Improving 3-day deterministic air pollution forecasts using machine learning algorithms

Christer Johansson[1,2], Zhiguo Zhang[3], Magnuz Engardt[2], Massimo Stafoggia[4], Xiaoliang Ma[3]

[1]Department of Environmental Science, Stockholm University, Stockholm, Sweden
[2]Environment and health administration, SLB-analys, Stockholm, Sweden
[3] KTH Royal Institute of Technology, Dept. of Civil and Architectural Engineering, Stockholm, Sweden
[4] Department of Epidemiology, Lazio Region Health Service, Rome, Italy

Correspondence to: Christer Johansson (christer.johansson@aces.su.se)

Abstract. As air pollution is regarded as the single largest environmental health risk in Europe it is important that communication to the public is up-to-date, accurate and provides means to avoid exposure to high air pollution levels. Long- as well as short-term exposure to outdoor air pollution is associated with increased risks of mortality and morbidity. Up-to-date information on present and coming days' air quality help people avoid exposure during episodes with high levels of air pollution. Air quality forecasts can be based on deterministic dispersion modelling, but to be accurate this requires detailed information on future emissions, meteorological conditions and process oriented dispersion modelling. In this paper we apply different machine learning (ML) algorithms – Random forest (RF), Extreme Gradient Boosting (XGB) and Long-Short Term Memory (LSTM) – to improve 1-, 2- and 3-day deterministic forecasts of PM_{10}, NO_x, and O_3 at different sites in Greater Stockholm, Sweden.

It is shown that the deterministic forecasts can be significantly improved using the ML models but that the degree of improvement of the deterministic forecasts depends more on pollutant and site than on what machine learning algorithm is applied. Deterministic forecasts of PM_{10} is improved by the ML models through the input of lagged measurements and Julian day partly reflecting seasonal variations not properly parameterised in the deterministic forecasts. A systematic discrepancy by the deterministic forecasts in the diurnal cycle of NO_x is removed by the ML models considering lagged measurements and calendar data like hour of the day and weekday reflecting the influence of local traffic emissions. For O_3 at the urban background site the local photochemistry is not properly accounted for by the relatively coarse Copernicus Atmosphere Monitoring Service ensemble model (CAMS) used here for forecasting O_3, but is compensated for using the ML models by taking lagged measurements into account. The machine learning models performed similarly well for the sites and pollutants. Performance measures like Pearson correlation, root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE), typically differed less than 30% between ML models. At the urban background site, the deviations between modelled and measured concentrations (RMSE errors) are smaller than uncertainties in the measurements estimated according to recommendations by the Forum for Air Quality Modeling (FAIRMODE) in the context of the air quality directives. At the street canyon sites modelled errors are higher, and similar to measurement uncertainties. Further work is needed to reduce deviations between model results and measurements for short periods with relatively high concentrations (peaks). Such peaks can be due to a combination of non-typical emissions and unfavourable meteorological conditions and may be difficult to forecast. We have also shown that deterministic forecasts of NO_x at street canyon sites can be improved using ML models even if they are trained at other sites. For PM_{10} this was only possible using LSTM.

An important aspect to consider when choosing ML algorithms is the computational requirements for training the models in the deployment of the system. Decision tree-based models (RF and XGB) requires less computational resource than the deep learning model. Therefore, a random forest model is now implemented operationally in the forecasts of air pollution and health risks in Stockholm. Development of the tuning process and identification of more efficient predictors may make forecast more accurate.

Key words: Dispersion modelling, random forest, XGboost, LSTM, neural network, PM_{10}, O_3, NO_x, GAM

1 Introduction

According to the World Health Organisation (WHO) air pollution is one of the leading causes of mortality worldwide and is regarded as the single largest environmental health risk (Fuller et al., 2022). Acute effects of air pollution are due to short-term (e.g. daily) exposures that can lead to reduced lung function, respiratory infections and aggravated asthma (Lee et al., 2021). According to the European air quality directive, information on the air quality should be made available to the public. Public information regarding the expected health risks associated with current or the next few days concentrations of pollutants can be very important for sensitive persons when planning their outdoor activities.

There are different approaches to obtain information on the spatio-temporal variation of air pollutant concentrations - from simple statistical models to advanced process-oriented models. Gaussian plume models are widely used in urban areas for estimating impacts on atmospheric concentrations from different emission sources and for health risk assessments (Munir et al., 2020; Johansson et al., 2009; Orru et al., 2015; Johansson et al., 2017). Eulerian chemical transport models that describe emission, transport, mixing, and chemical transformation of trace gases and aerosols such as e.g. CHIMERE, EMEP and MATCH are part of the Copernicus Atmosphere Monitoring Service (CAMS, atmosphere.copernicus.eu/) to predict air pollution over Europe (Horàlek et al., 2019). The uncertainties in the output of the deterministic models include uncertainties in the input, such as emissions, model algorithms and parameterisations. In urban areas detailed knowledge of the emissions is crucial, and there may be important non-linear relationship between the concentration of contaminants and emission. Another method widely used to obtain spatio-temporal estimates of air pollutant concentrations without detailed knowledge of emissions is Land use regression (Hoek et al., 2008).

Application of machine learning models (ML) to predict outdoor air quality is getting more and more popular (Rybarczyk and Zalakeviciute, 2018; Iskandaryan et al., 2020). Studies have used ML to predict both hourly and daily average concentrations of particulate matter (PM) as well as gaseous air pollutants using meteorological and traffic data (e.g. Quadeer et al., 2020; Di et al., 2019; Thongthammachart et al., 2021; Kamińska, 2019; Chuluunsaikhan et al., 2021; Doreswamy et al., 2020; Castelli et al., 2020; Stafoggia et al., 2020; Stafoggia et al., 2019). In addition, a combination of ML, LUR, dispersion modelling, ground-based and satellite measurements have been used to obtain temporally and spatially distributed concentrations (Shtein et al., 2020; Staffogia et al., 2019; Brokamp et al., 2017; Di et al., 2019). Forecasting air pollution concentrations in a longer-term horizon such as a day or several days have been investigated by e g Kleinert et al. (2022) for O_3. Some studies have also combined deterministic models and ML in forecasting air pollution levels of a few hours/days in the future (e g Hong et al., 2022), but mostly for one single pollutant at the time.

In this paper we demonstrate how ML can help improve the accuracy of 1-, 2- and 3-day deterministic forecasts of particulate matter (PM_{10}, particles with an aerodynamic diameter less than 10 µm), nitrogen oxides (NO_x) and ozone (O_3) for urban background and street canyon sites in Stockholm, Sweden. The deterministic forecast utilises the CAMS ensemble model to account for non-local sources (long-range transport). A Gaussian model is applied over the urban area of Stockholm accounting for local emissions and a street canyon model (OSPM) to account for the effect of buildings on the dispersion of local traffic emissions along the roads in the central area of the city. We compare three different machine learning algorithms; two based on decision trees (random forest and XG Boost) and one neural network model (LSTM). Important questions addressed are also if there are systematic differences in performance depending on different pollutants and different sites.

2 Methods

2.1 Air pollution measurements

Input data for ML modelling are taken from four monitoring stations in central Stockholm, including one urban background site (Torkel Knutssonsgatan, hereafter called UB or urban) and 3 street canyon sites (Hornsgatan HO, Folkungagatan FO and Sveavägen SV). They are all located in central Stockholm (Figure 1). Detailed descriptions of measurement methods and sites are provided in Appendix A.

Data from the UB site covers approx. 1000 days (10 April 2019 through 31 December 2021). As the OSPM-model became operational at a later date, the street canyon data extends over 500 days (5 August 2020 through 31 December 2021). All the data was collected at 1-hour intervals, and the details are shown in Table 1.

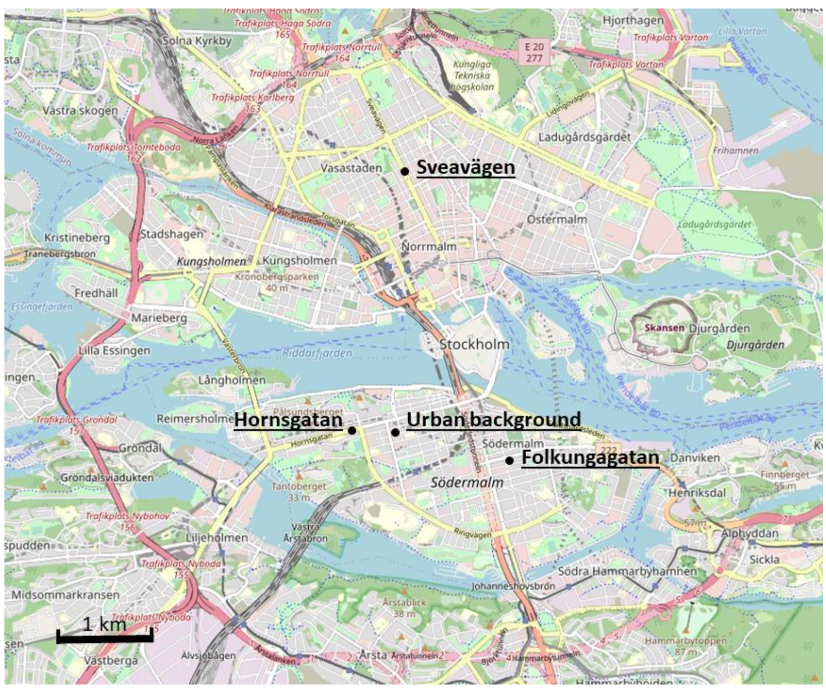


Figure 1. Map of central Stockholm showing locations of the urban background site and the street canyons traffic sites. Base map credits: © OpenStreetMap contributors.

Table 1. Details of the datasets.

Name	Start Time	End Time	Amount (Time interval=1 hour)
urban background	04/10/2019 00:00:00	12/31/2021 23:00:00	23927
Folkungagatan	08/05/2020 00:00:00	12/31/2021 23:00:00	12335
Hornsgatan	08/05/2020 00:00:00	12/31/2021 23:00:00	12335
Sveavägen	08/05/2020 00:00:00	12/31/2021 23:00:00	12335

The measurement data with a missing rate of less than 5% and missing values are replaced with mean values of available data in the neighbourhood according to the respective autocorrelation properties.

2.2 The Stockholm air quality forecast system

Three different dispersion models are used to forecast concentrations considering emissions and dispersion at European, urban and street scale (Figure 2). The CAMS ensemble model, part of the Copernicus program was used to obtain forecasts of long-range transported air pollution from outside of the Greater Stockholm model domain. Previous assessments have found the ensemble model to be the more accurate than any individual model part of CAMS (Meteo-France, 2017; Marècal et al., 2015). CAMS regional ensemble forecasts are published once a day and each forecast covers 96 hours (4 days). Forecasted concentrations representative of background air, hour by hour, are extracted from a location outside the greater Stockholm domain. All regional models constituting the CAMS ensemble includes physical and chemical schemes dealing with gas phase chemistry, heterogeneous chemistry, aerosol size distribution, aqueous phase chemistry, dry deposition, sedimentation, mineral dust, sea salt, wet deposition, etc. An evaluation of the CAMS regional ensemble forecast in Stockholm has been performed by Säll (2018).

The contributions to concentrations due to local emissions in the metropolitan area were performed on a 100 m resolution using a Gaussian dispersion model part of the Airviro system (https://www.airviro.com/airviro/). In this modelling domain (Greater Stockholm, 35 by 35 km) individual buildings and street canyons are not resolved but treated using a roughness parameter (Gidhagen et al., 2005). The Gaussian model is fed with meteorological forecasts from the Swedish Meteorological and Hydrological Institute (SMHI). A diagnostic wind model is used to account for influences of variations in topography and land-use on the dispersion parameters input to the Gaussian model. For details regarding uncertainties and validation of local modelling see Johansson et al. (2017).

Finally, the Operational Street Pollution Model (OSPM, Berkowicz, 2000), driven by forecasted meteorology from SMHI is applied for the street canyon sites. It has been applied earlier at Hornsgatan in Stockholm in a number of modelling studies (e.g. Krecl et al., 2021; Ottosen et al., 2015). NO_x and PM_{10} are modelled on all scales, whereas O_3 is only forecasted by the CAMS ensemble model.

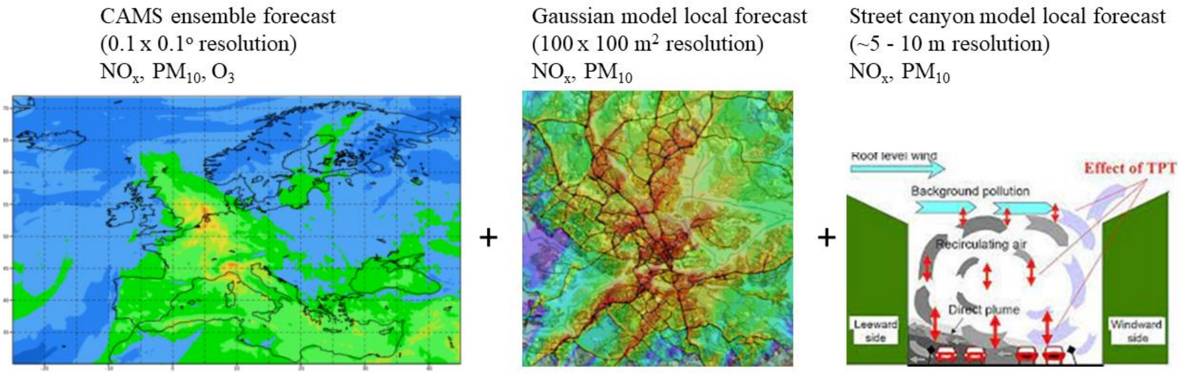


Figure 2. Illustration of the deterministic modelling from European scale at a resolution of 0.1° by 0.1° (ca 11 km × 6 km), via urban scale (100 m resolution over an area of 35 by 35 km) down to the street canyon sites. The CAMS ensemble forecast map example is taken from https://atmosphere.copernicus.eu/ (accessed 1 Feb 2023). The map with the Gaussian model local forecast example is output from the Airviro system (https://www.airviro.com/airviro/, accessed 1 Feb 2023) used in Stockholm. The illustration of a street canyon site is taken from https://www.wikiwand.com/en/Operational_Street_Pollution_Model (accessed 1 Feb 2023).

For the urban scale model domain a detailed emission database is used as input for the local dispersion modelling. The database and its applications and comparisons between modelling and measurements are described in SLB (2022). The total emissions from road traffic are based on emission factors for different vehicle types including passenger cars (diesel, gasoline, gas), buses (diesel, ethanol), light duty trucks <3.5 ton (diesel and gasoline) and heavy duty trucks >3.5 ton (diesel). Exhaust emission factors of NO_x and particles are based on HBEFA version 3.3 (Keller et al., 2017) depending on vehicles Euro class. The emission factors per vehicle category were weighted according to the national Swedish Transport Administration vehicle registry, but the vehicle composition taken from national vehicle registry has been shown to be similar to the local fleet using real world number plate recognition measurements at Hornsgatan in campaigns during 2009 (Burman and Johansson, 2010) and 2017 (Burman et al., 2019). For more details, see also Krecl et al., (2017). Non-exhaust emissions of PM due to wear of brakes, tyres and roads are calculated using the NORTRIP model (Denby et al., 2013) forced by the forecasted meteorology from SMHI. Information on shares of studded winter tyres is obtained from manual counting every week during the winter at different locations in the city centre and along highways outside of the city. Road traffic emissions are calculated for all roads with more than 3000 vehicles per day. Other emission sources included in the local emissions database include shipping, private and municipal heating (including burning of waste). More information about the Stockholm air quality forecast system is provided in Engardt et al. (2021).

2.3 Meteorological forecasts

As an integral part of the Stockholm air quality forecast system, meteorological forecasts for a point in central Stockholm are downloaded every morning from SMHI (https://www.smhi.se/data/oppna-data) and MET Norway (https://docs.api.met.no/doc/). The meteorological forecasts extend over 10 days and are a combination of output from a number of regional and global numerical weather prediction models. The combination is based on statistical adjustments as well as manual edits. The meteorology is initially used to drive the models of weather-dependent PM emissions and the urban- and street canyon air quality modelling. The forecasted meteorological data are, finally, also used as predictors in the ML algorithms as detailed below.

2.4 Machine learning models

As already mentioned in the introduction two decision tree based machine learning models, RF and XGB, and one deep learning model, LSTM are applied. In addition, an ensemble learning approach based on a General Additive Model (GAM), aggregating the above three learning models, is also applied to further optimise the results.

One essential challenge in this study is to forecast hourly concentrations for the coming one day, two days and three days based on historical air pollution measurement and other available information as inputs. This indicates that the essential statistical prediction involves time series prediction for multiple time steps, for example, 72 time steps for three days prediction. It is known that a sequence-to-sequence time series prediction, implemented using LSTM or other recurrent neural networks, provides a straightforward and rolling-over computational schemes. Nevertheless, training a machine learning model with multiple outputs requires much more computational effort, but often leads to inferior prediction accuracy compared to relatively simple models with only a single output dedicated for predicting output of a certain time step. Therefore, this study chooses, instead of more complex machine learning structure, multiple single-output machine learning models for forecasting different air pollutants for *k*=1 day, 2 day and 3 day interval:

$$\hat{\rho}_{i,j}(d,t) = \text{mlearn\_model}\left(\tilde{\rho}_{i,j}(d-k,t), \bar{\rho}^{S}_{i,j}(d-k,t), \check{\rho}_{i,j}(d,t), W(d,t), C(d,t)\ \right)$$

where $\hat{\rho}_{i,j}(d,t)$ is predicted concentration value of the pollutant j for day d and time t at the location i, and $\tilde{\rho}_{i,j}(d,t)$ is the corresponding real measurement; $\bar{\rho}^{S}_{i,j}(d,t)$ uses a set S to represent several statistical measures, including maximum, minimum, 25% quantile and 75% quantile of the measured concentration data during the past 24 hours until $\tilde{\rho}_{i,j}(d,t)$, and the measurement dataset can be represented by a set, i.e. $\{\tilde{\rho}_{i,j}(d,t), \tilde{\rho}_{i,j}(d,t-1), \tilde{\rho}_{i,j}(d,t-2) \ldots .\}$. $\check{\rho}_{i,j}(d,t)$ is the one day predicted concentration value using deterministic physical model. $W(d,t)$ represents the weather condition predicted for day d and time t.

Figure 3 demonstrates the prediction horizon and lagged information horizon for the case of one day prediction. To build consistent statistical machine learning models with a fixed rolling horizon, a new measurement point at current time (d, t) will lead to an additional prediction for one day ahead, i.e. the predicted value at (d+1,t). In the case, the measurement statistics $\bar{\rho}^{S}_{i,j}(d,t)$ will be based on one day preceding measurement data of (d, t), resulting in a lagged rolling horizon described in the figure.

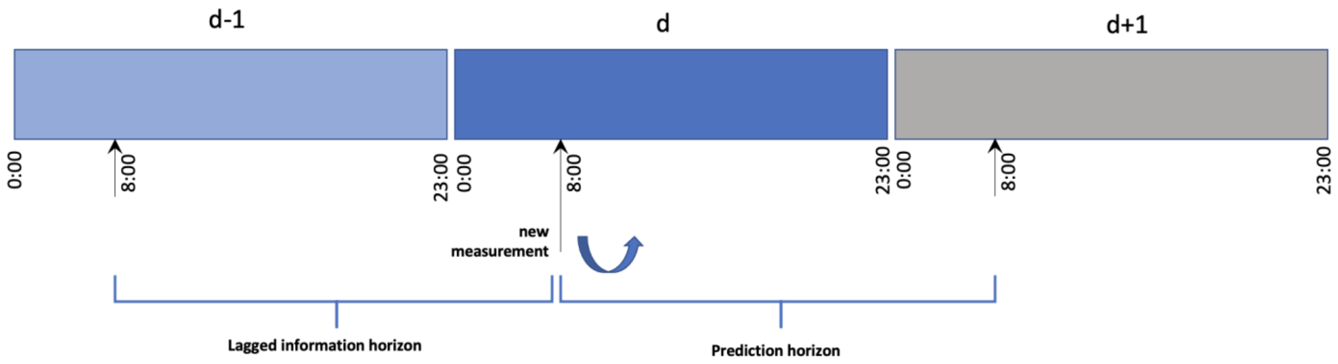


Figure 3. Illustration of the machine learning modelling framework for one-day prediction based on available datasets.

This study has applied both LSTM and two conventional supervised learning models, RF and XGB, as the essential machine learning cores to carry out supervised learning using the same input and output training dataset. In fact, an ensemble approach based on all three models is also applied to predict air quality for different days. The conventional models require nontrivial effort to prepare input feature data as they don't fit as easily with time series data as RNN. To make a fair comparison with both types of models, LSTM model in this case is only based on the same type of input as other two models. It is well known that LSTM can learn the temporal correlation of different ranges. Nevertheless, this study applies the data to a simple LSTM structure, without taking advantages of its full potential. In principle, the measurement data at (d, t) may provide hourly update of predicted values within the prediction horizon i.e. from (d,t+1) to (d+1,t). Nevertheless, it is our future work to extend the model structure and improve prediction using latest real-time information.

In addition to the measured air pollution time series data itself, the forecasted meteorological conditions for the prediction day *d* (or *d+1* or *d+2*) and calendar information such as weekday, hour etc. are also applied as input features. Moreover, the air pollutant concentrations predicted by the deterministic models is also used as inputs to the ML models. Figure 4 summarizes the methodological framework of machine learning and associated computational experiments for air pollution prediction.

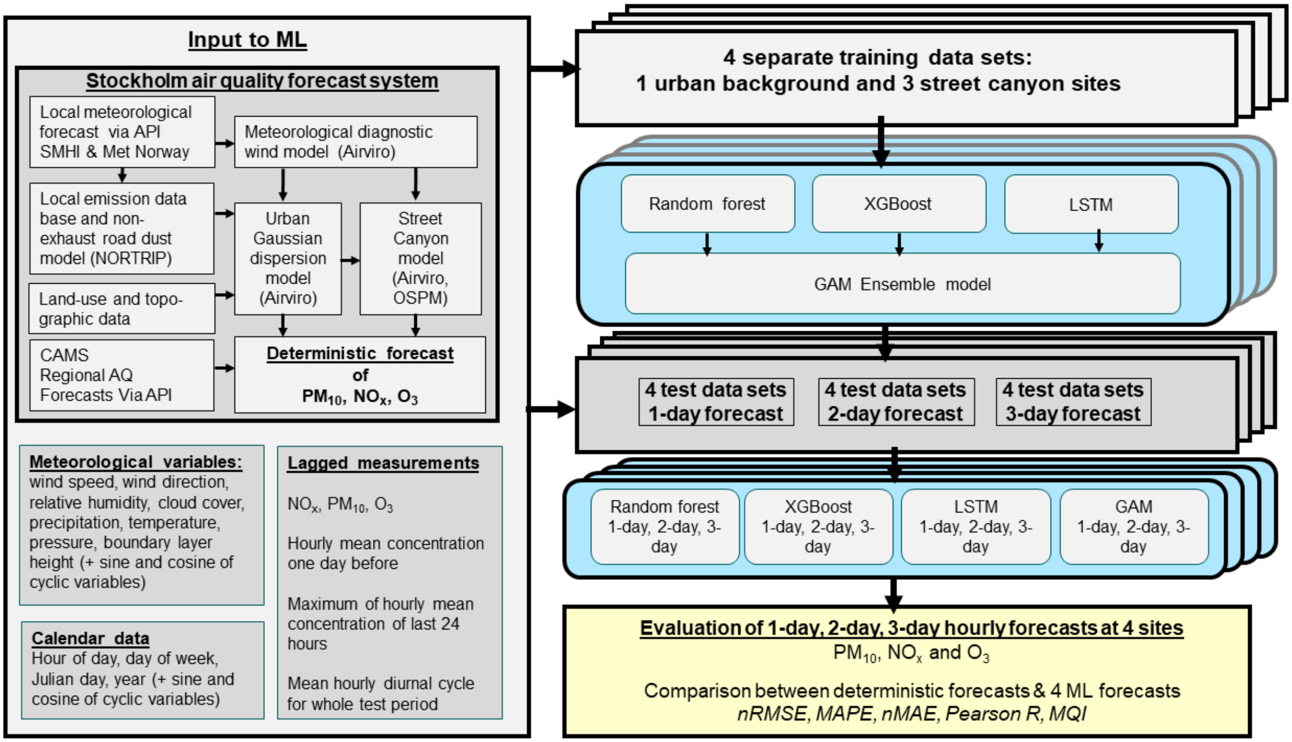


Figure 4. Illustration summarising input data for modelling 1-, 2- and 3-day forecasts of PM_{10}, NO_x and O_3 using the 4 models.

The input includes the deterministic forecasts of PM_{10}, NO_x and O_3, to evaluate how much the deterministic forecasts can be improved by the ML algorithms. In the computational experiments, data-driven forecasting models are trained for one urban background site and three street canyon sites separately, and different machine learning models are trained and tested separately for predicting various air pollution concentrations coming 1-day (0 – 24 h), 2-day (25 – 48 h) and 3-day (48 – 72 h) periods. The dataset is split along the time axis into non-overlapping training, validation, and test data in a ratio of 16:4:5.

Due to the temporal correlation of the air pollutant concentrations, the principal assumption of cross-validation is not satisfied. To preserve the time-dependent property, "*TimeSeriesSplit*" was chosen as the cross-validation strategy. In the k_{th} split, it turns the first k folds as the training set, and the $(k+1)_{th}$ fold as the test set. The value of parameter k is set as 5.

Table 2 presents detailed explanation of the essential input features that are applied in the computational experiments. All machine learning models are implemented in *python* using existing machine learning libraries including "*scikit-learn*" and "*tensorflow*" (also implemented using "*pytorch*") for conventional machine learning models and deep learnings models respectively. The detailed implementation can be referred to the code provided.

Table 2. Measured and forecasted air pollutant concentrations used as input data (features) in the ML modelling of pollutant concentrations at the urban background site (UB) and at the street canyon sites (SC). NO_x and PM_{10} are modelled at both UB and SC. Ozone is only modelled at UB. For periodic input data, using sine and cosine values can remove discontinuities and create consistent distance measures, thereby improving model accuracy.

Category	**Short names**	**Description**
Deterministic features	NO_x**_nday_local** PM_{10}**_nday_local** **n=1, 2, 3**	Deterministic 1-day, 2-day and 3-day forecast of contributions from local emissions based on urban scale Gaussian modelling
	NO_x**_nday_regional** PM_{10}**_nday_regional** O_3**_nd_regional** **n=1, 2, 3**	Deterministic 1-day, 2-day and 3-day forecast of contributions based from non-local emissions based on CAMS ensemble model (regional background)
Autocorrelation features	NO_x**_lagXX** PM_{10}**_lagXX** O_3**_lagXX** **XX = 24, 48, 72**	XX hour lagged air pollutant concentrations based on autocorrelation and prediction time span.
Statistical features	NO_x**_Sta_dXX** PM_{10}**_Sta_dXX** O_3**_Sta_dXX** **Sta=avg., median, min, max, Q1, Q3** **XX = 24, 48, 72**	Average, median, minimum, maximum, quantiles 1 and quantiles 3 of lagged air pollutant concentrations in rolling XX hour periods.
Time features	**Time** **Time**_sin **Time**_cos **Time= year, julianday, month, weekday, day, hour**	Julian day of the year (1, 2, 3, … 365), sine and cosine of 2*pi*day/365. Day of the week (1, 2, 3, … 7), sine and cosine of 2*pi*day/7. Hour of the day (0, 1, 2, … 23), sine and cosine of 2*pi*hour/24. Year Month Day
Meteorological features	wind_direction wind_direction_cos wind_direction_sin	Wind direction[0, 360) at 10 m in central Stockholm, sine and cosine of (2*pi/360)*wind direction
	pressure; temperature; precipitation; cloudiness	Pressure (10 m); Temperature (10 m)
	wind_speed	Wind speed (10 m)
	relative_humidity	Relative humidity
	boundary_layer_height	Boundary layer height for central Stockholm

While hyperparameter optimisation may improve the model performance, the improvement is limited in our test experiment in comparison to the gain over deterministic model. The following configurations are applied for the ML models:

- The two tree-based models use the default parameters of "*scikit-learn*".
- The LSTM model consists of two layers of LSTM, each with 100 neurons, and passed through a fully connected layer before the output. The activation function was a "*tanh*".
- The LSTM model was trained by Adam optimizer. The batch size is set as 72. The initial learning rate is 0.01 and is automatically adjusted using "*ReduceLROnPlateau*" with the parameter patience set to 10, i.e., training is stopped when the loss of the validation set is detected as not decreasing for 10 consecutive epochs.

After the model training process, feature importance is ranked for tree-based models and LSTM models using the mean decrease in impurity (Breiman, 2001) and gradient-based methods (Baehrens et al., 2010), respectively. It should be noted that the gradients in neural networks depend on both input and output data, the feature importance for the LSTM model was computed as the average of feature gradient obtained from all samples in the test set.

2.5 Statistical performance indicators

Several common performance metrics have been selected for comparing the prediction results of different machine learning models including Pearson correlation (r) and normalised error measures: mean average error (MAE), mean absolute percentage error (MAPE) and root mean squared error (RMSE). These measures have also been recommended for air quality model benchmarking in the context of the Air Quality Directive 2008/50/EC (AQD) by Janssen and Thunis (2022).

Mean absolute error:

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

where $\hat{y}_i$ is the predicted value of the i-th sample, and y_i is the corresponding true value for total n samples.

Root Mean Square Error:

$$RMSE(y,\hat{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

MAE and RMSE were normalised by diving by the mean of the measured concentrations, hereafter called nMAE and nRMSE.

Mean absolute percentage error:

$$MAPE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i|}$$

Pearson correlation coefficient:

$$\mathrm{r}(y,\hat{y}) = \frac{\sum_{i=1}^{n} (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}\sqrt{\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}}_i)^2}}$$

The model quality indicator (MQI):

In order to properly assess model quality it is necessary to consider measurement uncertainty. In the FAIRMODE community, the modelling quality indicator (MQI) is used to assess if a model fulfils certain objectives (Janssen and Thunis, 2022). It is defined as the ratio between the model bias at a fixed time (i), quantified by the RMSE, and a quantity proportional to the measurement uncertainty as:

$$MQI(i) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}}{\beta\sqrt{\frac{1}{n}\sum_{i=1}^{n} U(y_i)^2}} = \frac{RMSE}{\beta RMS_U}$$

$U(y_i)$ is the expanded 95th percentile measurement uncertainty and β is a coefficient of proportionality (Janssen and Thunis, 2022). The value of β determines the stringency of the MQI and is set equal to 2, allowing thus deviation between modelled and measured concentrations as twice the measurement uncertainty. The uncertainty of the measurements (RMS_U) was calculated for the mean of the measurement concentrations as:

$$U(y_i) = U_r(RV)\sqrt{(1-\propto^2)^2(y_i^{\,2}) + \propto^2 RV^2}$$

Here $U_r(RV)$ and $\propto$ are parameters that depend on pollutant and RV is a reference value, here taken to be 200, 50 and 120 µg m^{-3}, corresponding $U_r(RV)$ was 0.24, 0.28 and 0.18 and $\propto$ was 0.25, 0.20, 0.79 for NO_2, PM_{10} and O_3 respectively (Janssen and Thunis, 2022). In our case we have calculated NO_x, not NO_2, but we used the same settings of the parameters for NO_x as recommended for NO_2. It should be noted that another important source of error when comparing model results with measurements is associated with the spatial representativeness of a measurement station for comparison with the model. This is due to the mismatch between the model grid resolution and the location of the monitoring station. But in this paper we are mainly interested in comparing the results of the deterministic model with the results using the different ML models together with the deterministic model output.

3 Results

The focus of this paper is to compare the deterministic forecasts of NO_x, PM_{10} and O_3 with the forecasts based on the different machine learners which also include the deterministic forecasts as input variables (features). As described above we have made deterministic and ML forecasts for hourly mean concentrations for the coming 72 hours, based on 1-day, 2-day and 3-day meteorological forecasts for one urban background site (NO_x, PM_{10} and O_3) and three street canyon sites (NO_x and PM_{10}). We also compare results separately for the urban background site and the street canyon sites.

3.1 Urban background

3.1.1 Importance of features - urban background

The relative importance of different features depending on model (RF, XGB and LSTM), pollutant (PM_{10}, NO_x, O_3) and forecast period (1-day, 2-day and 3-day) is shown in plots in Appendix B. It should be noted that the local deterministic models (Gauss and OSPM) use the same meteorological data to forecast concentrations, so when the meteorological variables are important features for the ML models, it indicates that the deterministic models don't capture all processes related to those variables. In summary regarding importance of features for urban background:

NO_x**.** Lagged 24-hour mean concentrations, calendar data, wind speed and local deterministic forecasts are among the top-10 most important variables, but it can be noted that the deterministic forecast is not the most important feature for any model. Of the calendar features hour is most important reflecting the importance of regular, diurnal variations in traffic emissions.

PM_{10}**.** The regional deterministic forecast is the most important feature for PM_{10} forecasts, for all models and for all forecast days. Also lagged measurements, both average, minimum and maximum concentrations is important. Of the calendar features the seasonal variation is reflected in the importance of the Julian day. For LSTM also precipitation is important, which likely reflects the dependence of suspension of dust on surface wetness not being captured by the deterministic forecasts.

O_3**.** For O_3 the models shows very similar characteristics when comparing relative importance of different features. The regional deterministic forecasts is the dominant feature for all forecast days. Also lagged measured maximum concentrations is of some importance. The relative humidity is important, likely reflecting that O_3 concentrations are typically higher during dry, clear sky conditions, which may not be completely captured by the deterministic forecasts.

3.1.2 Comparison between deterministic forecasts and ML models - urban background

Figure 5 shows an example of the temporal variations in September 2021 in the forecasts with deterministic modelling and GAM in comparison to the observations. Similar plots are also given for individual models in Figure C1. The plots were made using the Openair R package (Carslaw and Ropkins, 2012). For all pollutants the ML models tend to improve the variability in the observed concentrations compared to the deterministic forecasts, but there are significant deviations. For O_3 the minimum concentrations observed is often not forecasted so well and for PM_{10} the highest concentrations is not captured by the models.

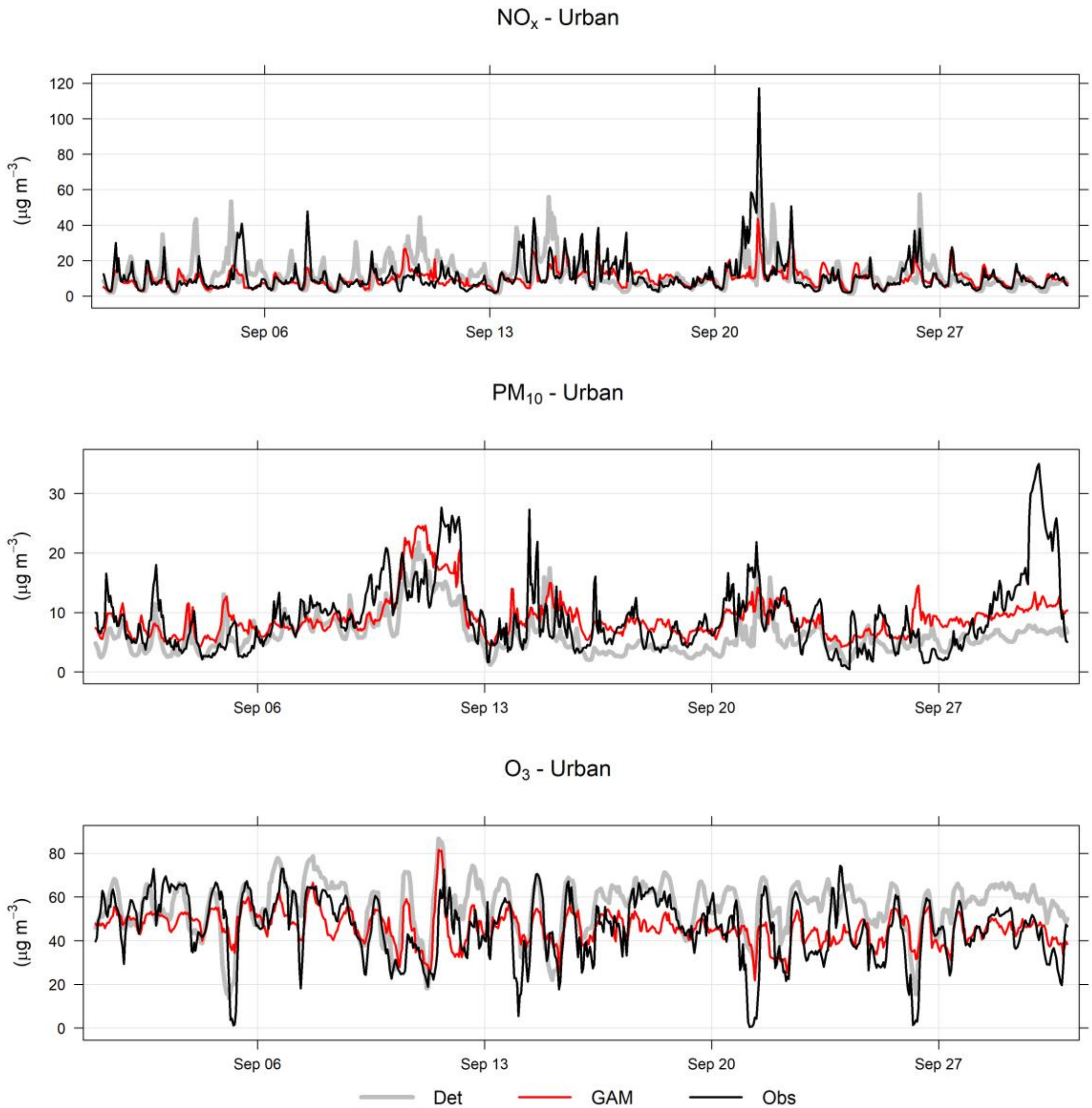


Figure 5. Temporal variations in hourly mean NO_x, PM_{10} and O_3 concentrations at the urban background site during September 2021 based on observations, deterministic forecasts and GAM. Mean of 1-, 2- and 3-day forecasts.

Figure 6 shows example of deviations from observations of forecasted NO_x, PM_{10} and O_3 for all models illustrating that during some hours all models systematically show large absolute deviations from the observed mean concentrations. Sometimes the hours with large deviation for NO_x coincide with deviations for PM_{10} indicating some specific meteorological situation or common source that cause this deviation.

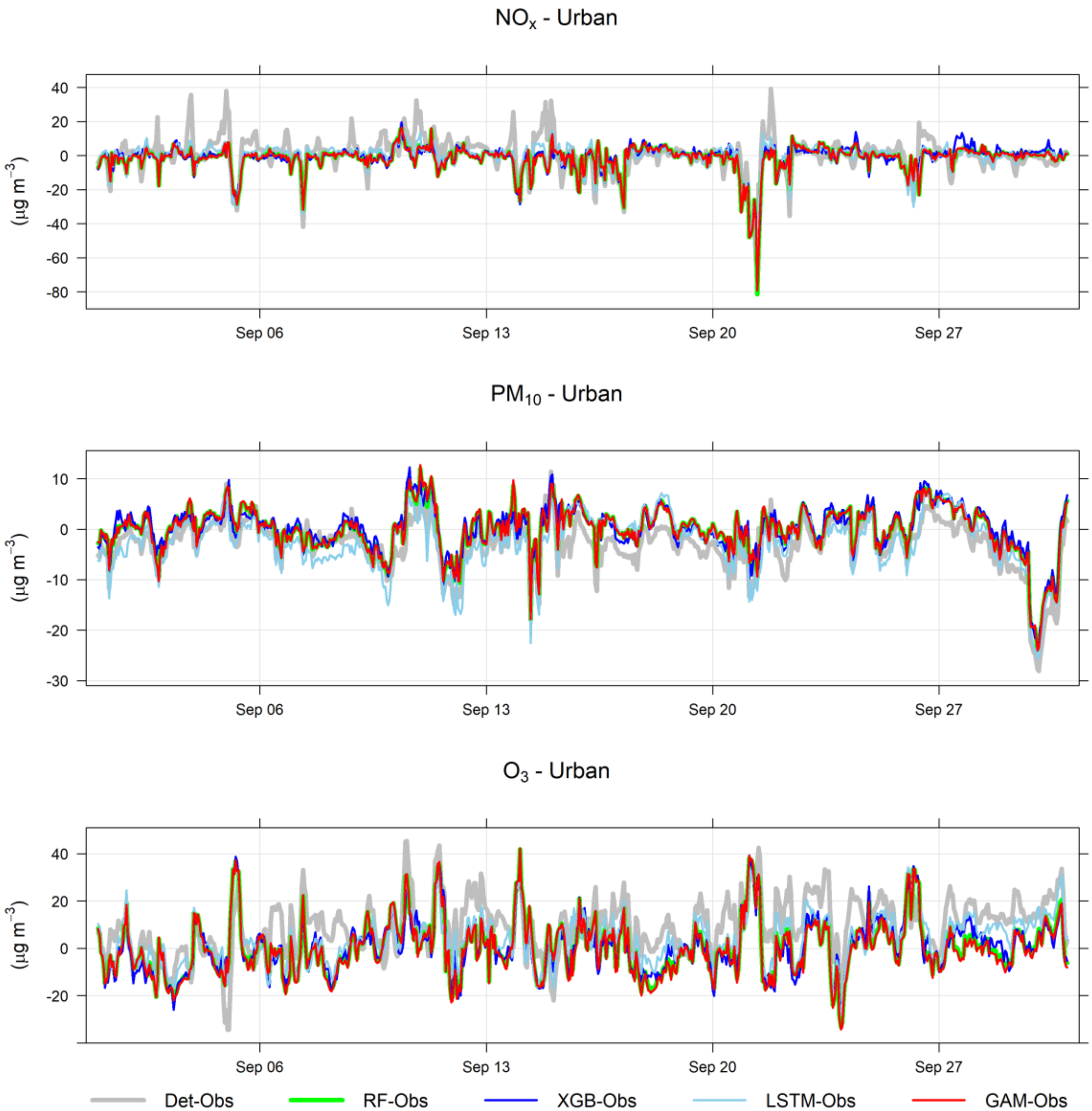


Figure 6. Absolute deviations of forecasted NO_x, PM_{10} and O_3 concentrations from observed (Obs) concentrations based on mean of 1-, 2- and 3-day forecasts for September 2021. All data are hourly mean concentrations.

Figure C2 shows systematic deviations between the observed mean diurnal variations and the deterministic forecast. This is significantly improved using the ML models, especially for NO_x and O_3. For O_3 the deterministic forecast systematically overestimates the concentrations which is mainly due to the fact that the chemical destruction of O_3 in the city centre is not properly accounted for by the regional CAMS model. For NO_x the concentrations calculated by the deterministic model are systematically shifted one hour compared to the observed concentration and this is likely associated with errors in parameterisation of traffic emissions, which is the most important source of NO_x in Stockholm. For PM_{10} concentrations modelled by the deterministic model are too low during the night compared to observations, but this is corrected using RF and XGB, but not using LSTM.

As can be seen in Table 3 and Figure 7 most of the statistical performance measures are improved compared to the deterministic forecasts of NO_x and PM_{10} using different ML models. For NO_x Pearson correlation (r) increases from 0.35-0.39 with deterministic forecasts to between 0.49 and 0.70 when ML models are used. MAPE, nRMSE and nMAE decreases for all models and all forecast days. For PM_{10} Pearson r increases from 0.50-0.53 with deterministic forecasts to between 0.50 and 0.74 when ML models are used. nRMSE and nMAE decreases for forecast days, but for MAPE results are not so consistent – MAPE increases slightly with XGB, RF and GAM, while it decrease for 1-day and 2-day forecasts using LSTM. For O_3 there are small improvements looking at Pearson r and MAPE, nRMSE and nMAE decreases. The Pearson correlation for O_3 is already relatively high and errors relatively small with the deterministic CAMS modelling.

Figure 7 presents mean of 1-day, 2-day and 3-day statistical performances as ratios of ML to deterministic forecasts. This shows that NO_x is consistently improved using all ML models for all statistical performance indexes, whereas for PM_{10} and O_3 there are improvements in nRMSE and nMAE, but MAPE. Overall, the difference in performance between different models is small, less than 30%, but larger when comparing different pollutants.

Table 3. Comparison of 1-, 2-, 3-day deterministic and ML forecasts for NO_x, PM_{10} and O_3 for the urban background site. r = Pearson correlation, MAPE = mean absolute percentage error, nRMSE = normalised rootmean square error and nMAE = normalised mean absolute error. All data are based on hourly mean values. Best performances are bold.

NO_x	r			MAPE			nRMSE			nMAE		
	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day
Det	0.39	0.38	0.35	69%	65%	67%	130%	124%	116%	63%	61%	61%
XGB	0.49	0.53	0.54	42%	44%	48%	118%	114%	114%	44%	45%	47%
RF	0.54	0.57	0.60	**37%**	38%	37%	115%	112%	111%	**41%**	**41%**	**41%**
LSTM	**0.70**	**0.69**	**0.66**	50%	59%	54%	**99%**	**99%**	**101%**	43%	47%	46%
GAM	0.50	0.55	0.58	**37%**	**37%**	**37%**	117%	114%	112%	42%	42%	42%

PM_{10}	r			MAPE			nRMSE			nMAE		
	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day
Det	0.53	0.50	0.50	54%	56%	**59%**	81%	85%	87%	47%	48%	50%
XGB	0.71	**0.65**	0.56	61%	64%	69%	58%	64%	69%	41%	**44%**	47%
RF	**0.74**	**0.65**	**0.60**	55%	74%	78%	56%	**63%**	**66%**	**39%**	45%	**46%**
LSTM	0.71	0.57	0.50	**47%**	**54%**	60%	62%	73%	79%	42%	49%	53%
GAM	0.73	0.64	0.59	55%	76%	80%	**56%**	64%	67%	**39%**	46%	47%

O_3	r			MAPE			nRMSE			nMAE		
	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day
Det	0.74	0.71	0.69	**45%**	49%	**50%**	31%	32%	32%	24%	25%	25%
XGB	0.75	0.71	0.67	47%	51%	53%	25%	26%	27%	**19%**	**20%**	21%
RF	**0.76**	0.69	0.71	47%	54%	52%	**24%**	26%	26%	**19%**	21%	**20%**
LSTM	**0.76**	**0.74**	**0.74**	46%	**47%**	51%	**24%**	**25%**	**25%**	**19%**	**20%**	**20%**
GAM	0.75	0.66	0.69	47%	55%	52%	**24%**	27%	27%	**19%**	22%	21%

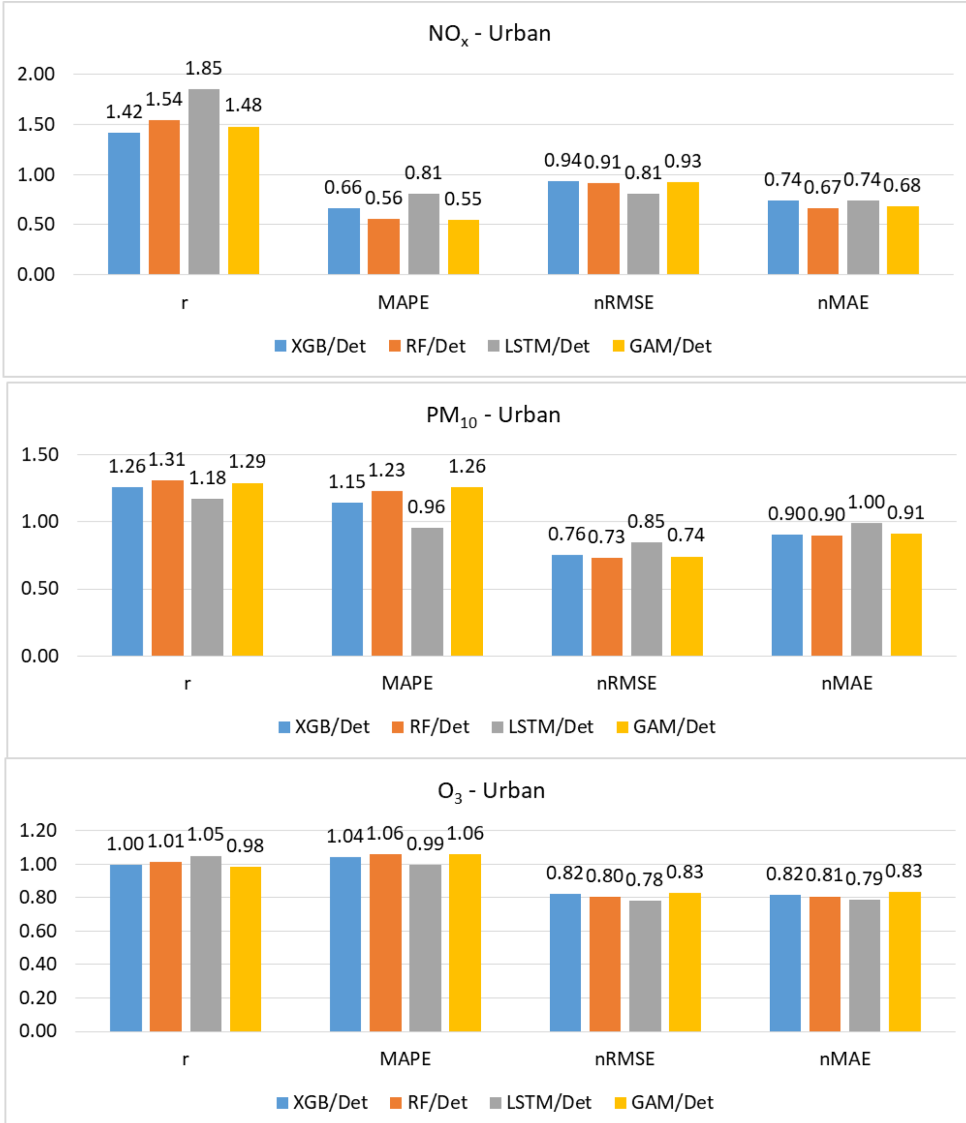


Figure 7. Ratios of statistical performances for ML models versus the deterministic hourly forecasts for the urban site. Mean of 1-day, 2-day and 3-day forecasts.

For the general public it is important to receive information on future pollution episodes with high concentrations. The plots in Figure D1 shows that statistical performances for all models is worse when concentrations higher than when the mean value is analysed. Pearson r is somewhat higher for PM_{10} and O_3, but not when RF and XGB is used for NO_x. MAPE is reduced for PM_{10} and NO_x but not for O_3. The nRMSE is both higher and lower with ML models compared to the deterministic model, while, finally, nMAE is lower for NO_x and PM_{10} using RF and XGB, but not for PM_{10} using LSTM.

As can be seen in Figure 8 all MQI are below 100% indicating that deviations between model results and measurements are smaller than the estimated uncertainties in the measurements. It can also be seen that LSTM is somewhat more efficient in reducing MQI, from 68% to 60% for NO_x and O_3 from 40% to 29%, while RF and XGB provides no improvement for NO_x, but both PM_{10} and O_3 shows slightly lower MQI with RF and XGB compared to the deterministic forecast.

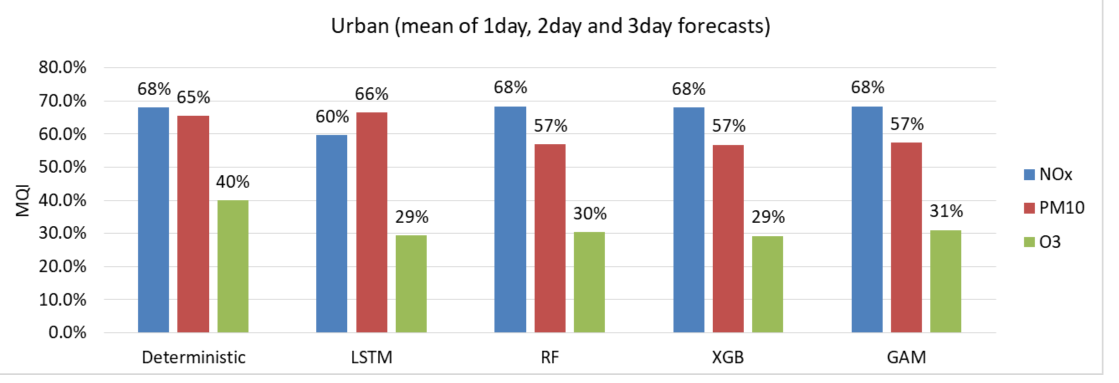


Figure 8. MQI based on hourly mean concentrations for the whole test period for NO_x, PM_{10} and O_3 of the urban site. Mean of 1-, 2- and 3-day forecasts.

3.2 Street Canyon sites

3.2.1 Importance of features - street canyon sites

For the street canyon sites the relative importance of different features is different for PM_{10} and NO_x and also somewhat different depending on ML model and street (see figures in Appendix E). There are, however, some typical features that tend to be more important. For PM_{10} Julian day, lagged measurements and deterministic forecasts are mostly among the top 5 most important features using RF and XGB, while precipitation is an important feature using LSTM. For NO_x deterministic forecasts, hour of the day and weekday are the most important, while lagged measurements are less useful for the ML models. The importance of calendar data for NO_x likely reflects importance of diurnal and weekday variations in traffic emissions not correctly captured by the deterministic forecast. Julian day likely reflects seasonal variations in non-exhaust emissions of PM_{10} and precipitation reflects the importance of street wetness for suspension of road dust. Even though there are variations it is difficult see any systematic difference in the features between ML for the different street sites.

3.2.2 Comparison between deterministic forecasts and ML models - street canyon sites

Comparisons between the hourly temporal variations in observations and forecasts of NO_x with the GAM model in September 2022 are shown in Figure 9 and for all models in Appendix F. One can see that the deterministic forecast tend to overestimate concentrations of NO_x during daytime especially for Sveavägen and this is corrected when ML modelling is being applied. Corresponding plots for PM_{10} are shown in Figure 10. In this case the GAM overestimates concentrations on Folkungagatan and Hornsgatan during the end of September, but performs well otherwise, whereas the deterministic forecast overestimates PM_{10} on Sveavägen and Hornsgatan during the first half of the month.

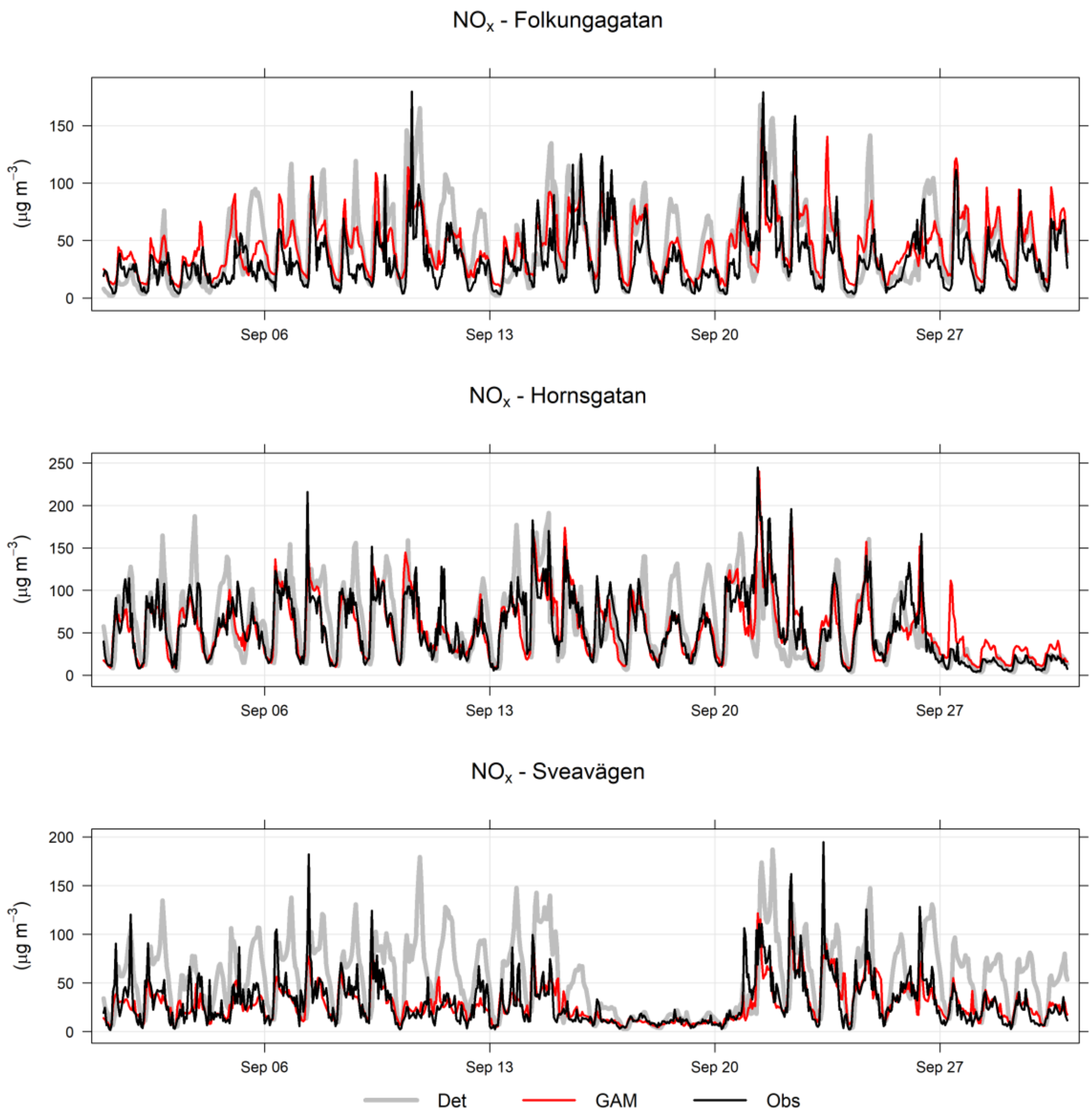


Figure 9. Temporal variations in hourly mean NO_x concentrations at the street canyon sites during September 2022 based on observations (red) and 1-day forecasts based on deterministic modelling (blue) and GAM (green).

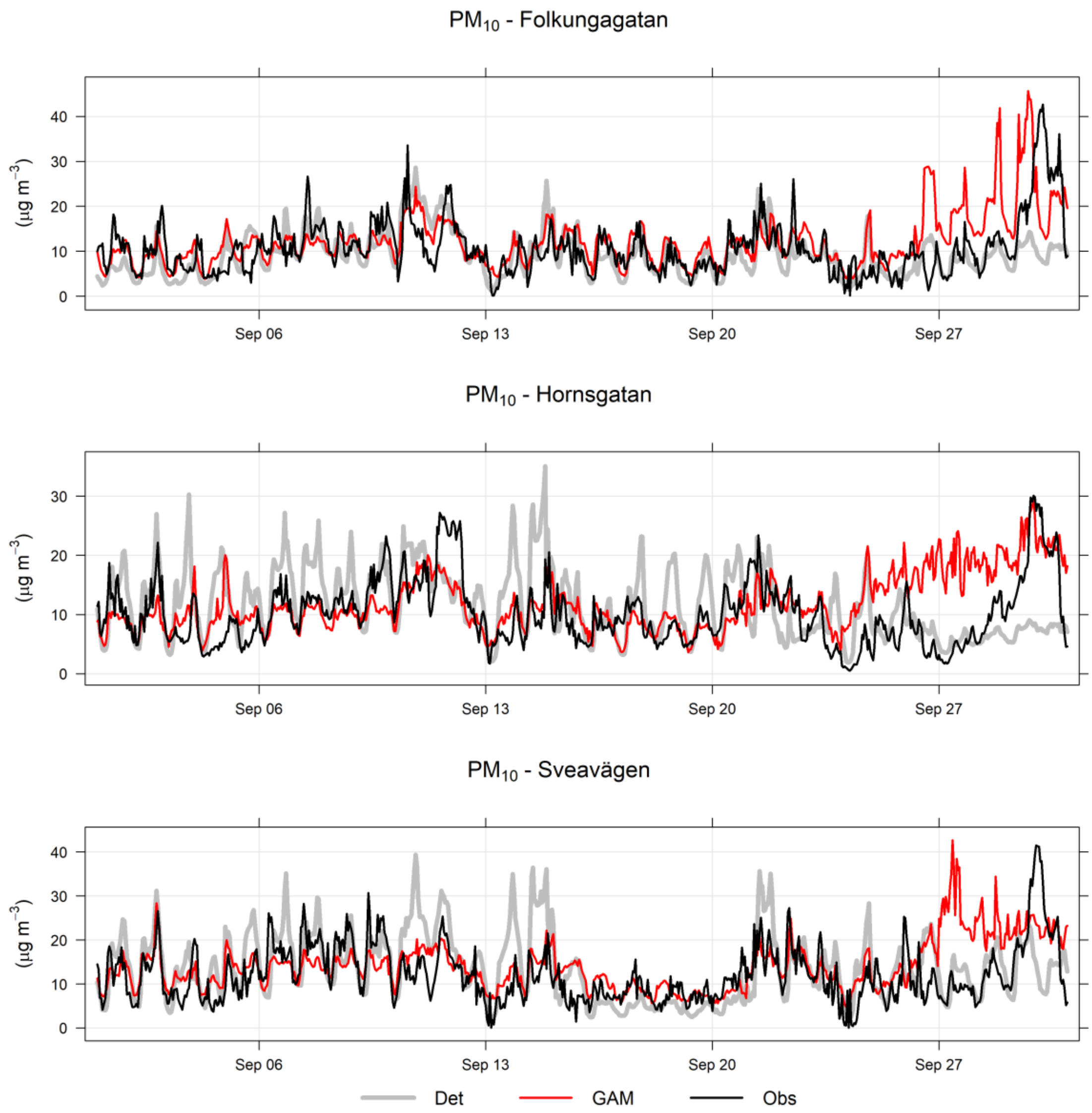


Figure 10. Temporal variations in hourly mean PM_{10} concentrations at the street canyon sites during September 2022 based on observations, deterministic modelling and GAM forecasts. Mean of 1-, 2- and 3-day forecasts.

The improvement of the temporal variations of NO_x and PM_{10} is well illustrated by comparing the mean diurnal variations in observations with deterministic modelling and using the ML models, GAM shown in Figure 11 and all models shown in figures in Appendix G. For all street sites, both NO_x and PM_{10} concentrations shows systematic deviations from observations using deterministic modelling, but this is corrected using the ML models, especially for NO_x. The tendency that the LSTM model is not as good to capture variations in PM_{10} at the urban site is also seen here for the street canyon sites.

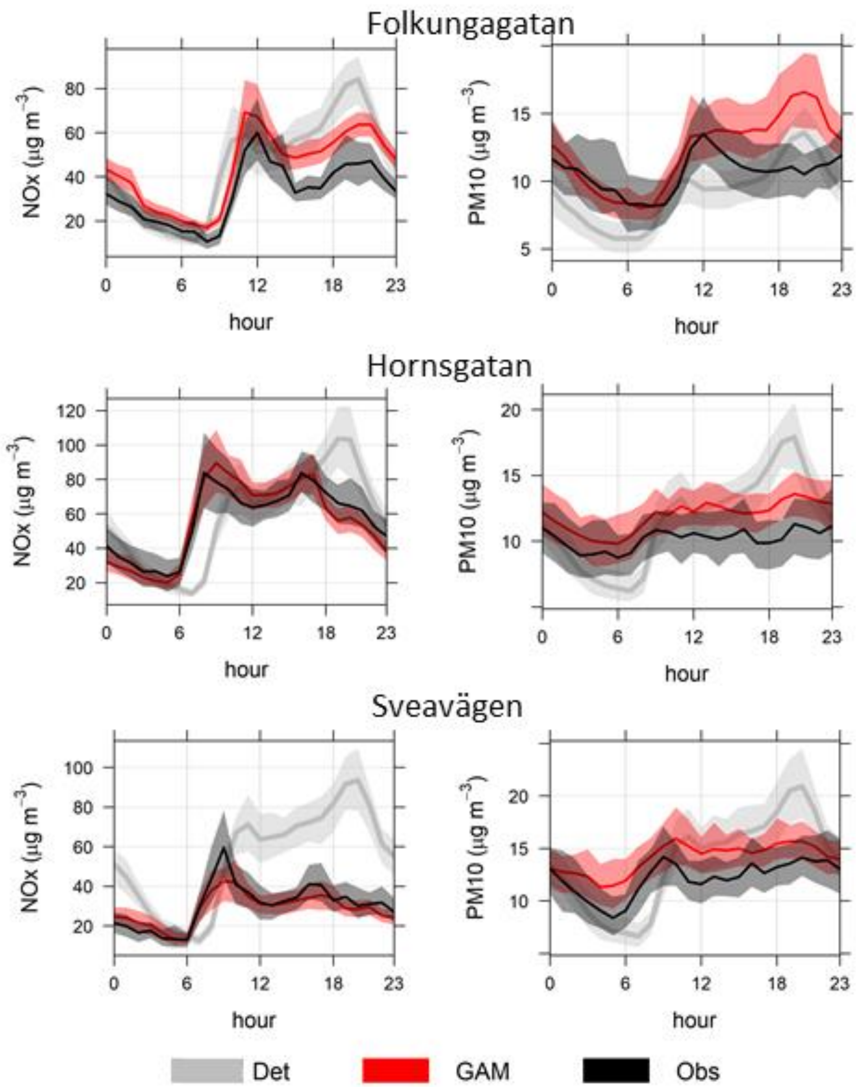


Figure 11. Mean diurnal variations in hourly mean observations, deterministic and GAM forecasts of NO_x and PM_{10} for the street canyon sites. Mean of 1-, 2- and 3-day forecasts.

For all streets statistical performance of NO_x forecasts are improved using the ML models as shown for all forecasts in Table 4. The improvement in terms of Pearson correlation (r), MAPE, nRMSE and nMAE is very similar for the ML models but differ between streets, with forecasts for Hornsgatan showing higher r and lower relative errors compared to the other streets.

Table 4. Comparison of 1-, 2-, 3-day deterministic and ML forecasts for NO_x for the street canyon sites. r = Pearson correlation, MAPE = mean absolute percentage error, nRMSE = normalised rootmean square error and nMAE = normalised mean absolute error. All data are based on hourly mean values.

Folkungagatan												
	r			MAPE			nRMSE			nMAE		
	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day
Det	0.48	0.49	0.47	107%	118%	120%	108%	109%	106%	72%	73%	73%
XGB	0.65	0.64	0.63	67%	73%	**76%**	74%	75%	75%	47%	50%	**50%**
RF	**0.66**	**0.65**	**0.65**	**64%**	73%	81%	**71%**	**74%**	77%	**45%**	49%	53%
LSTM	0.64	0.61	0.62	65%	**60%**	79%	72%	**74%**	**74%**	46%	**46%**	**50%**
GAM	**0.66**	**0.65**	**0.65**	65%	75%	81%	73%	75%	77%	46%	51%	53%
Sveavägen												
	r			MAPE			nRMSE			nMAE		
	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day
Det	0.46	0.53	0.44	159%	161%	163%	137%	136%	134%	99%	98%	97%
XGB	0.69	0.68	0.66	59%	57%	59%	68%	69%	71%	41%	41%	41%
RF	**0.73**	**0.73**	**0.73**	**51%**	**51%**	50%	**65%**	**65%**	**65%**	**37%**	**38%**	**37%**
LSTM	0.71	0.69	0.66	58%	60%	64%	68%	69%	71%	41%	41%	43%
GAM	0.72	0.71	0.71	52%	**51%**	**49%**	**65%**	67%	66%	38%	39%	**37%**
Hornsgatan												
	r			MAPE			nRMSE			nMAE		
	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day
Det	0.53	0.56	0.55	80%	69%	73%	82%	79%	80%	55%	52%	54%
XGB	**0.80**	**0.81**	0.81	45%	45%	44%	52%	**51%**	**50%**	32%	**32%**	32%
RF	0.79	0.79	0.81	**42%**	**43%**	**43%**	52%	53%	**50%**	**31%**	**32%**	**31%**
LSTM	0.77	0.76	0.76	48%	51%	51%	57%	57%	56%	36%	36%	36%
GAM	**0.80**	0.80	**0.82**	**42%**	**43%**	**43%**	**51%**	**51%**	**50%**	**31%**	**32%**	**31%**

Figure 12 clearly illustrates the improvements of all statistical performance indexes for NO_x at all street canyon sites and for ML models. The errors (MAPE, nRMSE, nMAE) are reduced by between 30% and 60% and the Pearson correlation coefficients increase by between 30% and 50%.

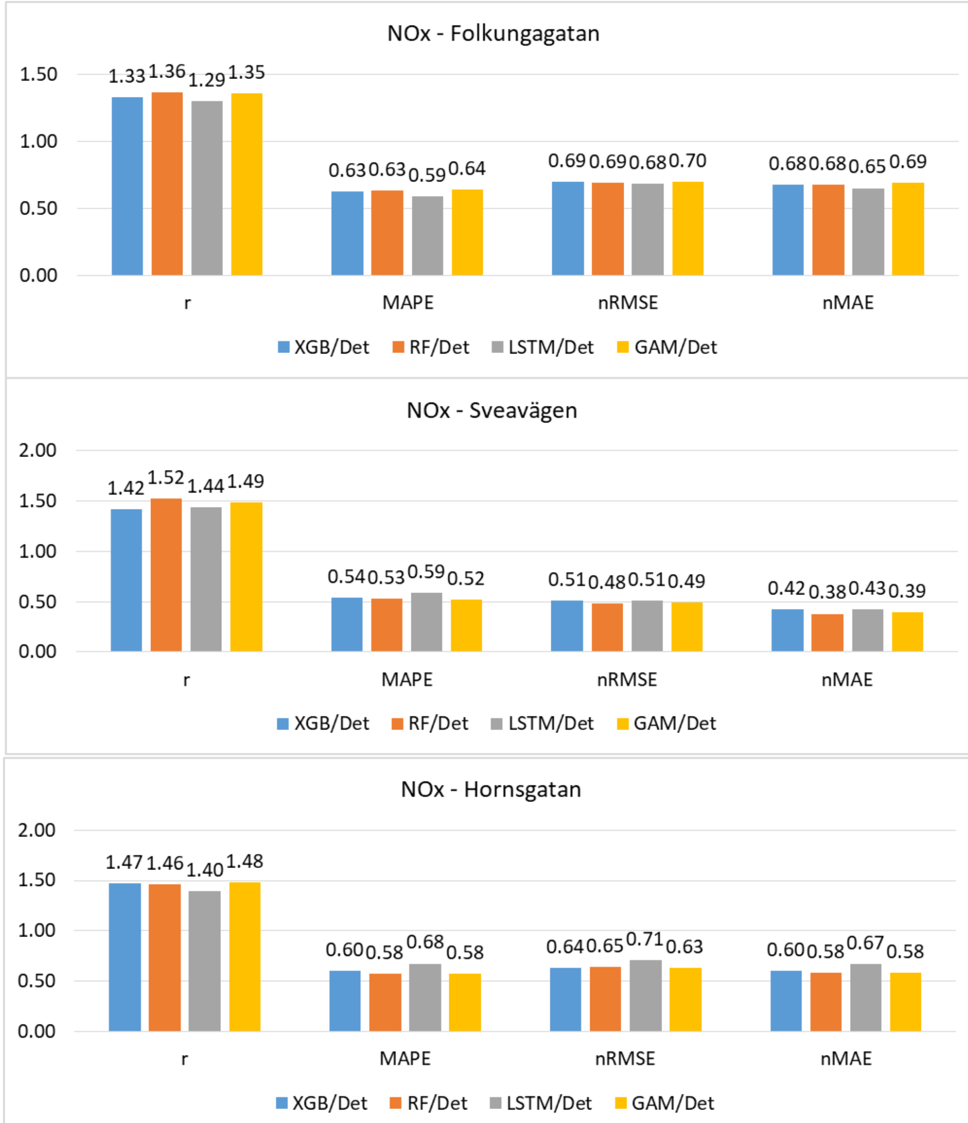


Figure 12. Ratios of statistical performances for ML models versus the deterministic hourly forecasts for NO_x at the street canyon sites. Mean of 1-day, 2-day and 3-day forecasts.

Comparison between the statistical performance measures for ML models and deterministic forecasts for PM_{10} shows somewhat variable results depending on statistical measure, street and ML. Person r values increase slightly in most cases and the normalised RMSE and MAE are lower for most ML models and streets, but not always, while MAPE often increase using the ML models (Table 5 and Figure 13). Errors measured as nRMSE decrease by between 10% and 30%, whereas errors measured as MAPE mostly increase slightly and nMAE is about unchanged. Pearson r increase at Folkungagatan for all ML models (10% - 30%) but show somewhat varying results for Sveavägen and Hornsgatan.

Table 5. Comparison of 1-, 2-, 3-day deterministic and ML forecasts for PM_{10} for the street canyon sites. r = Pearson correlation, MAPE = mean absolute percentage error, nRMSE = normalised rootmean square error and nMAE = normalised mean absolute error. All data are based on hourly mean values.

Folkungagatan												
	r			MAPE			nRMSE			nMAE		
	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day
Det	0.32	0.30	0.34	121%	**112%**	**119%**	115%	116%	115%	56%	57%	56%
XGB	0.41	0.30	0.34	122%	134%	121%	85%	102%	83%	**52%**	63%	54%
RF	0.36	0.39	**0.41**	134%	121%	129%	89%	82%	**75%**	**52%**	**52%**	**49%**
LSTM	**0.47**	**0.43**	0.34	**102%**	115%	141%	**82%**	**77%**	83%	58%	53%	58%
GAM	0.37	0.34	0.39	132%	123%	127%	88%	95%	77%	**52%**	57%	50%
Sveavägen												
	r			MAPE			nRMSE			nMAE		
	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day
Det	0.42	0.40	0.40	98%	**100%**	**95%**	92%	92%	92%	55%	**56%**	54%
XGB	0.42	0.31	0.45	122%	124%	109%	76%	92%	73%	51%	58%	**49%**
RF	0.49	0.27	0.40	113%	125%	114%	**67%**	99%	74%	**45%**	57%	50%
LSTM	**0.51**	**0.49**	**0.46**	**90%**	106%	109%	**67%**	**67%**	**68%**	47%	48%	**49%**
GAM	0.45	0.28	0.41	115%	121%	111%	71%	93%	75%	46%	**56%**	**49%**
Hornsgatan												
	r			MAPE			nRMSE			nMAE		
	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day	1-day	2-day	3-day
Det	0.40	0.36	0.30	81%	**80%**	**87%**	113%	116%	118%	59%	60%	62%
XGB	0.46	0.30	**0.37**	84%	103%	91%	89%	110%	**89%**	**56%**	67%	59%
RF	0.42	0.21	0.33	85%	115%	94%	91%	130%	90%	57%	73%	59%
LSTM	**0.49**	**0.40**	0.34	**77%**	84%	93%	**82%**	**85%**	**89%**	**56%**	**59%**	64%
GAM	0.45	0.25	0.34	84%	107%	92%	88%	114%	**89%**	**56%**	68%	**58%**

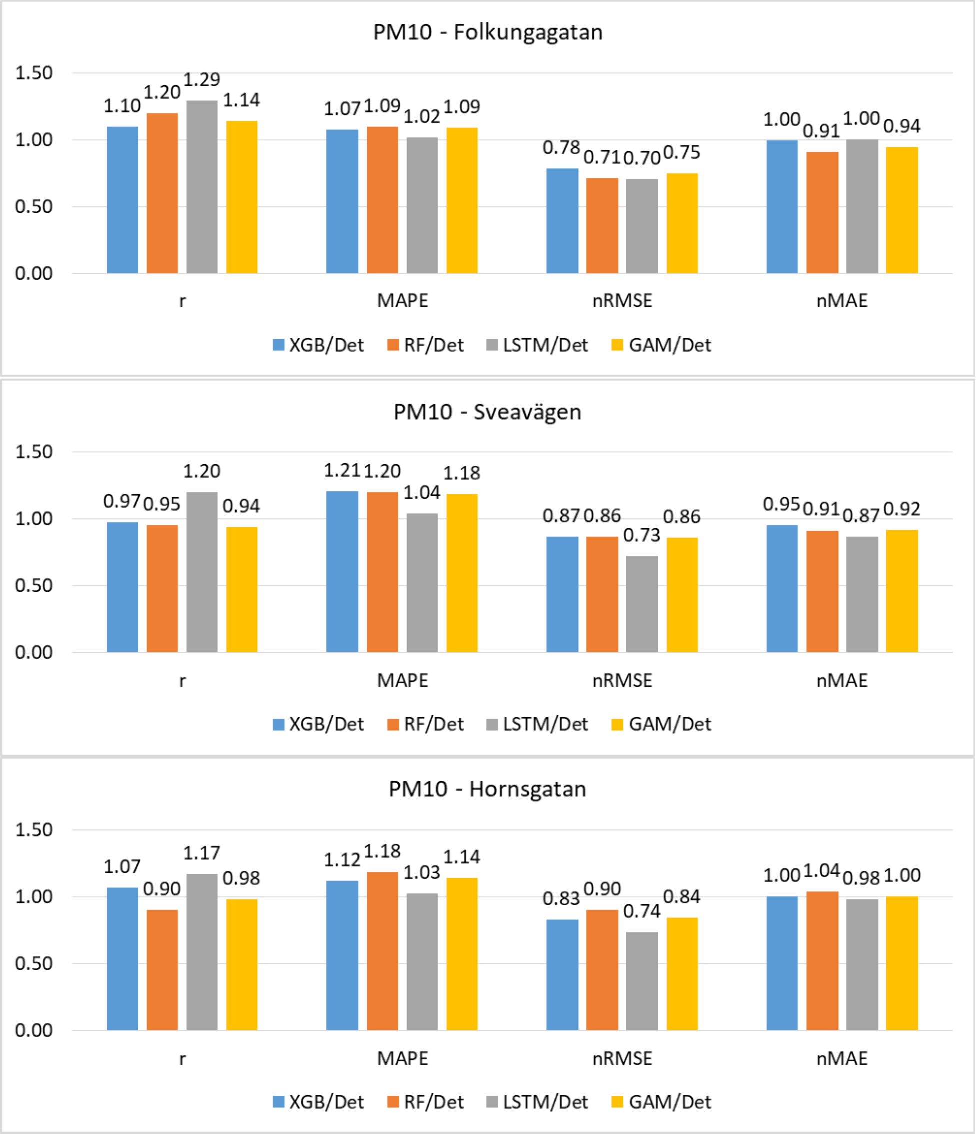


Figure 13. Ratios of statistical performances for ML models versus the deterministic hourly forecasts for PM_{10} at the street canyon sites. Mean of 1-day, 2-day and 3-day forecasts.

As pointed out before it is important to assess statistical performance measures for periods with high concentrations. Similar to what is seen for the urban site the statistical performances for all models are much worse for the hourly mean concentrations that are higher than the mean values and the pattern is also similar for the different streets.

3.2.3 MQI street canyon sites

Figure 14 illustrates that deviations between model results and measurements compared to the uncertainties of the measurements for all pollutants and street canyon sites. For NO_x relative uncertainties decreases using the ML models compared to the deterministic forecast, while for PM_{10} results varies, but there is no systematic improvement using ML models compared to the deterministic model.

Figure 14. MQI for NO_x and PM_{10} forecasts at street canyon sites. Mean values for 1-, 2- and 3-day forecasts.

3.3 Generalisation of street canyon modelling

Until now, the model performance is evaluated using training and testing data from three single sites respectively. In Stockholm as well as in other cities most of the streets do not have any monitoring station. This is of course due to resource constraints but also associated with the fact that the EU Air Quality Directives regulates the number of monitoring sites required in a city depending on the level of air pollution and number of inhabitants. The monitoring stations should provide information for both areas where the highest concentrations of air pollutants occur and other areas that are representative of the exposure of the general population. Less resources is required if this information can be achieved by accurate enough modelling.

We therefore analyze the generalization capacities of the models, with the expectation that we can achieve certain prediction performance of one site without having any measurement data. Computational experiments were carried out through cross-validation, which combines training and testing data coming from different measurement sites. For the street canyon sites, four combinations of training datasets were applied to evaluate the generalization abilities of different ML models.

Figure 15 shows mean of 1-day, 2-day, and 3-day forecasted NO_x concentrations for the three street canyon sites based on training the models on the other streets. It shows that the forecast is improved compared to the deterministic forecast for Hornsgatan and Sveavägen, but not so much for Folkungagatan. For Hornsgatan the correlation is 0.55 using the deterministic forecast whereas the ML models gives correlations between 0.61 and 0.67 and all errors decrease slightly using the ML models. For Sveavägen the correlation is 0.48 using the deterministic forecast whereas the ML models gives correlations between 0.62 and 0.63 and here all errors decrease substantially using the ML models. But for Folkungagatan the ML models show different results. Correlations are similar or even decreases, whereas errors mostly decreases depending on ML applied.

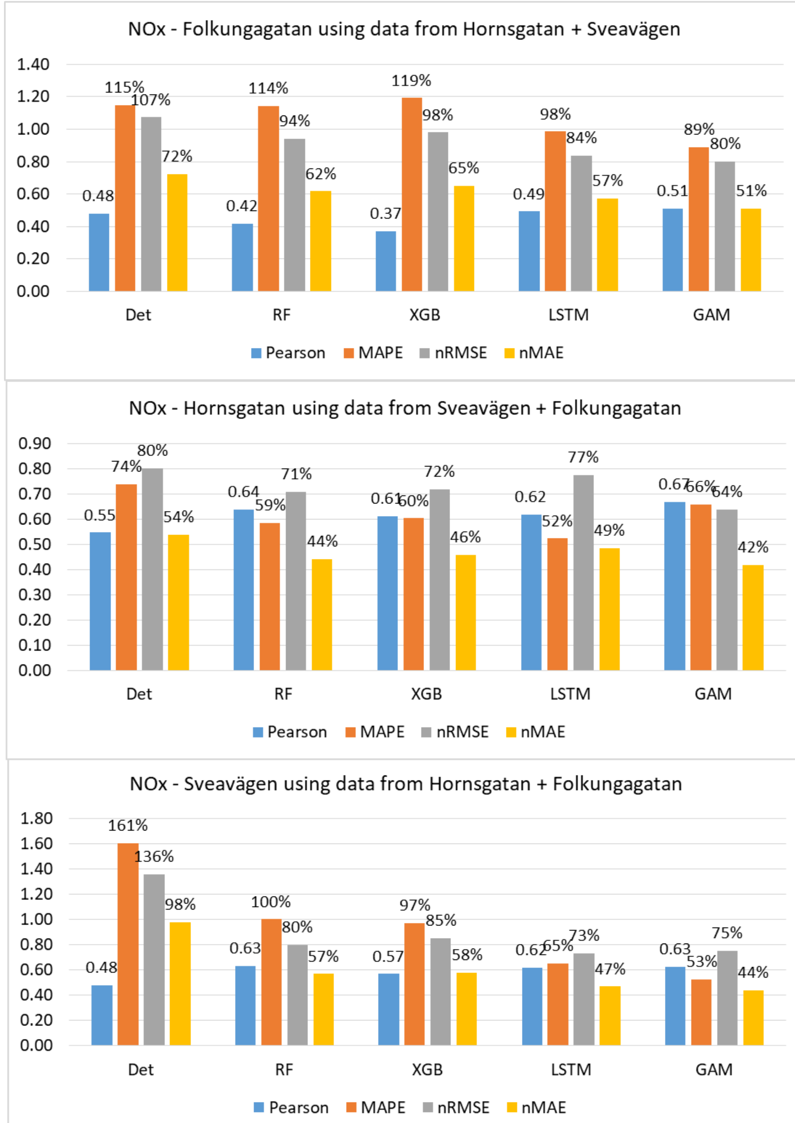


Figure 15. Statistical performances of NO_x forecasts for the streets when the ML models are trained using only data from the other streets. Mean of 1-day, 2-day, and 3-day forecasts.

Figure 16 shows mean of 1-day, 2-day, and 3-day forecasted PM_{10} concentrations for the three street canyon sites based on training the models on the other streets. It can be seen that it is not possible to find any systematic improvement of the deterministic forecast for the streets using RF and XGB compared to the deterministic forecasts. But with LSTM correlations increase slightly and errors decrease at all streets compared to the deterministic forecasts.

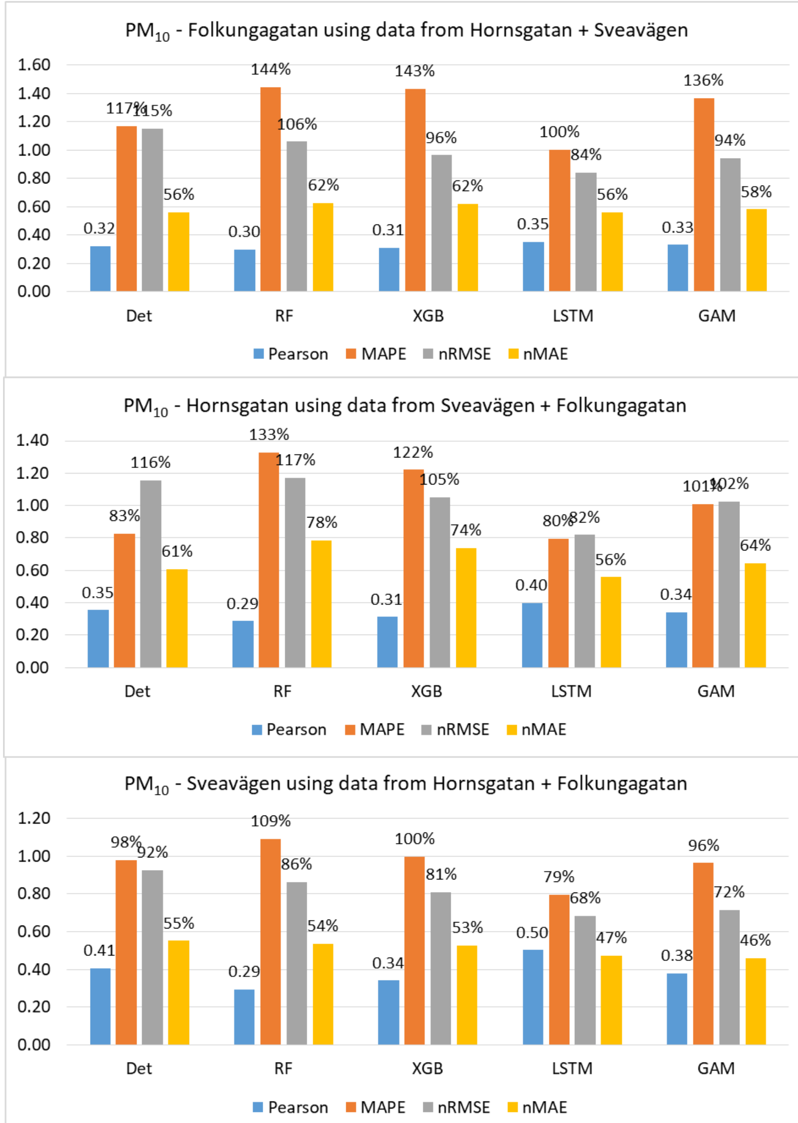


Figure 16. Statistical performances of PM_{10} forecasts for the streets when the ML models are trained using only data from the other streets. Mean of 1-day, 2-day, and 3-day forecasts.

4 Discussion

The performance of the ML models are quite similar for the different sites and forecast days. But there are large differences in improvements for different pollutants. In general, our results indicate that ML models are more effective in improving NO_x than PM_{10} and O_3. For PM_{10} the ML models show slight improvement in r but not much improvements in relative errors. This difference in improvement is likely associated with the different processes controlling the concentrations, such as different sources: NO_x concentrations being mainly due to vehicle exhaust emissions which shows regular variations from one day to the next depending on day of the week and time of day, while PM_{10} is mainly due to road dust emissions controlled by a combination of variations in vehicle volumes and meteorological conditions that affect suspension of coarse particles from street surfaces (e g Denby et al., 2013a; Johansson et al., 2007; Krecl et al., 2021). Road dust is accumulated on the road surfaces during wet road surface conditions and suspended by vehicle induced turbulence during dry conditions (Denby et al., 2013a).

The improvement of the forecasts of NO_x with ML is partly driven by the calendar, hour, day of the week and to some degree also Julian day, but different features appear as important for RF compared to XGB. For PM_{10} the seasonal variation described by Julian day is the most important feature at the street canyon sites, both for RF and XGB. This indicates that the deterministic forecasts is not capable at describing impacts of meteorology and road dust emissions on PM_{10}, even though parameterisations of these processes are included in the deterministic modelling system. The total mass generated by road wear is a key factor for PM_{10} emissions and these emissions are strongly controlled by surface moisture conditions and this is taken into account by the NORTRIP model. But as pointed out by Denby et al (2013b) there are periods where surface wetness is not well modelled and it is not known if this is the result of input data, e.g. precipitation, or of the model formulation itself.

It is clear that the deterministic forecast of O_3 underestimates concentrations at the urban site due to the fact that the local emissions of NO_x influencing the photochemistry is not properly considered by the CAMS model, but this is corrected using the ML models. Despite this the deterministic forecast is the most important feature for both RF and XGB but also lagged measured mean and maximum O_3 concentrations improve the deterministic forecasts.

Despite the fact that the configurations and traffic situations are quite similar for the street canyon sites, the improvements of the deterministic forecasts using ML differs. For NO_x forecasts on Hornsgatan are more accurate (lower errors and higher r) than for the other two sites, while for PM_{10} there is no obvious difference between the sites.

The overall model quality according to the recommendations by the Forum for Air Quality Modeling (FAIRMODE) in the context of the air quality directives, is improved using the ML models resulting in uncertainties that are significantly smaller than the measurement uncertainties for all pollutants. But the forecasts of the highest concentrations including episodes with high concentrations, is not systematically improved for all pollutants and all statistical performance measures using the ML models.

We have shown that the statistical performances of the deterministic forecasts for concentrations of NO_x at the street canyon sites can be improved using the ML models. But for PM_{10} only LSTM showed systematic improvements at all sites. So again this accentuates the importance of testing the models not only for one pollutant. Further work is needed to improve deterministic forecasts of PM_{10} based on the training of ML models at a few monitoring stations. As discussed above the situation in Stockholm is different from cities in central and southern Europe since the road dust contribution is very large. It might be that results for PM_{10} is different in other cities, but we have not found any publication on this matter.

4.1 Comparison of different ML models

Several studies have compared performance of different machine learners for predicting air quality (Zaini et al., 2021). Assessing forecasts of PM_{10} and PM2.5 concentrations, Czernecki et al. (2021) found that XGB performed the best, followed by random forests and an artificial neural network model, while stepwise regression performed the worst in four Polish agglomerations. Likewise, Joharestani et al. (2019) found XGB to performed best of three ML models (XGB, RF and a deep learning algorithm), in predicting PM2.5 in Tehran (Iran). On the contrary, LSTM was shown to outperform XGBoost for forecasting hourly PM2.5 concentrations (Qadeer et al., 2020), similar to what was shown by Chuluunsaikhan et al (2021). Cai et al. (2009) obtained more accurate predictions of CO concentrations using artificial neural network modelling compared to using multiple linear regression and the deterministic California line source dispersion model. On the other hand Shaban et al. (2015) concluded that a tree based algorithm (M5P) outperformed artificial neural network modelling when comparing forecasts of different pollutants in Qatar. There may be many reasons for the different results presented in the literature, including different types of input data, different atmospheric conditions and source contributions governing the concentrations. Also different statistical measures of performance has been used. This makes it hard to draw general conclusions regarding which model to use. However, we find that other factors may be more important to consider than type of model – such as sources of pollutants and influence of photochemistry, characteristic of the site resulting in different features being of varying importance depending on pollutant type of location. In this context RF and XGB can provide useful output on the importance of features that is not possible using LSTM.

Another more practical aspect to consider when comparing the ML models is the complexity and computer resources required for training the models. In AQ literature, deep learning models such as standard LSTM and other Recurrent Neural Networks (RNNs) have been explored for their prediction capacities. However, most of the studies have adopted complex neural network structures, such as models of multiple outputs that mainly give convenience for data processing and automated feature handling. Nevertheless, training even a simple LSTM model is computationally much more expensive than the two conventional machine learning models, i.e. the decision tree based models (RF and XGB) in our case. In fact, we have to resort to the high performance machine (The Swedish Berzelius High-performance Computer) to reduce the computational time. For the current practice in our real air quality prediction system, we prefer the two tree-based models over LSTM. But this doesn't deny the possibility that well-designed deep learning models may replace the conventional machine learning models being adopted in the AQ system in near future, especially when the amount of data increases.

5 Conclusions

We have applied different machine learning algorithms to improve 1-, 2- and 3-day deterministic forecasts of NO_x, PM_{10} and O_3 concentrations for a number of locations in Stockholm, Sweden. It is shown that degree of improvement of deterministic forecasts depend more on pollutant and monitoring site than on what ML algorithm is applied. Deterministic forecasts of NO_x

are improved at all sites, using all models. Pearson correlations increase by up to 80% and errors are reduced by up to 60%. For PM_{10} more variable results are seen likely reflecting the more complicated processes controlling the road wear emissions which constitute a large fraction of PM_{10}. For O_3 at the urban background site deviation between deterministically modelled absolute level is correct using the ML models, nRMSE and nMAE is reduced by on average around 20%, but there is almost no improvement in the correlation and MAPE.

We have shown that it is possible to improve deterministic forecasts of NO_x at street canyon sites, based on training ML models at other sites. But when tested for PM_{10} only LSTM showed some improvements of the statistical performances compared to the deterministic forecast of PM_{10}.

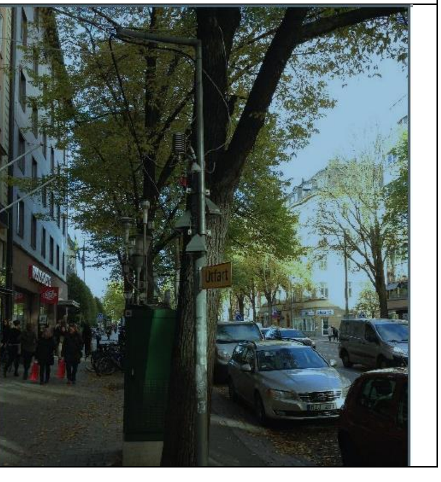
A strength of our study is that we compare forecasts based on several pollutants and base our forecasts on a combination of deterministic models (that are based on the underlying physicochemical mechanisms responsible for the emissions and dispersion of the pollutants) and 3 different machine learning algorithms with additional variables such as measurement data, calendar data and meteorological data. And this method is evaluated at different sites and for different pollutants during several months with different meteorological conditions.

There is still room for improvements of this work like e g fine tuning of the models, including and excluding features, expanding to other sites and making maps of spatial concentrations over a larger area.

6 Appendix A. Description of measurement methods and sites.

All measurement methods are approved for monitoring according to the EU air quality directive for NO_x, O_3 and PM_{10}. PM_{10} was measured either using an optical particle counter (Hornsgatan: OPC, Grimm EDM 180-MC) or Tapered Element Oscillating Microbalance (Sveavägen, Folkungagatan and Urban: TEOM model, 1400AB, Rupprecht & Patashnik, Co). NO_x was measured using chemiluminescence (AC32M, Environnement S.A.) and O_3 was measured by UV absorption (O342M, Environnement S.A.).

Table A1. Description of monitoring sites.

Site name	Description	Traffic volume	Photo
Hornsgatan	Street canyon site. Measurements of NO_x and PM_{10} on north side of street, 3 m above ground. Street width 24 m and building height 24 m.	23 000 veh/day (4% heavy duty vehicles). Vehicle composition measured during 4 week campaigns using automatic number plate recognition.	
Sveavägen	Street canyon site. Measurements of NO_x, PM_{10} on west side of street, 3 m above ground. Street width 33 m and building height 24 m.	21 000 veh/day (7% heavy duty vehicles).	

Folkungagatan	Street canyon site. Measurements NO_x, PM_{10} on west side of street, 3 m above ground. Street width 24 m and building height 24 m.	12 000 veh/day (18% heavy duty vehicles).	
Torkel Knutssongatan	Urban background. Measurements of NO_x, PM_{10}, ozone and meteorology on top of a 20 m high building.	Ca 13 000 vehicles on Hornsgatan road 250 m N of site.	

7 Appendix B Importance of features – urban background

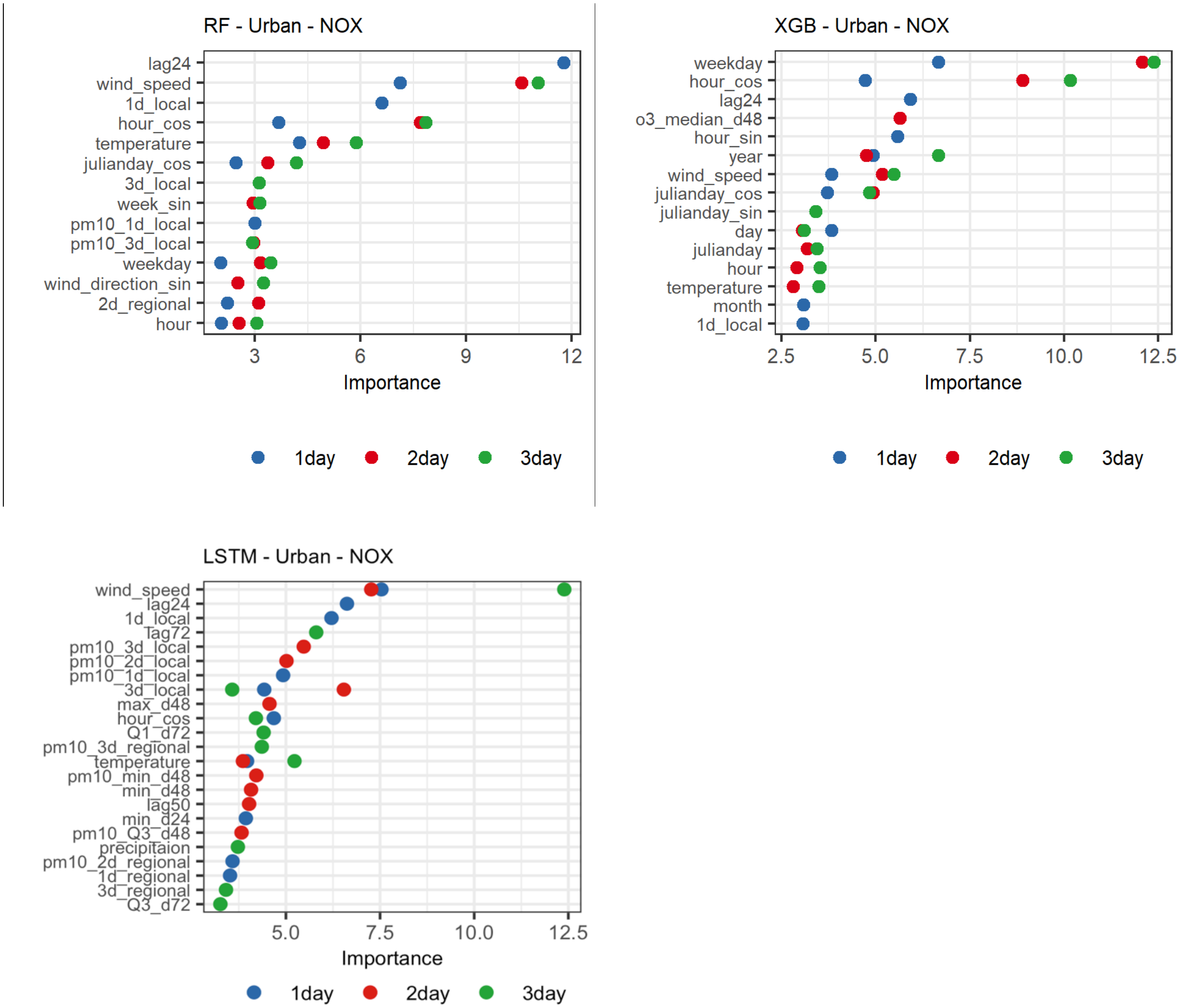


Figure B1.Most important features for NO_x forecasts using XGB, RF and LSTM at the urban site

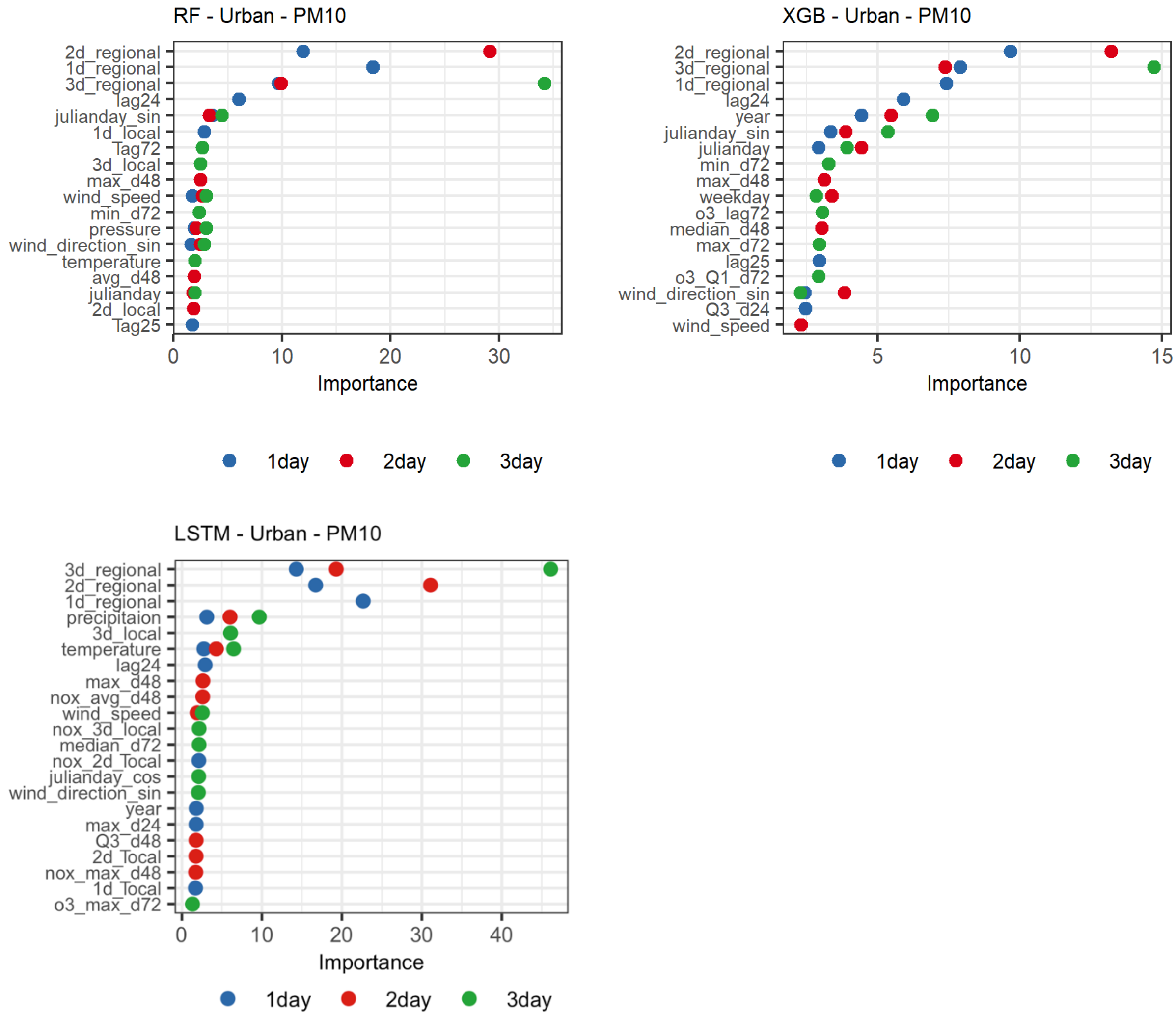


Figure B2. Most important features for PM_{10} forecasts using XGB, RF and LSTM at the urban site.

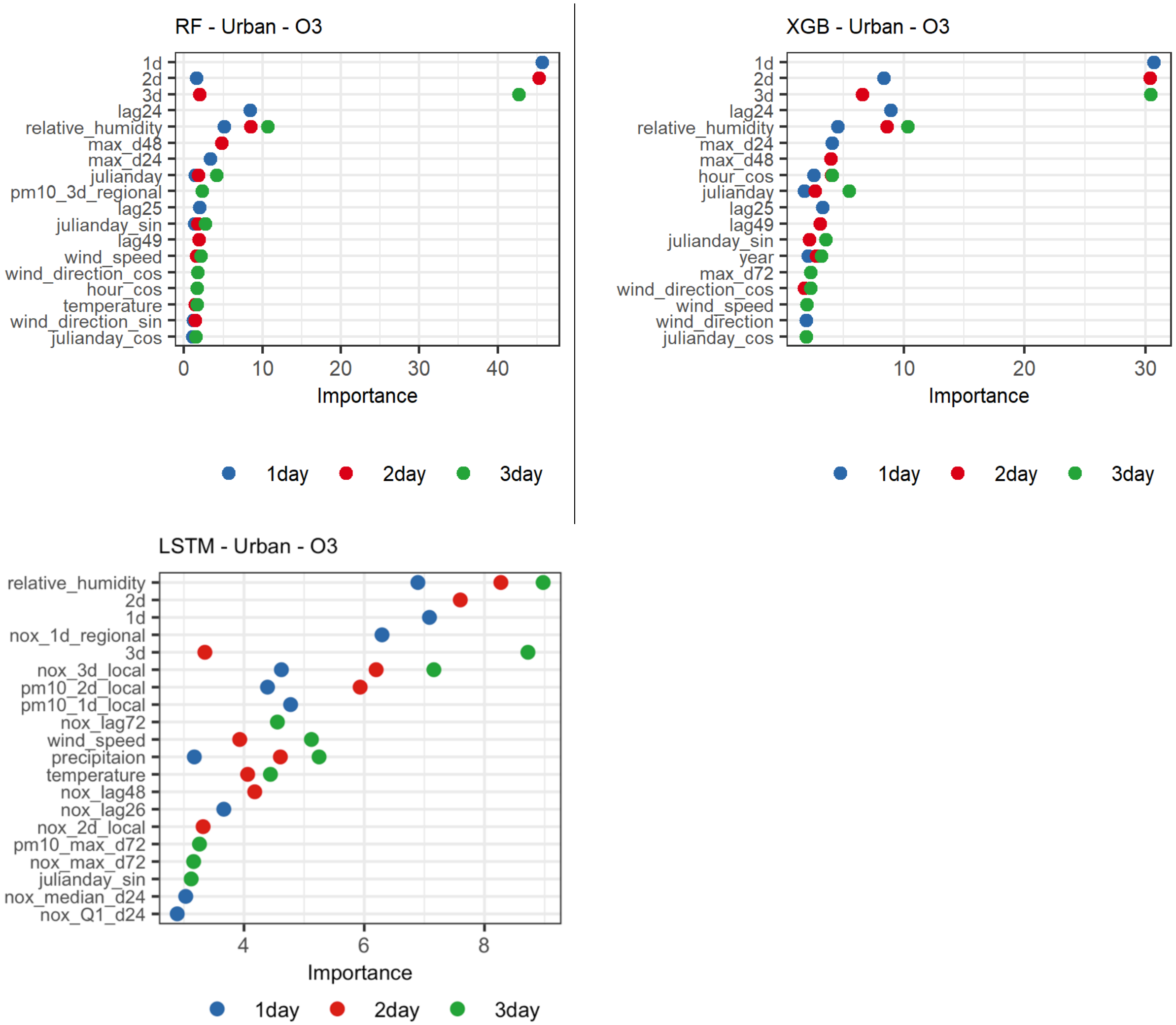


Figure B3. Most important features for O_3 forecasts using XGB, RF and LSTM at the urban site.

8 Appendix C. Temporal variations in hourly mean O_3, NO_x and PM_{10} concentrations at the urban background

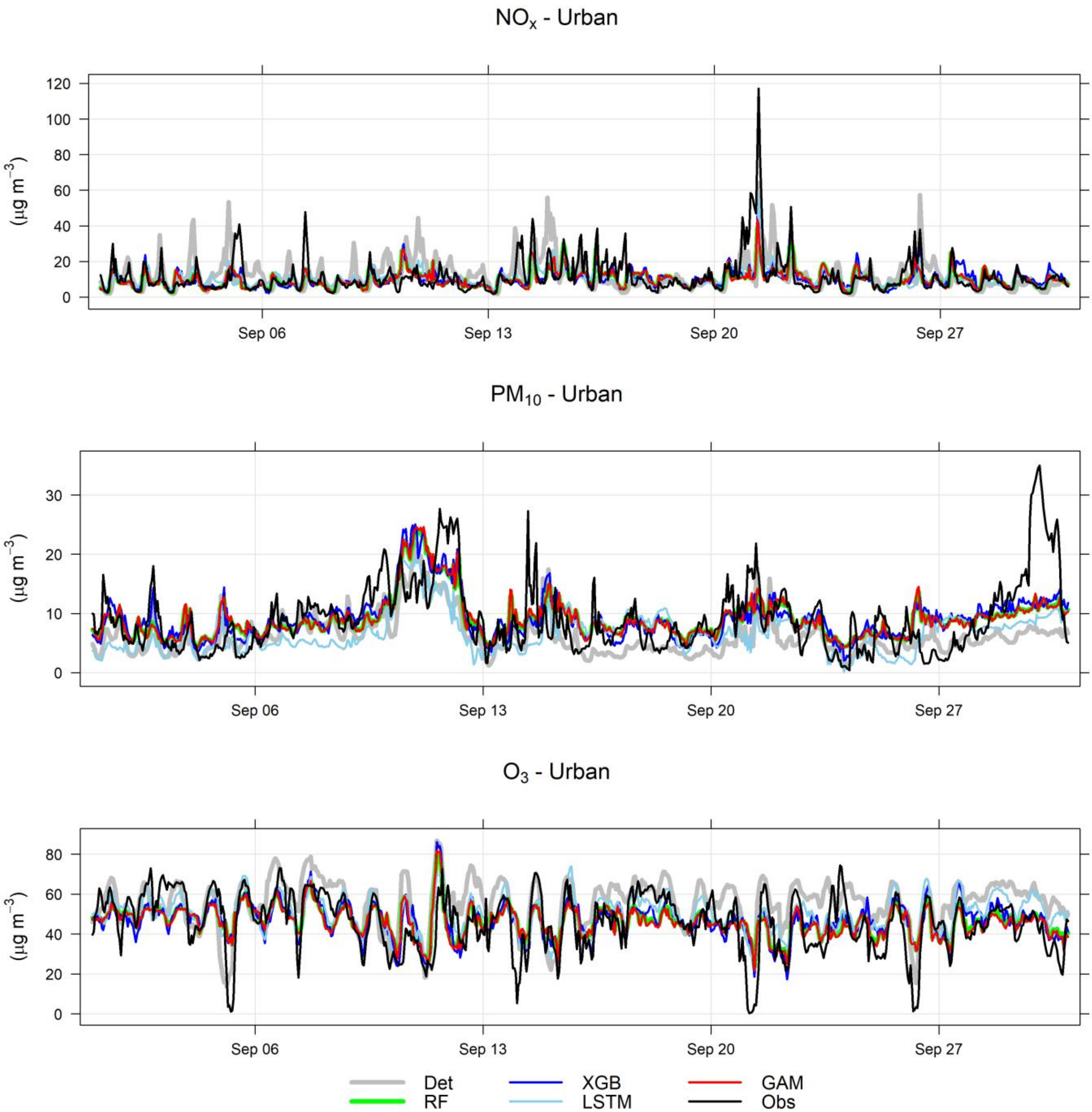


Figure C1. Temporal variations of deterministic and ML forecasted NO_x, PM_{10} and O_3 concentrations together with corresponding measured concentrations at the urban background site for September 2021. Mean of 1-, 2- and 3-day forecasts.

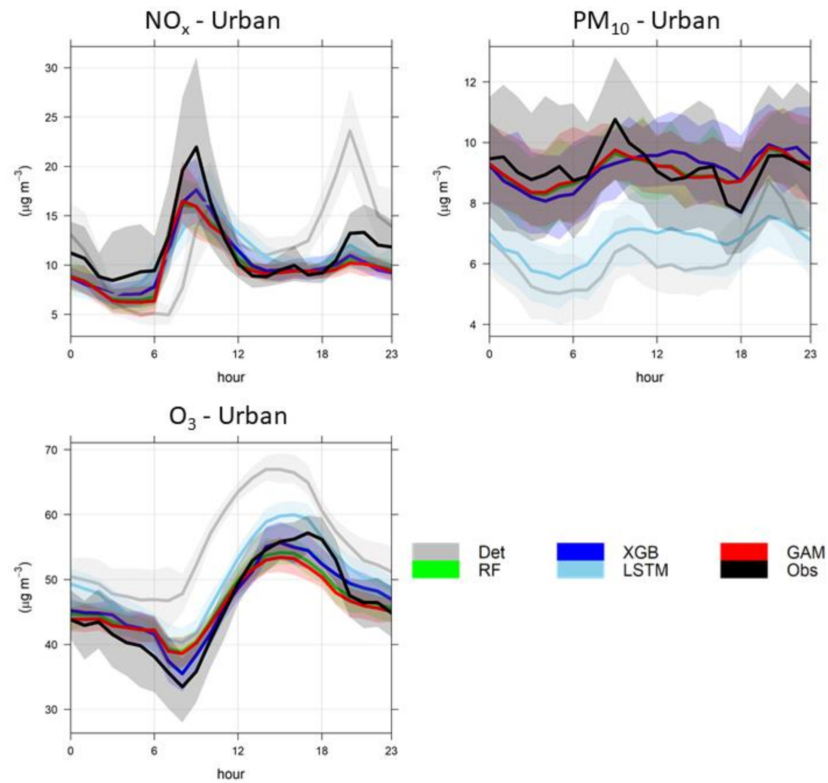


Figure C2. Mean diurnal variations in measured and forecasted concentrations of NO_x, PM_{10} and O_3 at the urban site. Mean of 1-, 2- and 3-day forecasts for August – December 2021.

Appendix D. Statistical performance measures for forecasts higher than the hourly mean concentrations at the urban site.

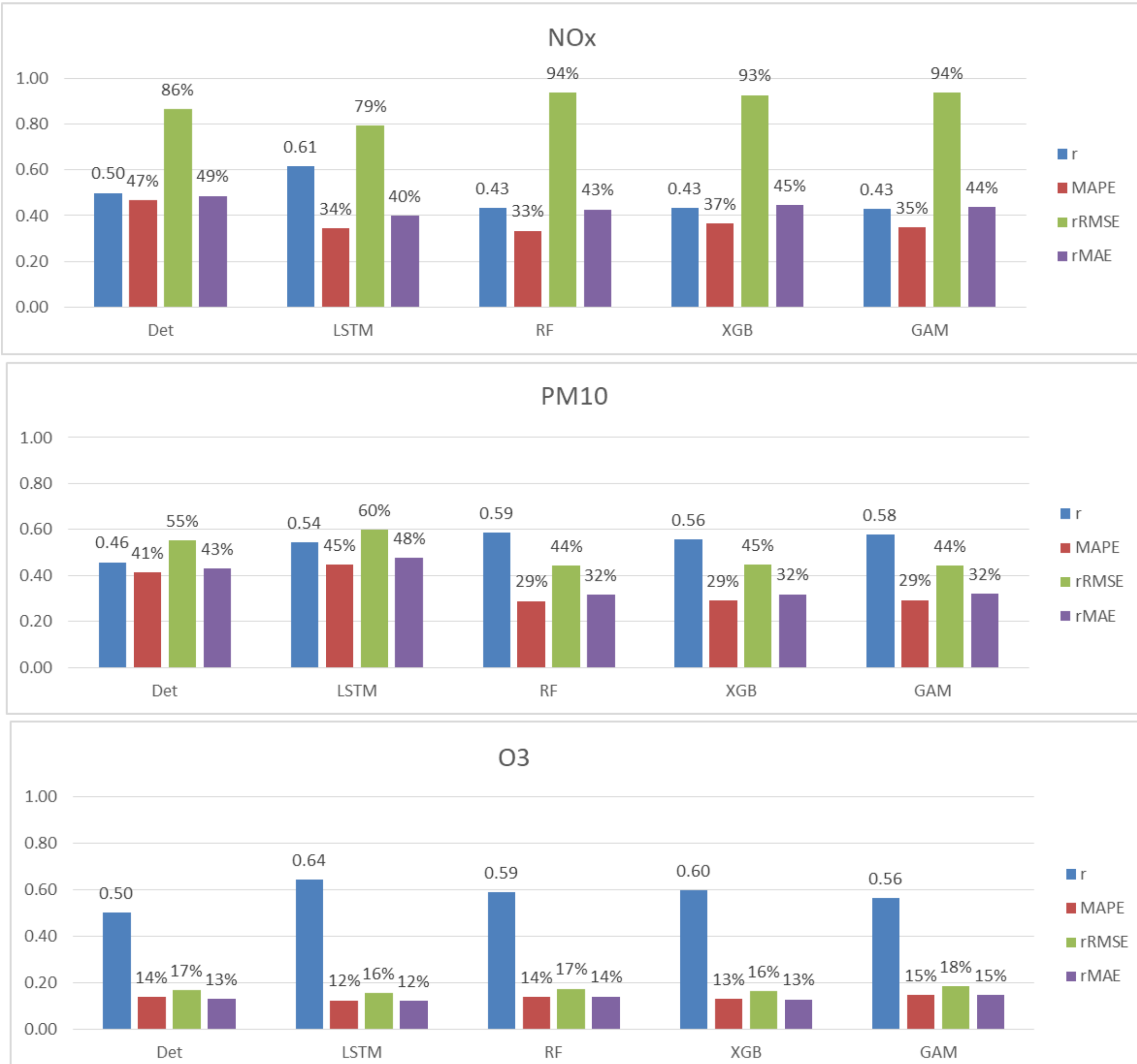


Figure D1. Statistical performance measures for concentrations of NO_x, PM_{10} and O_3 higher than the hourly mean value at the urban site. Mean of 1-, 2- and 3-day forecasts.

9 Appendix E. Importance of features – Street canyon sites

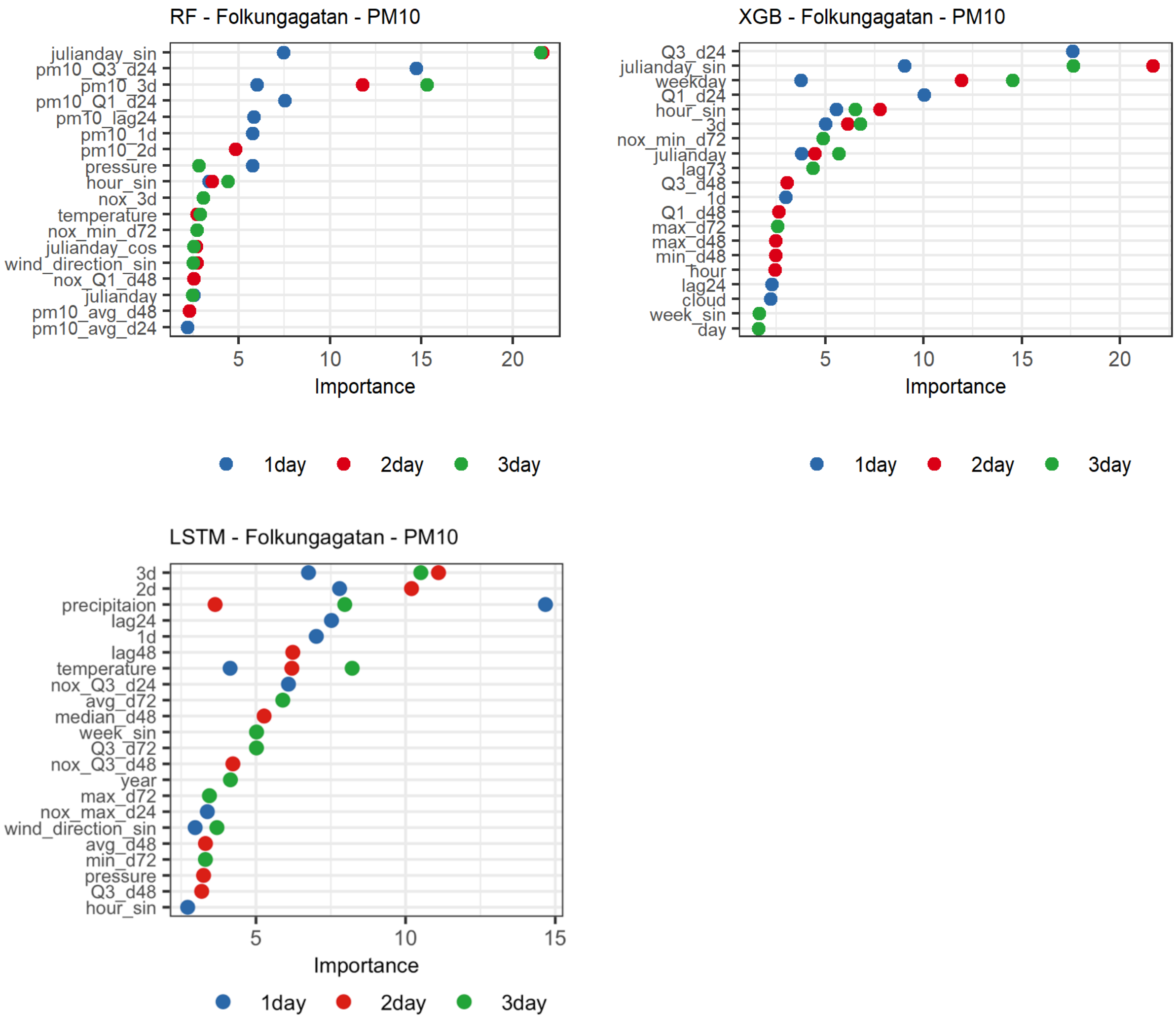


Figure E1. Most important features (%) for PM_{10} forecasts using RF, XGB and LSTM at Folkungagatan.

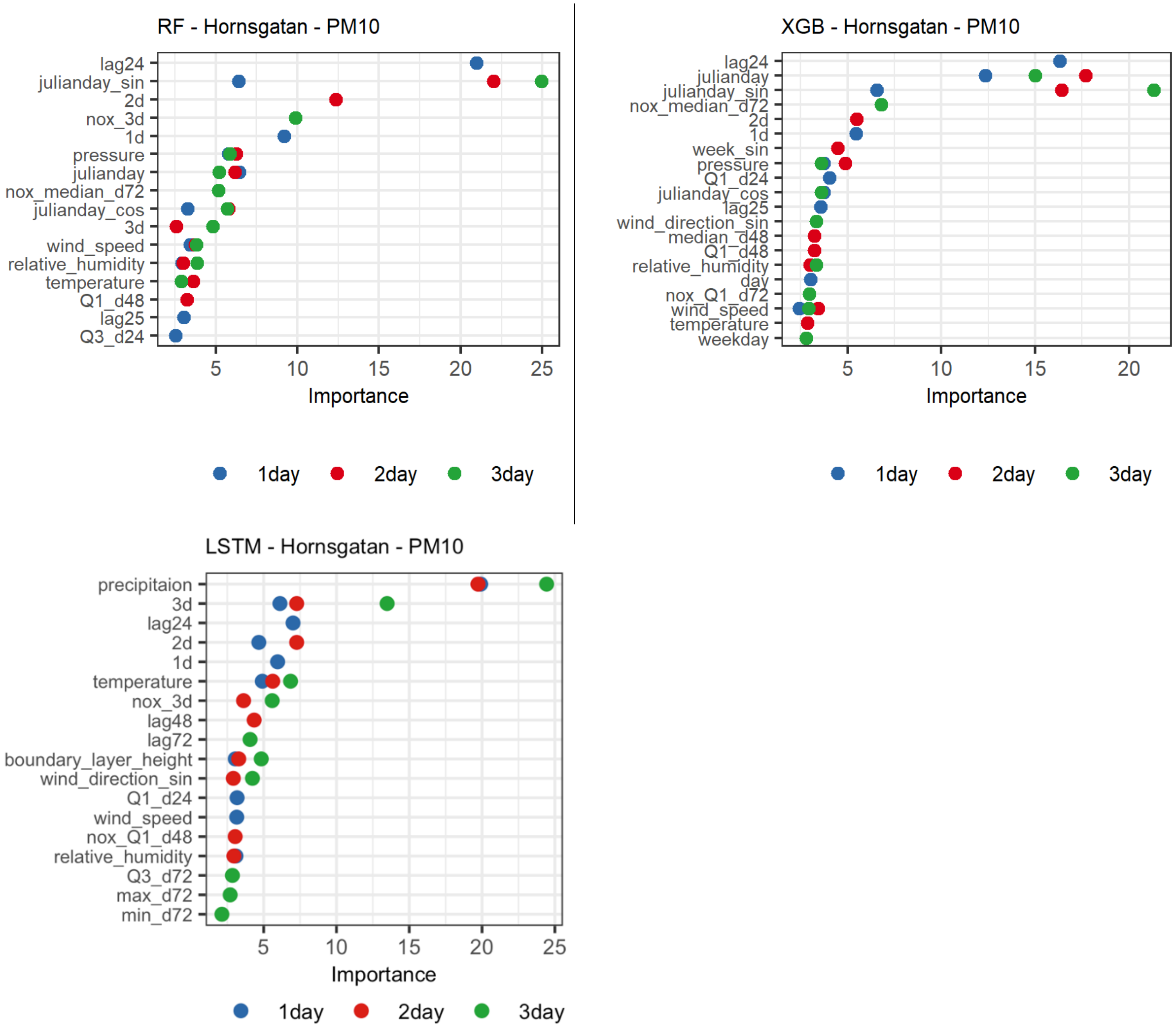


Figure E2. Most important features (%) for PM_{10} forecasts using RF, XGB and LSTM at Hornsgatan.

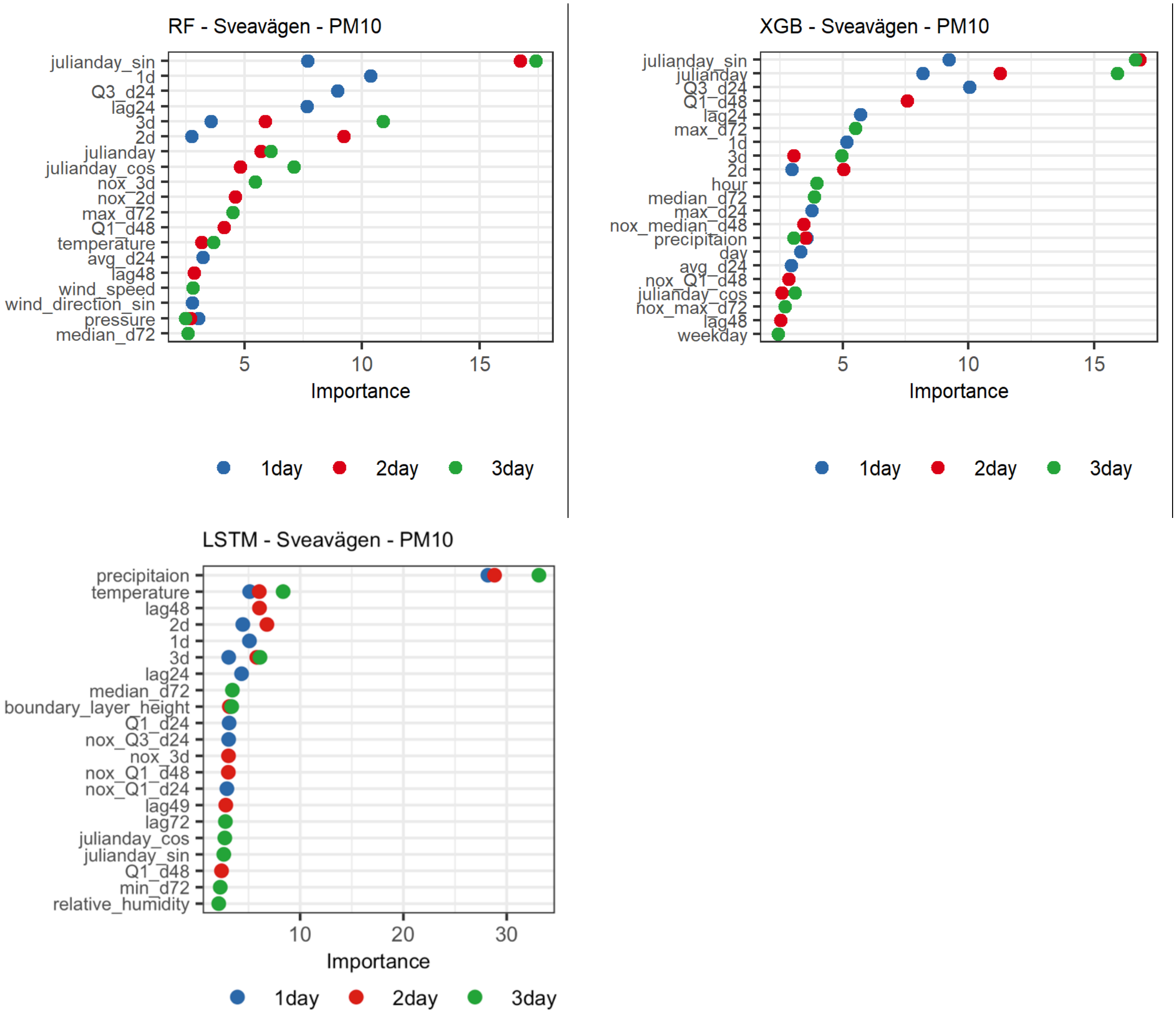


Figure E3. Most important features (%) for PM_{10} forecasts using RF, XGB and LSTM at Sveavägen.

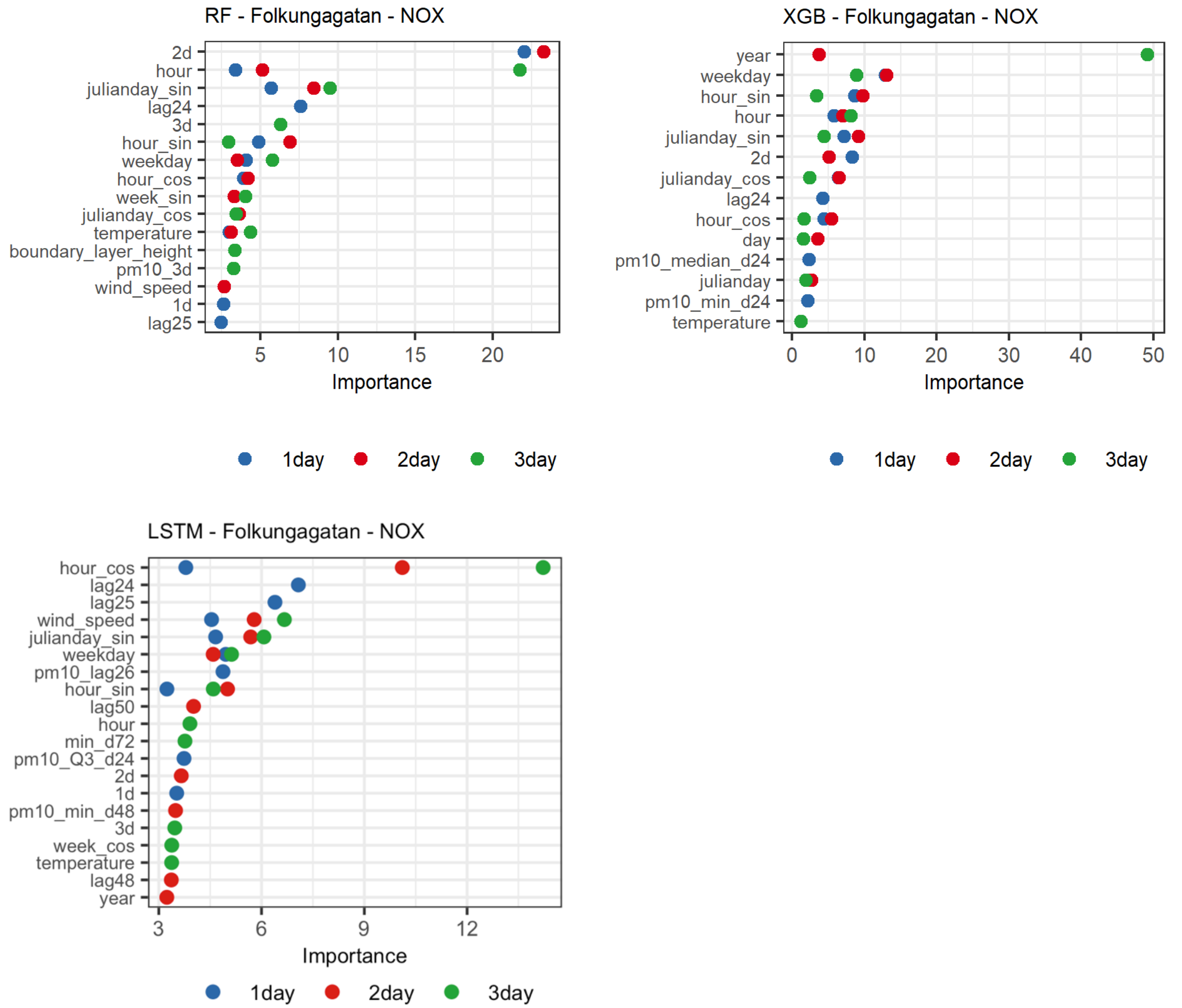


Figure E4. Most important features (%) for NO_x forecasts using RF, XGB and LSTM at Folkungagatan.

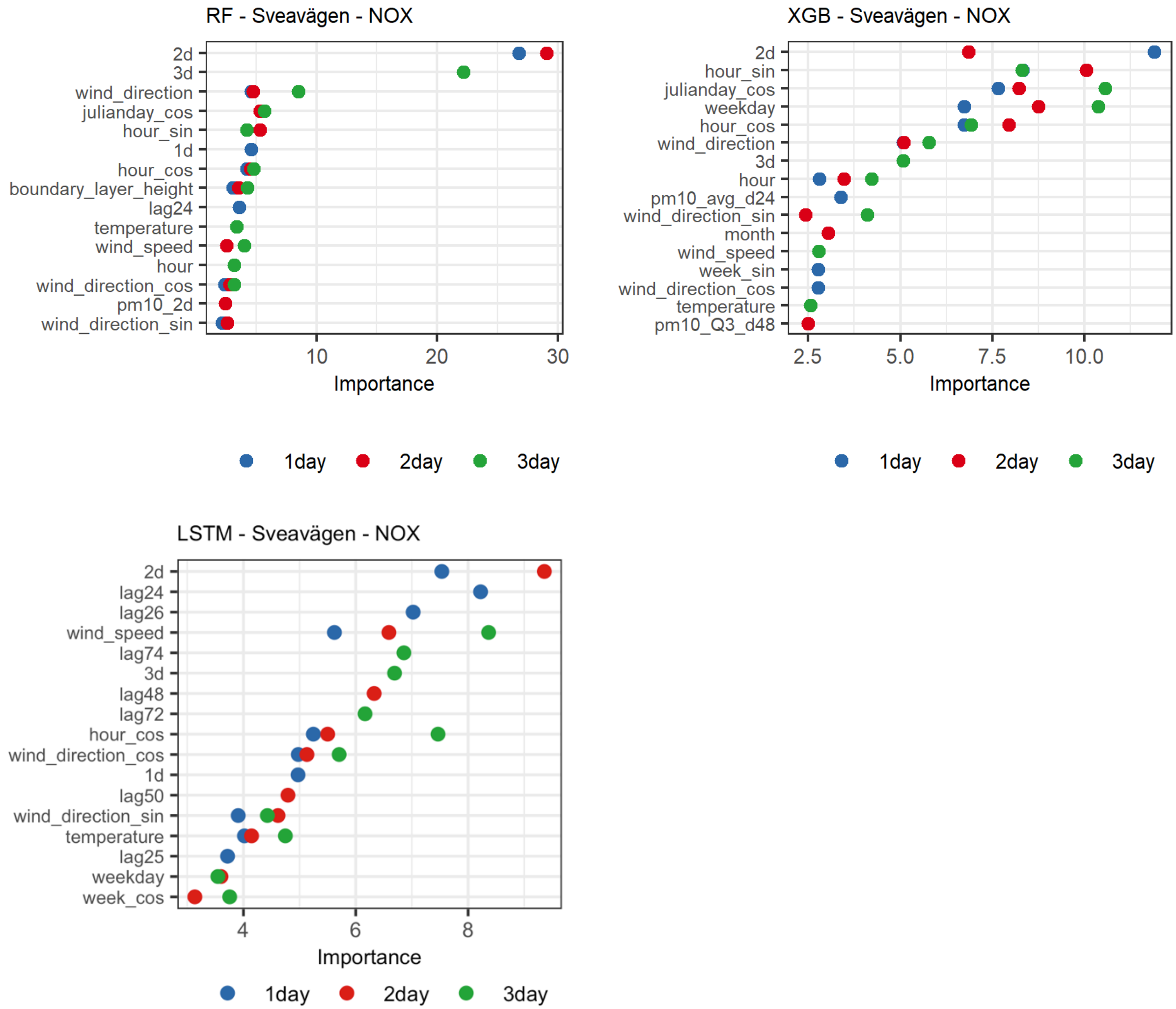


Figure E5. Most important features (%) for NO_x forecasts using RF, XGB and LSTM at Sveavägen.

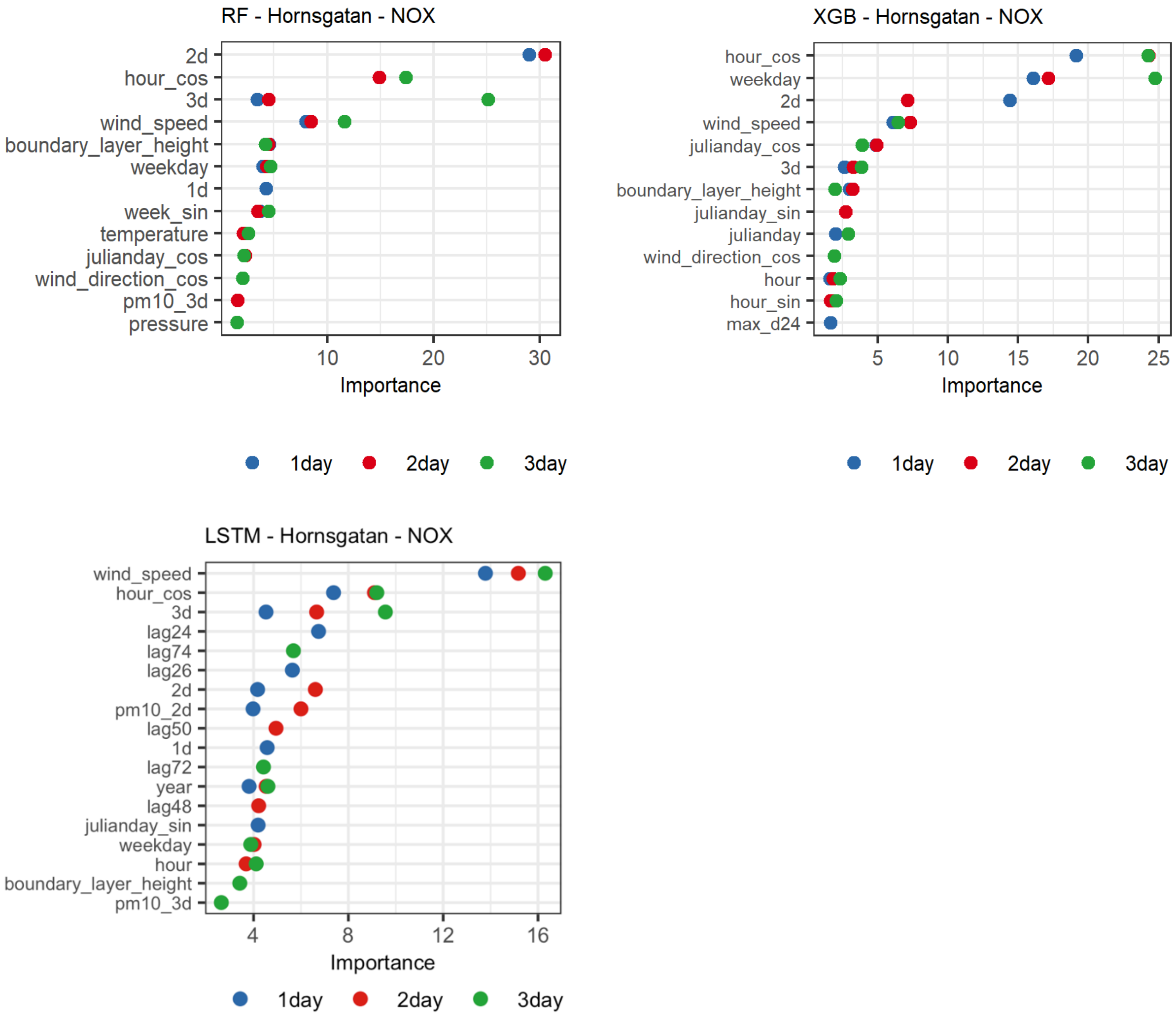


Figure E6. Most important features (%) for NO_x forecasts using RF, XGB and LSTM at Hornsgatan.

10 Appendix F. Temporal variations in hourly mean O_3, NO_x and PM_{10} concentrations at the street canyon sites

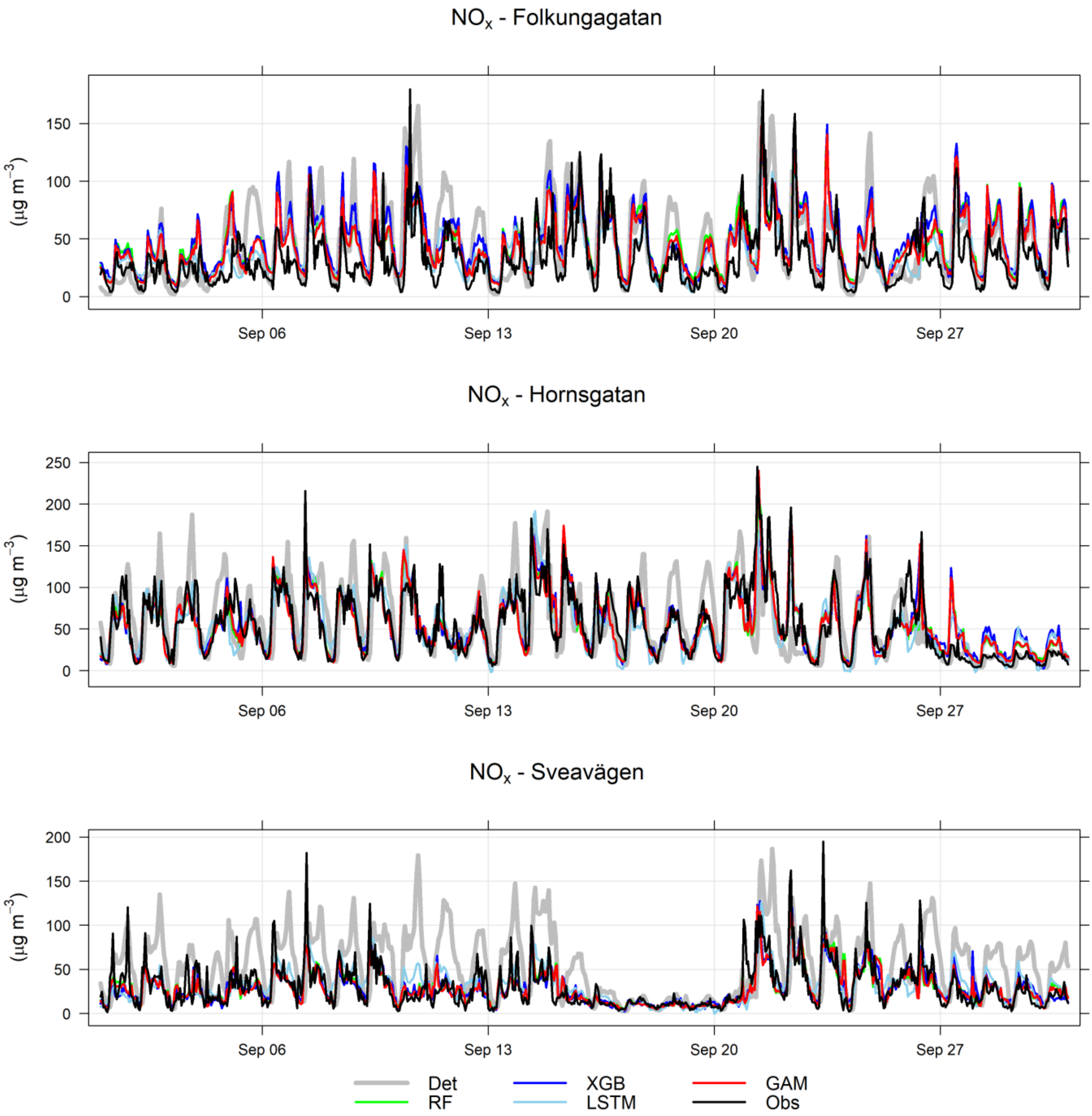


Figure F1. Temporal variations of hourly deterministic and ML forecasted NO_x concentrations together with corresponding measured concentrations at street canyon sites for September 2021. Mean of 1-, 2- and 3-day forecasts.

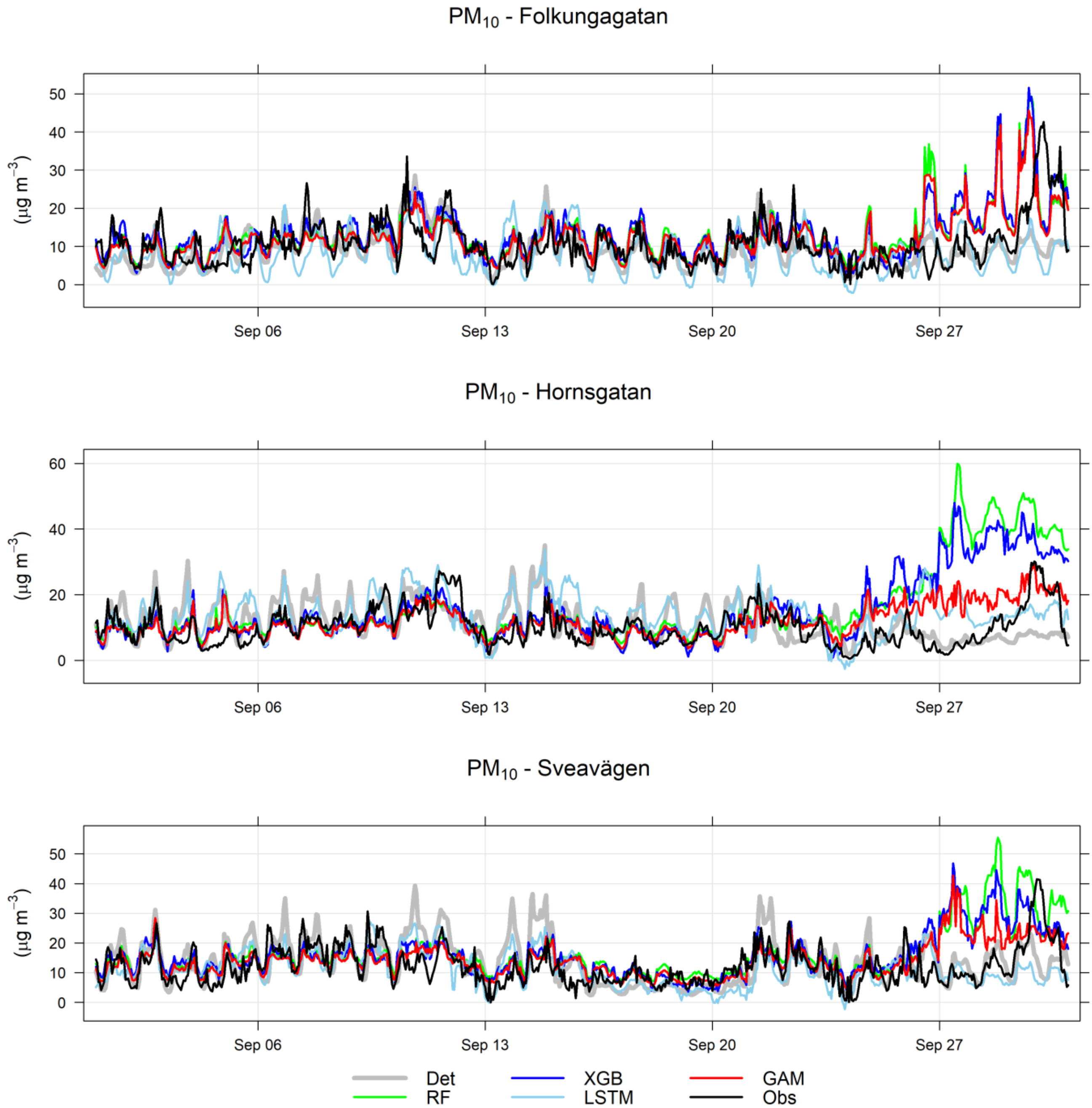


Figure F2. Temporal variations of hourly deterministic and ML forecasted PM_{10} concentrations together with corresponding measured concentrations at the street canyon sites for September 2021. Mean of 1-, 2- and 3-day forecasts.

11 Appendix G. Mean diurnal variations in hourly mean observations, 1-day, 2-day and 3-day deterministic and ML forecasts of NO_x and PM_{10} for the street canyon sites.

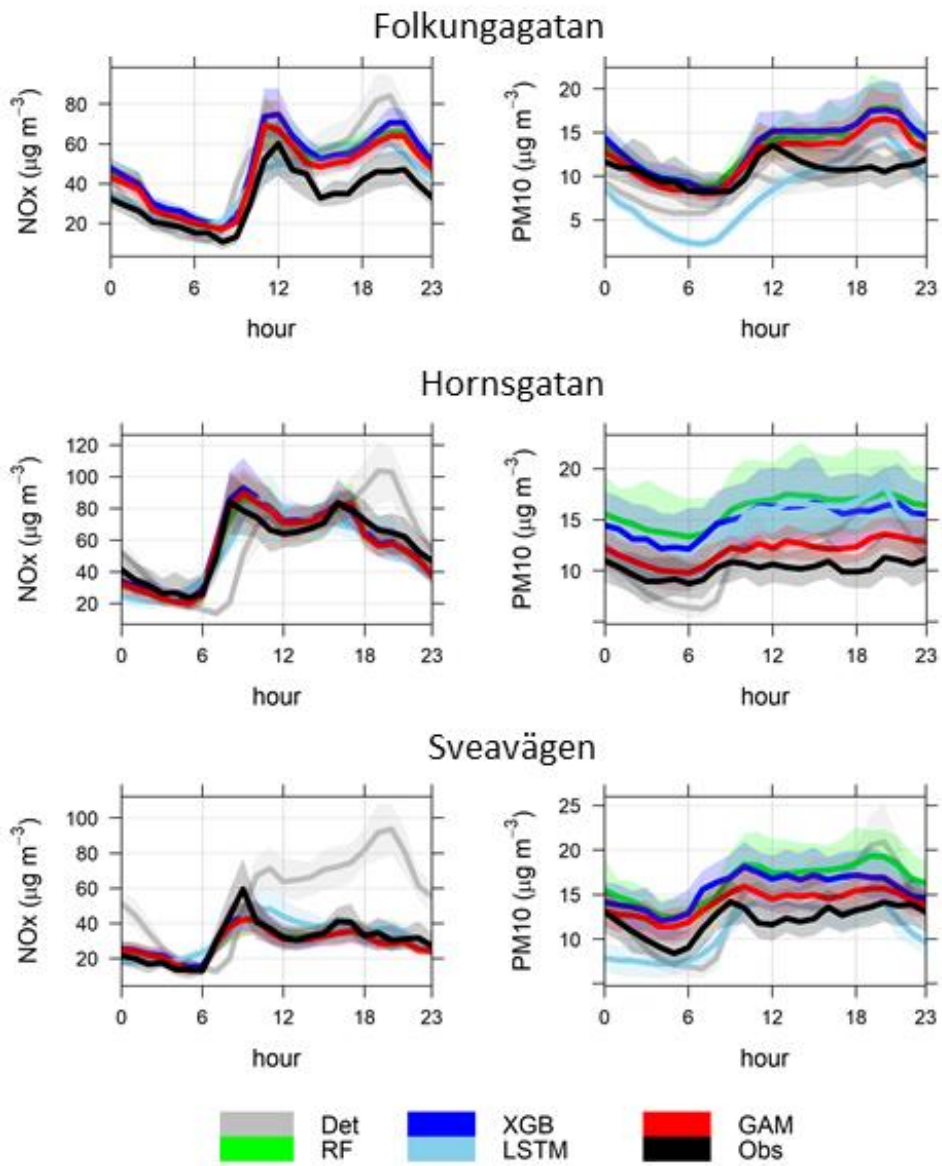


Figure G1. Mean diurnal variations in measured and forecasted concentrations of NO_x and PM_{10} at the street canyon sites. Mean of 1-, 2- and 3-day forecasts for August – December 2021.Shaded areas are 95% confidence intervals.

12 Appendix H. Statistical performance measures for forecasted hourly mean concentrations higher than the mean values at the street canyon sites.

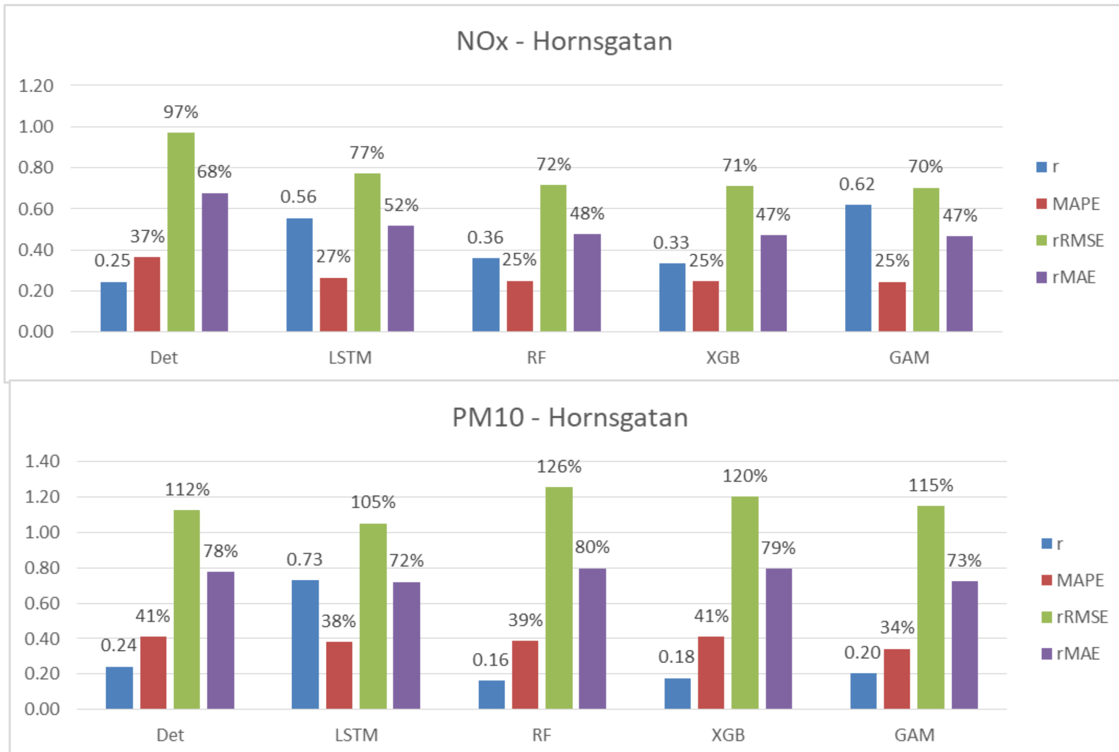


Figure H1. Statistical performance measures for forecasted NO_x and PM_{10} hourly mean concentrations higher than the mean values at Hornsgatan. Mean of 1-, 2- and 3-day forecasts.

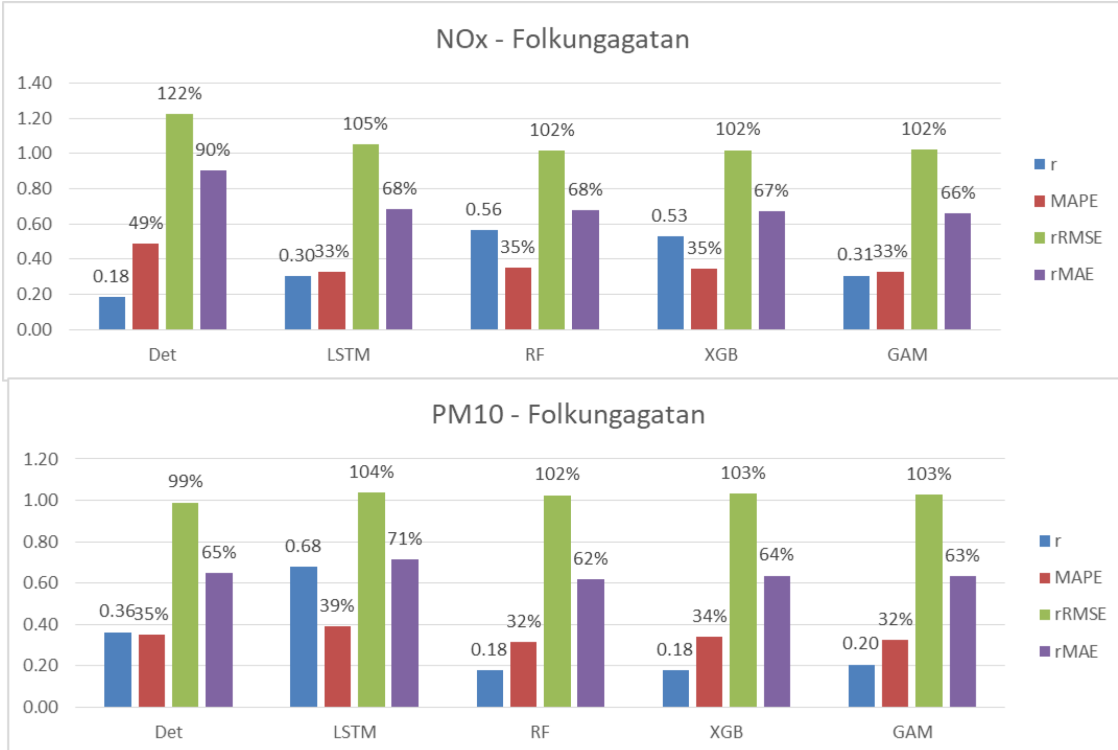


Figure H2. Statistical performance measures for forecasted NO_x and PM_{10} hourly mean concentrations higher than the mean values at Folkungagatan. Mean of 1-, 2- and 3-day forecasts.

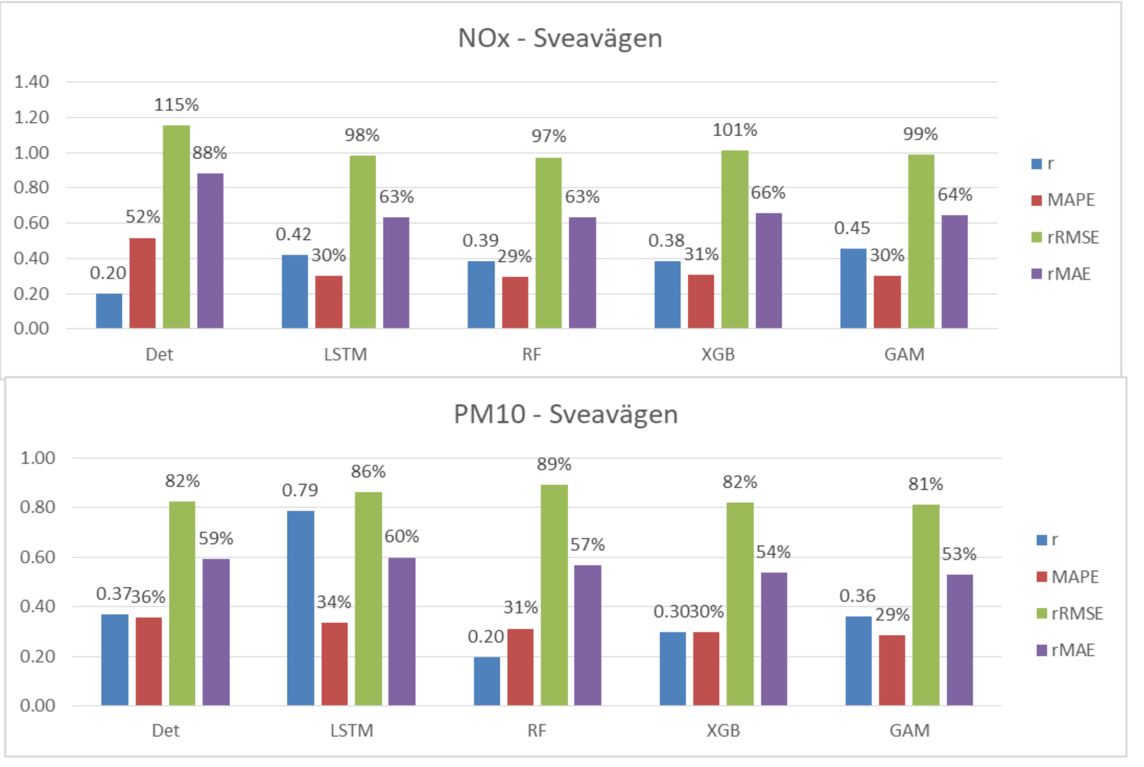


Figure H3. Statistical performance measures for forecasted NO_x and PM_{10} hourly mean concentrations higher than the mean values at Sveavägen. Mean of 1-, 2- and 3-day forecasts.

Code/Data availability: Python codes and data are available here: https://zenodo.org/record/7576042#.Y9k3AXbMK71 .

Author contribution: ME has been responsible for the deterministic modelling and providing with monitoring data and meteorological forecasts. ZZ and XM has been responsible for the ML modelling and statistical calculations. CJ, XM and ME initiated and planned the project. All authors have contributed to analysing data and writing of the manuscript.

Competing interests: The authors declare that they have no conflict of interest.

Acknowledgements: Financial support: The project was funded by ICT – The next generation and Digital future at KTH Royal Institute of Technology (contract VF 2021-0082).

References

Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K., Muller, K-R.: How to Explain Individual Classification Decisions. Journal of Machine Learning Research 11, 1803-1831, 2010.

Berkowicz, R.: OSPM - A parameterised street pollution model, Environmental Monitoring and Assessment, 65, 323-331, 2020.

Breiman, L.: "Random forests." Machine learning 45, 5-32, 2001.

Brokamp, C., Jandarov, R., Rao, M.B., LeMasters, G., Ryan, P.: Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. Atmos. Environ, 151, 1–11, 2017.

Burman, L., Johansson, C.: Emissions and Concentrations of Nitrogen Oxides and Nitrogen Dioxide on Hornsgatan Street, Evaluation of Traffic Measurements during Autumn 2009 (In Swedish Only). SLB Report 7. https://www.slb.nu/slb/rapporter/pdf8/slb2010_007.pdf, 2010.

Burman, L., Elmgren, M., Norman, M.: Fordonsmätningar på Hornsgatan år 2017. https://scholar.google.com/scholar_lookup?title=Fordonsm%C3%A4tningar%20P%C3%A5%20Hornsgatan%20%C3%85r%202017&author=L.%20Burman&publication_year=2019, 2019.

Cai, M., Yin, Y., Xie, M.: Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. Transport Research Part D. 14, 32-41. doi:10.1016/j.trd.2008.10.004, 2009.

Carslaw, D.C. and K. Ropkins.: Openair — an R package for air quality data analysis, Environmental Modelling & Software, 27-28, 52–61, 2012.

Castelli, M., Clemente, F.M., Popovič, A., Silva, S. and Vanneschi, L.: A Machine Learning Approach to Predict Air Quality in California. Hindawi, Complexity, Article ID 8049504, 23 pages, https://doi.org/10.1155/2020/8049504, 2020.

Chuluunsaikhan, T., Heak, M., Nasridinov, A., Choi, S.: Comparative Analysis of Predictive Models for Fine Particulate Matter in Daejeon, South Korea. Atmosphere, 12, 1295. https://doi.org/10.3390/atmos12101295, 2021.

Czernecki, B., Marosz, M., Jędruszkiewicz, J.: Assessment of Machine Learning Algorithms in Short-term Forecasting of PM_{10} and PM2.5 Concentrations in Selected Polish Agglomerations. Aerosol and Air Quality Research. 21, 200586, https://doi.org/10.4209/aaqr.200586, 2021.

Denby, B. R., Sundvor, I., Johansson, C., Pirjola, L., Ketzel, M., Norman, M., Kupiainen, K., Gustafsson, M., Blomqvist, G., Omstedt, G.: A coupled road dust and surface moisture model to predict non-exhaust road traffic induced particle emissions (NORTRIP). Part 1: road dust loading and suspension modelling Atmos. Environ., 77, 283-300, 2013a.

Denby, B. R., Sundvor, I., Johansson, C., Pirjola, L., Ketzel, M., Norman, M., Kupiainen, K., Gustafsson, M., Blomqvist, G., Omstedt, G.: A coupled road dust and surface moisture model to predict non-exhaust road traffic induced particle emissions (NORTRIP). Part 2: surface moisture and salt impact modelling Atmos. Environ., 81, 485-503, 2013b.

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L.J., Schwartz. J.: An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environment International 130, 104909, 2019.

Doreswamy, Harishkumar K S., Yogesh, K.M., Gad, I.: Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models. Procedia Computer Science 171, 2057–2066, 2020.

Engardt, M., Bergström. S. and Johansson, C.: Luften du andas - nu och de kommande dagarna. Utveckling av ett automatiskt prognossystem för luftföroreningar och pollen. SLB 36:2021, 33 pp. (In Swedish). https://www.slbanalys.se/slb/rapporter/pdf8/slb2021_036.pdf, 2021.

Fuller, R., Philip J Landrigan, Kalpana Balakrishnan, Glynda Bathan, Stephan Bose-O'Reilly, Michael Brauer, Jack Caravanos, Tom Chiles, Aaron Cohen, Lilian Corra, Maureen Cropper, Greg Ferraro, Jill Hanna, David Hanrahan, Howard Hu, David Hunter, Gloria Janata, Rachael Kupka, Bruce Lanphear, Maureen Lichtveld, Keith Martin, Adetoun Mustapha, Ernesto Sanchez-Triana, Karti Sandilya, Laura Schaefli, Joseph Shaw, Jessica Seddon, William Suk, Martha María Téllez-Rojo, Chonghuai Yan.: Pollution and health: a progress update. Lancet Planet Health, 6, e535–47, https://doi.org/10.1016/S2542-5196(22)00090-0, 2022.

Gidhagen, L., Johansson, C., Langner, J., Foltescu, V. L.: Urban scale modeling of particle number concentration in Stockholm. Atmospheric Environment 39, 1711–1725, 2005.

Hoek, G., Beelen, R.,, de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D.: A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ, 42, 7561-7568, doi:10.1016/j.atmosenv.2008.05.057, 2008.

Hong, H.; Choi, I.; Jeon, H.; Kim, Y.; Lee, J.-B.; Park, C.H.; Kim, H.S.: An Air Pollutants Prediction Method Integrating Numerical Models and Artificial Intelligence Models Targeting the Area around Busan Port in Korea. Atmosphere 13, 1462, 2022.

Horálek, J., Hamer, P., Schreiberová, M., Colette, A., Schneider, P., Malherbe, L.: Potential use of CAMS modelling results in air quality mapping under ETC/ATNI. Eionet Report – ETC/ATNI 2019/17, ISBN 978-82-93752-21-9, 2019.

Iskandaryan, D., Ramos, F. and Trilles, S.: Air Quality Prediction in Smart Cities Using Machine Learning Technologies based on Sensor Data: A Review. Appl. Sci. 2020, 10, 2401, doi: 10.3390/app10072401, 2020.

Janssen, S. and Thunis, P.: FAIRMODE Guidance Document on Modelling Quality Objectives and Benchmarking (version 3.3), EUR 31068 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-52425-0, doi:10.2760/41988, JRC129254, 2022.

Johansson, C., Norman, M., Gidhagen, L.: Spatial & temporal variations of PM_{10} and particle number concentrations in urban air. Environ. Monit. Assess. 127, 477–487, 2007.

Johansson, C., Burman, L., Forsberg, B.: The effects of congestions tax on air quality and health. Atmos. Environ. 43, 4843-4854, 2009.

Johansson, C., Eneroth, K., Lövenheim, B., Silvergren, S., Burman, L., Bergström, S., Norman, M., Engström Nylén, A., Hurkmans, J., Elmgren, M., Brydolf, M., Täppefur, M.: Luftkvalitetsberäkningar för kontroll av miljökvalitetsnormer (with summary in English). SLB 11:2017 ver 2. https://www.slbanalys.se/slb/rapporter/pdf8/slb2017_011.pdf, 2017.

Johansson, C. Lövenheim, B.; Schantz, P.; Wahlgren, L.; Almström, P.; Markstedt, A.; Strömgren, M.; Forsberg, B.; Nilsson Sommar, J.: Impacts on air pollution and health by changing commuting from car to bicycle. Sci. Total Environ. 584-585, 55-63, 2017.

Joharestani, M.Z., Cao, C., Ni, X., Bashir, B. Talebiesfandarani, S.: PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. Atmosphere, 10, 373; doi:10.3390/atmos10070373, 2019.

Kamińska, J. A.: A random forest partition model for predicting NO_2 concentrations from traffic flow and meteorological conditions. Science of the Total Environment 651, 475–483, 2019.

Keller, M., Hausberger, S., Matzer, C., Wüthrich, P., Notter, B.: HBEFA 3.3. Update of NO_x Emission Factors of Diesel Passenger Cars- Background Documentation. https://www.hbefa.net/e/documents/HBEFA33_Documentation_20170425.pdf, 2017.

Kleinert, F., Leufen, L. H., Lupascu, A., Butler, T., and Schultz, M. G.: Representing chemical history in ozone time-series predictions – a model experiment study building on the MLAir (v1.5) deep learning framework, Geosci. Model Dev., 15, 8913–8930, https://doi.org/10.5194/gmd-15-8913-2022, 2022.

Krecl, P., Harrison, R.M., Johansson, C., Targino, A.C., Beddows, D.C., Ellermann, T., Lara, C. and Ketzel, M.: Long-term trends in nitrogen oxides concentrations and on-road vehicle emission factors in Copenhagen, London and Stockholm. Environmental Pollution, 290, 118105, 2021.

Krecl, P., Johansson, C., Targino, A.C.,, Ström, J., Burman, L.: Trends in black carbon and size-resolved particle number concentrations and vehicle emission factors under real-world conditions. Atmospheric Environment, 165, 155-168, 2017.

Lee, Y.-G.; Lee, P.-H.; Choi, S.-M.; An, M.-H.; Jang, A.-S.: Effects of Air Pollutants on Airway Diseases. Int. J. Environ. Res. Public Health 18, 9905, 2021.

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., Kadygrov, 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, ¡., van Velthoven, P., van Versendaal, R.,Vira, J. and Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production. Geoscientific Model Development, Volume 8, issue 9, 2777–2813, 2015.

Meteo-France for Copernicus.: 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, 2017.

Munir, S., Mayfield, M., Coca, D., Mihaylova, L.S. and Osammor, O.: Analysis of Air Pollution in Urban Areas with Airviro Dispersion Model—A Case Study in the City of Sheffield, United Kingdom. Atmosphere 11, 285; doi:10.3390/atmos11030285, 2020.

Olstrup, H., Johansson, C., Forsberg, B., Åström, C.: Association between Mortality and Short- Term Exposure to Particles, Ozone and Nitrogen Dioxide in Stockholm, Sweden. Int J Environ Res Public Health, 16, 6, 1028-1042, 2019.

Orru, H. Lövenheim, B. Johansson, C. Forsberg, B.: Estimated health impacts of changes in air pollution exposure associated with the planned by-pass Förbifart Stockholm. J Expo Sci Environ Epidemiol, 1-8, 2015.

Ottosen, T.-B. and Kakosimos, K. E. and Johansson, C. and Hertel, O. and Brandt, J. and Skov, H. and Berkowicz, R. and Ellermann, T. and Jensen, S. S. and Ketzel, M.: Analysis of the impact of inhomogeneous emissions in the Operational Street Pollution Model (OSPM). Geoscientific Model Development, 8, 3231—3245, 2015.

Qadeer, K., Rehman, W.U., Sheri, A.M., Park, I., Kim, H.K. and Jeon, M.: A Long Short-Term Memory (LSTM) Network for Hourly Estimation of PM2.5 Concentration in Two Cities of South Korea. Appl. Sci. 2020, 10, 3984, doi:10.3390/app10113984, 2020.

Rybarczyk, Y. and Zalakeviciute, R.: Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review. Appl. Sci. 2018, 8, 2570, doi:10.3390/app8122570, 2018.

Säll, B.: Evaluation and validation of Copernicus Atmosphere Monitoring Service regional ensemble forecast of air pollutants and birch pollen in the Stockholm region. Master thesis report 30 HP (MO9001). Department of Meteorology, Stockholm university, 2018.

Shaban, K., B.,Kadri, A., and Rezk, E.: Urban Air Pollution Monitoring System With Forecasting Models. IEEE sensors Journal, 16, 2598-2606, 2016.

Shtein, A., Kloog, I., Schwartz, J., Silibello, C., Michelozzi, P., Gariazzo, C., Viegi, G., Forastiere, F., Karnieli, A., Just, A.C. and Stafoggia, M.: Estimating Daily PM2.5 and PM_{10} over Italy Using an Ensemble Model. Environmental Science & Technology 54, 120-128 DOI: 10.1021/acs.est.9b04279, 2020.

SLB, Methods for calculating air pollution concentrations in relation to the limit values. Report in Swedish with summary in English. Environment and Health Administration of Stockholm, SLB analys, Box 8136, 104 20 Stockholm, Sweden, Report nr. 50:2021. https://www.slbanalys.se/slb/rapporter/pdf8/slb2021_050.pdf, accessed 30 November, 2022.

Stafoggia, M., Johansson, C., Glantz, P., Renzi, M., Shtein, A., de Hoogh, K., Kloog, I., Davoli, M., Michelozzi, P., Bellander, T.: A Random Forest Approach to Estimate Daily Particulate Matter, Nitrogen Dioxide, and Ozone at Fine Spatial Resolution in Sweden. Atmosphere, 11, 239, 1-19, 2020.

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., de Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I., Schwartz, J.: Estimation of daily PM_{10} and PM2.5 concentrations in Italy, 2013-2015, using a spatiotemporal land-use random-forest model. Environ. Int., 124, 170−179, 2019.

Thongthammachart, T., Araki, S., Shimadera, H., Eto, S., Matsuo, T. and Kondo, A.: An integrated model combining random forests and WRF/CMAQ model for high accuracy spatiotemporal PM2.5 predictions in the Kansai region of Japan. Atmospheric Environment 262, 118620, 2021.

Zaini, N., Ean, L.W., Ahmed, A.N., Malek, M.A.: A systematic literature review of deep learning neural network for time series air quality forecasting. Environmental Science and Pollution Research, https://doi.org/10.1007/s11356-021-17442-1, 2021.