HAQT - Helsinki metropolitan Air Quality Testbed

D3-2: Impact and benefit of HAQT network on ENFUSER AQ modelling system

Lasse Johansson

13/05/2019

Table of Contents

Summary

Using the FMI-ENFUSER model the benefits of having several affordable AQ sensors to complement the existing measurement network were evaluated. In the evaluation data for the year 2018 were used, including meteorological data from multiple sources, regional scale modelled background and modelled emission inventories.

Based on the study the complementary sensors are able to deliver usable information on the air quality in many ways that the significantly more costly reference-level AQ stations do on the modelling perspective. The hourly variability of sensor data is high and reliability is significantly lower, yet with emphasis being put to longer term data fusion the emission source appointment patterns emerge from the data. The added benefits for the sensor data in HAQT were seen to be the largest for PM10 and the smallest for $PM_{2.5}$. The successful use for sensor data in this study can be considered as a strong result; with further sensor technology and modelling development the combination of measurement stations and complementary sensor networks should provide the most cost effective means for extending the measurement network.

In general, in Helsinki region specifically there is a limited potential for obtaining modelling performance increases just by increasing the amount of measurement locations since the existing reference quality network is already quite extensive. However, the current operational modelling system has limited measurement network for PM_{10} and O_3, in which the additional sensor network can assist.

During the study it was concluded that several pre-processing steps need to be made before the volatile sensor data can be effectively utilized in real-time modelling. Beneficial impacts on the modelling performance were observed when acceptance limits were used to filter out clearly unusable data. Secondly, the modelling system can be used to automatically calibrate sensor offset correction, which was seen to be a near mandatory step based on the simulations. Finally, a significant temporal drifting of the sensor measurement quality during the study period was observed for the gaseous components. During the HAQT project, the data fusion algorithm of ENFUSER was extended to quantify the reliability of sensors over time, and in the future information of this kind could be used for the effective management of the sensor network.

The addition of $PM_{2.5}$ sensor data were seen to reduce the modelling accuracy in Helsinki area. The reason for this result is most likely the dependency of measurement quality and ambient meteorological conditions, which causes cross-correlated errors to be incorporated in the sensor measurements. In its current state the data fusion algorithm is not able to function properly in such cases.

This report contains unpublished material that are planned to be used for a peer reviewed journal article.

Modelling system

The FMI-ENFUSER is an operational, adaptive local-scale dispersion model used in Helsinki region as a part of HAQT project. The model utilizes real-time measurement data to fine tune emission factors and dispersion modelling parameters. The modelling system has been designed to predict hourly pollutant concentrations (SO_2, CO, NO NO_2, O_3, $PM_{2.5}$, and PM_{10}) and air quality index (AQI) in the area with a resolution of 13 x 13 meters. Due to the computational limitations that a real-time operational AQ system presents, the dispersion modelling is performed with a combination of Gaussian puff and Gaussian plume dispersion methodology. In the local scale (0-300m distance from emission sources, Gaussian plume modelling is used while the local urban morphology is taken into account with simplistic corrections (e.g., buildings, street canyons, vegetation). For emission contributions originating farther away dispersion modelling is performed using Gaussian puff modelling, in which the individual emission puffs have been made to follow wind trajectories according to the available NWP data (given by HIRLAM, GFS or ECMWF). The long-range transportation of pollutants is taken into account by nesting the local dispersion modelling on regional scale chemical transport model FMI-SILAM.

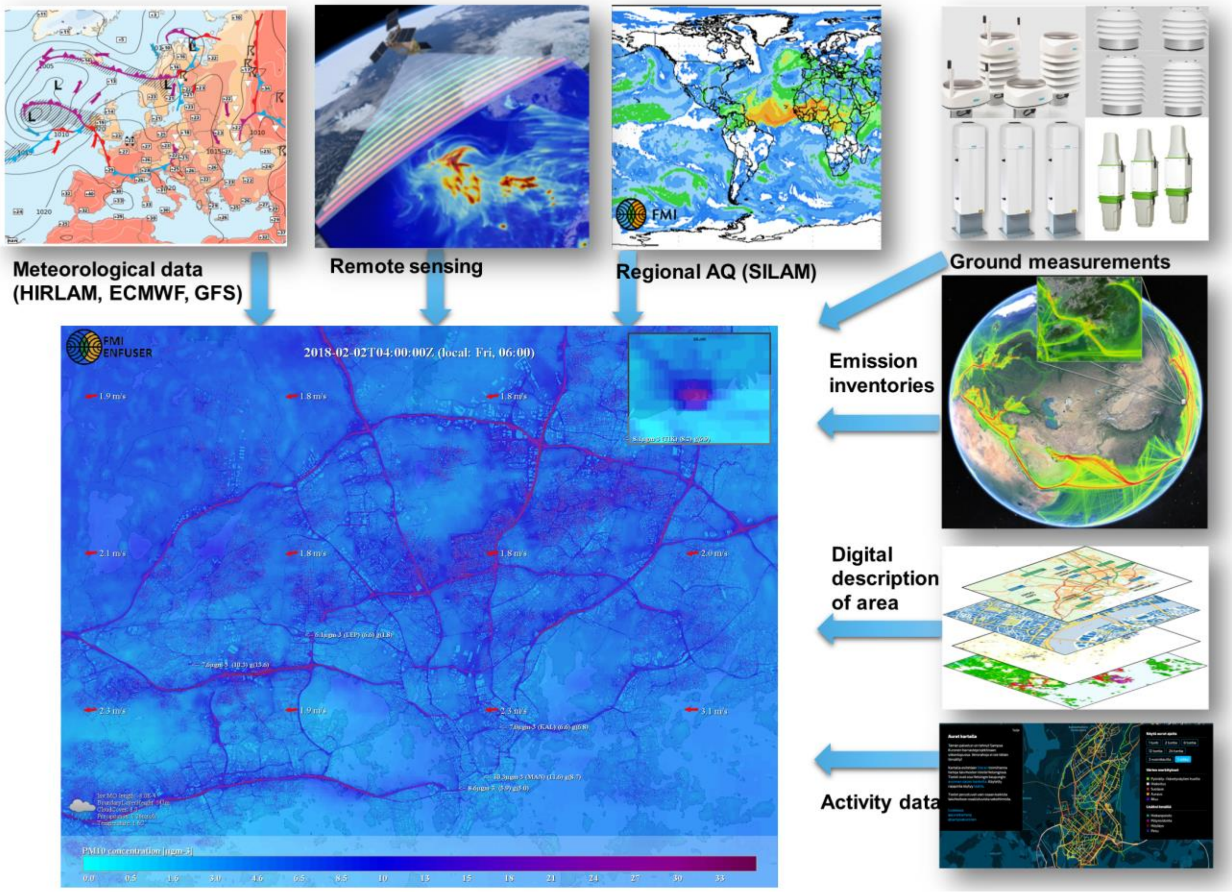

Figure 1: The core components of FMI-ENFUSER modelling system, which utilizes latest sensor observations, meteorological data and regional air quality forecast in the production of high resolution AQ heatmaps for the current and future situation in urban locations. GIS-datasets and emission inventories together with archived concentration time series are used for the calibration of the model.

The model relies mostly on public, open access information with global coverage and as such makes it possible to install and use the model abroad. The model is also currently being installed in Nanjing, China as a part of Nanjing Air Quality Testbed (NAQT, Business Finland).

In this document we focus on the modelling system for Helsinki. In this local configuration the background concentrations in which the local scale modelling is being nested comes from FMI-SILAM chemical transport model. Local measured pollutant concentrations are being extracted from FMI Open Data portal, which provides hourly average concentrations for the selected pollutant species; depending on species there are 5 to 12 measurement sites in the modelling area. Traffic emissions are modeled based on OpenStreetMap road data linked with hourly traffic flow estimates given by local authorities (HSL,HSY). Emission inventory for private households is also given by the local authorities, which describes $PM_{2.5}$ emissions using a multi-component representation (fireplaces, saunas and heating). Local power plants (for which there are only a few) have been dealt as elevated point sources. Shipping emissions are given by FMI-STEAM shipping emission model for the years of 2017 and 2018. Main sources of meteorological information are the HIRLAM model and approx. 10 road weather measurement stations in the area. Traffic congestion data are given by HERE.com traffic application interface (API). For improved road dust modelling, information on road maintenance actions with GPS-tracking is used. Additionally, forecasts and measurements for road weather and moisture are also being used (Digitraffic.fi). The modelling system can also utilize SENTINEL-5p TROPOMI remote sensing data; however, this capability is targeted mainly for Asian model installations in which the local emission sources are less known. More information on the modelling system can be found from https://en.ilmatieteenlaitos.fi/environmental-information-fusion-service. The operative modelling system AQ predictions can be seen from ilmanlaatukartta.hsy.fi. The model data that is updated each hour for a time period of 24h can be accessed from https://en.ilmatieteenlaitos.fi/open-data-manual-accessing-data.

In HAQT project 16 Vaisala AQT sensors were also installed in the modelling area. ENFUSER has real-time access to these sensor data via Vaisala NM10 service; however, for this study an offline collection of sensor measurements for duration of 10 months was delivered. The purpose of this study was to investigate how the addition of these sensor data affect the ENFUSER modelling performance. The sensor data are highly volatile and often regarded as unreliable with respect to reference-quality AQ measurement data, and an effective utilization of such sensor input requires data fusion, which is described in detail in Section 2.

Uncertainty considerations in Helsinki

Based on the previous modelling results (2017) and statics it is possible to distinguish the most notable source of error in the modelling for ENFUSER running in Helsinki metropolitan area. In general, the local emission sources in the area are relatively well known and the limitations of the used information sources used to describe the emissions source are also known. The used dispersion methodology is designed to be computationally effective but this effectiveness comes with a cost when dispersion modelling is applied over complex urban terrain. For many of the identified sources for error, increase in the amount of AQ measurements in the modelling area will facilitate model development that can result in better model predictions. However, these improvements to modelling accuracy do not occur automatically and requires further model development.

Uncertainties for emission sources

The most notable local emission source is the vehicular traffic, and therefore the prediction errors for traffic flows in the road network can explain most of the observed errors (especially for NO_2 and O_3). The OpenStreetMap data gives a detailed and accurate description of the network itself, but there is significant amount of uncertainty associated to the hourly vehicle flow estimates for cars, trucks and buses on individual roads. The most effective way to reduce this source of error is to obtain access to real time service provider for vehicular traffic. It should be noted that local measurements can reveal more appropriate vehicular patterns in the vicinity of the measurement locations as well as provide general emission factors for the whole traffic emission source category. However, even a dense network of measurements will not be sufficient for assessing the actual vehicle flows for cars, buses heavy vehicles.

During spring, one of the strongest influences to air quality originates from the resuspension of road dust (PM_{10}). However, the modelling of road dust particle generation and resuspension is very challenging due to the complex interaction of a) meteorological factors, b) road surface conditions, c) traffic flows, d) dispersion/resuspension of particles and e) road maintenance actions. An increase in the amount of measurement data can be used to improve the modelling of street dust, however, the improvements are not automatic and may involve extensive model development using e.g., neural networks.

Uncertainties in dispersion modelling

FMI-ENFUSER is a local scale urban dispersion model, but it is currently not yet possible to perform real obstacle resolving modelling in the complex urban landscape operatively. As such, the way large building blocks and vegetation affect the dispersion is not realistic in many cases, which produces a prediction error in complex urban environment. As the number of urban measurement locations increase it is possible to reduce this error source by learning from the wind speed/ wind direction / stability conditional error statistics that are provided automatically for each measurement location. This learning process is not automatically and requires manual model development based on the collected measurement data.

FMI-ENFUSER currently has a limited support for chemical processes due to the fact that the dispersion modelling is performed in a relatively small area while it is nested on a regional scale chemical transport model. As such the lack of chemical processes is not a significant source of error in Helsinki region. That being said, NO_x-ozone chemistry needs to be included in the modelling in a local scale and a comprehensive network of measurements for NO, NO_2 and O_3 will make model development possible.

One source of model prediction error comes from the modelling of emissions from strong elevated point emission sources such as power plants. For these emission sources the locations are well known which makes it possible to utilize to learning mechanisms provided by the model for e.g., temporal patterns of emission factors based on the measurement data. However, with strong point sources there is a frequent risk for frequent strong overestimations and severe under predictions in the measurement locations, which can make the emission source adjustment to behave erratically; as an example, in a measurement location a strong contribution from a power plant can be predicted but ultimately missed by the measurement device if the meteorological conditions are slightly different than was being estimated.

Data fusion with measurement data

Unlike many other urban air quality models, the ENFUSER utilizes measurements during operative near-real-time modelling. The purpose for the use of measurement data is to provide continuous feedback on model predictions at the measurement locations, which makes it possible to adapt the dispersion modelling to fit the measurement evidence. The adjustments made within this process operate on a defined set of local emission source categories. Gradually, the adjustments made hour after hour over a longer learning period can be converted into improved knowledge on the local emission sources and their characteristics. This learning process has been illustrated in Figure 2.

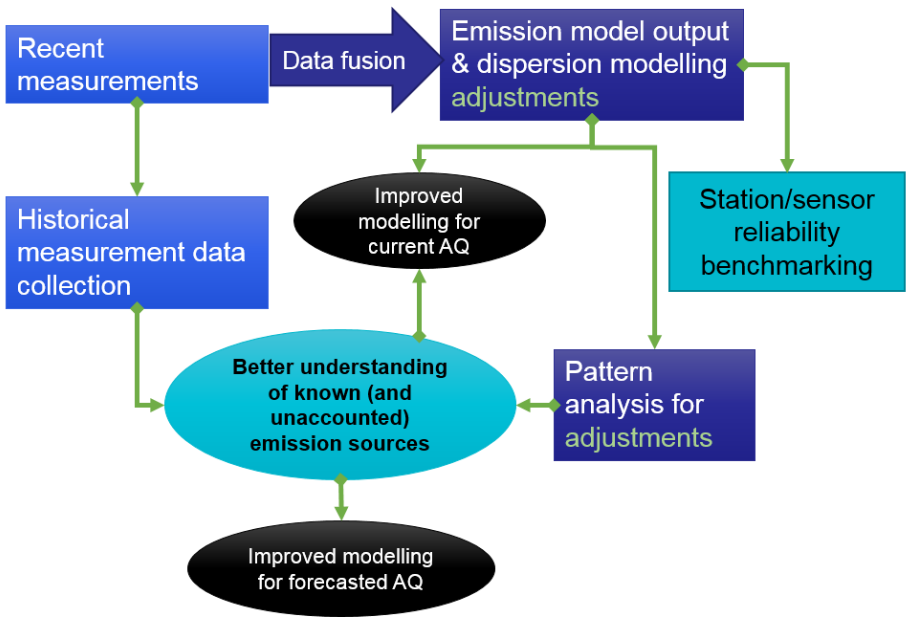

Figure 2: The learning process of Enfuser data fusion. Based on the model predictions at measurement locations, data fusion is used to adjust dispersion modelling parameters and local emission source outputs. Over a longer time period, statistical analysis of the hourly adjustments can be used to reveal patterns that facilitate model improvements, especially for the used local emission inventories. As a by-product, measurement device reliability benchmarking over time is available.

As a consequence of this emission source adjustment, direct emission factors need not to be assigned for these local emission source mappings initially. It is possible to start with an 'initial guess' and let the learning process arrive at a conclusion based on the measurement evidence. Naturally this approach works the better the more measurement locations there are in the modelling area. In this study we have used this option to assess whether or not the complementary sensors provide the same conclusion for selected emission source categories (such as traffic) as the reference quality AQ stations have previously done during the baseline study for 2017. Additionally, the learning mechanism that uses measurement data is also able to alert in case there are unknown emission sources nearby the measurement locations, such as industrial sources, which in turn makes it possible to develop the emission source modelling further.

In its simplest form, when high-quality measurement data are used and the local emission sources are fairly known, the learning process works as a way to fine-tune local emission source modelling. The hourly adaptation can reveal differences in actual and assumed vehicular patterns in the city, or the meteorological conditions observed during the adaptation can reveal how, e.g., the ambient temperature affects household $PM_{2.5}$ emissions. In Helsinki modelling area the learning mechanism has proven to be valuable for detecting the relationship between road surface conditions and the observed coarse-particle concentrations. Additionally, the observed model prediction errors in urban street canyon measurement stations have been highly usable for model development when the prediction errors as a function of wind direction have been assessed.

More commonly, the measurement devices that serve AQ data to the model are heterogeneous in quality and there are multiple poorly known emission sources in the modelling area. Additionally, the regional scale modelling that provides the background for the local scale modelling has its own sources for prediction errors. Taking these factors into account the data fusion process becomes much more complicated, since the data source reliability must be considered during the fusion process. As is shown in Figure 2, the reliability of independent information sources is continuously being monitored and stored for later uses.

Data fusion algorithm

ENFUSER is a dispersion model that models separately the dispersion of emissions originating from different sources such as traffic, shipping, power plants and private households. One crucial difference to common Gaussian dispersion models is that ENFUSER makes the computations in reverse with source appointment; this means that for a selected location the concentration contributions from nearby locations are scanned whereas commonly in dispersion modelling each emission source's contributions are being dispersed in the surrounding area. The end result is similar in both cases, but this formulation makes it possible to select a single location and assess the pollutant concentration components (contributions for a set of modelled emission sources) in that location. Technically this process is performed using a directed emission source footprint area shown in Figure 3.

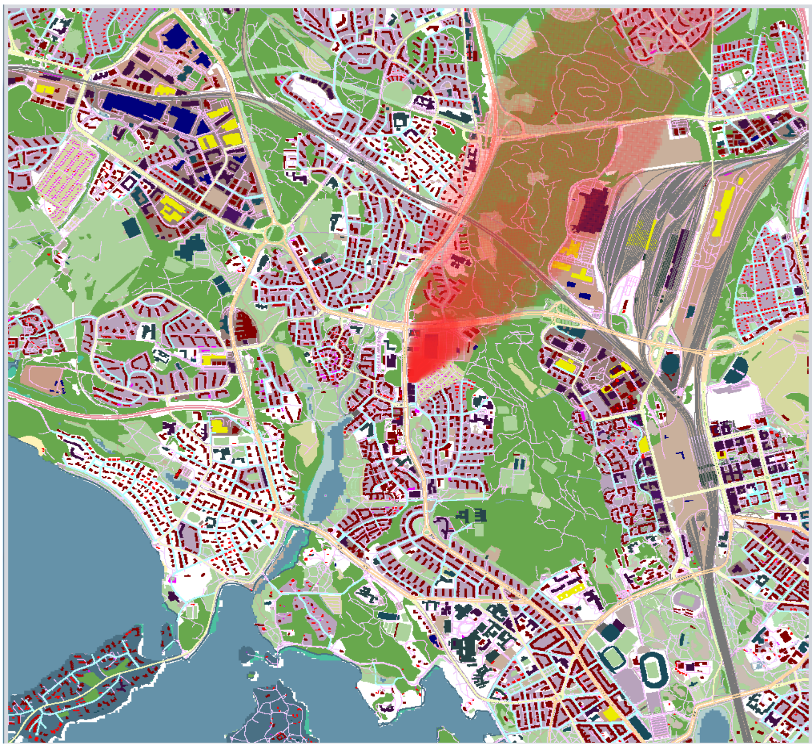

Figure 3: Example of an emission source fingerprint area in selected location (at the centre of the image) when the wind direction is from North-East. Brighter red color signifies higher relative weight per square meter. The footprint area consists of thousands of evaluation cells.

In the data fusion process this emission footprint analysis is performed for each measurement data point. It should be noted that the emission footprint computations are strongly affected by meteorological conditions as well as the surrounding environment. The observed pollutant concentration is explained as the sum of concentration components, which also contains the background concentration given by the regional scale model at the measurement location.

Once this component representation is available for each measurement the task for the data fusion process is to adjust the strengths[1] of modelled emission sources so that the modelling error is minimized collectively while taking into account the measurement device quality and the measurement location characteristics. To put it simply, for lower quality sensors and also for high quality measurement stations that naturally show high hourly variability (e.g., urban street canyons) higher prediction errors are allowed.

[1] The adjustments are applied multiplicatively using a defined allowance range, for example [0.5,2]. For background concentration component the adjustment is an additive correction without range limitations.

Thorough mathematical formulation of the data fusion process is not presented in this paper. However, there are a couple of topics that need to be described: First, in case the measurement quality was ignored it would be possible to optimize the emission source adjustment using a simple 'least sum of squared errors' (LSSE) approach. Now that we consider the measurement data point unreliability (p_i, that is a dynamic variable that can change hourly) the sum of squared errors is affected by an additional weight term that scales the measurement error independently for each measurement point.

$$\min\left\{f(\mathbf{p})\sum\frac{1}{p_i}\,(c_i - c_{Mi}(\boldsymbol{a}))^2\right\} \tag{1}$$

Where p_i represents the dynamic measurement unreliability (a penalty factor for which a lower values means higher measurement quality), c_i is the *i*-th measurement available in data fusion, $c_{Mi}(\boldsymbol{a})$ is the model prediction at the measurement location and $f(\mathbf{p})$ is a counter-balancing function that increases strongly as p_i increases. The role of $f(\mathbf{p})$ is to force the algorithm to search for a trade-off solution that achieves low prediction errors with low amount of weight reductions. In this formulation the solution consists of two vectors: the emission source adjustment vector (**a**) that affects c_{Mi} and the measurement point penalty vector (**p)**.

Finding an analytical solution to the optimization problem of this kind is very difficult (if not impossible) and therefore an optimum is being searched numerically in a discrete space using a gradient descend approach. Shortly put, the gradient descend method is an iteration where in each step the state of vector $\boldsymbol{a}$ is updated by assessing the optimal direction and adjustment amount so that the function in Eq. 1 decreases the most in each step. However, the vector **p** needs to be optimized during each of these steps, and we simply use a separate gradient descend computation for **p** using the current state of vector **a**.

The gradient descend approach has been described in Figure 4. In the main iteration - that starts from an arbitrary initial state - the emission source adjustment vector (**a**) is updated in a series of discrete steps until the function in Eq. 1 can no longer be improved[2]. Depending on the emission source category, we define adjustment ranges, e.g., for traffic sources the optimal solution is being searched with an adjustment span of $[0.5, 0.6, \ldots, 1.5]$. In a similar manner we allow a limited amount of discrete values for each $p_i \in [1,2,4,8,16]$, and as was described above the optimal configuration of **p** is assessed separately for each main iteration step. It should be noted that even with a small number of observations and emission source categories the amount of possible unique combinations with **a** and **p** exceeds billions, and this amount increases exponentially as a function of measurement points. The selected two-phased discrete descend algorithm that progresses towards the steepest improvement direction provides a fast solution even with a large amount of data points.

[2] When the function in Eq. 1 can no longer be improved it is possible that only a local minimum has been found for Eq 1. To avoid algorithm termination at local minimum, the steepest descend direction is searched again with larger step sizes, which makes it possible to "leap over" the local minimum.

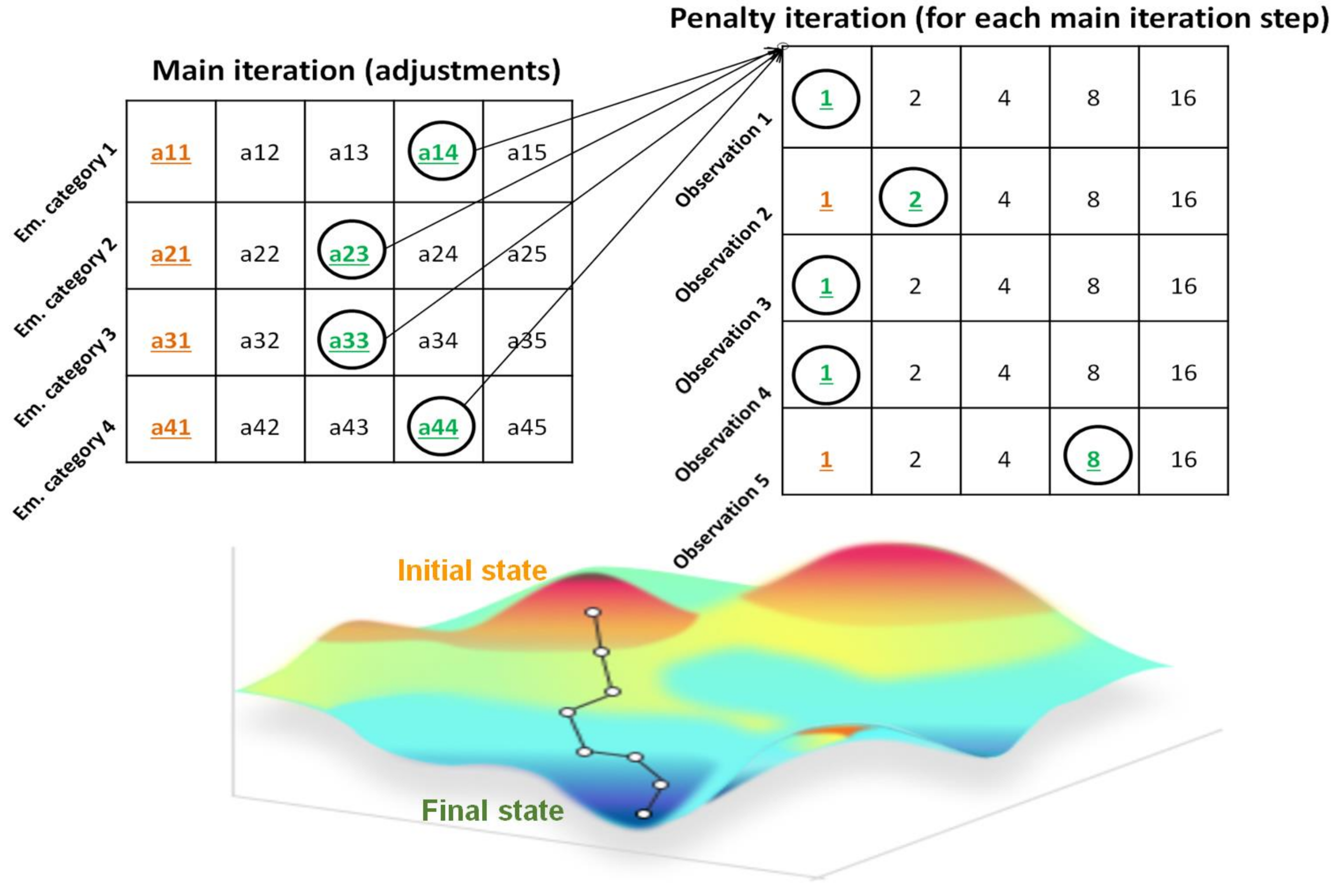

Figure 4: The two stage steepest descend method. The algorithm starts with an initial starting point (orange) vector 'a'. During each iteration the steepest descend direction for function described in Eq. 1 is identified. The iteration terminates when no improvement directions are found (circles).

While the use of the presented data fusion algorithm makes it possible to use most recent measurements in the dispersion modelling, it provides multiple learning mechanisms for the modelling system. First, for each measurement device the associated reliability over time can be monitored. Secondly, the system can learn to predict local emission outputs better based on the hourly adaptation factors over a longer evaluation period (Fig 2). The measurement source reliability measure interpretation, however, may require more thorough statistical analysis since it is computed with respect to model predictions and other measurement evidence; a low reliability measure in this case can also indicate structural modelling shortcomings at the measurement locations instead of a low measurement quality. Fortunately it is often possible to distinguish these two when the modelling performance (and penalty associations) is analysed for temporal and meteorological patterns.

Preparing sensor data for data fusion

As described in the Section 1.1 (Uncertainties) the modelling performance can be improved with a number of approaches that all benefit from having a large amount measurement data available. Affordable sensors can be deployed in large quantities and in theory the high hourly volatility should not make the learning mechanisms (Section 2) unusable for sensor data when the emerging long-term patterns dominate the hourly occurrences.

It is to be expected that sensor data sources will be associated with a higher penalty (lower weight) than reference quality measurement data and this way the data fusion algorithm is able to utilize the sensor data without compromising the overall modelling quality. However, one source of error that the data fusion algorithm is not able to deal with automatically is the structural bias of sensor data.

For example, when multiple sensors incorrectly signal strongly elevated concentrations this kind of error may be propagated to the modelling output since the errors are correlated; in such cases the final interpretation may be that the erroneous high concentrations become more ‘real’ than the evidence given by the rest of the measurement data. As such it is beneficial to apply sensor data corrections before they are given to the data fusion algorithm, and especially consider the possible sources for cross-correlated structural bias.

During the initial tests it became apparent that many of the sensors’ data incorporated strong structural offsets (e.g., consistently too low or too large values). These were especially noticeable for O3 and NO2 measurements and the strength of these offsets changed over time during the study year. For several sensors there were periods of time when the sensor provided abnormally high measured concentration that exceeded the highest observed reference quality measurements with an order of magnitude. A pollutant species -specific data acceptance range can be given to ENFUSER and to filter out these abnormal occurrences it was used with the following settings[3]:

- NO_2 must be between 0 and 250µgm-3
- Coarse PM must be between 0 and 250µgm-3
- $PM_{2.5}$ must be between 0 and 80µgm-3
- O_3 must be between 0 and 160µgm-3

After data filtering with the define limitations the sensor data observations still contained clear offsets that can be seen as structural bias, suggesting that a recalibration can increase the performance significantly when sensor data is being utilized in modelling. As it turns out, the ENFUSER model can be used to apply automatic bias corrections to sensors without ‘forcing’ the sensor data to fit model predictions. Based on all the reference-quality measurement data available as well as the regional modelled background, ENFUSER is able to estimate raw predictions for the hourly sensor measurements. Further, most pollutant concentrations exhibit a clear diurnal pattern; there exist certain times of day when concentrations are the lowest and usually during this time of day the source of pollutants are not local but mostly background concentrations. By exploiting this fact it is possible to arrange the diurnal measurement- and model prediction data in a way that effectively reveals structural bias (Figure 5).

[3] The acceptance ranges have been set to be very tolerant for concentrations commonly observed in the Helsinki area. The probability of filtering out valid data is therefore low.

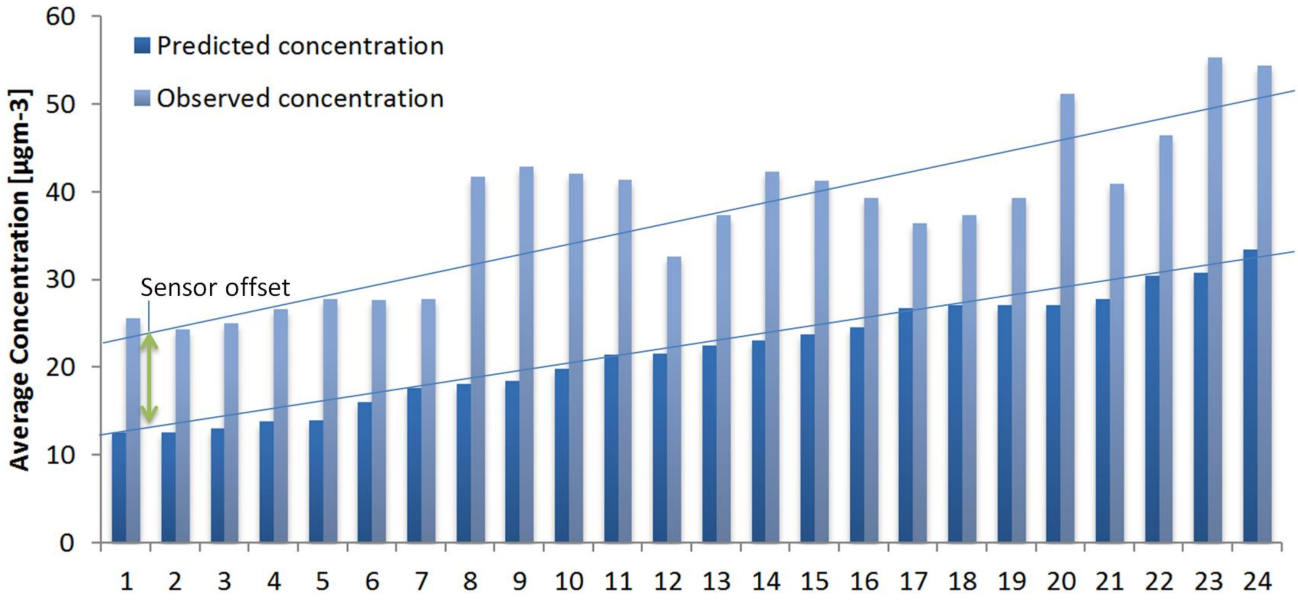

Figure 5: Sensor offset evaluation based on ranked diurnal patters of predicted and measured (observed with a sensor) concentrations.

The process of automatically applying offset corrections for sensor data was used for each sensor on a monthly basis in all the studies presented in the Results section. The results without offset corrections are not presented in this paper but it can be said that this pre-processing of sensor data significantly improves the quality of ENFUSER model runs with sensor data. It should be noted that this method in theory could be used as well for span/scaling correction factors for the sensor data. However, such span corrections are not being applied in this study since these could impair the value of sensor data in cases where the emission models are systematically under/over-estimated near the sensor location. In other words, it is possible that the sensor is accurate while the model, and in such cases an automatic span correction can force the sensor data characterize incorrect measurement data.

The process of automatic offset evaluation is described in Figure 5. On a monthly basis for example, model predictions are evaluated for each hourly sensor measurement and diurnal averages are computed for both the measurements and predictions. Then, the diurnal averages are sorted by prediction's absolute value. At the bottom of the list, the concentrations align well with background concentrations and a strong candidate for a correction offset can be found. It should be noted that for some pollutants (such as for O_3), the maximum values provide better estimates for correction offsets.

Once the sensor data has been checked for structural bias (offset) the data is more usable and can be efficiently included in the data fusion process. One key assumption in the data fusion algorithm is that the measurement devices can exhibit high variability and contain a significant amount of noise, but the measurement errors are not cross-correlated. One source of correlated errors that has not been discussed yet in this paper is the impact of meteorological conditions; we have assumed that the sensor's capability to measure is not strongly affected by meteorological conditions, which can have a similar detrimental effect on multiple measurement devices simultaneously. In the results (section 3) for $PM_{2.5}$ modelling results this issue is further discussed.

Results

As was described in Section 2, the successful utilization of less reliable sensor data depends on the performance of the data fusion algorithm. Since the data fusion methodology has been significantly expanded during the project, a brief concept test for the approach developed is necessary.

The described data fusion algorithm was tested before using it in HAQT model analyses. For this test the reference quality measurement data for 2017 (NO_2 concentrations) was used. To simulate a malfunctioning sensor, the data for an urban measurement station (Mannerheimintie) was replicated while the measurement timestamps were shifted by 12 hours. Then, a new virtual measurement information source was included to hold this replicated measurement data (named as 'fakeSensor') and for this virtual measurement site a remote location near Paloheinä, Helsinki, was selected. ENFUSER model was used to predict concentrations in 2017 while the data from this virtual sensor was being included.

In figure 6a, the running average credibility penalty ratings are shown for all information sources during this test. The running average shown is computed by using 50 consecutive modelling hours. A value of one corresponds to a situation where the data fusion algorithm applies no penalties to hourly data points for a given measurement device; whereas, a larger value shows that the data fusion algorithm applies credibility penalties to the information source. As can be seen from the figure, the 'fakeSensor' is heavily weighted down during the test, which shows that the data fusion algorithm is working as intended. Surprisingly, there is another urban background station that is also weighted down significantly. Further investigations revealed that this station in Kumpula was confirmed to be malfunctioning during the winter of 2017. Based on absolute concentrations the measurement errors were not significant (Fig 5b) but the data fusion process was able to weight these measurement data points down regardless.

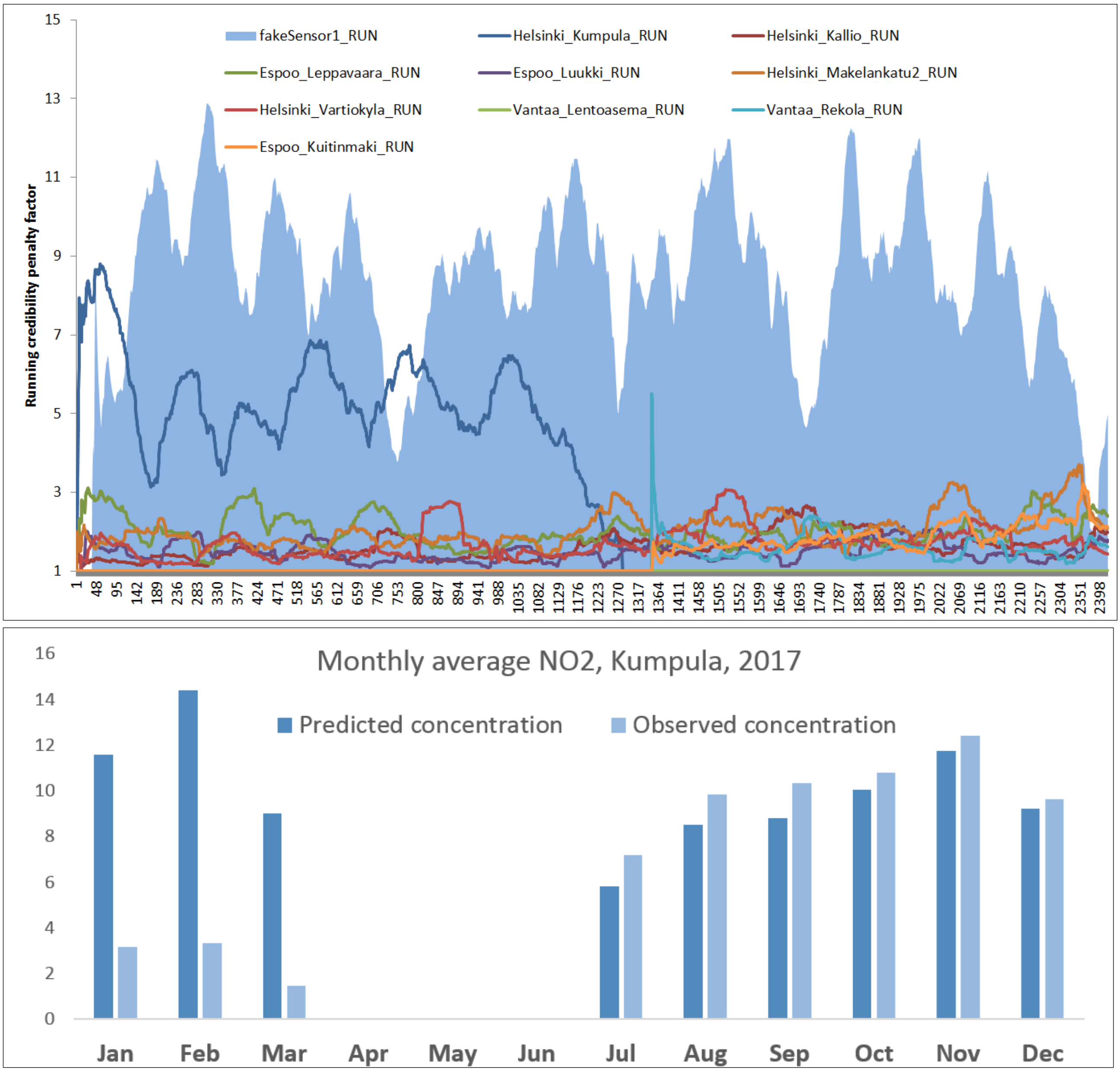

Figure 6a-b: Test results for data fusion algorithm used in 2017 while an additional measurement location was added to a remote location holding replicated measurement data for an actual urban measurement station. In a), the running average credibility penalty factor (see Fig 4) is shown and the horizontal axis describes the hour of year. In b) the monthly averages for a station in Kumpula is shown.

ENFUSER model runs with sensor data

ENFUSER model was used to predict NO_2, O_3, $PM_{2.5}$ and the coarse fraction of PM_{10} (abbreviated as 'PM_{coarse}') in Helsinki region between January and October in 2018, for which the main results are shown in Table 1. During this analysis period sensor measurement data from 16 Vaisala AQT's were available.

Technically, ENFUSER does not model PM_{10}, but the modelled pollutant concentrations for PM_{10} can be obtained as a post process by combining the modelled $PM_{2.5}$ and PM_{coarse} concentrations. The reason for this split of PM_{10} into two separate components is that the modelling of PM_{course} needs to be performed by taking into account the higher settling velocity of coarse particles than is used for $PM_{2.5}$. For data fusion, a virtual PM_{course} measurement is derived by subtracting $PM_{2.5}$ from PM_{10}, which are both present for each measurement location.

Table 1: Results for ENFUSER model runs with or without additional AQ sensor information. AVE: observed 10 month average range. MAE: mean absolute error for hourly concentrations. N: the amount of measurement stations (there are 16 sensors for each pollutant). 24h_FMAE: mean absolute error for forecasts 24 hours ahead. A value of 'x' means that the simulation or the statistical measure has not yet been produced at the time of writing.

		Reference stations				Sensors	
		AVE	**MAE**	**N**	**24h_FMAE**	**MAE**	**AVE**
NO2	**baseline**	6-33	5.8	12	8.9	16.6	18-40
	with sensors	-	5.7	-	8.6	15.9	-
	mostly sensors	-	7.3	-	9.0	14.9	-
O3	**baseline**	45-65	7.9	5	13.8	23.0	29-58
	with sensors	x	x	-	x	x	-
	mostly sensors	x	x	-	x	x	-
PMcoarse	**baseline**	5-16	5.5	9	7.3	7.2	5.5-18
	with sensors	x	5.4	x	x	7.0	x
	mostly sensors	x	6.4	x	7.4	6.8	x
PM2.5	**baseline**	5.7-9	1.8	11	4.0	5.5	6-11
	with sensors	x	x	-	x	x	-
	mostly sensors	-	3.4	-	4.7	4.2	-

In the table, three different modelling configurations are described: 'baseline', 'with sensors' and 'mostly sensors'. The baseline scenario is an ENFUSER modelling run without Vaisala AQT sensor data; however, the model predictions have still been provided for the sensor locations for comparison (see Table 2). The 'with sensor' corresponds to a scenario where all available data including sensors has been used. The final scenario, 'mostly sensors', corresponds to the most demanding situation where only 1-2 background stations have been used with the sensor data to predict pollutant concentration in the area.

The analysis has been done using a leave-one-out –validation, which means that one measurement device/station at a time, the information source has been removed from the pool of information, after which the model has been used to predict concentrations (including the use of data fusion algorithm) for the exact location for comparison. As such, the amount of modelling runs is significant for this study and consequently the model has been used to predict concentration for these selected discrete locations only, rather than for the whole Helsinki region[4]. Sensor data has been offset-corrected on a monthly basis, using the baseline configuration.

The modelling results for the baseline model runs are not described in detail in this paper. Shortly put, the most dominant emission source affecting local concentrations for NO_2, PM_{coarse} and O_3 is vehicular traffic (for O_3 the contribution is negative). For $PM_{2.5}$ the most notable local emission source are the small households, however the regional scale background dominates the measured and modelled concentration for $PM_{2.5}$. In coastal measurement location a small contribution from shipping is modelled based on the input data.

In general, the observed and modelled concentrations are fairly small for all species but the hourly variability is large especially for urban measurement sites (example shown in Figure 7). This causes the hourly mean absolute error (MAE) to be large with respect to the observed 10-month averages; in Helsinki the modelled concentrations have MAE-values approximately 30% of the average concentrations. Coarse particle concentrations have the highest relative prediction error with respect to averages. This is due to the fact that road dust concentrations are difficult to predict since these are affected by large number of meteorological factors, road maintenance actions and road surface properties.

[4] Using the model with discrete receptor points for a long analysis period also limits the accuracy of the modelling with respect to the modelling done by the operative ENFUSER in Helsinki area. The modelling has been limited to Gaussian Plume modelling (more dynamic Gaussian Puff modelling has not been used).

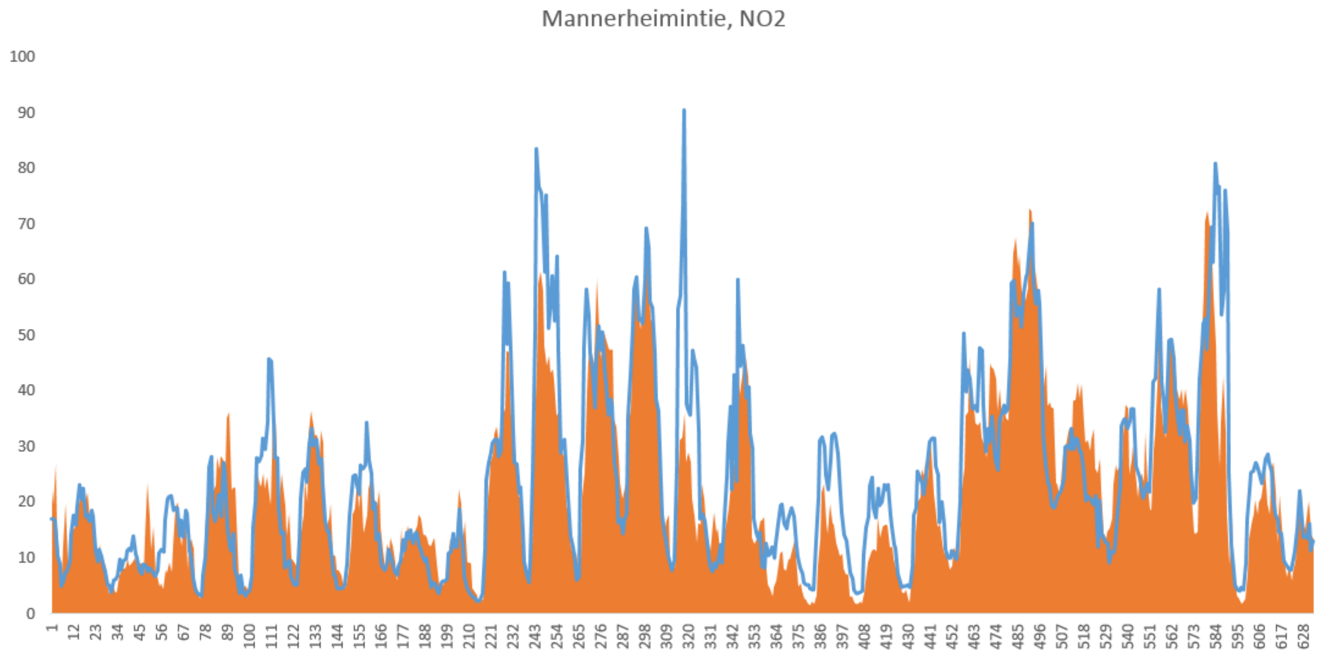

Figure 7: Example of modelling predictions (blue line) vs. observed concentrations (orange area) with a mean absolute error of approx. 6µgm-3 in an urban traffic measurement site (NO2, Spring 2018).

It can be seen from the results presented in Table 2 that the addition of sensors has mostly slight positive impacts on the modelling quality, $PM_{2.5}$ being an exception. For example, for NO_2 the mean errors for NO_2 are slightly smaller than is observed in the baseline model run, especially for the 24h forecast simulation. These slight modelling performance improvements come somewhat unexpectedly, since as it was discussed in Section 1.1 that the most effective ways to increase modelling performance in Helsinki area involve manual model development and higher quality input data. For NO_2 and coarse PM, the simulations done with only a few background stations being included (mostly sensors) suggest that a modelling system could function even with such limited amount of reference quality measurements. More specifically, in the "mostly sensors" simulation for NO_2 and PM_{coarse} the final emission factor assignment and temporal characteristics seem to align to the ones achieved in the base line modelling run.

When sensor data is compared against model predictions on an hourly basis, it is clear that the hourly average concentrations given by the sensor data has significantly lower reliability with respect to measurement stations. This lower reliability is reflected on high values for MAE (Table 2), higher data fusion weight penalty as well as in the correlation of measurements against model predictions. An example of the correlation of model predictions for a measurement station and a sensor has been shown
in Figure 8 for ozone. The MAE values for sensors can be seen to be smaller in the 'mostly sensors' scenarios than is observed for baseline scenarios. This outcome is likely to be due to cross-correlated errors for sensor data, i.e., the data fusion algorithm associates systematic error with decreased or elevated background concentrations resulting in an improved correlation for sensors specifically.

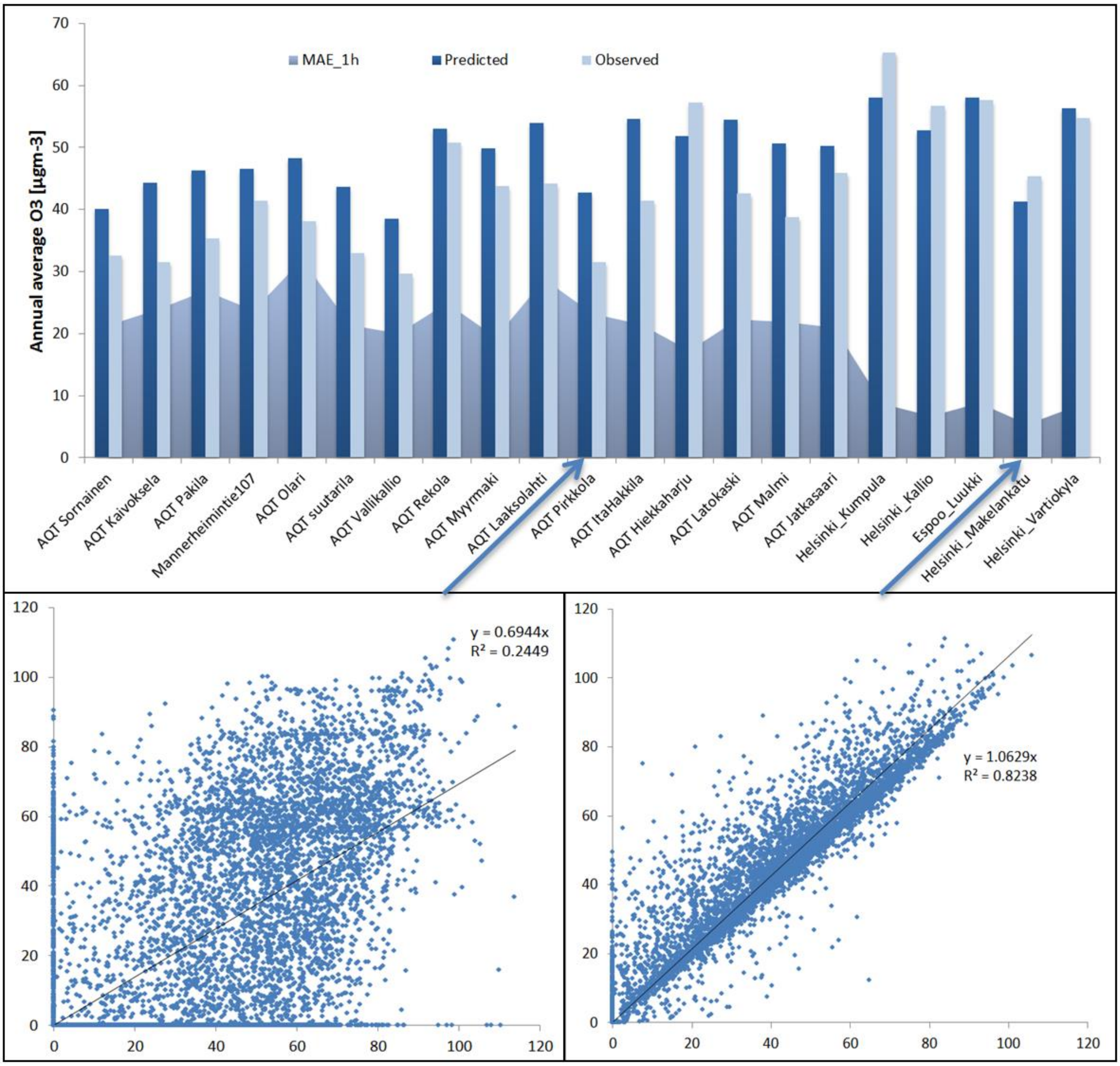

Figure 8: Above, the predicted and measured 10-month average O3 concentration for all measurement locations (left: AQT's, right: reference quality stations) are shown (bar plot). The average hourly mean absolute error (MAE_1h) is shown as an area plot. Below, example scatter plots (hourly measurements versus predictions during 1st Jan 2018 – 15th Oct 2018) for a selected AQT and a reference station have been shown.

As it was described in Section 2 the learning mechanism of the modelling system works on a longer term, relying on the emerging patterns using a large pool of measurements and adjustments made to local emission sources over time. Based on the simulations done with the sensor data it appears that the less reliable sensors provide usable information for this purpose in a cost-effective way. As an example, in Figure 9 selected conditional averages for NO_2 for a sensor has been shown. It can be seen from the figure that the predicted and measured concentrations are in agreement even though the hourly correlation of measurements versus predictions exhibit similar scattering as has been shown in Figure 8.

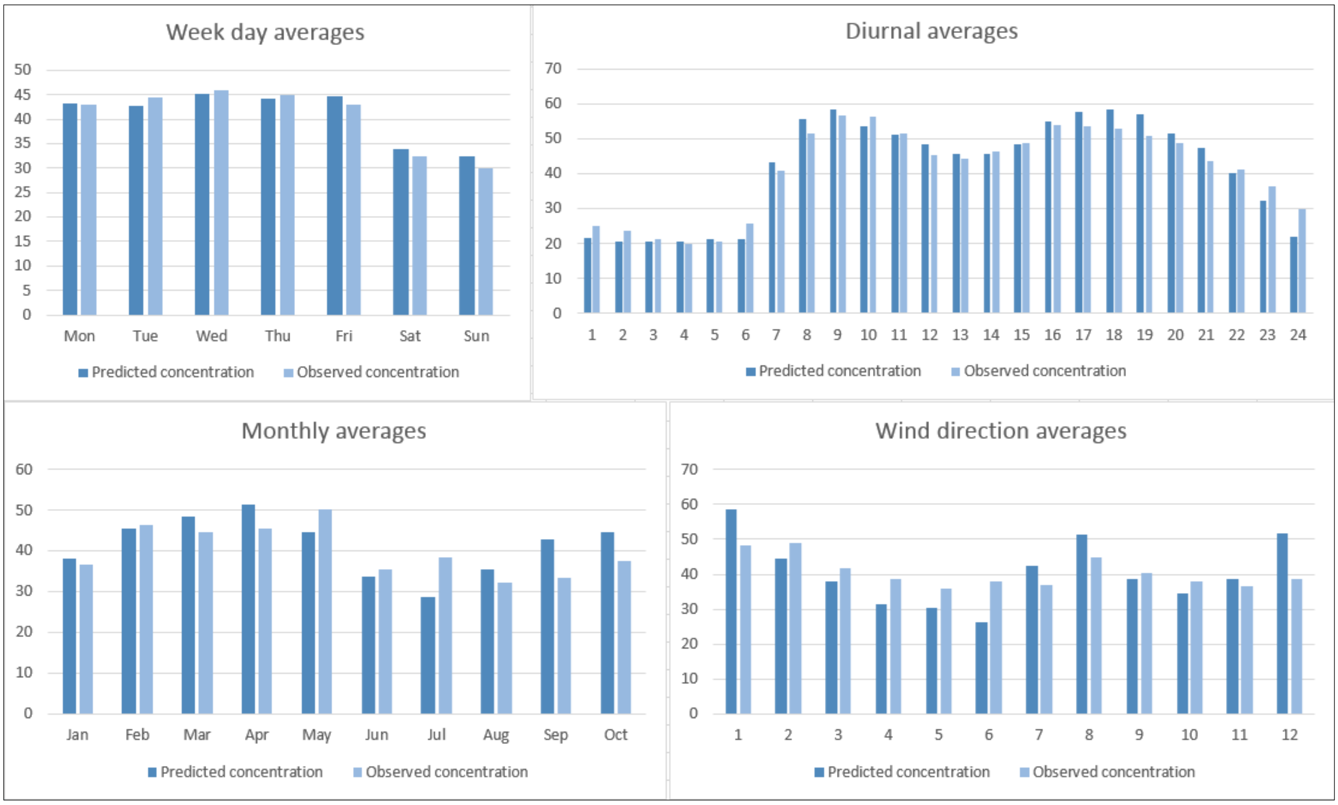

Figure 9: Vaisala AQT at Vallikallio, NO2.

Specifically for NO_2 and O_3 sensor evaluation the measurement performance (based on model estimates) seem to decrease steadily throughout the analysis period. This decrease is a known issue with the sensor electrochemical cells. For example, NO_2 sensor at Vallikallio shows high correlation during winter and early Spring. The monthly R2 coefficient starts from a respectable 0.74 but decreases steadily and has been lowered to 0.2 by August. As the monthly correlation decreases over time, so does the automatically set offset correction. Based on these findings, it may be possible to utilize the modelling system in detection malfunctioning sensors and thereby aid in the maintenance of the sensor network, while the sensors are providing measurements for the system simultaneously. This decreasing performance can also be observed with the progression of data fusion penalty weight assignment (Fig 5a).

As it has been discussed there is a limited potential in Helsinki area to gain automatic modelling performance improvements by increasing the amount of measurement data regardless of the measurement quality. However, for the modelling of PM10 (and coarse PM) the complementary sensor network is able provide valuable information that also translates to improved modelling performance directly. In the operational modelling system there are only a few measurement sites that provide information on background concentrations directly; the data for measurement sites at Espoo, Luukki, and Helsinki, Vartiokylä, are not available for the operational system and these measurement sources have been manually added for these modelling studies. It would seem that with the additional sensors being included the modelling system is able obtain a clearer picture on the local emission sources and their behaviour for course particles. In contrast, the addition of $PM_{2.5}$ sensor data in their current state seems to cause the opposite effect; the relative humidity seem to affect the measurement quality causing correlated errors, which in turn confuse the data fusion algorithm to yield worse results than without the additional sensor information.

Depending on pollutant species there was 5 to 12 measurement stations and 16 complementary AQ sensors available in this study. The increase in the amount of measurement locations had one more beneficial effect: the conditional measurement averages (especially as function of wind direction) were able to reveal certain short-comings in the dispersion modelling approach. For example, a sensor placed in Sörnäinen was located near a building block close to an otherwise open urban environment. The sensor data revealed that for certain wind directions the modelled and measured concentration were clearly in a disagreement. Data of this kind provides valuable information that can be used for further model development.

Conclusions

the benefits of having several affordable AQ sensors to complement the existing measurement network were evaluated Using the FMI-ENFUSER model. For this, a time period of 10 months during 2018 was selected for which hourly concentrations for $PM_{2.5}$, PM_{10}, O_3 and NO_2 were available for 16 sensor locations.

The sensor data are characterized by substantial hourly variability, making it difficult to utilize directly in air quality modelling. During the HAQT project, the data fusion algorithm in ENFUSER was improved so that such volatile AQ information could be included in the modelling. The resulting approach was seen to be successful, although the effective utilization of sensor data seems to require automatic calibration processes that are built-in the modelling system directly. This could be developed into a two-way coupling the modelling system and the sensor network: the sensors provide measurement data for the prediction of air quality and the modelling system simultaneously gives feedback that can be used for the effective maintenance of the sensor network. During the tests it was observed that the measurement quality for the sensor can decrease quite substantially over time and the modelling system has the means to detect such behaviour.

Based on the study the new sensors are able to deliver useful information on the air quality complementing the reference level AQ stations. The AQ sensors help in adjusting the temporal patterns and emission factors of various emission sources and based on measurement data fusion. In this sense, the sensors are able to provide more valuable information than their laboratory validations, co-location studies and correlations with model predictions would suggest.

For Helsinki area specifically, it is difficult to obtain modelling performance improvements by increasing the amount of measurement locations alone, as the most notable emission sources are well known in the area and, for example, more traffic site measurements are only able to enforce the already known shortcomings of the available traffic flow input. For the modelling of coarse particles and PM10 the addition of complementary sensor network was seen to provide direct improvements for the system. It was found out that the complementary sensors provide a cost effective way to offer more measurement locations of varying characteristics in the modelling system such that each of the additional locations have the potential to reveal shortcomings in the modelling approach, making it possible to develop the model further. In another region where the local emission sources are less known it is expected that the sensors that are coupled with ENFUSER could also reveal local unknown emission sources.

For $PM_{2.5}$ the effect of additional sensor data to modelling performance was seen to be negative and further sensor development is suggested for fine particles. The reason for this shortcoming is the effect on meteorological conditions to sensor measurements (relative humidity), which in turn can cause cross-correlated errors entering the data fusion process.

As a final conclusion, this study showed very promising results on the utilization of sensor measurements together with high quality measurement information. On the modelling system's perspective, a more cost-effective measurement network is likely to be obtained with a combination of stations and a large amount of sensors, rather than having a slightly larger amount of stations in the area. One potential yet limited measurement network configuration is the combination of a couple of background measurement stations and a large sensor network; in Helsinki, such a configuration was seen to function with a slightly reduced modelling accuracy with respect to the baseline study. A measurement network without reference quality stations, however, is not to be recommended since the option for automatic calibration for sensors is not available and the assessment of background concentrations becomes significantly more difficult.