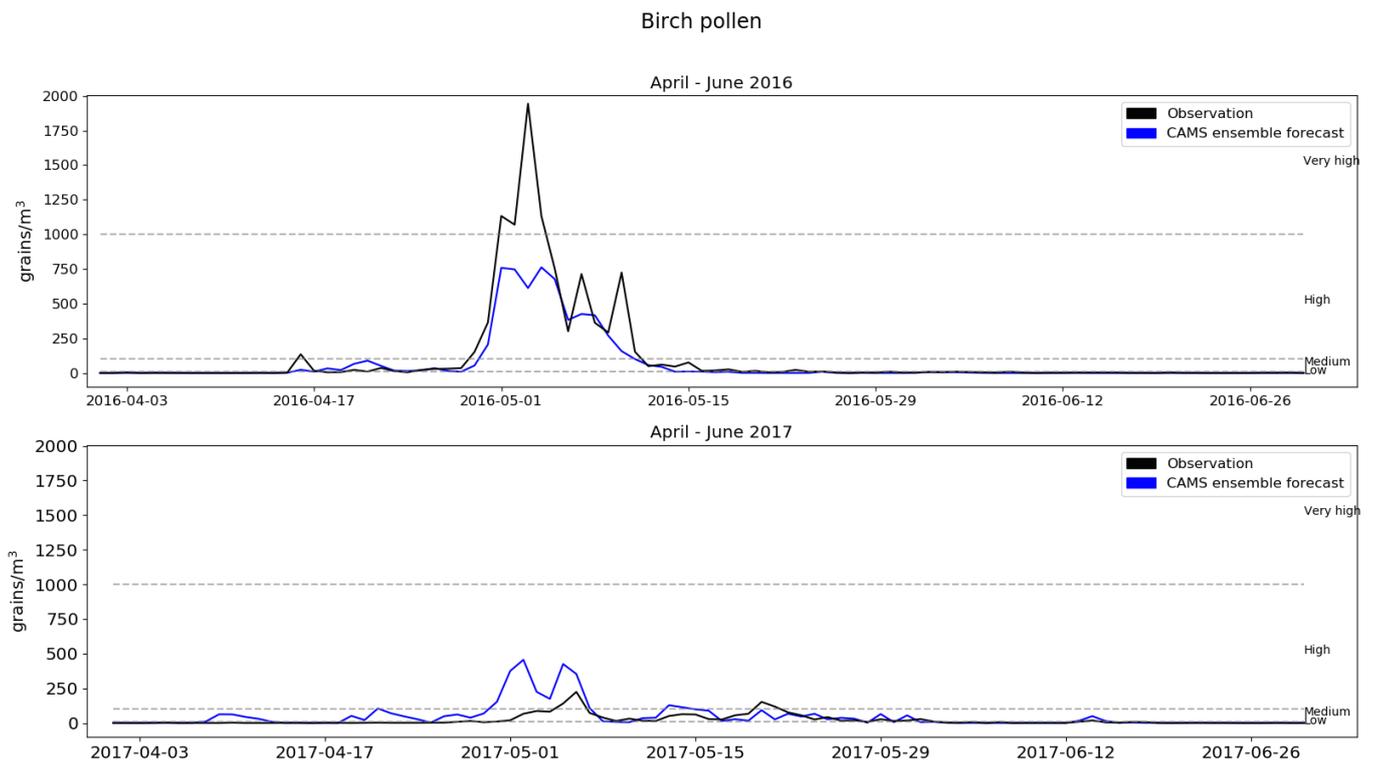


**Figure 22:** Hourly median of PM<sub>10</sub>-PM<sub>2.5</sub>, December, January and February. Y-axis displays median concentration of PM<sub>10</sub>-PM<sub>2.5</sub> in  $\mu\text{m}^{-3}$  and x-axis time with hourly resolution. Blue and black curves represent forecast and observations respectively. Dotted curves represent the urban background site at Torkel Knutssonsgatan and squared curves the regional background site at Norr Malma. Blue and black shades represent 25th/75th percentile of forecast and observation median respectively.

### 3.2.5 Birch pollen

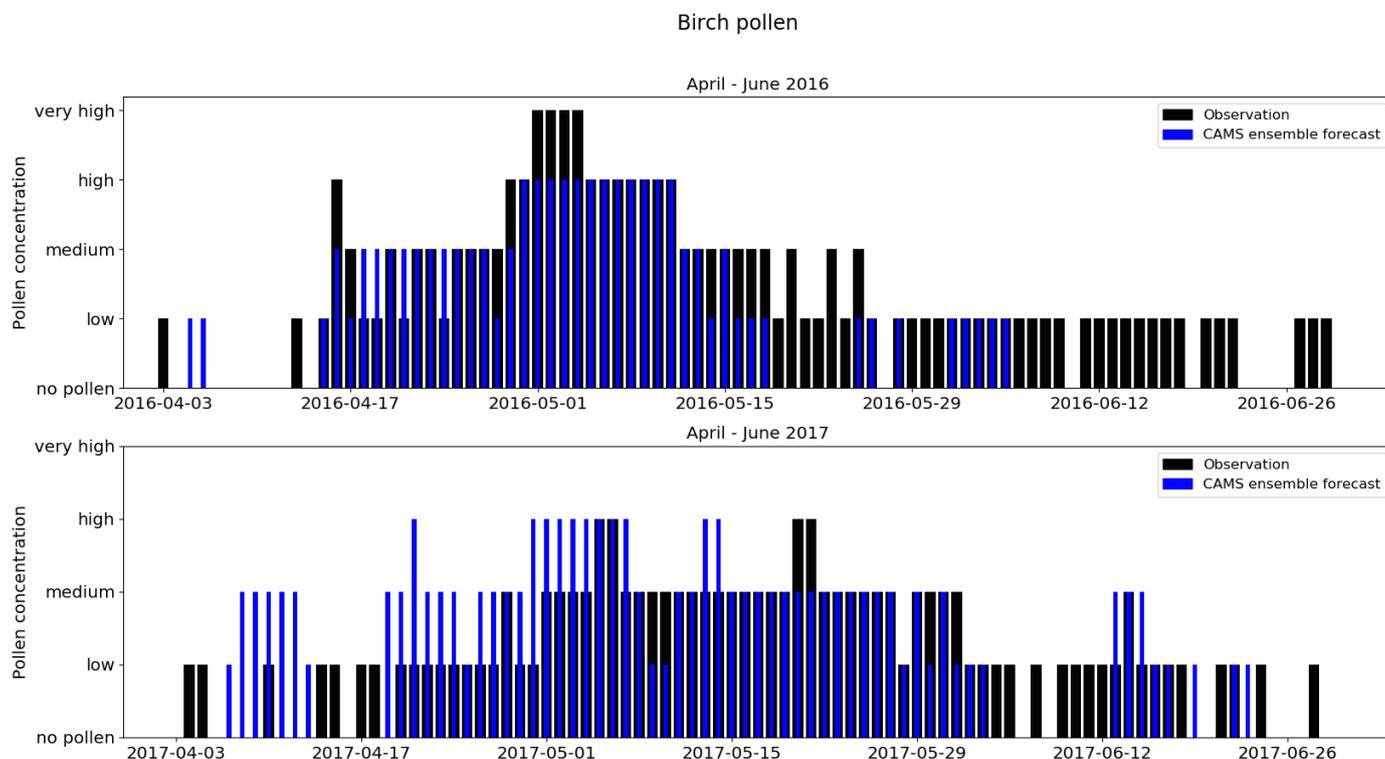
There were large differences in observed daily mean concentrations of birch pollen between 2016 and 2017 (Figure 23). Pollen season occurred roughly during the same time both years (Figure 23). The ensemble forecast captured the start of the intense pollen bloom quite well (Figure 23). Notable is that the first forecasted high peak in 2016 occurred just a few dates before the observed. The end of the pollen season was captured quite well 2017 but ends to early in 2016. Daily mean concentrations of pollen were strongly underestimated by the model in 2016, but overestimated in 2017. Hence, there does not seem to be a systematic over or underestimation, however two years might not be enough data to fully conclude that.

The ensemble forecast does not include quasi-biannual pollen blooming events (Sofiev *et al.*, 2017). Therefore, would forecasted variations instead be due to variations in meteorological conditions such as wind direction and speed as well as temperature. The forecasted pollen levels would be too low if the temperatures are underestimated, since temperature controls blooming start. Long range-transport of pollen by air masses also affect pollen concentrations and depending on dominating wind direction and boundary layer height at the blooming site during the pollen season. Levels of pollen would be higher or lower depending on if the air mass originates from the European continent or the Atlantic Ocean respectively. This is important for determining pollen concentrations in both forecast and observations. Thus, as long as the wind is accurately forecasted, the forecasted concentrations would be accurate, provided that blooming in the forecast has started. Therefore, temperature seem to be the factor with largest influence on birch pollen forecast accuracy, except for the lack of quasi-biannual pollen blooming events.



**Figure 23:** Time series of daily mean values of birch pollen during pollen season 2016 (top) and 2017 (bottom). Y-axis displays mean concentration of birch pollen in grains/m<sup>3</sup> and x-axis time with daily resolution. Blue and black curves represent forecast and observations respectively. Concentration levels are represented by dashed lines.

Low or medium levels of pollen were most frequent levels both years (Figure 24). There were some dates with very high pollen levels during 2016, but none during 2017. The ensemble forecast predicted the correct pollen level in 46 % of the dates. Furthermore, forecasted level was within one pollen level of the observed in 68 % of the dates when, both forecast and observation showed pollen concentrations were above zero or when both showed no pollen. In 24 % of the dates observed pollen concentrations were above zero when the forecast was no pollen and in 3 % of the dates the forecast was concentrations above zero when no pollen was observations. Binary statistics were used to asses pollen categories and a binary



**Figure 24:** Daily mean values of birch pollen divided into concentration categories during pollen season 2016 (top) and 2017 (bottom). Y-axis displays concentration categories of birch pollen in and x-axis time with daily resolution. Blue and black bars represent forecast and observations respectively.

statistics table was constructed (Table 5). No distinction between the concentration levels were made in Table 5. The POD scored 0.67 and FBI scored 0.75. FAR and FAM scores were calculated in the case of "yes" event being pollen level of high or very high and "no" event being below those levels. FAR score was then 39 % and FAM 22 %. These scores indicated that the forecast still has room fore improvement when it comes to predicting pollen concentrations. The current representation of the blooming season is perhaps over-simplified, since birch pollen forecast does not include quasi-biannual cycles. However, as the birch pollen concentration levels are defined over quite large intervals, the forecast still gives a good indication of the pollen concentration level even though the it does not capture the details. If a warning system would be based on the forecast it would give a false alarm more frequently than missing an alarm. These flaws could be problematic, since inaccurate pollen forecast used to inform the public can result in unnecessary, harmful exposure. However, if for example uncertainties are stated together with the pollen forecast, it would make a good indicator of which pollen level is to come.

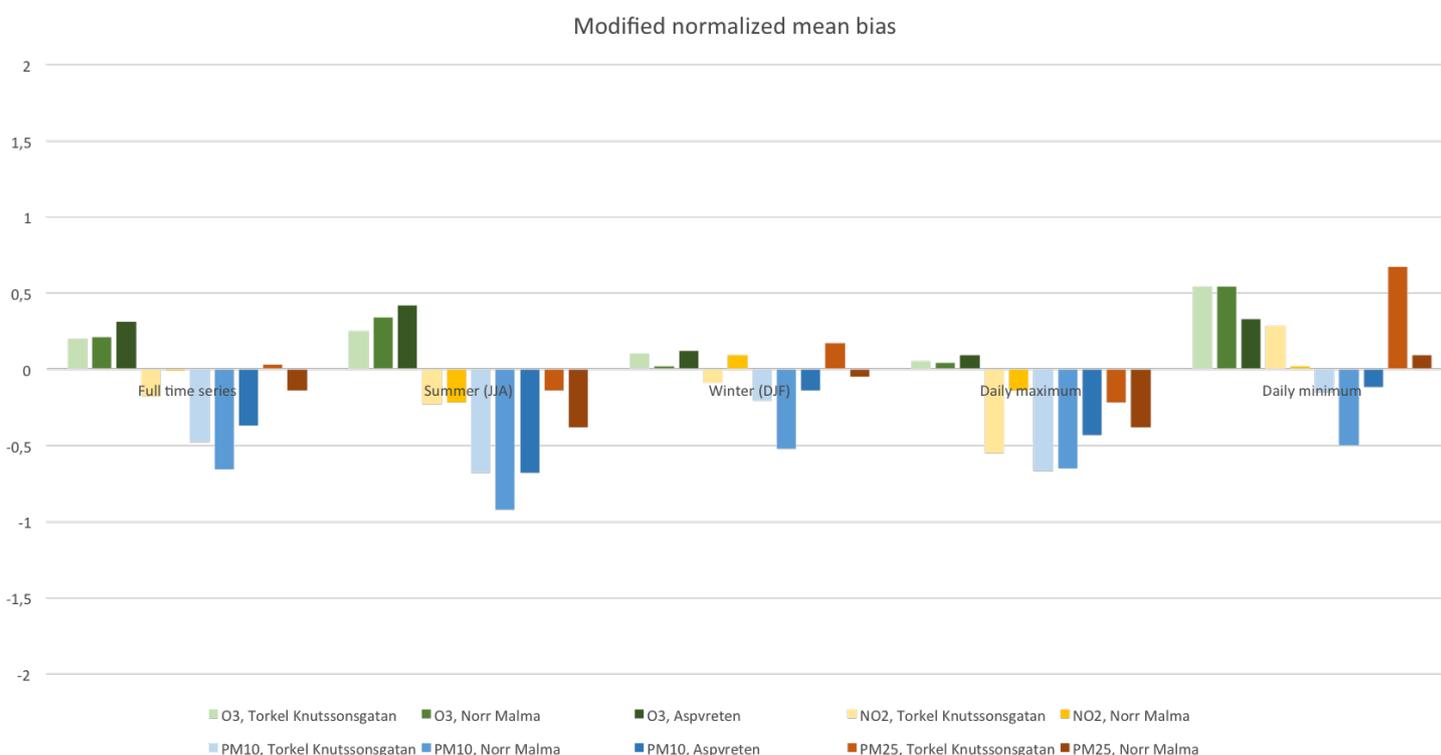
**Table 5:** Binary statistics table birch April – June 2016 & 2017. Total of hits, misses, false alarms, correct non-events and total "yes" and "no" events have been summarized according to Table 4.

	Event observed, yes	Event observed, no	Total
Event forecasted, yes	94	11	105
Event forecasted, no	46	31	77
Total	140	42	182

### 3.3 Statistical indicators

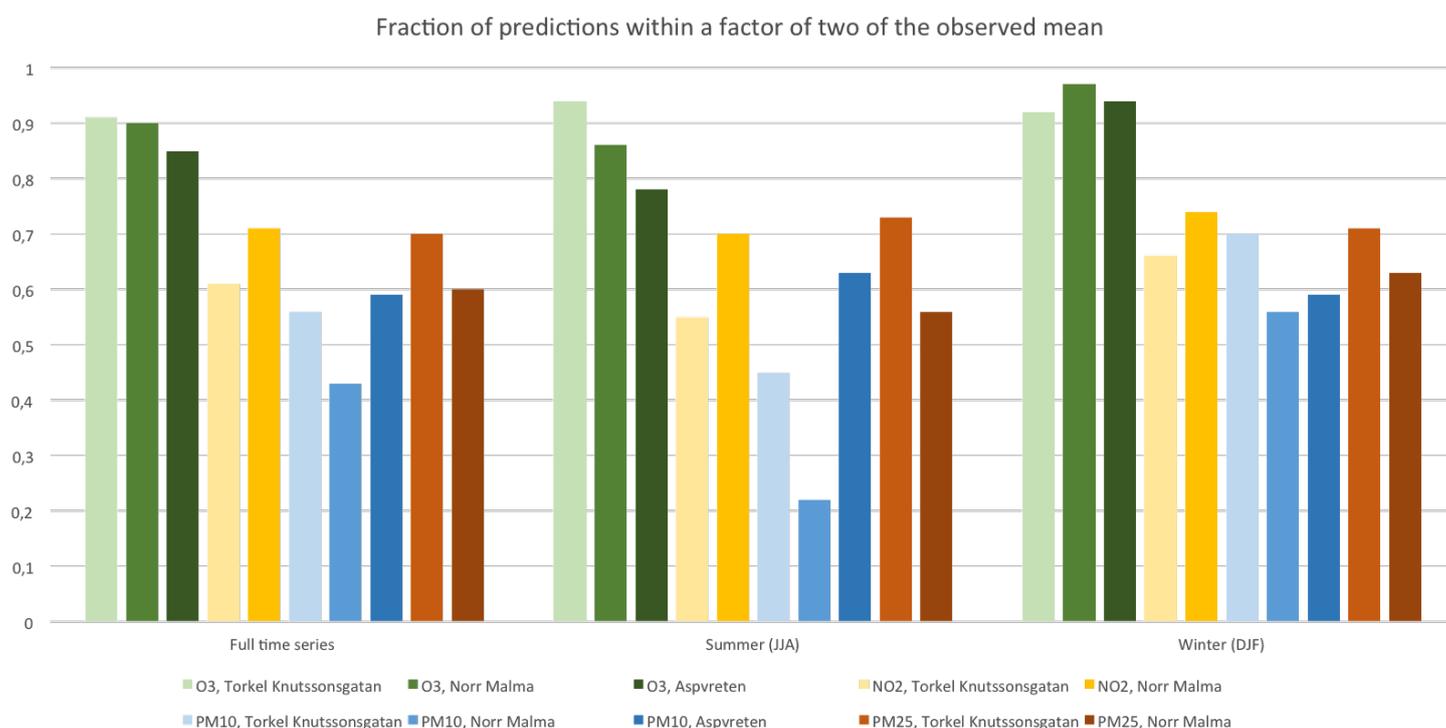
Since only the ozone data set had a normal distribution, statistical reliability of for the other pollutants is decreased as discussed in section 2.3 and the obtained scores for NO<sub>2</sub>, PM<sub>10</sub> and PM<sub>2.5</sub> are influenced by outliers. Nonetheless, statistical indicators for all pollutants were calculated according to the assumption of a normal distribution. Statistical indicators MNMB, fraction of predictions within a factor two of the observed mean, FGE, R and MQI were calculated for each pollutant for the whole time series, summer (JJA), winter (DJF). MNMB, FGE and R were also calculated for daily maximum and minimum for the whole time series.

MNMB scores for JJA showed that ozone was overestimated by the forecast at all sites and the other pollutants were underestimated at all sites (Figure 25). PM<sub>10</sub> had the largest MNMB scores at all sites, ranging between -0.68 and -0.92, being largest at the regional background site at Norr Malma. All other MNMB scores were lower than  $\pm 0.5$  (Figure 25). MNMB scores for DJF were lower for the majority of the pollutants. Ozone was again overestimated by the forecast at all sites as was NO<sub>2</sub> and PM<sub>2.5</sub> at the regional background site. All MNMB scores except for PM<sub>10</sub> at the regional background site were lower than  $\pm 0.5$ . The over- and underestimations of the respective pollutants corresponds to what have been seen in the results of the time series mean and medians. MNMB scores of PM<sub>10</sub> were largest at the regional background site at Norr Malma both seasons. Meaning that the magnitude of concentration underestimations was largest there. However, the reason behind the underestimations does not seem to have been the same for both seasons. PM<sub>10</sub> consists of coarse particle matter and PM<sub>2.5</sub>. MNMB score of PM<sub>2.5</sub> at the regional background site was relatively large during summer and small during winter. Whereas MNMB scores of PM<sub>10</sub> were relatively large both seasons. This indicates that during JJA, discrepancy in the forecast of PM<sub>2.5</sub> contributed more to the PM<sub>10</sub> MNMB score than during DJF. During DJF, it could be uncertainties related to coarse particulate matter (PM<sub>10</sub> - PM<sub>2.5</sub>) which contributed to the MNMB score. As previously stated, discrepancies in the forecast during JJA could be related to uncertainties in the model parameterizations of chemical processes and wet deposition whereas during DJF the uncertainty lies in fact that the models cannot resolve local contributions of coarse particles.



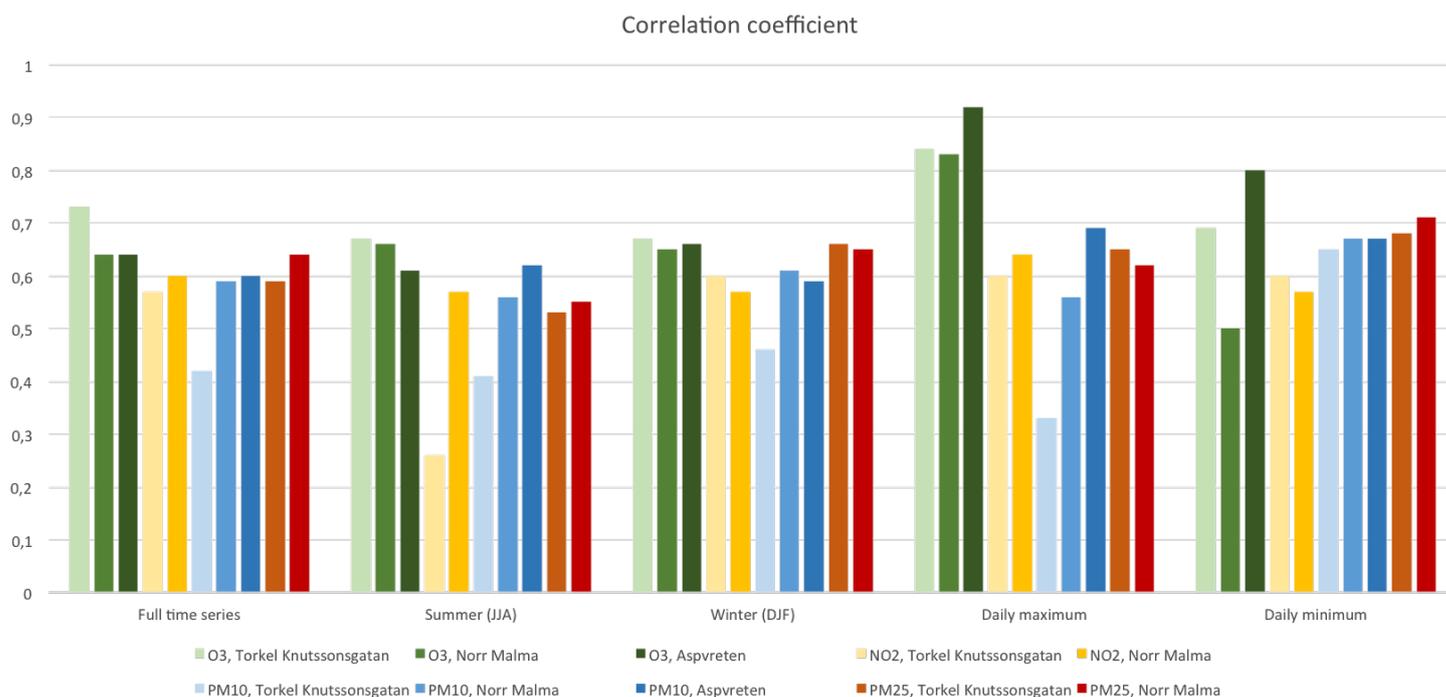
**Figure 25:** MNMB for the pollutants at each sites at various time periods. Y-axis displays dimensionless indicator score and x-axis each pollutant and time period. Green colors represent O<sub>3</sub>, yellow NO<sub>2</sub>, blue PM<sub>10</sub> and red PM<sub>2.5</sub>. Periods shown are full time series, JJA, DJF, daily maximum and daily minimum.

PM<sub>10</sub> at the regional background site during JJA scored lowest fraction of predictions within a factor two of observed mean. Ozone had very high fraction of predictions within a factor two of observed mean at all sites. Fraction of predictions within a factor two of observed mean was generally higher during winter than summer. Fraction of predictions of NO<sub>2</sub> within a factor two of observed mean were higher at the regional background site than the urban during both DJF and JJA. The opposite was true for PM<sub>2.5</sub> (Figure 26). This indicates that the spread of forecasted NO<sub>2</sub> concentrations is larger at the urban background site whereas the spread of forecasted PM<sub>2.5</sub> concentrations is larger at the regional background site. This is an indication of forecast accuracy comparing a pollutant which is strongly affected by local emissions (NO<sub>2</sub>) and a pollutant which is not as influenced by local emissions but by long-range transport (PM<sub>2.5</sub>). The main source for NO<sub>2</sub> is local emission from fossil fuel burning in cars, factories and for residential heating while the main source of PM<sub>2.5</sub> is long-range transport together with a relatively small local contribution from fossil fuel burning. NO<sub>2</sub> display small spread at the regional background site and large at the urban during JJA whereas the spread for PM<sub>2.5</sub> is quite large at both sites, but largest at the urban background site. This indicates that for NO<sub>2</sub> it is local emissions which contribute to uncertainties in the forecast. Long-range transport and input emissions seem to be accurately represented since the spread at the regional background site is small. Whereas for PM<sub>2.5</sub> the spread, which is relatively large even though the contribution from local emission is very small, indicates that limitations in the forecast parameterizations of chemical processes, such as secondary aerosol formation, contribute to uncertainty in determining concentrations of PM<sub>2.5</sub>.



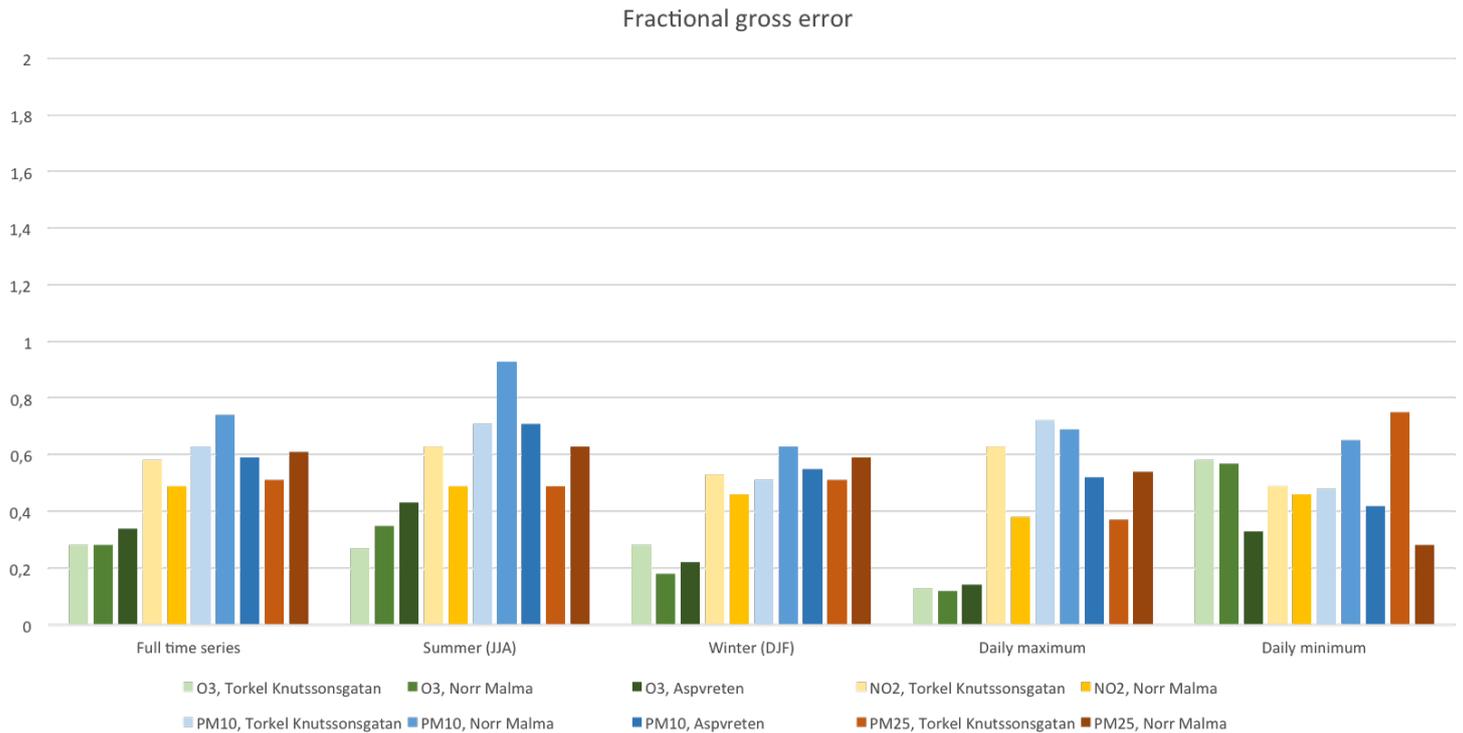
**Figure 26:** Fraction of predictions within a factor 2 of observed mean for the pollutants at each sites at various time periods. Y-axis displays dimensionless indicator score and x-axis each pollutant and time period. Green colors represent O<sub>3</sub>, yellow NO<sub>2</sub>, blue PM<sub>10</sub> and red PM<sub>2.5</sub>. Periods shown are full time series, JJA, DJF, daily maximum and daily minimum.

R did not vary much for ozone comparing seasons or sites (Figure 27). However, R did vary comparing daily maximum and minimum. This corresponds to what have observed in the daily mean results, where it was established that the forecast in much less accurate at predicting low compared to high concentrations of ozone (as discussed in section 3.2.2). R for  $\text{NO}_2$  did not vary much with season for the regional background site, but was substantially lower during JJA for the urban background site (Figure 27). The forecast better captured the behavior of the variations during DJF than JJA at the urban site. This can also be seen in the figures of  $\text{NO}_2$  hourly median (section 3.2.1). R varies for  $\text{PM}_{10}$ , being lower at the urban background site compared to the regional sites, both during JJA and DJF (Figure 27). Again, both coarse particulate matter and  $\text{PM}_{2.5}$  contribute to error in the  $\text{PM}_{10}$  forecast. The relative contribution of uncertainty from  $\text{PM}_{2.5}$  was larger during JJA and from DJF coarse particle matter contribute with a relatively larger uncertainty (Figure 27).



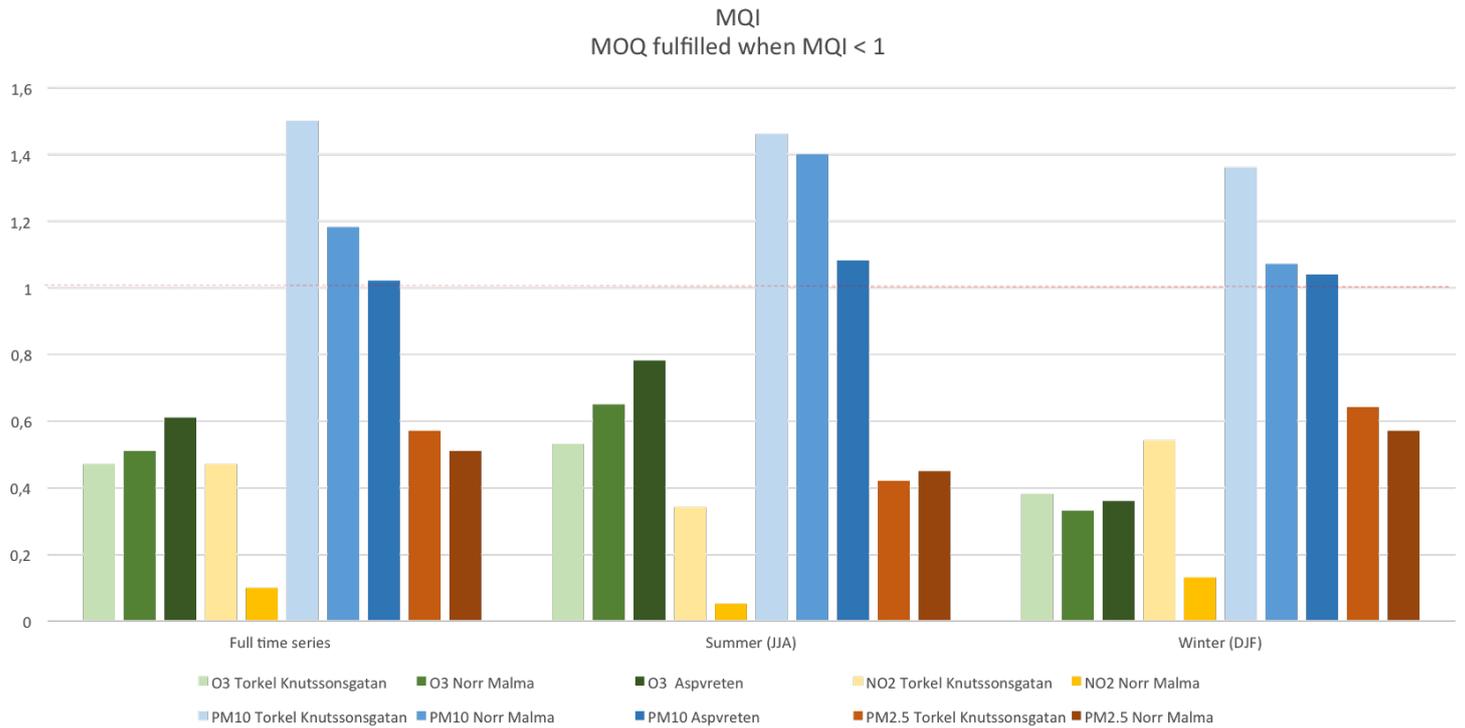
**Figure 27:** R for the pollutants at each sites at various time periods. Y-axis displays dimensionless indicator score and x-axis each pollutant and time period. Green colors represent  $\text{O}_3$ , yellow  $\text{NO}_2$ , blue  $\text{PM}_{10}$  and red  $\text{PM}_{2.5}$ . Periods shown are full time series, JJA, DJF, daily maximum and daily minimum.

FGE scores were generally small for ozone, except for the daily minimum. FGE scores for  $\text{NO}_2$  were small at the regional background site and slightly larger at the urban site. FGE scores for  $\text{PM}_{10}$  were largest during JJA, particularly at the regional background site at Norr Malma (Figure 28). This could be due to the fact that the site is situated at a crop field where harvest takes place during JJA leading to coarse particulate matter. FGE did not differ much between JJA and DJF for  $\text{PM}_{2.5}$  (Figure 28). It is however slightly smaller at the urban background site compared to the regional. This agrees with the fraction of predictions within a factor two of observed mean.



**Figure 28:** FGE for the pollutants at each sites at various time periods. Y-axis displays dimensionless indicator score and x-axis each pollutant and time period. Green colors represent  $O_3$ , yellow  $NO_2$ , blue  $PM_{10}$  and red  $PM_{2.5}$ . Periods shown are full time series, JJA, DJF, daily maximum and daily minimum.

MQI was below 1 for all pollutants except for  $PM_{10}$  (Figure 29). Thus, were MQO:s fulfilled for all except  $PM_{10}$ . This means that the ensemble forecast fulfilled the objectives set by FAIRMODE. The MQI scores could have been more exact, had there been constants available for the Stockholm region to estimate measurement uncertainty with. The statistical indicators should be considered with some caution since only the ozone data set had a normal distribution. However, the distribution of the forecast data sets were similar to the distribution of each corresponding observation data set. This indicates that the offset from a form a normal distribution is similar in both forecast and observations for each pollutant. The statistical indicators provide information on how well the forecast captures variation occurring in observations and how large discrepancies between the two were. Therefore, the statistical indicators still provide useful information despite the fact the assumed normal distribution was not found in all data sets. If the distribution had differed much between forecast and observation for each pollutant, the statistical indicators would not have been useful. The fact that the forecast in fact does produce a distribution similar to that of the observation is a positive result regarding forecast quality. The forecast, can however not, as previously discussed reproduce high and low extreme values seen in the observation.

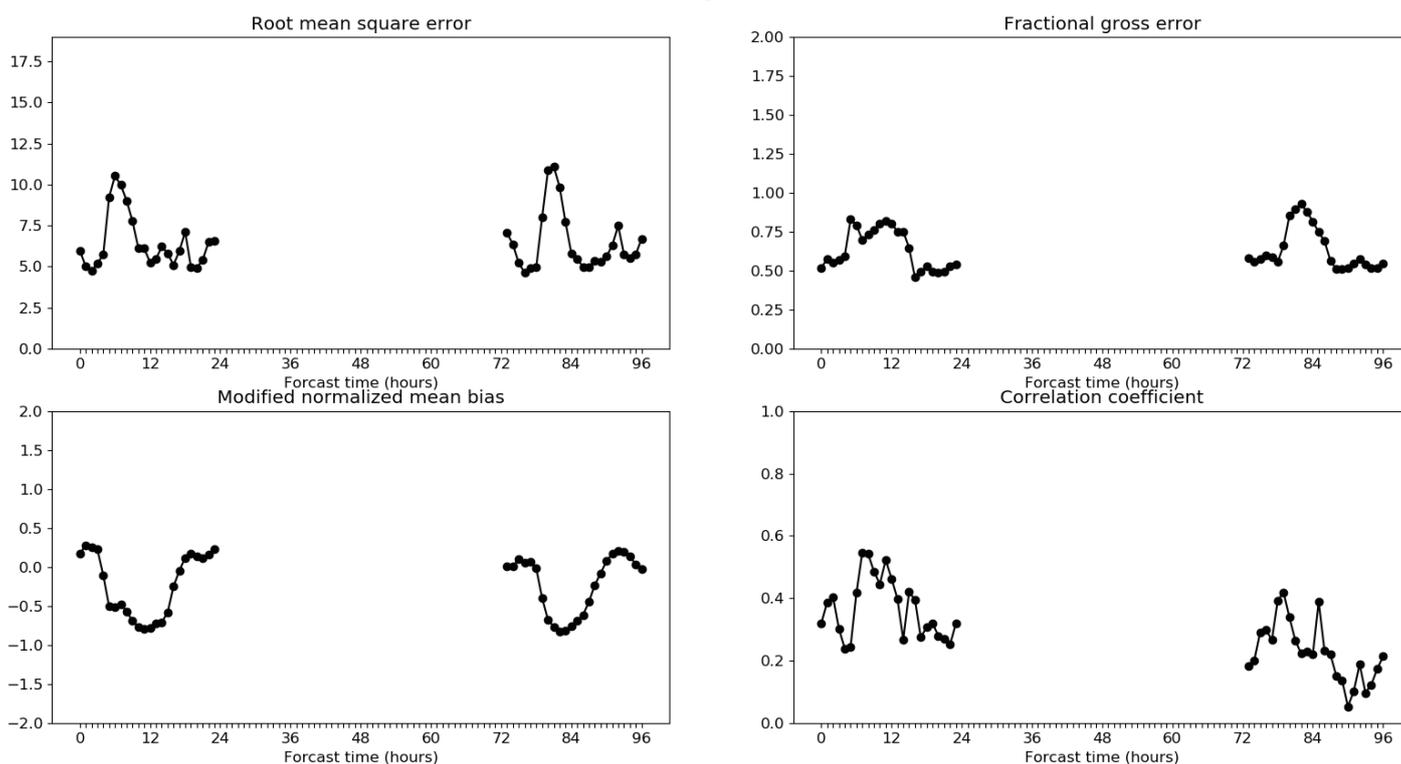


**Figure 29:** MQI for the pollutants at each sites at various time periods. Y-axis displays dimensionless indicator score and x-axis each pollutant and time period. Green colors represent O<sub>3</sub>, yellow NO<sub>2</sub>, blue PM<sub>10</sub> and red PM<sub>2.5</sub>. Periods shown are full time series, JJA, DJF, daily maximum and daily minimum. The red horizontal line indicates the maximum value for which MQO is fulfilled.

### 3.4 Forecast accuracy as a function of forecast time

Forecast accuracy decreased over time approximately of equal magnitude for all pollutants. Therefore, only one example graph is shown in this report (Figure 30). The accuracy decrease over time, simply because model uncertainties grow with forecast time leading to a large spread in individual models and therefore in the ensemble forecast output. This does, as previously mentioned, lead increased inaccurate in the forecast. This was seen for all pollutants as R decreases and MNMB, FGE and RMSE slightly increased comparing the first 24 hours with that last of the forecasts. Thus, indicating that the linear relationship between forecast and observation was weakened and that the errors were increased. R was most affected by forecast time. This indicates that the forecast time was sufficiently short as to not result in a spread in the models which decreased the quality of the forecast severely.

NO<sub>2</sub> statistical indicators, hourly mean JJA  
Torkel Knutssongatan 2016 - 2017



**Figure 30:** Statistical indicators RMSE (top left), FGE (top right), MNMB (bottom left), R (bottom right) of NO<sub>2</sub> urban background site for day 1 and day 4 of the forecast, during JJA. Y-axis displays dimensionless indicator score and x-axis time with hourly resolution.

Forecast time does not seem to be a major source of uncertainty in the ensemble forecast. Instead, errors due to for example local emissions or limitations in the model parameterization were of the same magnitude in the last 24 hours of the forecast as in the first 24. For example, the forecast accuracy of NO<sub>2</sub> as was reduced at the urban background site, during daytime when local emissions from traffic strongly influence NO<sub>2</sub> concentrations at the site (Figure 30). Another example, the forecast's inability to predict low ozone connections during night time at the regional background sites resulted in an error of equal magnitude both day one of the forecast and day four (Figure not shown).

### 3.5 Health risks

Initiatives like CAMS are key in order to increase understanding of the current and future air quality. CAMS encourages organizations and companies to utilize the data collected by CAMS in various ways which they feel useful. The regional ensemble forecast could be used as background concentrations in an information system which forecasts air quality in the Stockholm region in the days to come. Such an information system could be of great value since poor air quality contributes to health issues in Stockholm. The inhabitants of the city could utilize the information regarding the current air quality when planning daily activities and if the forecasts are accurate and recommendations are followed harmful exposure to air pollution could be avoided. If such an information system were to be implemented, it is important to convey obtained information to the public in a manner which is clear and unambiguous. If the ensemble forecast is to be used to inform the the public its uncertainties must be stated as not to provide a false sense of safety to people using it.

According to Johansson *et al.* (2013), pollen and PM<sub>10</sub> are most important with regards to the health risk index since the two are the reasons for largest increase in hospital admissions when concentration of them are high. CAMS ensemble forecasts of birch pollen and PM<sub>10</sub> were the least accurate as previously

discussed. This must be kept in mind if the forecasts are to be used to inform the public about the air quality situation in the days to come. It is however important to note that this health risk index is based on emergency visits regarding respiratory issues. A health risk index could instead be based on for example premature deaths. Birch pollen would most likely not be of importance in such a health risk index. Out of all pollutants regarded in this study all but ozone were generally underestimated by the forecast. This would be problematic, as it would make the air quality seem better than what it is. This could put people, with respiratory disease in particular, at unnecessary risk. If the forecasts would be used as a public information system these flaws would have to be taken into account and compensated for.

## 4 Summary and Conclusions

This study evaluated and validated CAMS regional ensemble forecast of four pollutants and birch pollen in the Stockholm region. Forecast data was obtained from CAMS database and observation data from Stockholm municipality database for three sites in the Stockholm region. The data was analyzed using a number of statistical indicators. Forecast accuracy over the year was evaluated by comparing different seasons. Forecast accuracy depending on variations from local emissions was examined by comparing urban and regional sites. CAMS regional ensemble forecast could be an important tool used to inform the general public about the current and coming air quality situation in the Stockholm region.

The accuracy of the ensemble forecast of  $\text{NO}_2$  was higher at the regional background site than at the urban.  $\text{NO}_2$  concentrations are strongly affected by local emissions at the urban site, but not at the regional. Thus, the difference in accuracy was probably due to local emissions during traffic rush hours at the urban background site, not fully resolved by the forecast. The accuracy of  $\text{NO}_2$  forecasts did only differ depending on season at the urban background site, capturing the general behavior of diurnal cycles more accurately during winter compared to summer. However, the forecasted magnitude of concentrations were less accurate during winter. This could be due to that local emissions signatures becomes clearer during winter, when the relative importance of local emissions is higher.

The ensemble forecast accurately predicted the general behavior of the annual cycle of ozone. Furthermore, forecasted high concentrations of ozone corresponded well to the observations. Forecast accuracy did not vary much depending on urban or regional location. This could be due to that the residence time of ozone results in well-mixed concentrations of ozone. The ozone forecast was more accurate during winter compared to summer. This could be due to that photochemical reactions are less active during winter, reducing the importance of chemical processes in the individual models and resulting in more accurate forecasts. The main shortcoming in the forecast was prediction of low ozone concentrations, particularly at the regional background sites, which were overestimated by the forecast with approximately  $20 - 40 \mu\text{g m}^{-3}$ . This result could be due to the fact that the horizontal and vertical grid is coarse when compared to the point observation data. This could lead to that local meteorological situations are not resolved by the forecast. The results indicated that meteorological conditions during summer night are least accurately predicted by the forecast. Furthermore, the forecast models land surface cover could also contribute to this. Results indicate that the land surface cover is not represented in a satisfactory manner in the forecast models.

For  $\text{PM}_{2.5}$ , the ensemble forecast was inaccurate without any clear seasonal or location related patterns visible explanation. Inaccuracies could possibly be related to factors such as too effective wet deposition, too effective mixing due to the forecast grid size and uncertainties in the representation of secondary aerosol formation. The horizontal and vertical grid size did probably also contribute to forecast inaccuracy of  $\text{PM}_{2.5}$ .

For coarse mode particles ( $\text{PM}_{10} - \text{PM}_{2.5}$ ), forecast accuracy of the ensemble forecast was higher at the regional background site compared to the urban. This was probably due to that these particles depend largely on local emissions due to short residence time, leading to local deposition. The concentrations were however systematically underestimated by the forecast at both sites during JJA and DJF. This could be due to an overestimation of dilution and the fact that the coarse grid size and outdated input emissions data.

The forecast accuracy of birch pollen was affected by shortcomings in the birch pollen forecast construction. The lack of quasi-biannual cycles in the forecast was probably the main source of inaccuracy, which has also been seen in a previous study (Sofiev *et al.*, 2015). However, the forecast predicted pollen level within one category level of the observed level in 68 % of the dates. The start of the pollen season was well predicted within a few dates. Thus, the ensemble forecast does give an indication of the pollen concentration, but it cannot be considered to be more than an indicator in the present form.

The ensemble forecast passed MQO for all pollutants, except for PM<sub>10</sub>. Note however that statistical indicators must be reviewed with some caution since only ozone displayed a normal distribution. However, statistical indicators suggest how the forecast vary in relation to observations. Thus the statistical indicators were useful despite of their limitations. PM<sub>10</sub> and birch pollen concentrations are important with regards to the health risk index and the fact that the forecast does not fulfill the MQO for PM<sub>10</sub> and that birch pollen forecasts are only good as an indicator must be taken into account if the ensemble forecast is to be used in an information system for the general public. For example, a forecast model with higher resolution could be used as a supplement to the ensemble in order to better predict urban concentrations of PM<sub>10</sub> and uncertainties with the pollen forecast must be clearly stated if it is to be released to the public.

Generally, it was found that local emissions contributed most to inaccuracy in the results for NO<sub>2</sub> and coarse mode particles. Inaccuracy due to local emissions was also found for ozone. However, some variations due to local emissions seemed to be better represented by the forecast than others. For example, daytime peaks of NO<sub>2</sub> during winter were better captured by the model than corresponding termination of ozone production by NO<sub>x</sub> concentrations higher than 100 ppt(v). The difference in accuracy was likely due to limitations in the representation of chemical processes within the models computing termination of ozone production, resulting in larger uncertainties than for NO<sub>2</sub> represented by direct emissions. Annual large scale trends were captured by the forecast but extreme values were not.

The ensemble forecast construction has its limitations. Each of the seven individual models must be post-processed, adding uncertainty to the forecast output. The harmonization also reduces accuracy of individual features in the models. For example, the vertical resolution is not the same in the individual models. Thus, the accuracy of the concentration at surface level of each model will vary, and the ensemble forecast is of those a median. However, as previously shown by CAMS (Meteo-France, 2017) the accuracy of processes in each individual model vary depending on e.g. time of year, location and pollutant. Thereby the resulting ensemble is overall more accurate, since neither of the individual models constantly performs better than the others in a processes. It is important to be aware of the limitations following the forecast construction since they can explain errors in the forecast output. If CAMS data is to be used to describe e.g. the background air in the Stockholm region, it is important to be aware of the forecast limitations described in this study. Improvements of the forecast could be made in a number of areas, e.g. surface coverage, traffic input data and vertical profiles. Particularly forecast of ozone and PM<sub>2.5</sub> could probably be improved considerably if model surface conditions were more accurately represented. Input data of traffic flow and local emission sources could be updated in order to improved model accuracy, particularly in the urban region. Implementing a function in the individual models simulating how pollutant concentrations vary with height and not only one value for the concentration of each pollutant in the lowest vertical level could lead to that individual models capture how the pollutants vary with height more accurately. The decrease in forecast accuracy with time does not seem to be a major source of uncertainty. Results indicate that 96 hours is sufficiently short as to not create a large spread in the individual models. Rather it is features such as local emissions, local meteorological variations, the construction of the ensemble forecast and chemical parameterizations which contribute most to forecast inaccuracy.

This study gives an initial overview of the regional ensemble forecast accuracy in the Stockholm region. The data collection period is sufficiently long as to be able to detect seasonal and diurnal cycles but limited to variations which are specific to a certain year. A shorter period, covering less than a year might not have been enough. A longer data period would have been interesting to analyze. Then there would have been a possibility to detect improvements over time in the forecast. Data was however not available, as previously mentioned. An alternative approach could have been to only consider shorter time periods and to examine whether or not the forecast captures particular events. This could be an interesting next step in the evaluation and validation of the regional ensemble forecast. Further studies

can be conducted in order to obtain a more detailed understanding of the forecast. For example, it would be interesting to examine the effect if the forecast are more accurate when compared to observations of the year 2011, which match the year of input emission data. However, there are no forecasts available from 2011 to compare to observations. A comparison of the forecasts included in this study to observations from 2011 not lead to meaningful results since it would be difficult to distinguish between the effects from emissions and from difference in meteorology between forecast and observation. Furthermore, it would be interesting to examine how the forecast captures extreme events of the pollutants and also to examine meteorological parameters such as precipitation and stability affect the forecast in more detail.

It is important to be aware of the limitations in the CAMS ensemble forecast if it is to be used as background concentrations of air pollution in an information system to the public. Some improvements might be necessary, particularly for PM and birch pollen. Furthermore, local emissions would need to be better modeled or compensated for. This could be achieved with an updated emission data set. The forecast is sufficiently good at capturing the general behavior of the air pollutants in the Stockholm region, but not in the city center where the influence of local emissions is larger. However, provided that the forecast limitations are taken into consideration, and that a forecast model with higher resolution is used as a supplement in the urban region, the ensemble forecast could be used as basis for background emission in general public information regarding the air quality situation in the Stockholm region.

## Acknowledgements

I want to thank my supervisors Christer Johansson and Caroline Leck who supported me through the project with relevant feedback and good insights. I also want to thank Magnuz Engard who also supported me in the project. Furthermore, I want to thank SLB-analys for letting me do my work in their office. Thanks also to the palynological lab at the Natural history museum for sharing pollen data to this study. I also want to thank Jörg Gumbel for his advice on how to write a report and on how to make a presentation. Lastly, I want to thank Rasmus Gustafsson, Malin Boman and Stina Karlsson.

This thesis contributes to the ERA-PLANET ([www.era-planet.eu](http://www.era-planet.eu)) trans-national project SMURBS ([www.smusbs.eu](http://www.smusbs.eu)) (Grant agreement n. 689443) under the EU horizon 2020 framework program.

## References

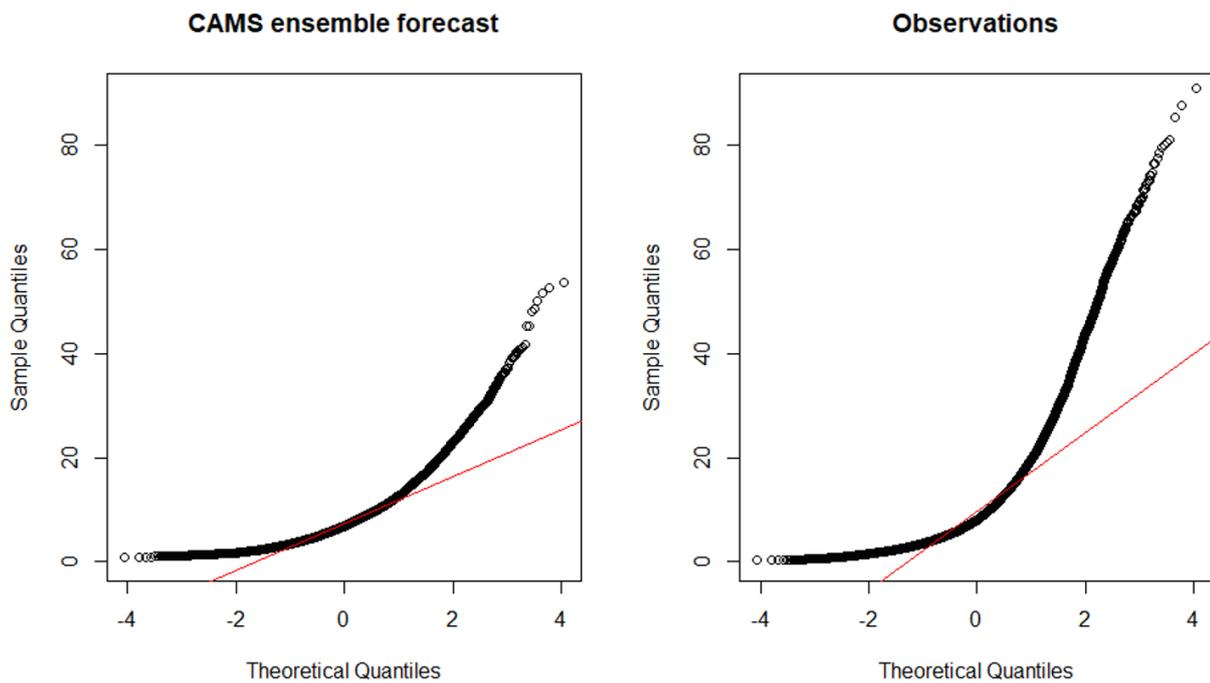
- ACES. 2017. SMURBS, project info. Retrieved 2018-10-05. Available at: <https://www.aces.su.se/research/projects/smurbs/>
- Alpfjord Wylde, H. 2018. Email source, Reference laboratory for urban air modeling, SMHI
- Areskoug, H. 2018. Email source, Department of Environmental Science and Analytical Chemistry, Stockholm University.
- Areskoug, H., Johansson, C., Alesand, T., Hedberg, E., Ekengrena, T., Vesely, V., Wideqvist, U., Hansson, H.-C. 2004. Concentrations and sources of PM10 and PM2.5 in Sweden INSTITUTE OF APPLIED ENVIRONMENTAL RESEARCH
- CAMS. 2017. VERIFICATION PLOTS: DOCUMENTATION. Retrieved 2018-09-05. Available at: [https://www.regional.atmosphere.copernicus.eu/doc/USER\\_GUIDE\\_VERIFICATIO\\_STATISTICS.pdf](https://www.regional.atmosphere.copernicus.eu/doc/USER_GUIDE_VERIFICATIO_STATISTICS.pdf)
- Carlsaw D.C., Ropkins K. 2012. openair – An R package for air quality data analysis. Environmental Modelling & Software. Environmental Research Group, King's College London. [Software]. Retrieved 2018-10-02. Available at: <http://davidcarlsaw.github.io/openair/>
- Copernicus. 2016. Atmosphere Monitoring Service. Retrieved 2018-12-04. Available at: [https://www.copernicus.eu/sites/default/files/documents/Copernicus\\_AtmosphereMonitoring\\_Feb2017.pdf](https://www.copernicus.eu/sites/default/files/documents/Copernicus_AtmosphereMonitoring_Feb2017.pdf)
- Ekeholm, A. 2018. Email source, Palynological laboratory, Sweden Natural History Museum.

- Gustafsson, M., Lindèn, J., Tang, L., Forsberg, B., Orru, H., Åström, S. and Sjöberg, K. 2018. Quantification of population exposure to NO<sub>2</sub>, PM<sub>2.5</sub> and PM<sub>10</sub> and estimated health impacts
- Janssen, S., Wesseling, J., Viaene, P., Georgieva, E., Thunis, P. and Cuvelier, K. 2018. Guidance Document on Modelling Quality Objectives and Benchmarking. Retrieved 2018-09-05. Available at: [http://fairmode.jrc.ec.europa.eu/document/fairmode/WG1/Guidance\\_MQO\\_Bench\\_vs2.2.pdf](http://fairmode.jrc.ec.europa.eu/document/fairmode/WG1/Guidance_MQO_Bench_vs2.2.pdf)
- Johansson, C., Törnqvist, L., Forsberg, B., Meister, K., Åström, C., Robertsson, L., Engardt, M., Alm Kübler, K., Ekeboom, A., Östensson, P. 2013. Kartbaserad prognosinformation till astmatiker och andra känsliga grupper om luftkvaliteten. *SLB-analys*
- Johansson, C. 1993. Luftföroreningar på Svenska Högarna och Landsort
- Jönsson, O., Andersson, C., Forsberg, B., and Johansson, C. 2013. Air pollution episodes in Stockholm regional background air due to sources in Europe and their effects on human population *Boreal environment research*, **Volume 18**, 280–302.
- Latvala, S. 2018a. Luftföroreningar och dess effekter Retrieved 2018-12-03. Available at: <https://www.naturvardsverket.se/Sa-mar-miljon/Klimat-och-luft/Luftfororeningar/>
- Latvala, S. 2018b. Fakta om partiklar i luft Retrieved 2018-12-03. Available at: <https://www.naturvardsverket.se/Sa-mar-miljon/Klimat-och-luft/Luftfororeningar/Partiklar/>
- Marècal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chèroux, F., Colette, A., Coman, A., Curier, R. L., Denier van der Gon, H. A. C., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillè, E., Josse, B., Kadyrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, J., van Velthoven, P., van Versendaal, R., Vira, J. and Ung, A. 2015. A regional air quality forecasting system over Europe: the MACC-II daily ensemble production. *Geoscientific Model Development*, **Volume 8**, issue 9, 2777–2813.
- Meteo-France for Copernicus. 2017. Regional Production, Description of the operational models and of the ENSEMBLE system. Retrieved 2018-11-20. Available at: [https://atmosphere.copernicus.eu/sites/default/files/2018-02/CAMS50\\_factsheet\\_201610\\_v2.pdf](https://atmosphere.copernicus.eu/sites/default/files/2018-02/CAMS50_factsheet_201610_v2.pdf)
- Naturhistoriska riksmuseet. 2017. Sägör vi prognoserna. Retrieved 2018-10-20. Available at: <http://pollenrapporten.se/ompollen/sagorviprognoserna.4.314e02dd13d69872ec097.html>
- Naturhistoriska riksmuseet. 2016. Halter Retrieved 2018-09-20. Available at: <http://pollenrapporten.se/ompollen/halter.4.314e02dd13d69872ec079.html>
- Norman, M. and Lövenheim, B. 2016. Kvalitetssäkringsprogram för mätningar och beräkningar av luftföroreningar. SLB
- Olstrup, H., Johansson, C., Forsberg, B., Tornevi A., Ekeboom, A. and Meister, K. 2019. Article A Multi Pollutant Air Quality Health Index (AQHI) Based on Short-Term Respiratory Effects in Stockholm, Sweden *International Journal of Environmental Research and Public Health*, **Volume 2019** 16, 105.
- Pleijel, H. 1999. Marknära ozon - ett hot mot växterna. Naturvårdsverkets förlag
- Python Software Foundation. 2018. Python version 3.6. Beaverton, USA. [Software]. Retrieved 2018-09-03. Available at: <https://www.python.org/>
- R development core team. 2017. The R Project for Statistical Computing . R: A language environment for statistical comp. R Foundation for statistical computing, Vienna, Austria. [Software]. Retrieved 2018-10-02. Available at: <https://www.r-project.org/>
- Robertsson, L. Personal communication. SMHI
- RStudio team. 2015. Rstudio: Integrated Development for R. RStudio, Inc., Boston, MA. [Software]. Retrieved 2018-10-02. Available at: <http://www.rstudio.com>

- Sabelström, H. 2018. Luftövervakning Retrieved 2018-12-03. Available at: <https://www.naturvardsverket.se/Miljoarbete-i-samhallet/Miljoarbete-i-Sverige/Uppdelat-efter-omrade/Luft/Luftovervakning/>
- Sabelström, H. 2017a. Partiklar (PM<sub>2.5</sub>) i regional bakgrund Retrieved 2018-10-25. Available at: <https://www.naturvardsverket.se/Sa-mar-miljon/Statistik-A-O/Partiklar-PM25-halter-i-luft-regional-bakgrund-arsmedelvarlden/>
- Sabelström, H. 2017b. Beskrivning av mätstationstyper för luft. Retrieved 2018-10-25. Available at: <https://www.naturvardsverket.se/Miljoarbete-i-samhallet/Miljoarbete-i-Sverige/Uppdelat-efter-omrade/Luft/Luftovervakning/Beskrivning-av-matstationstyper/>
- SMHI. 2014a. Kvävedioxid. Retrieved 2018-12-03. Available at: <http://www.smhi.se/reflab/om-luftfororeningar/luftfororeningar/ozon-1.19670>
- SMHI. 2014b. Ozon. Retrieved 2018-12-03. Available at: <http://www.smhi.se/reflab/om-luftfororeningar/luftfororeningar/ozon-1.19670>
- SMHI. 2014c. Partiklar. Retrieved 2018-12-03. Available at: <http://www.smhi.se/reflab/om-luftfororeningar/luftfororeningar/partiklar-1.19671>
- Sofiev, M., Ritenberga, O., Albertini, R., Arteta, J., Belmonte, J., Geller Bernstein, C., Bonini, M., Celenk, S., Damialis, A., Douros, J., Elbern, H., Friese, E., Galan, C., Oliver, G., Hrga, I., Kouznetsov, R., Krajsek, K., Magyar, D., Parmentier, J., Plu, M., Prank, M., Robertson, L., Steensen, B. M., Thibaudon, M., Segers, A., Stepanovich, B., Valdebenito A. M., Vira, J., and Vokou, D. 2017 Multi-model ensemble simulations of olive pollen distribution in Europe in 2014: current status and outlook *Atmospheric science and physics*, **Volume 17**, 12341–12360.
- Sofiev, M., Berger, U., Prank, M., Vira, J., Arteta, J., Belmonte, J., Bergmann, K.-C., ChÈroux, F., Elbern, H., Friese, E., Galan, C., Gehrig, R., Khvorostyanov, D., Kranenburg, R., Kumar, U., MarÈcal, V., Meleux, F., Menut, L., Pessi, A.-M., Robertson, L., Ritenberga, O., Rodinkova, V., Saarto, A., Segers, A., Severova, E., Sauliene, I., Siljamo, P., Steensen, B. M., Teinmaa, E., Thibaudon, M. and Peuch, V.-H. 2015. MACC regional multi-model ensemble simulations of birch pollen dispersion in Europe *Atmospheric science and physics*, **Volume 15**, 8115–8130.
- Stockholm stad. 2018. Säsongsnederbörd. Retrieved 2018-12-20. Available at: <http://miljobarometern.stockholm.se/klimat/klimat-och-vaderstatistik/sasongsnederbord/>
- Targino, A.C.; Krecl, P.; Johansson, C.; Swietlicki, E.; Massling, A.; Coraiola, G.C.; Lihavainen, H. 2013. Deterioration of air quality across Sweden due to transboundary agricultural burning emissions. *Boreal Environment Research*, **Volume 18**, 19 – 36.
- Wallace, J. M., Hobbs, P. V. 2006. Atmospheric Science (Second Edition) An Introductory Survey Academic press, USA, 483pp.
- World health organization. 2018a. Ambient air pollution: Health impacts Retrieved 2018-12-03. Available at: <http://www.who.int/airpollution/ambient/health-impacts/en/>
- World health organization. 2018b. Ambient air pollution: Pollutants Retrieved 2018-12-03. Available at: <http://www.who.int/airpollution/ambient/pollutants/en/>
- Wilks, D. S. 2011. Statistical Methods in the atmospheric sciences. Third edition, Elsevier Inc., USA, 676pp.

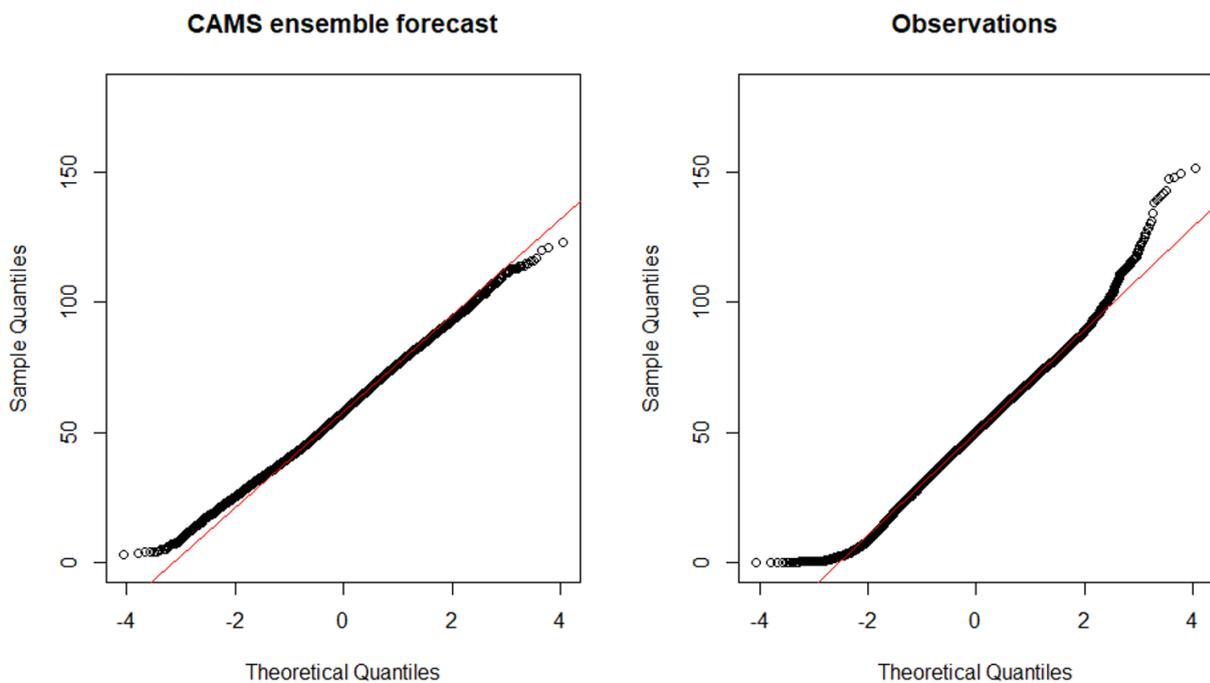
## 5 Appendix I: QQ-plots

### Q-Q plot NO<sub>2</sub>, Torkel Knutssonsgatan



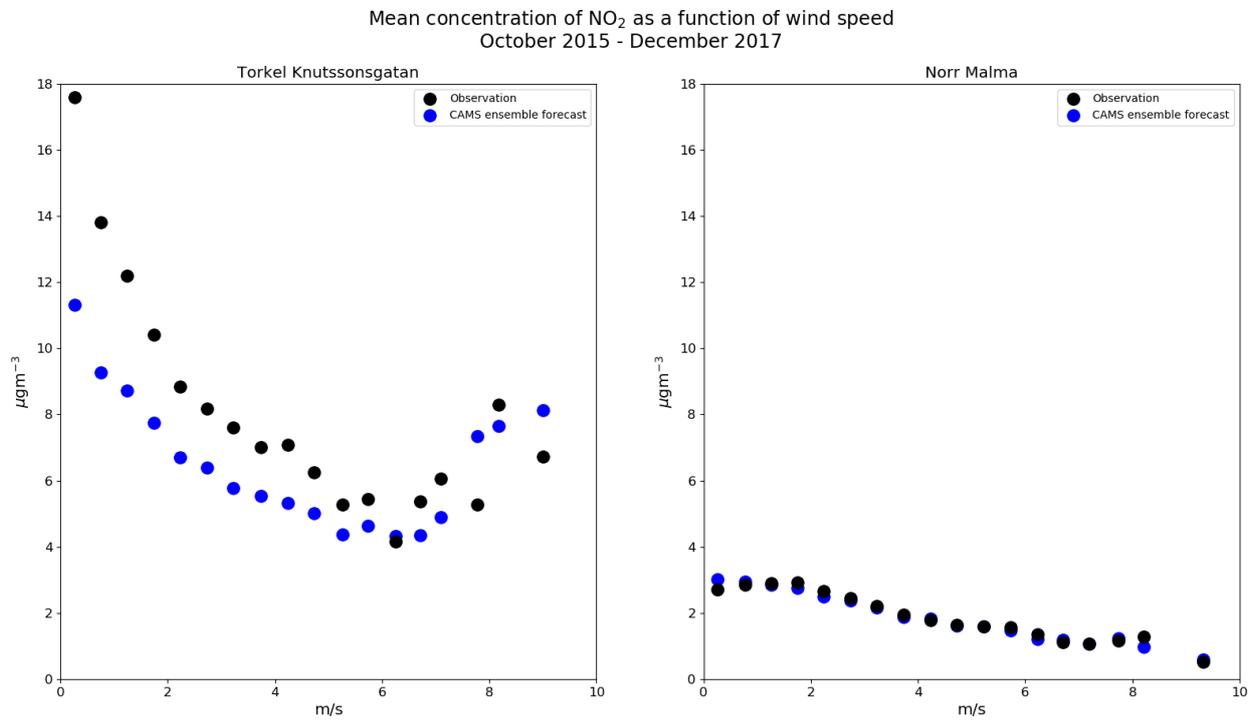
**Figure 31:** QQ-plot of NO<sub>2</sub> at the urban background site forecast (left) and observation (right). The red line represent a theoretical normal distribution. Sample quantiles on the y-axis and theoretical quantiles on the x-axis.

### Q-Q plot O<sub>3</sub>, Torkel Knutssonsgatan

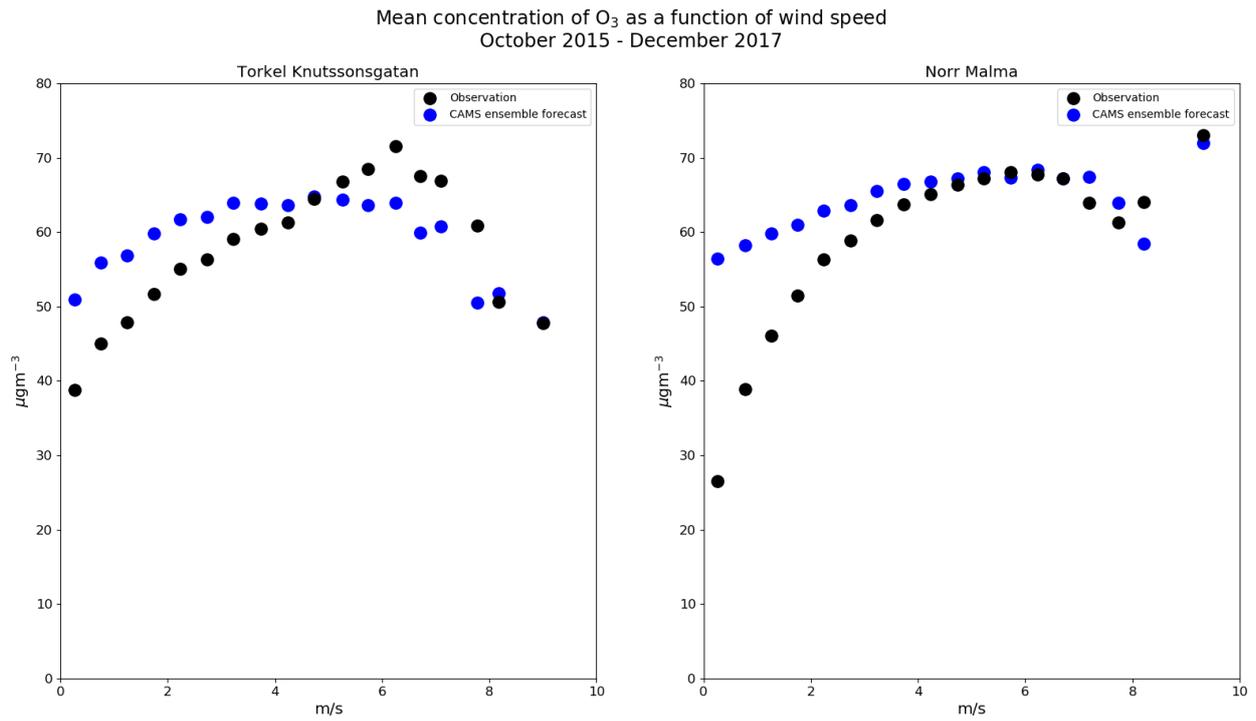


**Figure 32:** QQ-plot of O<sub>3</sub> at the urban background site forecast (left) and observation (right). The red line represent a theoretical normal distribution. Sample quantiles on the y-axis and theoretical quantiles on the x-axis.

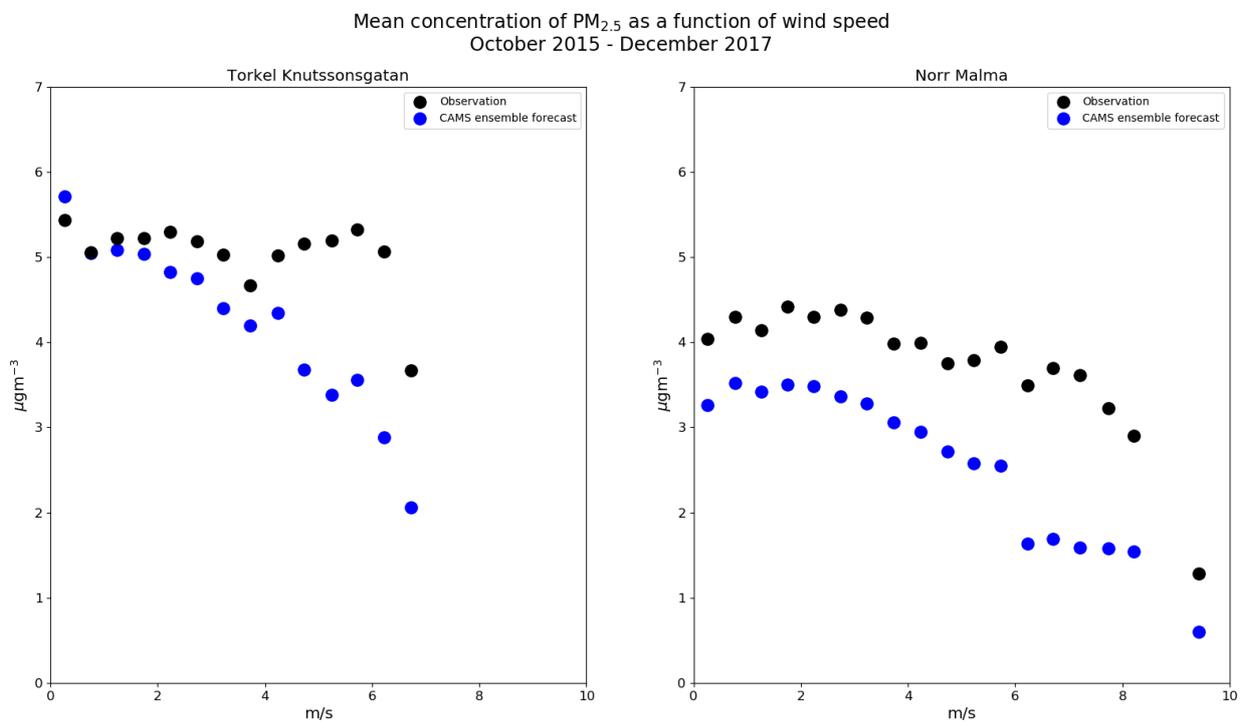
## 6 Appendix II: Pollutant concentration as a function of wind speed



**Figure 33:** Concentration of NO<sub>2</sub> as a function of wind speed. Y-axis displays mean concentration  $\mu\text{m}^{-3}$  and x-axis mean wind speed in m/s. Blue and black dots represent forecast and observations respectively.



**Figure 34:** Concentration of O<sub>3</sub> as a function of wind speed. Y-axis displays mean concentration  $\mu\text{m}^{-3}$  and x-axis mean wind speed in m/s. Blue and black dots represent forecast and observations respectively.



**Figure 35:** Concentration of PM<sub>2.5</sub> as a function of wind speed. Y-axis displays mean concentration  $\mu\text{m}^{-3}$  and x-axis mean wind speed in m/s. Blue and black dots represent forecast and observations respectively.