Review of amt-2019-410: Final Response

High-resolution mapping of urban air quality with heterogeneous observations: a new methodology and its application to Amsterdam

I want to thank both reviewers for their useful comments. It resulted in a revised manuscript which puts more emphasis on the added value of low-cost sensors by including results for different assimilation configurations. Also, a sensitivity study on traffic emissions have been added, and the accuracy and limitation of the method are better discussed. To put more emphasis on actual measurements governing the spatial interpolation, the subsection on observations is now put forward as an independent section. Also, new figures have been added to better support the content. Results which are considered of importance but distracting from the main argument have been moved to Supplemental Material. The Discussion section and Conclusion section have been joined together and rewritten.

Below is my response to the reviewers. In blue the original comments, and in black the answers. It is followed by a marked-up version of the manuscript, indicating the changes made.

Response to reviewer #1

I favor the publication of the manuscript in AMT after the following two main issues and the specific comments have been carefully addressed and the manuscript has been substantially revised.

1. Due to the modularity of the modelling framework it is necessary that the accuracy and limitations of the individual components are determined and described in more detail and a little bit more critical. For example, the used dispersion model is computationally fast but it seems that it has clear limitations when applied in an urban environment as e.g. building geometries are not an input. Similarly, the emissions used as input for the dispersion model are derived in a simple way that similar data is probably available also for many other cities but the accuracy of the derived emission inventory seems by far not to be optimal. I think that the gain in accuracy of a spatially variable pollutant field by assimilating measurement data strongly depends on the model's capability to resolve small scale structures. It should be made more clear in the manuscript if measurements adjust local deviations in emission source activity or only general model deficiencies. In the latter case the assimilation of measurements does not necessarily lead to throughout improved results in a local environment around the sensor.

The revised manuscript elaborates on the comments above. The assimilation corrects both general model deficiencies and local deviations. The leave-one-out validation results show that in most cases the assimilation improves the accuracy, by reducing the local modelling bias, and improving the precision. This is better shown by including histograms of Observation minus Forecast (OmF) and Observations minus Analysis (OmA) for all validation locations. The potential shortcomings of the method are better discussed. Further reduction of the bias will be the subject of future studies, concentrating on (1) improving the modelling of traffic emissions, (2) including street canyon effects, (3) improving NO_X chemistry (ratio NO_2/NO_X and lifetime), (4) improving the modelling of the model covariance.

2. The validation part should be extended. The model uses a proxy for residential emissions as input data. Residential emissions probably have a distinct seasonality. Therefore, I strongly recommend to use an additional winter period to validate the modelling framework and to analyse the resulting alpha_pop. Actually, the simulation of a whole year would be best. No low-cost sensor data are required for this analysis.

I have added an analysis of the seasonality of the emission proxy calibration in Section 4.1. It compares a summer period (July 2016) with a winter period (January 2017). The results show that the average emissions do not agree well with the expected emissions from a bottom-up inventory. The regression analysis of the dynamic calibration finds the best linear combination of traffic proxies and residential proxies which explains the observations. Unlike traffic, the diurnal cycle for the residential contribution is shaped by the regression analysis. The seasonal analysis shows that the fitted diurnal cycle for the residential sector not only describes the cycle of the residential emissions, but also compensates for changing NO_2/NO_X ratios over the day due to changing photochemistry and temperature. Also, due to collinearity, part of the traffic emissions can be explained

by population density. Therefore, the found emission factors (and the corresponding sectoral emissions) should be considered as "effective" rather than real, i.e. factors which best describe the observations under the given model assumptions.

Specific comments

Introduction (section 1) or Setting up an urban air quality model (section 2): For completeness, a short discussion/list of other dispersion models that could be used as an alternative to AERMOD should be included.

Added to the end of the introduction in Section 3: "Note that any other dispersion model can be used in the Retina methodology, as long as it is capable of simulating concentrations from individual emission sectors on an arbitrary receptor mesh."

Traffic emissions (section 2.2.1): The author writes that traffic emissions are the dominant source of NO2 in the Amsterdam domain. Hence, traffic emissions are an important input factor for the simulation of the NO2 concentration. The interpolation of the vehicle counts for arbitrary locations based on the counting sites using IDW is practical but, at least for urban roads, possibly limitedly accurate as network characteristics are neglected. When I think of parallel roads in close distance, IDW would assign them similar vehicle counts, but in reality the true counts can be very different. An analysis/description/discussion of the accuracy of the resulting traffic input data should be added.

Indeed, the question remained how well this IDW interpolation describes the traffic flow differences found for nearby roads of the same road type. This is now assessed by two different approaches (presented in the Supplementary Material, and mentioned at the end of Section 3.2.1): a leave-one-out validation to study the error in local traffic flow estimations, and a concentration validation study of dispersion simulations done under different traffic scenarios. The results show that for this counting network IDW predicts the traffic volume within a 50% error margin at most locations. The model simulations show that using inferior traffic data is partly compensated by the calibration dynamics, at the expense of less pronounced concentration gradients.

Population data (section 2.2.2): The indication of the magnitude of different contributions (heating, cooking, others) to the total residential emissions in Amsterdam would be helpful. For the all-season applicability of the model: does the population database also include the spatial distribution of employees to account for heating emissions of office buildings? And for the selected period: Are heating emissions substantial in this summer period?

This is now assessed in newly added Section 4.1, where the diurnal cycles of sectoral emissions are analysed for summer and winter. There is no clear answer to the reviewer's questions, as the NOx emissions used by the model after calibration should be regarded as "effective" rather than real, i.e. best describing the spatial concentration patterns under given model assumptions. The effective residential emissions contain an unquantified contribution which compensates for the simplified NOx chemistry assumptions.

Calibrating the model (section 3): First, lines 217 to 219 are unclear for me. After reading these sentences I was confused if cj(t) in Eq. 6 is the measured NOX concentration. But it is not, correct?

The confusion arises from the fact that *P* represents proxies for NO_x emissions, while *c* represents the observed NO_2 concentrations. The conversion from NO_x to NO_2 is implicitly done by dispersion *f*, using an NO_2/NO_x ratio from the Ozone Limiting Method.

In Gaussian dispersion modelling there is a linear relation between emission strength and concentration, but this linearity breaks down when conversion from NO_x to NO_2 is included. This is now better formulated by eliminating lines 217 to 219 and changing the description of Eq. (6) to: "(...) with f_{ij} describing the dispersion of a unit emission from *i* to *j*, including the conversion from NO_x to NO_2 from the OLM. Eq. (6) is assumed to describe a linear relation between emission and concentration, although strictly speaking the variable NO_2/NO_x ratio introduces a weak nonlinearity."

Second, can you comment on how worse the model is performing when residential emissions are omitted? I guess that in the selected summer period heating emissions are nearly zero. Are the estimated two-hourly alpha_pop plausible and can you show that the temporal pattern of the values are related e.g. to cooking emissions? In Figure 3b, the contribution of residential emissions to the NO2 is surprisingly large given the fact that residential emissions are only 1/3 of the traffic emissions (stated in section 2).

This is now assessed in Section 4.1, where also the collinearity between the traffic and residential proxies is mentioned.

Assimilation of observations (section 4): The described algorithm is applied by using the pollutant concentrations transformed into the log-space. Here, one has to be aware that the distribution of the pollutant concentration at a particular location is not equal to the measurement error that is required for the algorithm in this section. The measurement error is described in the manuscript as being dependent on the concentration (section 2.3). So, the reasoning of this transformation is not correct. However, I suppose that the transformation of the measurements into the log-space has a positive effect on the stability of the results as the modeling framework becomes less sensitive to (less frequent) measurements of high concentrations by reducing their impact. The transformation into the log-space is fine, but the respective paragraph should be reformulated.

The reviewer is right. Changed "*The analysis is therefore done in log-space ($z_j = ln\ c_j$), which converts lognormal distributions to Gaussian, for which the Bayesian assumptions behind Equation 8-10 are valid*." to "*The analysis is done in log-space ($z_j = ln\ c_j$), stabilizing the results by reducing the impact of less frequent measurements of high concentrations.*"

Modelling the model error covariance matrix (section 4.1): The interpolation of the model error by IDW might result in an error field that is too smooth in the urban environment given that the model is limitedly capable to represent small-scale structures (e.g. buildings). At least a comment should be included in the manuscript that points out this issue.

It is impossible to assess the model error at all locations and under all conditions when the "ground truth" is only available at 15 locations. In my opinion, interpolation of the model error gives a good first impression of the local model performance

away from validation locations. However, the reviewer is right that small-scale structures provoked by the local built-up area might not be well represented in the model. Inclusion of the street canyon effect, together with refinements in covariance modelling, is subject of further study. It will reduce local bias in the modelling, but also suppress the introduction of bias by the assimilation. This is now pointed out in the discussion section.

Validation (section 5): The first paragraph is, as I understand, only an example where the modelling framework works well. It can be removed. Start with the second paragraph ("overall assessment"). Figure 6 can be presented directly after the overall assessment by discussing sites where the model performs well (NL49019) and where it performs less optimal (e.g. NL49002, NL49014). Figure 6 should include examples for both types. The time period of the used data in this Figure should correspond to table 3. Omit in the Figure the performance analysis of low-cost sensors but extend chapter 6. Moreover, add a file to the manuscript with supplementary materials where the scatter plots of the remaining, in the manuscript not presented air quality monitoring sites are shown analogue to Fig. 6 in order to provide the reader a clear picture of the model performance.

The reviewer's suggestions have been implemented. The first paragraph has been deleted. Figure 6 has been replaced by two validation examples, for a well performing location (NL49012) and a worse performing location (NL49014). The time series plots have been removed, regarded as redundant as the performance can also be read from the scatter plots. Bar plots with error distributions have been added to better illustrate the effect of the assimilation. The validation plots for all reference locations are included in the Supplementary Material.

Added value of low-cost sensors (section 6): The material presented in this chapter is only qualitative. The two average concentration maps presented are not validated and so the accuracy of their differences is unclear. The single example of the "Oude Schans" site is not sufficient to show that low-cost sensors add value. I have some questions here: What is the reasoning of largely adjusting the results of the dispersion model by low-cost sensors when there is also the option to improve the input data for the dispersion model?

To start with the last question: better input data is not always available. In this particular case there was no detailed information available about traffic flow and changing traffic patterns. To better assess the added value of low-cost sensors, additional results have been added from different assimilation configurations. The different assimilation scenarios show that low-cost sensor data assimilation improves the results locally, even in absence of reference data. Generally, the best results are obtained when both reference data and low-cost data are included. This is the configuration used to obtain the map showing the NO_2 reduction during the holiday period (Figure 11).

Is it possible to generate traffic input data for the dispersion model for each month based on the traffic counts you have access to?

Yes, it would be possible to generate monthly traffic data. This is done for instance in traffic scenario TS3, which can be found in the Supplementary Material. However, as both the holiday period and the closure of the tunnel started at half July and ended

half August, neither traffic data for July or August would be representative for the whole month. Anticipating the application for other cities where such detailed traffic data might not be available, it was decided to evaluate the system for a yearly averaged traffic "climatology".

Moreover, NO2 concentrations also depend on meteorology. What fraction of the differences in the monthly aggregated concentration fields is related to different weather conditions?

The influence of meteorological variability is now included in the discussion of the figure. It can be estimated from the NO_2 reduction found at rural stations NL49565 and NL49703. Average values drop from 16.60 and 12.57 ug/m^3 during 15 June - 15 July, to 15.53 and 11.55 ug/m^3 during 16 July - 15 August. Added to this section: "*Based on averaged NO_2 measurements at rural stations NL49565 and NL49703, the NO_2 reduction due to meteorological variability is estimated to be 7%*."

As I understand the main benefit of the low-cost sensors for the modelling framework is the increased spatial resolution of the measurements. Here, I miss some sensitivity analyses or similar material regarding measurement network design. What options exist when using this modelling framework in reducing traditional air quality monitoring sites and adding low-cost sensors? The accuracy of low-cost sensors is reported to be about 30%. Is this enough for adding substantial information?

These questions are now implicitly answered by the results of the different assimilation scenarios. Validation results are comparable when only observations of 14 low-cost sensors are assimilated (AS3), instead of observations at 3 reference sites (AS2). Even a notable improvement is visible in bias and RMSE at location NL49019, where the low-cost sensors are relatively nearby.

Technical comments

page 4, lines 102-103. How are the parallel distances of 75 and 125 m related to the grid? Maybe reformulate this sentence to make it clearer.

Changed to: "Receptor locations are chosen at every 75 m along the parallel curves with 25 m distance to the road, and at every 125 m along the parallel curves with 50 m distance to the road."

page 5, lines 143-144. Refers traffic "climatology" to counting sites?

Changed "For each location" to "For each counting site".

page 7, line 197. Mijling (2018) instead of Mijling (2017)?

Changed to Mijling (2018)

page 7, lines 204-205. Pik, Pki: keep consistent.

Adapted

page 7, lines 213-214. I would not say that b(t) is observed. It is rather the output to another modelling system.

Agreed, changed to "*Note that both background concentrations b(t) and local concentrations $c_j(t)$ are taken from external data*"

page 9, line 266. Section 3 instead of Section 2.3?

c refers to the measurements of the reference network, described (according to the new numbering) in Section 2.

page 10, line 285-286. "Isotropic" is the wrong word here as it is not isotropic.

True. Changed to "*The correlation of model errors between different locations is parametrized with a downwind correlation length L_{dw} and a crosswind correlation length L_{cw}.*"

page 11, line 336. Change to "lower accuracy".

Changed accordingly

Figure 1. Add units to the x and y axes.

Units are mentioned in the figure caption.

Figure 2. What do the depicted lines show? Sample week, yearly aggregation? Add more information.

Caption changed to "*Weekly cycle of highways and urban roads at counting locations, aggregated from hourly data from 2016. The morning and evening rush hours on working days are clearly visible for highways. Urban traffic has, apart from lower volume, less distinct peaks. The thick lines show the median of traffic flow for both road types.*"

Figure 3a. Add north arrow and scale in one of the four Figures. The meaning of the three dots should be explained already in the caption of Figure 3a.

Scale bar and arrow added. The figure caption now explains the three dots.

Figure 3b. Indicate in the Figure in an appropriate manner the weekday the dates refer to. 2016-07-07 –> Thu, 2016-07-07. In Figure 2 the distinct traffic pattern is shown. It is interesting if there is a clear relation between traffic and NO2 in the modelling results.

Labels on x-axis of Figure 3b changed.

Figure 5. Add units for x and y axes and scale in all Figures. Remove the first "and" in the second sentence of the caption.

Figure updated, scale bars added.

Figure 7. Add scale in both Figures. Moreover, the visibility of the points could be better in all the presented maps.

Figure 8. Add location of IJ-tunnel and of the historic center.

A new panel has been added showing the location of reference stations and low-cost sensors, the location of the IJ-tunnel, and the scale.

Table 3. Indicate more precisely the date period the analysed measurements refer to.

Table caption changed to "*Validation results at reference locations, June 1 - August 31, 2016*"

Response to reviewer #2

One of my main concerns relates to the lack of detail in some of the sections, but particularly in the section that is supposed to demonstrate the added value from assimilating low-cost sensor data (Sec 6). Given that multiple previous studies have already successfully assimilated regular stations observations using OI (e.g. Tilloy et al. 2013), one of the more novel aspects of this study is the assimilation of low-cost air quality sensors. The manuscript goes through great lengths of building up a dispersion model system with simplified emissions as well as an OI assimilation scheme, but the added value from low-cost sensor networks is covered in just a few lines towards the end without much detailed analysis. I think the manuscript could be a lot stronger and have more impact if a more comprehensive analysis of this were carried out in Section 6.

The work of Tilloy et al. is now referenced in the manuscript. This work is different in the sense that it presents a flexible urban dispersion model, it presents an alternative covariance model, and it studies the added value of low-cost sensors (details below). The section on the added value of low-cost sensor data has been extended with a sensitivity study of different network configurations.

Secondly, Section 2.2.1 on traffic emissions (which are crucially controlling NOx/NO2) left me scratching my head at times. I realize that the system is designed to be portable and thus the necessary input data should be kept to a minimum, but I wonder if some of the simplifications taken here are defensible. In particular spatial interpolation of traffic monitoring sites seems to me a quite crude approximation that introduces significant uncertainties in the modelling.

Spatial interpolation of traffic flow might indeed seem a crude way of solving the lack of traffic data. However, part of the introduced inaccuracy is avoided by distinguishing between two road types, highway and urban roads. Both having different diurnal patterns and vehicle counts, the traffic flow interpolation of one road type does not affect the other road type.

The validity of the approach is now assessed in two different manners, presented in the Supplementary Material, and mentioned at the end of Section 3.2.1: a leave-one-out validation to study the error in local traffic flow estimations, and a concentration validation study of dispersion simulations done under different traffic scenarios. The results show that for this counting network IDW predicts the traffic volume within a 50% error margin at most locations. The model simulations show that using inferior traffic data is partly compensated by the calibration dynamics, at the expense of less pronounced concentration gradients.

Further, what about distinguishing different types of vehicles? Regular cars versus heavy trucks? Euro 4/5/6 emissions categories? These things can have a very significant impact on the modelling results for NO2 and I wonder if some more care in setting up the modelling would not be beneficial in the long run? This is particularly a concern in the sense that OI should technically only be used when model and observations are unbiased against each other and ignoring certain high polluting vehicle classes could introduce potentially damaging biases.

The reviewer is right. We consider only two emission factors, one for highway traffic and one for urban traffic. The emission factors, however, are estimated from analysis against observations in the calibration phase, which implicitly compensates for

i.e. different fleet composition. Further refinement of the traffic model is desirable (e.g. based on the COPERT database), and will definitely improve local model performance. However, introducing a detailed traffic model, including fleet composition, was considered outside the scope of this work.

At the very least, the author should discuss these potential issues and lay out future steps to resolve these problems. In the best case scenario, it would be good to see some sensitivity studies testing the modelled NO2 sensitivity to inclusion of these different classes.

New validation results in the Supplementary Material show that the current method of interpolated traffic flow predicts the traffic volume within a 50% error margin at most locations. Better results are obtained when more counting locations are available, or when they are selected strategically around crossings and access roads. It is now remarked in the Discussion & Conclusion section that improved traffic emission models should take local differences in local fleet composition into account.

Thirdly, I feel that the manuscript could benefit from some more detail on how the error characteristics of the observations were derived. Estimating uncertainties from reference instruments and low-cost sensors on its own is a difficult subject and the paper does not provide the reader with information on how these were estimated or how such uncertainties were then transformed into error characteristics suitable for ingestion in the OI scheme. Such a discussion should be included in the paper and I believe this would strengthen the authors conclusions.

The section on observations has been elaborated hereupon. The accuracy of reference instrumentation is determined following the EN 14211 standard (now referenced in the section on observations), which includes all aspects of the measurements method: uncertainties in calibration gas and zero gas, interfering gases, repeatability of the measurement, derivation of NO_2 from NO_X and NO, and averaging effects. The error in the low-cost sensors is determined by side-by-side comparison against reference instruments. Details can be found in the referenced Mijling et al. (2018) paper.

Finally, given that 90% of the paper deals with modelling and data assimilation, I do find the choice of AMT for this manuscript slightly puzzling and I feel that the paper would probably be better suited for a journal more focused on modelling or general air quality issues. However since the editor has accepted the paper to AMTD I assume that the material is considered suitable for the journal.

AMT was chosen because the study describes an observation-driven framework which can be used for processing air quality measurements of different sources. Therefore, it was felt that it is of interest to the air quality measurement community. As the involved data assimilation sits on top of both observation and model, the current work would neither fully qualify for an air quality modelling journal. To put more emphasis on the importance of observations, Section 2.3 on observations is now put forward in the manuscript as an independent section.

DETAILED COMMENTS

L15: "Retina" - why is it called this? Include the full name if this is an acronym.

The algorithm's name reflects its ambition to produce high-resolution imagery; it is not an acronym. A simple name was preferred above a badly constructed or unpronounceable acronym.

L16/17: how are these percentages to be interpreted? Would be good to mention here how accuracy is defined. Something as simple as "... a typical accuracy (defined here as [...]) of 39%" or similar

Added "(defined here as the ratio between the root means square error and the mean of the observations)"

L23: "enhanced understanding of reference measurements". Please [missing]

L38: "adding value". I suggest you give an example of what you consider as adding value to the measurements or otherwise better write "exploiting the measurements"

Changed to "exploiting the measurements"

L39: In single-author papers it looks quite odd to use plural terms such as "our" and"we". Consider revising.

I replaced the inappropriate use of *pluralis majestatis* by the passive tense.

L43: I would add here that it depends on the mapping resolution and the pollutant. The required sampling density increases with the desired spatial resolution of the map. Furthermore, NO2 with its very sharp spatial gradients will always require a much denser network than for example mapping PM2.5 with its relatively smooth spatial gradients.

Added: "To obtain high-resolution information *of air pollutants with sharp concentration gradients*, (...)". From the first paragraph it is clear that is especially the case for NO2 concentrations, "which can vary considerably from street to street".

L60: The introduction/background section is missing a reference and a discussion of Tilloy et al (2013) (https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/jgrd.50233),who have essentially done the same as this paper (OI of point-based observations into an urban-scale AQ model).

The paper of Tilloy is now referenced twice.

In the Introduction: "*Tilloy et al. [2013] use the 3-hourly output of a well-developed implementation of the AMDS Urban dispersion model in Clermont-Ferrand, France, to assimilate in-situ NO2 measurements at 9 reference sites in an optimal interpolation scheme. With a leave-one-out validation they show a strong reduction in root mean square error of the time series after assimilation.*"

In Section 5.1 on the modelling of the error covariance matrix: "*Tilloy et al. [2013] choose to model the covariances depending on the road network. Error correlations are assumed to be high on the same road or on connected roads. For background locations, the correlation decreases fast in the vicinity of a road, while the error correlation between two background locations*

remains significant across a larger distance. The error covariances are kept constant in time, and taken independent of traffic conditions."

L61: Again, the name "Retina" comes a bit out of the blue. You should probably introduce here what the acronym stands for.

Retina is not an acronym, please see the answer above.

L70: I would be a bit careful with the term "calibration" in this context, given that it has a very specific meaning for measurements (both reference as well as sensors). Maybe reword or describe a bit more thoroughly what happens in this step.

I prefer to stick to *calibration* here, as I think that within the context of model calibration it is sufficiently clear that it refers to adjusting model parameters to best match the evaluation criteria.

L98: A reference to AERMOD would be useful here.

AERMOD has already been introduced and referenced shortly above (Cimorelli et al., 2004). This particular version of AERMOD has no specific scientific reference.

L99: "local equidistant coordinate system" - at this point you might as well give the actual projection you used. Presumably something UTM-like?

Clarified to: "*All coordinates are reprojected in a custom oblique stereographic projection (EPSG:9809) around the city center coordinate, such that the coordinate system can be considered equidistant at the urban scale.*"

L100: "road-following grid". This is used as if this a commonly known term, which in my opinion it is not. So first of all you might want to introduce this term a bit more carefully by saying something like "we use a road-following grid, which is essentially....". Secondly, to me it sounds a bit weird to use the term "grid" in this context, when you are basically talking about a spatially irregular and scattered set of receptor points with higher density along road links. I think the term grid should be reserved for a somewhat regular arrangement of cells.

The occurrences of *road-following grid* are replaced by *road-following mesh*.

L148-152: I realize that the goal of this paper is not to build the world's best model so a certain amount of simplification is expected, but interpolating traffic flow using IDW seems to be an incredibly crude method. How can this method possibly work?Between two loop counters there will likely be many road segments that either have much more or much less traffic than at the observation sites, so I fail to understand how simply interpolating here can lead to useful results. I think this section needs more detail on how this is carried out and a robust demonstration that the chosen methods are meaningful.

This concern is shared with Reviewer #1. In the revised article, the traffic model is assessed in two different approaches, presented in the Supplementary Material, and mentioned at the end of Section 3.2.1: a leave-one-out validation to study the

error in local traffic flow estimations, and a concentration validation study of dispersion simulations done under different traffic scenarios. The results show that for this counting network IDW predicts the traffic volume within a 50% error margin at most locations. The model simulations show that using inferior traffic data is partly compensated by the calibration dynamics, at the expense of less pronounced concentration gradients.

L164: Don't they assimilate UTD data? In that case it wouldn't be a day old but just a few hours (maybe better write "up to a day old" or so).

Clarified to: "*The analysis of the ensemble is based on the assimilation of up-to-date (UTD) air quality observations provided by the European Environment Agency (EEA).*"

L230: This should be Figure 3b? Also, I think this Figure should be discussed a bit more (maybe in the discussion section?) for example with respect to potential reasons for the difference between model and observations, particularly for the highway location.

Reference to Figure 3b now included. Differences between model and observations are now mentioned in the discussion section and mainly attributed to its inability to resolve all small-scale structures provoked by local built-up area, and sketchy traffic emission modelling. The latter is especially true for highway location NL49007, which is very near this strong source.

L235: I recommend to remove the term "geostatistical" here. While OI is mathematically very similar to kriging-based techniques (and it can in fact be shown that it provides identical results to kriging if the same inputs are used) it is not traditionally considered a part of the field of geostatistics. Geostatistics was developed in the mining and earth resources community (Matheron et al.), whereas OI was developed within the meteorological community (Gandin 1965).

Rephrased to "the interpolation technique of choice here is".

L241: Again, I would not use the term "grid" for what is essentially a set of irregular, scattered receptor points.

Grid replaced by *mesh*.

L265: I think it would be good to mention here that Statistical Interpolation/OI is essentially the same assimilation scheme (just a different mathematical framework) asprevious kriging-based approaches. The main advantage of OI over geostatistics (but also an added complexity) is that one has detailed manual control over the Pb covariance matrix, which allows for a more comprehensive specification of the area of influence for each contributing observation.

This valuable remark has been included in the beginning of the section, where OI is first introduced.

L287: extend -> extent (or maybe magnitude?); also reflect -> reflects

Corrected

L335-340: I think this section should be either left out entirely or expanded upon significantly. As it is currently it does not represent a robust demonstration that low-cost sensors add value to the system, since the effect has only been shown at a single site and not been analysed in detail. Demonstrating that the information from low-cost sensors can improve urban-scale air quality modelling is clearly a very worthwhile goal but this short section reads unfortunately more like an afterthought than a proper analysis.

This section has been expanded with new material: an additional analysis and discussion for 5 different assimilation scenarios. The different assimilation scenarios show that low-cost sensor data assimilation can improve the results locally, even in absence of reference data.

L354: I think it would also be worthwhile noting here that, while CAMS is definitely useful for providing background conditions and initial conditions, for NO2 the CAMS forecasts can be very misleading when interpreted at the local scale. The predicted diurnal cycle can often deviate substantially from that observed at urban AQ stations.

This valuable remark has been added in the discussion.

L356: Agreed. And in addition the higher resolution from a dispersion model is also much more appropriate than CAMS for applications such as exposure estimates etc.

L374: "Traffic data tend to be harder to obtain". That is very true (and maybe even an understatement) and is one of the most limiting factors in running local-scale dispersion models at random locations. Given that traffic is typically the most important source for NOx I have the suspicion that even the comparatively portable Retina methodology is likely to fail when no such traffic data is available at all. I think it would be worthwhile discussing here that at some point, if nearly all the crucial input data to Aermod is either of low quality or entirely missing, the resulting forecasted concentration fields will be so bad that any type of sophisticated data assimilation of observations is no longer very meaningful.

A traffic degradation study is now included in the Supplemental Material. From this can be seen that the calibration is partly capable of compensating for inferior traffic data by relating traffic emissions more to population density, at the expense of higher RMSE and lower correlation. In general, degraded input data and imperfections in the dispersion modelling will deteriorate the system's capability to resolve local structures; it will lower the effective spatial resolution of the simulations. In its extreme it will only describe the blurry urban background pollution contribution added to the rural background. Oppositely, with improved input data and atmospheric modelling, the effective resolution will improve. This insight has been added to the discussion.

L396: It might be more detailed, but is it really much better? This section is too qualitative to draw much of a conclusion. As I said above, I think the manuscript would benefit from a more robust analysis along these lines.

This is now better addressed in the section on the added value of low-cost sensors.

Figure 2: The caption should indicate more clearly that the thin lines represent the traffic at individual stations.

The caption is rewritten.

Figure 3a: These maps would benefit from some basic cartographic elements, e.g. a background map from OpenStreetmap similarly to Figure 7/8, scale bar, coordinates etc.

A scale bar and North arrow have been added. A cartographic background has not been added as the domain can now be inspected by comparison with new Figure 1.

Figure 5: "Units are in meters" - not all of them. I recommend to either label all axes properly or to have a more thorough caption describing the various elements of this busy Figure in more detail. It would also be helpful to have labels for each subplot (a,b, c..) so that they can be better referred to in the caption.

The representation has been improved by removing distance units on the axes by scale bars, by better formulating plot titles, and adding grid lines. The figure caption has been reformulated.

Figure 6: I think it is a bit confusing that the time series is only for 8 days, whereas the scatter plots show an entire month of data. Why not show the time series also for the entire period? If it is a visualization issue, it could be plotted over multiple rows. Similar to my earlier comments I also think that the analysis here would benefit from looking at more than just a single station.

Figure 6 has been replaced by two validation examples, for a well performing location (NL49012) and a worse performing location (NL49014). The time series plots have been removed, regarded as redundant as the performance can also be read from the scatter plots. Bar plots with error distributions have been added to better illustrate the effect of the assimilation. The validation plots for all reference locations are included in the Supplementary Material.

Figure 8: "IJ-tunnel" Should be marked on the map since non-locals will not be familiar with this.

A new panel has been added showing the location of reference stations and low-cost sensors in the central area, and the location of the IJ-tunnel.

High-resolution mapping of urban air quality with heterogeneous observations: a new methodology and its application to Amsterdam

Bas Mijling[1]

[1]Royal Netherlands Meteorological Institute (KNMI), Postbus 201, 3730 AE, De Bilt, The Netherlands

Correspondence to: Bas Mijling (mijling@knmi.nl)

Abstract. In many cities around the world people are exposed to elevated levels of air pollution. Often local air quality is not well known due to the sparseness of official monitoring networks, or unrealistic assumptions being made in urban air quality models. Low-cost sensor technology, which has become available in recent years, has the potential to provide complementary information. Unfortunately, an integrated interpretation of urban air pollution based on different sources is not straightforward because of the localized nature of air pollution, and the large uncertainties associated with measurements of low-cost sensors. ~~In this study, we present~~This study presents a practical approach to producing high spatio-temporal resolution maps of urban air pollution capable of assimilating air quality data from heterogeneous data streams. It offers a two-step solution: (1) building a versatile air quality model, driven by an open source atmospheric dispersion model and emission proxies from open data sources, and (2) a practical spatial interpolation scheme, capable of assimilating observations with different accuracies.

The methodology, called Retina, has been applied and evaluated for nitrogen dioxide (NO_2) in Amsterdam, the Netherlands, during the summer of 2016. The assimilation of reference measurements results in hourly maps with a typical accuracy (defined as the ratio between the root means square error and the mean of the observations) of 39% within 2 km of an observation location, and 53% at larger distances. When low-cost measurements of the Urban AirQ campaign are included, the maps reveal more detailed concentration patterns in areas which are undersampled by the official network. It is shown that ~~During~~ during the summer holiday period, NO_2 concentrations drop about 10%~~ due to reduced urban activity~~. The reduction is less in the historic city ~~center~~centre, while strongest reductions are found around the access ways to the tunnel connecting the northern and the southern part of the city, which was closed for maintenance. The changing concentration patterns indicate how traffic flow is redirected to other main roads.

Overall, ~~we~~ it is shown that Retina can be applied for an enhanced understanding of reference measurements, and as a framework to integrate low-cost measurements next to reference measurements in order to get better localized information in urban areas.

1 Introduction

Due to growing urbanization in the last decades, more than half of the world's population lives in cities nowadays. Dense traffic and other human activity, in combination with unfavourable meteorological conditions, often cause unhealthy air pollution concentrations. Over 80% of the urban dwellers are forced to breathe air which does not meet the standards of the

World Health Organization (WHO, 2016). In 2015, an estimated 4.5 million people died prematurely from diseases attributed to ambient air pollution (Lelieveld et al., 2018). Good monitoring is important to better understand the local dynamics of air pollution, to identify hot spots, and to improve the ability to anticipate events. This is especially relevant for nitrogen dioxide (NO_2) concentrations, which can vary considerably from street to street. NO_2 is, apart from being a toxic gas on its own, an important precursor of particulate matter, ozone, and other regional air pollutants. Observations from a single location are not necessarily representative for a larger area. Unfortunately, urban air quality reference networks are usually sparse or even absent due to their high installation and maintenance costs. New low-cost sensor technology, available for several years now, has the potential to extend an official monitoring network significantly, even though the current generation of sensors have significant lower accuracy (WMO, 2018).

However, ~~adding value to the~~exploiting these measurements (either official or unofficial), apart from publishing the data as dots on a map, is not straightforward. ~~Our~~ Here, the aim is to make better use of the existing measurement networks to get the best description of hourly urban air quality, and to create value from low-cost measurements towards a Level 4 product, according to the classification proposed by Schneider et al. (2019)

To obtain high-resolution information of air pollutants with sharp concentration gradients, a very sparse observation network needs to be accompanied by a valid high-resolution air quality model, whereas a very dense network can do with simple spatial interpolations. The situation in most large cities is somewhere in between. There is often a reasonably large reference network present (10+ stations), sometimes complemented with an experimental network of low-cost AQ sensors. Assumptions about underlying unresolved structures in the concentration field are still needed, but this can be done with a simplified air quality model, using the available measurements to correct simulation biases where needed.

A popular approach in detailed mapping of air quality is land use regression modelling (LURM), see e.g. Beelen et al. (2013). LURM uses multiple linear regression to couple a broad variety of predictor variables (geospatial information such as traffic, population, altitude, land use classes) to the observed concentrations. It is typically used in exposure studies, which target long integration intervals by definition. Problems of over-fitting might arise when too many predictor variables are used. Alternatively, Denby (2015) advocates the use of less proxy data, and a model based on more physical principles. In his approach, the emission proxies are first (quasi) dispersed with a parameterized inverse distance function, before being coupled to observed concentrations in a regression analysis.

Mapping of air pollution for short time scales is challenging. Only a few scientific studies are published aiming at assimilation of near-real time observations in hourly urban concentration maps. Tilloy et al. (2013) use the 3-hourly output of a well-developed implementation of the AMDS Urban dispersion model in Clermont-Ferrand, France, to assimilate in-situ NO_2 measurements at 9 reference sites in an optimal interpolation scheme. With a leave-one-out validation they show a strong reduction in root mean square error of the time series after assimilation. Schneider et al. (2017) use Universal Kriging to combine hourly NO_2 observations of 24 low-cost sensors in Oslo, Norway, with a time-invariant basemap. The basemap is created from a yearly average concentration field calculated with an Eulerian/Lagrangian dispersion model on a 1 km grid,

downscaled to 100 m resolution. Averaged over reference locations, their study shows that hourly values compare well with official values, showing the potential of low-cost sensor data for complementary air quality information at these time scales.

In this paper~~, we present~~ presents a more advanced yet practical approach to map hourly air pollutant concentrations, named Retina. Its main system design considerations are:

- Observation driven
- Able to assimilate observations of different accuracy
- Potential near-real time application
- Versatile / portable to other domains
- Based on open data
- Reasonable computer power

~~Retina uses a two-stage approach. It runs an urban air quality model to account for hourly variability in meteorological conditions (described in Section 2) which is dynamically calibrated with recent measurements (Section 3). In the second stage it assimilates current measurements using statistical interpolation (Section 4). Section 5 presents the validation of the system, while Section 6 shows the added value when assimilating additional low-cost sensor measurements. Section 7 and 8 are reserved for discussion, conclusion and outlook.~~

The method is applied to Amsterdam, a city like many where NO_2 emissions are dominated by transport and residential emissions and where local exceedances of limit values are regularly observed. ~~The methodology is flexible enough to be applied to other cities, mainly because it is relatively easy to implement the urban model for a new domain.~~ Amsterdam is the most populous city in the Netherlands, with an estimated population of 863,000. Located at 52°22′N 4°54′E, it has a maritime climate with cool summers and moderate winters. Concentrations of NO_2 within the city vary considerably, being partly produced locally and partly transported from outside the city. Measurements of 2016 show that, compared with regional background values from the CAMS ensemble (see Section 3.2.3), urban background concentrations are on average around 50% higher, while at road sides NO_2 concentrations are about 100% higher.

Retina uses a two-stage approach. It runs an urban air quality model to account for hourly variability in meteorological conditions (described in Section 3) which is dynamically calibrated with recent measurements (Section 4). In the second stage it assimilates current measurements using statistical interpolation (Section 5). Section 6 presents the validation of the system, while Section 7 shows the added value when assimilating additional low-cost sensor measurements. The last section is reserved for discussion, conclusion and outlook.

~~2.~~3 Air quality measurements

The Public Health Service of Amsterdam (GGD) is the responsible authority for air quality measurements in the Amsterdam area. Within the domain used in this study their NO_2 network consists of 15 reference stations: 5 stations classify as road station, 5 as urban background station, 2 as industry, 2 as rural, and 1 undecided. Alternatingly, GGD operates a Teledyne API 200E and a Thermo Electron 42I NO/NO_X analyser, both based on chemiluminescence. A catalytic-reactive converter converts NO_2 in the sample gas to NO, which, along with the NO present in the sample is reported as NO_X. NO_2 is calculated as the

difference between NO_X and NO. ~~Laboratory calibration estimates the combined uncertainty of hourly NO_2 measurement at 3.7% (GGD, 2014).~~The accuracy of both type of reference instruments is estimated at 3.7% (GGD, 2014), following the EN 14211 standard, which includes all aspects of the measurements method: uncertainties in calibration gas and zero gas, interfering gases, repeatability of the measurement, derivation of NO_2 from NO_X and NO, and averaging effects.

Low-cost NO_2 measurements are taken from the 2016 Urban AirQ campaign (Mijling et al., 2018). Sixteen low-cost air quality sensor devices were built and distributed among volunteers living close to roads with high traffic volume for a 2-month measurement period, from 13 June to 16 August. The devices are built around the NO2-B43F electrochemical cell by Alphasense Ltd (Alphasense, 2018). The sensor generates an electrical current when the target gas diffuses through a membrane where it is chemically reduced at the working electrode. Better sensor performance at low ppb levels is obtained by using low-noise interface electronics. The sensor devices were carefully calibrated ~~in Mijling et al. (2017)~~using side-by-side measurements next to a reference station, solving issues related to sensor drift and temperature dependence (Mijling et al., 2018). After calibration, they are found to have a typical accuracy of 30%.

~~2~~ 3 Setting up a versatile urban air quality model

~~Amsterdam is the most populous city in the Netherlands, with an estimated population of 863,000. Located at 52°22′N 4°54′E, it has a maritime climate with cool summers and moderate winters. Concentrations of NO_2 within the city vary considerably, being partly produced locally and partly transported from outside the city. Measurements of 2016 show that, compared with regional background values from the CAMS ensemble (see Section 2.2.3), urban background concentrations are on average around 50% higher, while at road sides NO_2 concentrations are about 100% higher.~~

One of the largest unknowns when modelling urban air quality is a detailed, up-to-date emission inventory capable of describing the local contribution. For cities such as Amsterdam the local emissions are dominated by the transport and residential sector. This is confirmed by the EDGAR HTAP v2 emission inventory (Janssens-Maenhout et al., 2013), which estimates the contribution of NO_X emissions in a 20 × 33 km^2 (0.3 degree) area around the ~~center~~ centre being 62%, 20%, 12%, and 6% for the sectors transport, residential, energy, industry respectively. Especially the contribution of road transport is relevant, as its emissions are close to the ground in densely populated areas. ~~We will use t~~Traffic information and population density will be used as proxies for urban emission (see Section ~~2~~3.2.1 and ~~2~~3.2.2).

In contrast to the regional atmosphere, the urban atmosphere is more dominated by dispersion processes, while many chemical reactions are less important due to a relatively short residence time (Harrison, 2018). For the dispersion of the emission sources, ~~we use~~ the open source steady-state plume model AERMOD (Cimorelli et al., 2004) is used, developed by the American Meteorological Society (AMS) and United States Environmental Protection Agency (EPA). Based on the emission inventory and meteorology (see Section ~~2~~3.2.4), AERMOD calculates hourly concentrations of air pollutants. The concentration distribution of an emission source is assumed to be Gaussian both horizontally as vertically when boundary layer conditions

are stable. In a convective boundary layer, the vertical distribution is described by a bi-Gaussian probability density function. Note that any other dispersion model can be used in the Retina methodology, as long as it is capable of simulating concentrations from individual emission sectors on an arbitrary receptor mesh.

~~2~~3.1 AERMOD simulation settings

~~We use~~ AERMOD version 16216r is used with simulation settings summarized in Table 1, operating on a rectangular domain of 18 × 22 km^2 covering the municipality of Amsterdam for the most part. All coordinates are reprojected in a custom oblique stereographic projection (EPSG:9809) around the city centre coordinate, such that the coordinate system can be considered equidistant at the urban scale.~~All coordinates are reprojected to a local equidistant coordinate system with the city center as projection center.~~ Instead of using a regular grid, ~~we use~~ a road-following ~~grid~~ mesh (Lefebre et al., 2011) is used. This reduces the number of receptor points, while maintaining accurate description of strong gradients found close to roads. Receptor locations are chosen at every 75 m along the parallel curves with 25 m distance to the road, and at every 125 m along the parallel curves with 50 m distance to the road.~~Grid points are defined at 25 and 50 m distances perpendicular to roads, and at parallel distances of 75 and 125 m respectively.~~ The open spaces between these ~~grid~~ points are filled with a regular grid at 125 m resolution. Roads are modelled as line sources, while residential emissions are described as area sources. The dispersion is calculated for NO_X to avoid a detailed analysis of the rapid cycling between its constituents NO and NO_2. Afterwards, an NO_2/NO_X ratio is applied, depending on the available ozone (O_3), see Section ~~2~~3.1.1.

Memory usage of AERMOD for the Amsterdam domain is proportional to the total number of emission source elements (here 17,069 road fragments and 12,182 residential squares) and the number of receptor points in the road-following ~~grid~~ mesh (here 42,128). The calculation time for a single concentration field is around 10 minutes, but can be reduced to a fraction of this by parallelizing the code.

~~2~~3.1.1 Ozone chemistry and lifetime

Primary emissions of NO_2 (e.g. directly from the tailpipe) are only 5-10% of the total emitted NO_X (Sokhi et al., 2007). At short time scales, secondary NO_2 is formed by oxidation of NO with O_3, while this reaction is counterbalanced by photolysis converting NO_2 to NO. The reaction rate of the first reaction is temperature dependent, while the latter depends on the available sunlight. The NO_2/NO_X ratio has therefore an intricate dependence on temperature, radiation, and the proximity to the source (i.e. the travel time of the air mass since emission).

A practical approach to estimate this ratio is the Ozone Limited Method (OLM), as described in EPA (2015). The method uses ambient O_3 to determine how much NO is converted to NO_2. The dispersed (locally produced) NO_X concentration is divided into two components: the primary emitted NO_2 (here assumed to be 10%), and the remaining NO_X which is assumed to be all NO available for reaction with ambient O_3: $NO+O_3 \rightarrow NO_2+O_2$. If the mixing ratio of ozone (O_3) is larger than the 90% of (NO_X), than all NO is converted to NO_2. Otherwise, the amount of NO converted is equal to the available O_3, i.e. (NO_2) =

$0.1(NO_X) + (O_3)$. The reaction is assumed to be instantaneous and irreversible. The resulting NO_2 concentration is added to the NO_2 background concentration.

Removal processes of NO_X are modeled with an exponential decay. The chemical lifetime is in the order of a few hours. Liu et al. (2016) find NO_X lifetimes in a range from 1.8 to 7.5 h using satellite observations over cities in China and the USA. Given the size of ~~our~~ the domain and average wind speeds, its exact value is not of great importance here. Based on regression results~~, we choose~~ a practical value of 2 hour is chosen.

~~2~~3.2 Simulation input data

The dispersion simulation is driven by input data regarding emissions, background concentrations, and meteorology, listed in Table 2. All data, except for the traffic counts of inner city traffic, are taken from open data portals. The emission proxies are mapped in Fig. ~~1~~2.

~~2~~3.2.1 Traffic emissions

A recurrent problem when building urban air quality models is finding sufficiently detailed traffic emission information. Traffic emissions depend roughly on traffic flow and fleet composition, including engine technology. For many cities, unfortunately, this information is not available. Here a distinction is made between highways and primary roads, as both have a distinct traffic volume and weekly cycle. Differences in driving conditions and fleet composition are captured by assigning two different emission factors later on.

Road location data and road type definition data are taken from OpenStreetMap (OSM, 2017), which is a crowd-source project to create a free editable map of the world. ~~We separate~~A distinction is made between urban roads (labelled in OSM as "primary", "secondary", and "tertiary") and highways (labeled as "motorway" and "trunk"), as they have a distinct traffic pulse, fleet composition, and driving conditions. Road segments labeled as "tunnel" are not taken into account.

When the traffic flow q (in vehicles per hour) is known, the emission rate E for a road segment l can be written as

$$E = \alpha_{\text{veh}} q l \quad (1)$$

with emission factor α_{veh} representing the (unknown) NO_X emission per unit length per vehicle. Hourly traffic flow data is taken from 29 representative highway locations from the National Data Warehouse for Traffic Information (NDW, 2019), which contains both real-time and historic traffic data. For the urban traffic flow ~~we use~~, data from 24 inductive loop counters provided by the traffic research department of Amsterdam municipality is used. Due to its large numbers, traffic flow is relatively well predictable, especially when lower volumes during holiday periods and occasional road closures are neglected. For each ~~location~~ counting site ~~we construct~~ a traffic "climatology" is constructed, parametrized by hour and weekday, based on hourly data of 2016, see Fig. ~~2~~3.

Traffic counts correlate strongly between different highway locations, all showing a strong commuting and weekend effect. Urban traffic typically shows, apart from lower volumes, less reduction between morning and evening rush hours, a less pronounced weekend effect, and higher traffic intensities on Friday and Saturday night.

For locations **x** between the counting locations $\mathbf{x}_i$ the traffic flow $q(\mathbf{x})$ is spatially interpolated by inverse distance weighting (IDW):

$$q(\mathbf{x}) = \begin{cases} \frac{\sum_i w_i(\mathbf{x}) q_i}{\sum_i w_i(\mathbf{x})}, & \text{if } d(\mathbf{x}, \mathbf{x}_i) \neq 0 \text{ for all } i \\ q_i, & \text{if } d(\mathbf{x}, \mathbf{x}_i) = 0 \text{ for some } i \end{cases} \quad (2)$$

in which the weighting factors w_i depend on the distance d between **x** and the counting location $\mathbf{x}_i$:

$$w_i = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)^2} \quad (3)$$

Validation in the Supplementary Material shows that for this counting network IDW predicts the traffic volume within a 50% error margin at most locations. Better results are obtained when more counting locations are available, or when they are selected strategically around crossings and access roads. Model simulations show that using inferior traffic data is partly compensated by the calibration (Section 4), at the expense of less pronounced concentration gradients.

~~2~~3.2.2 Population data

Population density is considered to be a good proxy for residential emissions, e.g. from cooking and heating. Here, ~~we take~~ data is taken from the gridded population database of 2014, compiled by the national Central Bureau for Statistics (CBS, 2019) at a 100 m resolution. Each grid cell is offered to the dispersion model as a separate area source. To reflect the observation that residential emissions per capita are less when people are living closer to each other (Makido et al., 2012), the emission fluxes are taken proportional to the square root of the population density p:

$$E = \alpha_{\text{pop}} \sqrt{p} \quad (4)$$

~~2~~3.2.3 Background concentrations

As AERMOD only describes the local contribution to air pollution, ~~we add~~ background concentrations are added which are taken from the Copernicus Atmosphere Monitoring Service (CAMS) European air quality ensemble (Marécal et al., 2015). The CAMS ensemble consists of 7 regional models producing hourly air quality and atmospheric composition forecasts on a 0.1 × 0.1 degree resolution. The analysis of the ensemble is based on the assimilation of ~~1-day-old~~up-to-date (UTD) air quality observations provided by the European Environment Agency (EEA). Each model has its own data assimilation system.

In the CAMS product the local contributions are already present. To get a better estimate for regional background concentrations avoiding double counts, ~~we take~~ the lowest concentration found in a 0.3 × 0.3 degree area around the city for NO_2 is taken, together with~~, and~~ the mean concentration found in this area for O_3.

~~2~~3.2.4 Meteorological data

The dispersion of air pollution is strongly governed by local meteorological parameters, especially the winds driving the horizontal advection and the characterization of the boundary layer which defines the vertical mixing. Meteorology also affects the chemical lifetime of pollutants.

~~We use~~ AERMET (EPA, 2019) is used as a meteorological pre-processor for organizing available data into a format suitable for use by the AERMOD model. AERMET requires both surface and upper air meteorological data, but is designed to run with a minimum of observed meteorological parameters. Vertical profiles of wind speed, wind direction, turbulence, temperature, and temperature gradient are estimated using all available meteorological observations, and extrapolated using similarity (scaling) relationships where needed (EPA, 2018).

Hourly surface data from the nearby Schiphol airport weather station can be obtained from the Integrated Surface Database (ISD, see Smith et al. (2011)). ~~We retrofit o~~Observations of temperature, winds, cloud cover, relative humidity, pressure, and precipitation are retrofit to match the SAMSON data format (WebMet, 2019a) which is supported by AERMET. Upper air observations are taken from daily radiosonde observations in De Bilt (at 35 km from Amsterdam), archived in the Integrated Global Radiosonde Archive (IGRA) (Durre et al., 2006). ~~We convert p~~Pressure, geopotential height, temperature, relative humidity, dew point temperature, wind speed and direction are converted to the TD6201 data format (WebMet, 2019b) for each reported level up to 300 hPa.

~~2.3 Air quality measurements~~

~~The Public Health Service of Amsterdam (GGD) is the responsible authority for air quality measurements in the Amsterdam area. Within the domain used in this study their NO_2 network consists of 15 reference stations: 5 stations classify as road station, 5 as urban background station, 2 as industry, 2 as rural, and 1 undecided. Alternatingly, GGD operates a Teledyne API 200E and a Thermo Electron 42I NO/NO_x analyser, both based on chemiluminescence. A catalytic reactive converter converts NO_2 in the sample gas to NO, which, along with the NO present in the sample is reported as NO_x. NO_2 is calculated as the difference between NO_x and NO. Laboratory calibration estimates the combined uncertainty of hourly NO_2 measurement at 3.7% (GGD, 2014).~~

~~Low-cost NO_2 measurements are taken from the 2016 Urban AirQ campaign (Mijling et al., 2018). Sixteen low-cost air quality sensor devices were built and distributed among volunteers living close to roads with high traffic volume for a 2-month measurement period, from 13 June to 16 August. The devices are built around the NO2-B43F electrochemical cell by Alphasense Ltd (Alphasense, 2018). The sensor generates an electrical current when the target gas diffuses through a membrane where it is chemically reduced at the working electrode. Better sensor performance at low ppb levels is obtained by using low-noise interface electronics. The sensor devices were carefully calibrated in Mijling et al. (2017), solving issues related to sensor drift and temperature dependence. After calibration, they are found to have a typical accuracy of 30%.~~

~~3~~ 4 Calibrating the model

Using proxy data instead of real emission introduces the problem to find the emission factors which best relate the activity data to their corresponding emissions. Instead of using theoretical values or values found in literature, ~~we derive~~ effective values are derived which best fit the hourly averaged NO_2 observations of a network of *N* stations.

For a certain hour t, the emission of a source element i belonging to source sector k can be written as

$$E_{ik}(t) = \alpha_k P_{ik}(t) \quad (5)$$

in which ~~P_{ki}~~ P_{ik} represents the corresponding emission proxy. The contribution of this source to the concentration at a receptor location j is

$$c_{ijk}(t) = f_{ij}(t) E_{ik}(t) \quad (6)$$

with f_{ij} describing the dispersion of a unit emission from i to j, including the conversion from NO_X to NO_2 from the OLM. Eq. (6) is assumed to describe a linear relation between emission and concentration, although strictly speaking the variable NO_2/NO_X ratio introduces a weak nonlinearity. ~~We apply a~~A regression analysis is applied for a certain period, assuming that for each t the total NO_2 concentration c_j at station j can be described as a background field b and a local contribution consisting of a linear combination of the dispersed fields of K emission sectors:

$$c_j(t) = b(t) + \sum_{k=1}^{K} a_k \sum_{i}^{S_k} f_{ij}(t) P_{ik}(t) \quad (7)$$

S_k represents the number of source elements for an emission sector k. The second sum in this equation is calculated for every hour with the Gaussian dispersion model taking the meteorological conditions during t into account. Note that both background concentrations $b(t)$ and local concentrations $c_j(t)$ are ~~observed quantities~~taken from external data, see Section ~~2~~3.2.3 and 2~~.3~~. Considering a period of T hours, Eq. (7) can be interpreted as a matrix equation from which the emission factors a_k can be solved using ordinary least squares. Given the physical meaning of a_k, only positive regression results are allowed.

~~Note that our linearity assumption in Eq. (6) works best when relating NO_x emissions to NO_x concentrations. The non-linearity introduced by the variable NO_2/NO_x ratio (here determined within AERMOD by OLM) is assumed to be sufficiently weak for this assumption to remain valid.~~

In ~~our~~ this setup, the emissions are approximated by three sectors highway traffic, urban traffic, and population density (K=3). The resulting a_k do not necessarily represent real emission factors. Their values partly compensate for unaccounted emission sectors and unrealistic modelling (e.g. based on wrong traffic data or an incorrect chemical lifetime). In Retina ~~we update~~ a_k every 24 hours, based on observations of the preceding week (T=168). Doing so, the periodic calibration adjusts itself to seasonal cycles and episodes not captured by the climatologies (e.g. cold spells or holiday periods). To avoid reducing the predictability of the regression model too much (a_k dropping to zero), ~~we do not use~~not all reference stations are used for calibration, but only stations classified as roadside or urban background. For the Amsterdam network, N=11. The residential emissions are represented by the population density, which is a time invariant proxy. To allow for a diurnal cycle, the residential emission factor is evaluated for two-hour bins. This brings the total number of fitted emission factors to 14: one for highway traffic, one for urban traffic, and 12 describing the daily residential emission cycle.

Figure ~~3~~4 shows an example of the air quality simulation after the emission factors have been determined. The stacked colours in the time series of Fig. 4b show that the contribution from different emission sectors to local air pollution can strongly vary from site to site.

4.1 Diurnal and seasonal analysis of calibration results

It is important to realize that the numerous modelling assumptions prevent that the calibration realistically solves the underdetermined inverse problem of finding the underlying NO_X emissions based on the observed NO_2 concentrations. Instead, it evaluates how much NO_X must be injected into the model to explain the observed spatial NO_2 patterns (unbiased with respect to the calibration locations). To study the results of the regression analysis, a comparison was made between a summer month (July 2016, mean temperature 18.4 ºC) and a winter month (January 2017, mean temperature 1.6 ºC). Figure 5 shows the diurnal emissions for a 0.2 × 0.1 degree area, corresponding to the two grid cells of the EDGAR inventory covering the city centre.

Ideally, the emissions would be around the values found in the EDGAR inventory (6.23 and 7.18 10^{-10} kg $NO_X/m^2/s$ for Summer and Winter respectively), and a corresponding ratio between residential and transport emissions (8% and 48% for Summer and Winter respectively). Unlike traffic, however, the diurnal cycle for the residential contribution is not prescribed, but is shaped in the regression analysis. The seasonal analysis shows that its fitted diurnal cycle not only describes changing residential emissions, but also compensates for changing NO_2/NO_X ratios over the day (not included in the OLM) due to changing photochemistry and temperature. In daylight, the destruction of NO_2 by photolysis ($NO_2 + h\nu \rightarrow NO + O_3$) is strong, reducing the NO_2/NO_X ratio. At low temperatures, the formation of NO_2 from NO ($NO + O_3 \rightarrow NO_2 + O_2$) is slow, also reducing the NO_2/NO_X ratio. Also, due to collinearity, part of the traffic emissions will be explained by population density. Therefore, the found emission factors (and the corresponding sectoral emissions) should be considered as "effective" rather than real, i.e. as factors which best describe the observations under the given model assumptions.

~~4~~ 5 Assimilation of observations

As the air quality network is spatially undersampling the urban area, ~~we need to combine~~ the observations need to be combined with additional model information to preserve the fine local structures in air pollutant concentrations. The interpolation technique of choice here is ~~From the various geostatistical techniques available we choose~~ Optimal Interpolation (OI) (Daley, 1991), having the desired property that the Bayesian approach allows for assimilation of heterogeneous measurements with different error bars. At an observation location the model value is corrected towards the observation, the innovation depending on the balance between the observation error and the simulation error. The error covariances determine how the simulation in the surroundings of this location is adjusted. Note that OI is essentially the same assimilation scheme as kriging-based approaches. The main advantage here is that one has detailed manual control over the error covariance matrix, which allows for a more comprehensive specification of the area of influence for each contributing observation. Outside the representativity range (i.e. the correlation length) of the observations, the analysis relaxes to the model values.

Consider a state vector $\mathbf{x}$ representing air pollutant concentrations on the (road-following) receptor ~~grid~~ mesh ($n \approx 40{,}000$). Define $\mathbf{x}^b$ as the background, i.e. the model simulation. Observation vector $\mathbf{z}$ contains m measurements, typically 10– 100. Following the convention by Ide et al. (1997), the OI algorithm can now be written as:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\big(\mathbf{z} - H(\mathbf{x}^b)\big) \quad (8)$$

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-\mathbf{1}} \quad (9)$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b \quad (10)$$

Matrix **R** is the $m \times m$ observation error covariance matrix. As all observations are independent (the measurement errors are uncorrelated), **R** is a diagonal matrix with the measurement variances on its diagonal.

$\mathbf{P}^b$ is the $n \times n$ model error covariance matrix, describing how model errors are spatially correlated. The calculation of $\mathbf{P}^b$ is not straightforward; in Section ~~4~~5.1 an approximation is derived.

Operator H is the forward model, which maps the model state to the observed variables and locations. ~~We can simplify t~~The matrix calculations can be simplified by reserving the first m elements of the state vector for the observation locations, and the other $n - m$ elements for the road-following ~~grid~~mesh. The Gaussian dispersion model is evaluated "in-situ" at the observation locations. Avoiding reprojection or interpolation means that there are no representation errors associated with H. The simulations at the observation locations $\mathbf{z}^b$ can then be written as a matrix multiplication

$$\mathbf{z}^b = H(\mathbf{x}^b) = \mathbf{H}\mathbf{x}^b \quad (11)$$

in which **H** is an $m \times n$ matrix for which its first m columns form a unity matrix, while its remaining elements are 0.

Eq. (8) describes the analysis $\mathbf{x}^a$, i.e. how the observations **z** are combined (assimilated) with the model $\mathbf{x}^b$. It is a balance between the model covariance and the observation covariances, described by the gain matrix **K** in Eq. (9). **K** determines how strong the analysis must incline towards the observations or remain at the simulated values, to obtain the lowest analysis error variance, $\mathbf{P}^a$ in Eq. (10).

Note that Eq. (8)-(10) are analogous to the first step in Kalman filtering. The second step of the filter, propagating the analysis to the next time step, cannot be made here as the plume model solves a stationary state which is independent of the initial air pollutant concentration field. Also note that since ~~we will use~~ an approximated model error covariance matrix will be used, generally these equations do not lead to an optimal analysis, hence this approach is more correctly referred to as Statistical Interpolation.

Let vector **c** represent the observed NO_2 mass concentrations, as described in Section 2.~~3~~. The distribution of the air pollutant concentrations resembles better the lognormal distribution than the Gaussian distribution, as can be seen from the Q-Q plots in Fig. ~~4~~6. The analysis is done in log-space ($z_j = \ln c_j$), stabilizing the results by reducing the impact of less frequent measurements of high concentrations. ~~The analysis is therefore done in log space ($z_j = \ln c_j$), which converts lognormal distributions to Gaussian, for which the Bayesian assumptions behind Eq. (8) (10) are valid.~~ Once returning from the log domain, Eq. (8) can be rewritten as:

$$\mathbf{c}^a = \exp(\mathbf{x}^a) = \mathbf{c}^b \exp(\mathbf{K}\Delta\mathbf{z}), \quad \text{with innovation vector } \Delta\mathbf{z} = \mathbf{z} - \mathbf{z}^b \quad (12)$$

By doing the analysis in the log-domain the assimilation updates correspond to multiplication instead of addition: $\exp(\mathbf{K}\Delta\mathbf{z})$ represents the local multiplication factor with which the simulated concentration $\mathbf{c}^b$ is corrected. This means that the shape of the model field (e.g. strong gradients found close to busy roads) is locally preserved. Note that the error in z_j corresponds to

the relative error in c_j : $\mathrm{d}z = \mathrm{d}(\ln c)/\mathrm{d}c = \mathrm{d}c/c$. The observation error covariance matrix is therefore $\boldsymbol{R} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$, with σ_j the relative error corresponding to observation j.

4.1 Modelling the model error covariance matrix

For an optimal result in the data assimilation a realistic representation of the model covariance matrix $\mathbf{P}^b$ is essential. The model covariances influence the spatial representativity of the observations: when model errors correlate over larger distances, the assimilated observation will change the analysis over a longer range.

Tilloy et al. (2013) choose to model the covariances depending on the road network. Error correlations are assumed to be high on the same road or on connected roads. For background locations, the correlation decreases fast in the vicinity of a road, while the error correlation between two background locations remains significant across a larger distance. The error covariances are kept constant in time, and taken independent of traffic conditions.

However, $\mathbf{P}^b$ changes from hour to hour, mainly because varying meteorology changes the atmospheric dispersion properties. ~~We estimate~~Here, the model error covariance is estimated for each hour based on the spatial coherence of the simulated concentration field. The covariance between two grid locations $\mathbf{x}_i$ and $\mathbf{x}_j$ can be expressed as their correlation ρ and their standard deviations σ:

$$P_{ij}^b = \sigma_i \, \rho(\mathbf{x}_i, \mathbf{x}_j) \, \sigma_j \tag{13}$$

The model error σ can only be evaluated at locations of the reference network using time series analysis. These model errors are spatially interpolated to other grid locations using IDW, analogous to Eq. (2)-(3). The correlation of model errors between different locations is~~We assume the covariance to be isotropic (i.e. location independent) but inhomogeneous: we~~ parametrized ~~the covariance~~ with a downwind correlation length L_{dw} and a crosswind correlation length L_{cw}. The ~~extend~~ extent of the correlation lengths reflect the turbulent diffusion and transport of the Gaussian dispersed plumes for a specific hour.

From spatial analysis of the simulation data a heuristic model is derived which describes the dependence of the correlation on distance:~~we see that the correlation depends on distance with a heuristic model~~

$$\rho(d) = \exp\left(-\sqrt{d}\right), \tag{14}$$

with d the scaled distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ (expressed as x_{dw} and x_{cw} along the downwind and crosswind axes)

$$d = \sqrt{\left(\frac{x_{\mathrm{dw}}}{L_{\mathrm{dw}}}\right)^2 + \left(\frac{x_{\mathrm{cw}}}{L_{\mathrm{dw}}}\right)^2}, \tag{15}$$

such that all points on an ellipse with semi-major axis L_{dw} and semi-minor axis L_{cw} have the same correlations.

To fit the parameters L_{dw} and L_{cw} for a certain hour, ~~we select~~ 1000 sample locations are selected from the road-following ~~grid~~mesh. To represent both polluted and less polluted areas, the locations are selected such that their concentrations are homogeneously distributed over the value range, excluding the first and last 5 percentile. For this sample, correlation lengths L_{dw} and L_{cw} are fitted using Eq. (14) and (15).

Figure ~~5~~ 7 shows the results of this analysis for two different hours. For fields with low gradients (e.g. when traffic contribution is low at night), large values of L can occur. To prevent assimilation instabilities, the fitted values of L are limited to a maximum

of 10 km. During the 2016 summer months, longest correlation lengths are found for fields with low gradients. Average midnight values, when traffic contribution is low, are about 8 km. Correlation lengths are shortest during the morning rush hour (~1 km), increasing to 3 km during the late morning and afternoon. There is a wind dependency, as stronger winds stretch the pollution plumes, increasing correlation lengths. From the fit results ~~we find an~~the average ratio between L_{cw} and L_{dw} ~~of~~ is found to be 68%.

Once the covariance parameters are known, the covariance matrix elements are calculated with Eq. (13). Note that for the calculation of the gain matrix **K** there is no need to calculate the full $\mathbf{P}^b$ matrix. Instead~~ we calculate,~~ $\mathbf{P}^b\mathbf{H}^T$,~~ ~~ is calculated, which due to the structure of **H** this matrix product corresponds to the first m columns of the $n \times n$ matrix $\mathbf{P}^b$.

~~5~~ 6 Validation of simulation and assimilation

~~We validate the system for the period June 15 – August 15 in 2016 with hourly observations from station NL49019 (Oude Schans), located in the city centre and classified as an urban background site. We test (1) the ability to simulate the NO_2 concentrations at this location with the dispersion model, and (2) the effectiveness of the data assimilation when only measurements of the neighbouring reference stations are included in the assimilation (i.e. a leave-one-out approach). From the results in Fig. 6 we see that the simulation describes the hourly observations with an RMSE of 11.7 μg/m^3. The results improve considerably when the surrounding observations of the reference network are assimilated, taking advantage of the covariance between the observational information from nearby stations and the validation location. The error reduces to 7.6 μg/m^3, while the correlation improves from 0.52 to 0.82.~~

To assess the data quality across the domain, ~~we perform~~ a leave-one-out analysis is performed at all locations of the reference network for the period June 1 - August 31, 2016. The results are summarized in Table 3. Figure 8 illustrates two examples; plots for all validation locations can be found in the Supplementary Material. For the observation-free simulation (i.e. the model forecast) ~~we find~~ an average RMSE is found of 13.6 μg/m^3 and correlation of 0.57. When assimilating observations, the average RMSE drops to 10.4 μg/m^3 while the correlation increases to 0.78. Strong systematic underestimations of the simulation (characterized by a large negative bias) are observed at street locations NL49002, NL49007 and industrial locations NL49546, NL49704. These are most likely caused by unrealistic assumptions of local emissions of either traffic or industry. The strong positive bias found at NL49014, located in a city park separated from the nearby main road by a block of 4-storey buildings, might be explained by an incorrect simulation of air pollutants in the direct vicinity of these buildings.

The CAMS regional ensemble analysis compares well with the average of the urban background stations; the very low bias (-0.1 μg/m^3) corresponds with the fact that data of these stations are used in its analysis. (Note that ~~we use here~~ the CAMS values used here correspond~~ing~~ to the Amsterdam grid cell, not the 3~~×x~~3 minimum values used as background for the modelling.) On the other hand, it shows strong underestimations at street locations, as expected. It is here where the Retina simulation outperforms the low resolution results of CAMS.

From Table 3 ~~we~~ can be seen that the relative error in the model forecast (defined as the ratio between the RMSE and the mean of the observations) is around 58% on average. When assimilating, the error becomes dependent on the distance to the nearest observation locations. For sites having the nearest assimilated observation within 2 km distance, the average RMSE drops from 16.8 to 11.9 µg/m^3, corresponding to an average relative error of 39%. For sites where the nearest assimilated observation is further away than 2 km, the average RMSE drops from 10.8 to 9.1 µg/m^3, corresponding to an average relative error of 53%.

7~~6~~ Added value of ~~low~~ low-costs sensors

The previous analysis is purely based on high-quality reference measurements. In this section ~~we~~ is explored whether the statistical interpolation scheme can be used to derive useful information of low-cost measurements, despite their ~~larger in~~lower accuracy.

~~During the Urban AirQ campaign (see Section 2.3) sensor SD04 was mounted at 120 m distance from location NL49019. From the hourly time series in Fig. 6 it can be seen that including its sensor data in the assimilation leads to a better description of NO_2 concentrations at location NL49019 than when assimilating with reference data alone. The RMSE drops from 7.6 µg/m^3 to 4.7 µg/m^3, while the correlation improves from 0.82 to 0.92.~~

This is done by testing different assimilation configurations during the Urban AirQ campaign, from 15 June to 15 August 2016 (see Section 2). The campaign targeted a central area with 4 reference stations and 14 low-cost sensors (See Figure 10a). Validation is done for 5 different assimilation scenarios (AS):

- AS1: Assimilation of all reference measurements (leave-one-out);
- AS2: Assimilation of measurements from 3 central reference sites (leave-one-out);
- AS3: Assimilation of low-cost data only;
- AS4: Assimilation of measurements from 3 central reference sites (leave-one-out) and all low-cost data;
- AS5: Assimilation of all reference measurements (leave-one-out) and all low-cost data.

The results are summarized in Figure 9. As expected, results deteriorate when the number of reference locations in the assimilation are reduced from 14 (AS1) to 3 (AS2). The correlation decreases at all 4 validation locations. At NL49012, the RMSE increases significantly due to a positive jump in the bias. The lower analysis with respect to the observations is due to the absence of assimilation of high values at nearby street location NL49002, which enlarges the influence of lower observations found at urban background location NL49019. At NL49019, located in the middle of 3 assimilation locations, the RMSE does not change significantly. Apparently, the effect of assimilation of observations farther than the surrounding locations is small.

When only observations of 14 low-cost sensors are assimilated (AS3), instead of observations at 3 reference sites (AS2), there is a notable improvement visible in bias and RMSE at location NL49019. Here, the low-cost sensors are relatively nearby, the closest being sensor SD04 at 120 m distance. At the other validation locations, the low-cost sensor assimilation results in similar RMSE (i.e. within 1 µg/m^3), a comparable bias, but a slightly lower correlation.

The results can be further improved if both reference and low-cost sensor data are included (AS4 and AS5). At NL49019, the RMSE drops to 5.1 µg/m^3 (compared to 7.8 µg/m^3 when no low-cost data are included) while the correlation increases to 0.89. Again, there is no significant difference between including the three surrounding reference locations and including all reference locations. Also at street location NL49017 and urban background location NL49003, the inclusion of low-cost sensor data improves RMSE and correlation compared to assimilations with reference data only (AS4 vs. AS2, and AS5 vs. AS1). At location NL49012, the bias reduces considerably only when all reference data are included in the assimilation (AS1 and AS5). The different assimilation scenarios show that low-cost sensor data assimilation improves the results locally, even in absence of reference data. Generally, the best results are obtained when both reference data and low-cost data are included. Assimilation can reduce local model biases. However, unrealistically modelled covariances can lead locally to the introduction of an additional bias.

Next, a monthly averaged concentration map of Amsterdam is constructed with ~~we use~~ all reference data and all low-cost sensor data from the first half of the Urban AirQ ~~campaign to construct a monthly averaged concentration map of Amsterdam~~, see Fig. ~~7~~10b. The addition of the low-cost data lowers the assimilation results by several µg/m^3 in the undersampled area west of Oude Schans (NL49019), while the NO_2 increases with several µg/m^3 around the traffic arteries found south and east of this location (Fig. 10c). A large fraction of traffic on these roads uses the IJ-tunnel to cross the river. On a monthly basis, this tunnel is used by approximately one million vehicles.

The second half of the Urban AirQ campaign coincides with the start of the summer holiday period and the closure of the IJ-tunnel for maintenance. Comparison of the NO_2 concentration maps of both periods reveal interesting features (Fig. ~~8~~11). Based on averaged NO2 measurements at rural stations NL49565 and NL49703, the NO2 reduction due to meteorological variability is estimated to be 7%. ~~Overall, the~~ The overall drop in NO_2 concentrations in the central area, however, ~~drop~~is around 10% due to reduced traffic during the summer break. Notable exception is the historic city ~~center~~centre, where the NO_2 reduction is only a few percent, probably related to the steady economic activity driven by tourism. The strongest NO_2 reductions, around 15%, are found around the access ways of the IJ-tunnel. A few main roads (e.g. De Ruijterkade/Piet Heinkade and Ceintuurbaan) show less NO_2 reduction than average, apparently due to redirected traffic avoiding the tunnel.

~~8~~7 Discussion and Conclusions

~~The validation analysis in Section 5 confirms that the CAMS ensemble is a good predictor for hourly NO_2 concentrations found in the urban background. However, local effects can be better resolved when CAMS data is used for background concentrations in a dispersion model which is driven by proxies for traffic and residential emissions.~~

~~The Retina simulation setup shows that such a system can be built from open software and open data. Applied to Summer 2016 in Amsterdam, it reduces the relative error at street locations from 70% to 51%, mainly by reducing the negative bias from 18.2 to 5.3 µg/m^3. At urban background locations the dispersion model introduces often a positive bias, especially when traffic sources are nearby. This is probably related to the assumption of uniform surface roughness used by the dispersion~~

model to account for the urban structure. A better description of street canyons in the model is likely to reduce this bias. Also unrealistic assumptions about chemical lifetime of NO_2 (influencing the plume "length") might play a role here.

The Retina approach works best for areas where air pollution is dominated by transport and residential emissions. Significant inaccuracies will be found in areas dominated by local emissions (e.g. from industry, port and airport activity) which are not described adequately by the proxies. This can be addressed by including these sources explicitly in the dispersion modelling.

The mapping results improve considerably with the second Retina step when available observations are assimilated by the statistical interpolation scheme. In general, the error of the assimilation results depends on the accuracy of air quality model, the number of assimilated observations, the quality of observations, and the distance to the observation location. When assimilating measurements of the reference network, the relative error in NO_2 concentrations drop to 44% on average. The local error depends on the distance to the nearest observations: approximately 39% within 2 km of an observation site, increasing to 53% for larger distances. The typical correlation increases from 0.6 to 0.8.

Retina has been built on open data to facilitate a flexible application to other cities. The meteorology needed for AERMOD is taken from global data sets of ISD and IGRA. Road network information can also be obtained globally from OpenStreetMap. Traffic data tend to be harder to obtain. When no local data is available on diurnal and weekly traffic flow its patterns should be estimated. In the absence of local census data, population density data can be taken from the Global Human Settlement database (Schiavina et al., 2019), which has global coverage on a 250 m resolution. For application within Europe, the necessary background pollutant concentrations can be obtained from CAMS. For applications outside Europe other data sets have to be found.

For near real-time monitoring and forecasting of air quality the CAMS ensemble analysis must be changed for the ensemble forecast. Instead of observation based meteorology one should use data from local or global numerical weather prediction models e.g. from the National Centers for Environmental Prediction (the Global Forecast System, GFS; open data) or the European Centre for Medium-Range Weather Forecasts (ECMWF; not open data).

8 Conclusions

In this paper we have presented Retina, a practical approach to interpolating hourly urban air quality measurements. As air pollution gradients can be strong in the urban environment, it is essential to combine (sparse) measurements with an air quality model when aiming at street-level resolution. The first step of Retina consists of a simulation by a dispersion model which is driven by meteorological data and proxies for traffic and residential emissions. In the second step, observations of different accuracy are assimilated using a statistical interpolation scheme.

A reasonable approximation of the model covariance matrix is found by assuming the model covariance to be isotropic and by fitting correlation lengths along the downwind and crosswind axes for every hour. Finding a more realistic description of the model covariance matrix may further improve the assimilation results, which will be subject of future research.

~~Retina can be used for an enhanced understanding of reference measurements by deriving detailed observation-based concentration maps. The Bayesian assimilation scheme also allows us to improve the results by including low-cost sensor data, in order to get improved localized information. However, biases must be removed beforehand with careful calibration, as most low-cost air quality sensors suffer from issues like cross-sensitivity or signal drift, see e.g. Mijling et al. (2018).~~

~~The assimilation of low-cost sensor data from the Urban AirQ campaign reveals more detailed structure in concentration patterns in an area which is undersampled by the official network. The additional measurements correct for wrong assumptions in traffic emissions used in the apriori interpolation, and give better insight in how traffic rerouting (for instance due to closure of an arterial road) affects local air quality.~~

~~Apart from assessment of historic data such as in this study, Retina has been applied successfully for near-real time monitoring and forecasting of NO_2 in the cities of Amsterdam, Barcelona, and Madrid. Future work includes the application of Retina to other cities inside and outside of Europe, and the application of Retina to other pollutants such as particulate matter.~~

As air pollution gradients can be strong in the urban environment, it is essential to combine (sparse) measurements with an air quality model when aiming at street-level resolution. Retina is a practical approach to interpolating hourly urban air quality measurements. The first step consists of a simulation by a dispersion model which is driven by meteorological data and proxies for traffic and residential emissions. The model is daily calibrated with historic measurements. In the second step, observations of different accuracy are assimilated using a statistical interpolation scheme.

Validation analysis confirms that the European CAMS ensemble is a good predictor for hourly NO_2 concentrations found in the urban background. However, the CAMS data for NO_2 can be misleading when interpreted at the local scale, as the predicted diurnal cycle often deviates substantially from that observed at urban air quality stations. Local effects can be better resolved when CAMS data is used for background concentrations in a dispersion model which is driven by proxies for traffic and residential emissions.

The Retina simulation setup shows that such a system can be built from open software and open data. Applied to Summer 2016 in Amsterdam, it reduces the relative error at street locations from 70% to 51%, mainly by reducing the negative bias from 18.2 to 5.3 µg/m3. At urban background locations the dispersion model often introduces a positive bias, especially when traffic sources are nearby.

The mapping results improve considerably with the second Retina step when available observations are assimilated by the statistical interpolation scheme. When assimilating measurements of the reference network, the relative error in NO_2 concentrations drops to 44% on average. The local error depends on the distance to the nearest observations: approximately 39% within 2 km of an observation site, increasing to 53% for larger distances. The typical correlation increases from 0.6 to 0.8.

The Bayesian assimilation scheme also allows us to improve the results by including low-cost sensor data, in order to get improved localized information. However, biases must be removed beforehand with careful calibration, as most low-cost air quality sensors suffer from issues like cross-sensitivity or signal drift, see e.g. Mijling et al. (2018). The assimilation of low-

cost sensor data from the Urban AirQ campaign reveals more detailed structure in concentration patterns in an area which is undersampled by the official network. The additional measurements correct for wrong assumptions in traffic emissions used in the apriori interpolation, and give better insight into how traffic rerouting (for instance due to closure of an arterial road) affects local air quality.

Retina has been built on open data to facilitate a flexible application to other cities. The meteorology needed for AERMOD is taken from global data sets of ISD and IGRA. Road network information can also be obtained globally from OpenStreetMap. Traffic data tends to be hard to obtain. When no local data is available on diurnal and weekly traffic flow its patterns should be estimated. In the absence of local census data, population density data can be taken from the Global Human Settlement database (Schiavina et al., 2019), which has global coverage on a 250 m resolution. For application within Europe, the necessary background pollutant concentrations can be obtained from CAMS. For applications outside Europe other data sets have to be found.

In general, degraded input data and imperfections in the dispersion modelling will deteriorate the system's capability to resolve local structures; it will lower the effective spatial resolution of the simulations. In its extreme it will only describe the blurry urban background pollution contribution added to the rural background. Oppositely, with improved input data and atmospheric modelling, the effective resolution will improve, reducing local biases. This is the focus of future research.

Significant inaccuracies due to local emissions which are not described adequately by the proxies (e.g. from industry, port and airport activity) can be reduced by including these sources explicitly in the dispersion modelling. Small-scale structures provoked by the local built-up area will be better described by introducing the street canyon effect. The model will also benefit from a more detailed traffic emission model, based on more counting locations and aggregated from shorter time intervals. Ideally, such an emission model takes local differences in fleet composition also into account. Finally, simulations will gain accuracy with a more realistic NO_X chemistry, concerning the NO_X chemical lifetime (influencing the plume length) and the NO_2/NO_X ratio.

Overall, the error of the assimilation results depends on the accuracy of the air quality model, the number of assimilated observations, the quality of observations, and the distance to the observation location. A reasonable approximation of the model covariance matrix is found by assuming the model covariance to be isotropic and by fitting correlation lengths along the downwind and crosswind axes for every hour. Finding a more realistic description of the model covariance matrix will better suppress the introduction of bias by the assimilation, and will be subject to future research.

For near-real time monitoring and forecasting of air quality the CAMS ensemble analysis must be changed for the ensemble forecast. Instead of observation-based meteorology one should use data from local or global numerical weather prediction models e.g. from the National Centers for Environmental Prediction (the Global Forecast System, GFS; open data) or the European Centre for Medium-Range Weather Forecasts (ECMWF; not open data).

Apart from assessment of historic data such as in this study, Retina has been applied successfully for near-real time monitoring and forecasting of NO_2 in the cities of Amsterdam, Barcelona, and Madrid. Future work includes the implementation of other cities inside and outside of Europe, and the application of Retina to other pollutants such as particulate matter.

Data availability

Author contribution

All work in this study was carried out by the author.

Competing interests

The author declares that he has no conflict of interest.

Acknowledgements

The author wishes to acknowledge the people behind the data sources used in this study, most notably GGD Amsterdam (reference measurements), volunteers of the Urban AirQ campaign (low-cost measurements), NDW and Amsterdam Traffic Research Department (traffic data), contributors to OpenStreetMaps (road location and classification), CAMS (background concentrations), IGRA and ISD (meteorology), and CBS (population data). This research was supported by KNMI-DataLab and the EU-H2020 project AirQast.

References

Alphasense: Alphasense Data Sheet for NO2-B43F, April 2016, available at: http://www.alphasense.com/WEB1213/wp-content/uploads/2018/12/NO2B43F.pdf (last access: 1 September 2019), 2018.

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.-Y., Künzli, V, Schikowski, T., Marcon, A., Eriksen, K. T., Raaschou-Nielsen, O., Stephanou, Eu., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, Ch., Falq, G, Stempfelet, M., Birk, M., Cyrys, J., von Klot, S., Nádor, G., Varró, M. J., Dėdelė, A., Gražulevičienė, R., Mölter, A., Lindley, S., Madsen, Ch., Cesaroni, G., Ranzi, A., Badaloni, Ch., Hoffmann, B., Nonnemacher, M., Krämer, U., Kuhlbusch, T., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M., Bellander, T., Korek, M., Olsson, D., Strömgren, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., and de Hoogh, K.: Development of NO_2 and NO_X land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project, Atmospheric Environment, Volume 72, 2013, Pages 10-23, ISSN 1352-2310, doi:10.1016/j.atmosenv.2013.02.037, 2013.

CAMS: Copernicus Atmosphere Monitoring System, European-scale air quality analysis from model ensemble, https://atmosphere.copernicus.eu/data, last access: 25 October 2019, 2019.

CBS: Statistics Netherlands, Statistische gegevens per vierkant 2000-2014, https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/gemeente/ruimtelijke-statistieken, last access: 25 October 2019, 2016.

Cimorelli, A.J., S.G. Perry, A. Venkatram, J.C. Weil, R.J. Paine, R.B. Wilson, R.F. Lee, W.D. Peters, and R.W. Brode: AERMOD: A dispersion model for industrial source applications Part I: General model formulation and boundary layer characterization, J.Appl.Meteor. 44, 682-693, 2004.

Daley, R.: Atmospheric Data Analysis. Cambridge University Press, 1991.

Denby, B.R.: Mapping of NO_2 concentrations in Bergen (2012-2014). METreport No. 12/15, Norwegian Meteorological Institute, ISSN 2387-4201, 2015.

Durre, I., Vose, R. S., and Wuertz, D. B.: Overview of the Integrated Global Radiosonde Archive. Journal of Climate, 19, 53-68, 2006.

EPA: Technical support document (TSD) for NO_2-related AERMOD modifications, U.S. Environmental Protection Agency, EPA-454/B-15-004, July 2015, 2015.

EPA: AERMOD Model Formulation and Evaluation, EPA-454/R-18-003, April 2018, available at https://www3.epa.gov/ttn/scram/models/aermod/aermod_mfed.pdf, 2018.

EPA: User's Guide for the AERMOD Meteorological Preprocessor (AERMET), EPA-454/B-19-028, August 2019, available at https://www3.epa.gov/ttn/scram/7thconf/aermod/aermet_userguide.pdf, 2019.

GGD: Foutenbeschouwing (revisie) kalibratie van eerstelijns standard (In Dutch), Public Health Service of Amsterdam (GGD), GGD/LO 14-1134, July 2014, 2014.

Harrison, R. M.: Urban atmospheric chemistry: a very special case for study, Nature, npj Climate and Atmospheric Science, 20175, 2397-3722, doi:10.1038/s41612-017-0010-8, 2018.

Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified notation for data assimilation: Operational, sequential and variational. J. Met. Soc. Japan,75, 181–189, 1997.

IGRA: National Oceanic and Atmospheric Administration, Integrated Global Radiosonde Archive, version 2. https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive, last access: 25 October 2019, 2019.

ISD: National Oceanic and Atmospheric Administration, Integrated Surface Database, https://www.ncdc.noaa.gov/isd/data-access, last access: 25 October 2019, 2019.

Janssens-Maenhout, G., Pagliari, V., Guizzardi, D., and Muntean, M.: Global emission inventories in the Emission Database for Global Atmospheric Research (EDGAR) - Manual (I): Gridding: EDGAR emissions distribution on global gridmaps, JRC Technical Reports, 33 pp., doi:10.2788/81454, 2013.

Lefebvre, W., Fierens, F., Trimpeneers, E., Janssen, S., Van de Vel, K., Deutsch, F., Viaene, P., Vankerkom, J., Dumont, G. , Vanpoucke, C., Mensink, C., Peelaerts, W., and Vliegen, J.: Modeling the effects of a speed limit reduction on traffic-related elemental carbon (EC) concentrations and population exposure to EC, Atmos. Env., 45, pp. 197-207, doi:10.1016/j.atmosenv.2010.09.026, 2011.

Lelieveld J., Haines, A., and Pozzer, A.: Age-dependent health risk from ambient air pollution: a modelling and data analysis of childhood mortality in middle-income and low-income countries. Lancet Planet Health 2018; 2(7): e292-e300, 2018.

Liu, F., Beirle, S., Zhang, Q., Dörner, S., He, K., and Wagner, T.: NO_X lifetimes and emissions of cities and power plants in polluted background estimated by satellite observations, Atmos. Chem. Phys., 16, 5283–5298, doi:10.5194/acp-16-5283-2016, 2016.

Makido, Y., Dhakal, S., and Yamagata, Y.: Relationship between urban form and CO2 emissions: Evidence from fifty Japanese cities, Urban Climate, 2, pp. 55-67, doi:10.1016/j.uclim.2012.10.006, 2012.

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, 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., Jaumouillé, 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, Á., 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, Geosci. Model Dev., 8, 2777–2813, doi:10.5194/gmd-8-2777-2015, 2015.

Mijling, B., Jiang, Q., de Jonge, D., and Bocconi, S.: Field calibration of electrochemical NO_2 sensors in a citizen science context, Atmos. Meas. Tech., 11, 1297-1312, doi:10.5194/amt-11-1297-2018, 2018.

NDW: National Data Warehouse for Traffic Information, Historical data https://www.ndw.nu/pagina/en/78/database, last access: 25 October 2019, 2019.

OSM: OpenStreetMap contributors, Planet dump retrieved from https://planet.osm.org. https://www.openstreetmap.org, 2017.

Schneider, P., Castell, N., Vogt, M., Dauge, F. R., Lahoz,, W. A., and Bartonova, A.: Mapping urban air quality in near real-time using observations from low-cost sensors and model information, Environment International, 106, 234-247, doi:10.1016/j.envint.2017.05.005, 2017.

Schneider, P., Bartonova, A., Castell, N., Dauge, F. R., Gerboles, M., Hagler, G. S. W., Hüglin, Ch., Jones, R. L., Khan, S., Lewis, A. C., Mijling, B., Müller, M., Penza, M., Spinelle, L., Stacey, B., Vogt, M., Wesseling, J., and Williams, R. W.: Toward a Unified Terminology of Processing Levels for Low-Cost Air-Quality Sensors, Environmental Science & Technology 2019 53 (15), 8485-8487, DOI: 10.1021/acs.est.9b03950, 2019.

Schiavina, M., Freire, S., and MacManus, K.: GHS population grid multitemporal (1975, 1990, 2000, 2015) R2019A. European Commission, Joint Research Centre (JRC), doi:10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218, 2019.

Smith, A., Lott, N., and Vose, R.: The Integrated Surface Database: Recent Developments and Partnerships. Bulletin of the American Meteorological Society, 92, 704–708, doi:10.1175/2011BAMS3015.1, 2011.

Sokhi, R. S., Middleton, D. R., Luhana, L.: Review of methods for NO to NO_2 conversion in plumes at short ranges. Science Report: SC030171/SR2. Environment Agency, Bristol, United Kingdom, 2007.

Tilloy, A., V. Mallet, D. Poulet, C. Pesin, and F. Brocheton: BLUE-based NO2 data assimilation at urban scale, J. Geophys. Res. Atmos.,118, 2031–2040, doi:10.1002/jgrd.50233, 2013.

WebMET: SAMSON Surface Met Data, http://www.webmet.com/MetGuide/Samson.html, last access: 25 October 2019, 2019a.

WebMET: TD6201 - Upper Air Data for AERMOD, http://www.webmet.com/MetGuide/TD6200.html, last access: 25 October 2019, 2019b

WHO: Global Urban Ambient Air Pollution Database (update 2016). World Health Organization, https://www.who.int/airpollution/data/cities-2016/en/, last access: 25 October 2019, 2016.

WMO: Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications, WMO-No. 1215, A. C. Lewis, E. Schneidemesser, and R. E. Peltier, Eds., World Meteorological Organization, 2018.

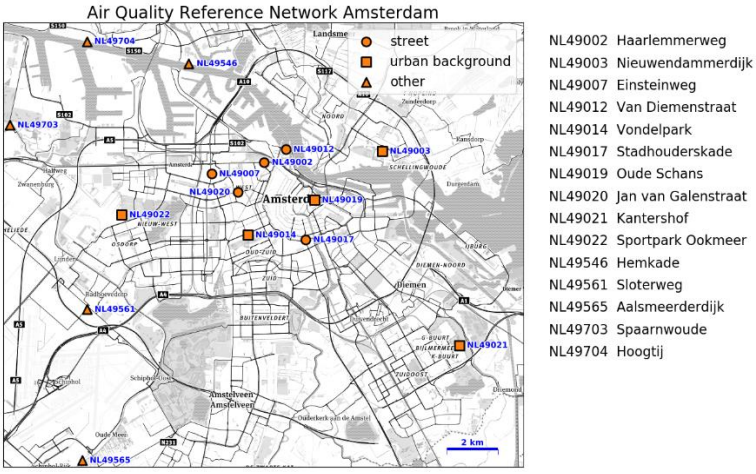


Figure 1: Air quality reference network of Amsterdam. (Basemap source: © Mapbox © OpenStreetMap contributors 2019. Distributed under a Creative Commons BY-SA License.)

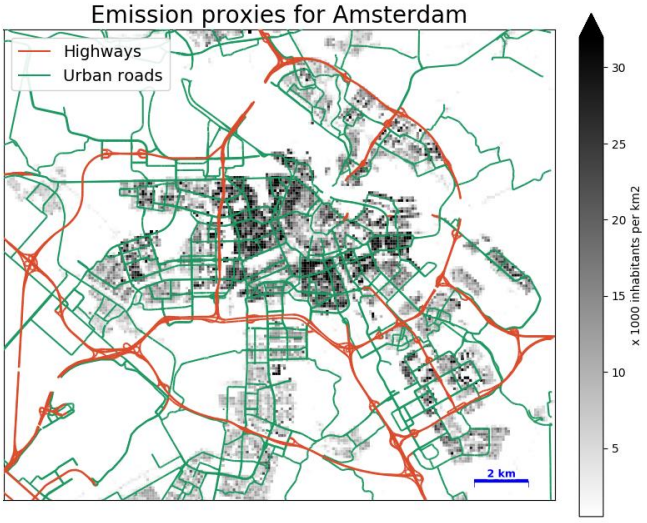


Figure ~~12~~: Map of the emission proxies used for the dispersion model. Red lines indicate the highways, green lines indicate the urban main roads. Grey colours indicate the population density. ~~The locations of the reference measurement sites are indicated with the yellow dots. Units on the axes are in meters.~~ (Road location data adopted from © OpenStreetMap contributors 2019. Distributed under a Creative Commons BY-SA License.)

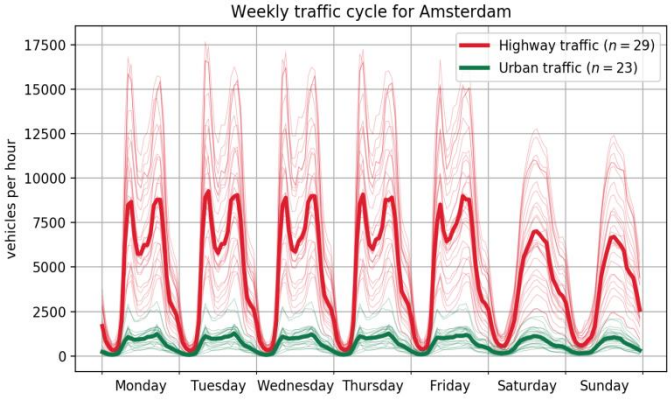


Figure ~~23~~: Weekly cycle of highways and urban roads at counting locations (thin lines), aggregated from hourly data from 2016. The thick lines show the median of traffic flow for both road types. ~~Weekly cycle of highways in the Amsterdam area, as opposed to the weekly cycle of urban roads.~~ The morning and evening rush hours on working days are clearly visible for highways. Urban traffic has, apart from lower volume, less distinct peaks. ~~The thick lines show the median of traffic flow for both road types.~~

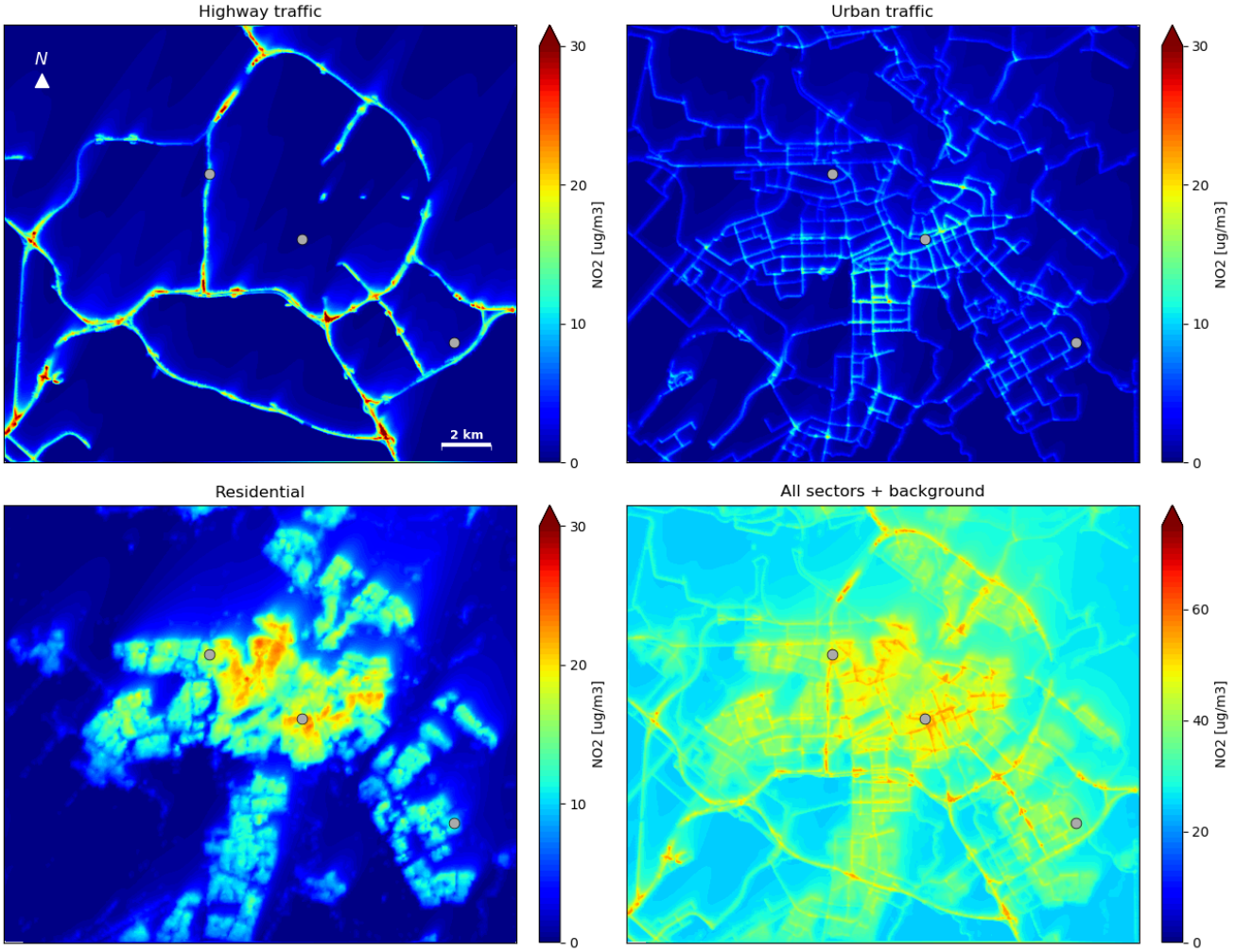


Figure 43a: Dispersion maps of NO_2 concentrations for each emission sector at 8 July 2016, 9:00. The lower right panel shows the linear combination which best fits the time series at the calibration sites. Wind is blowing from the southwest at 16 km/h. The grey dots indicate an urban background location, a street location, and a highway location.

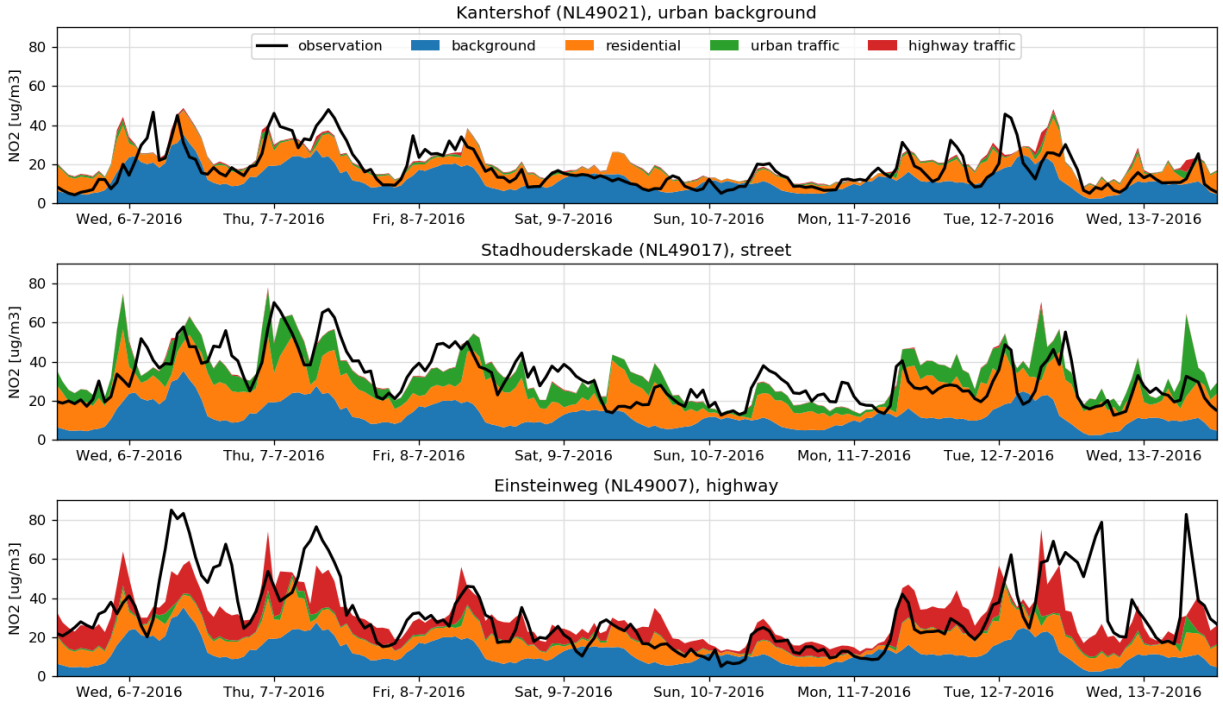


Figure ~~3b~~4b: Comparison of observed and simulated NO_2 time series for ~~three different sites (marked with grey dots above): an~~the urban background location, ~~a~~ the street location, and ~~a~~ the highway location. The colours indicate the simulated contribution of the three source sectors and the background.

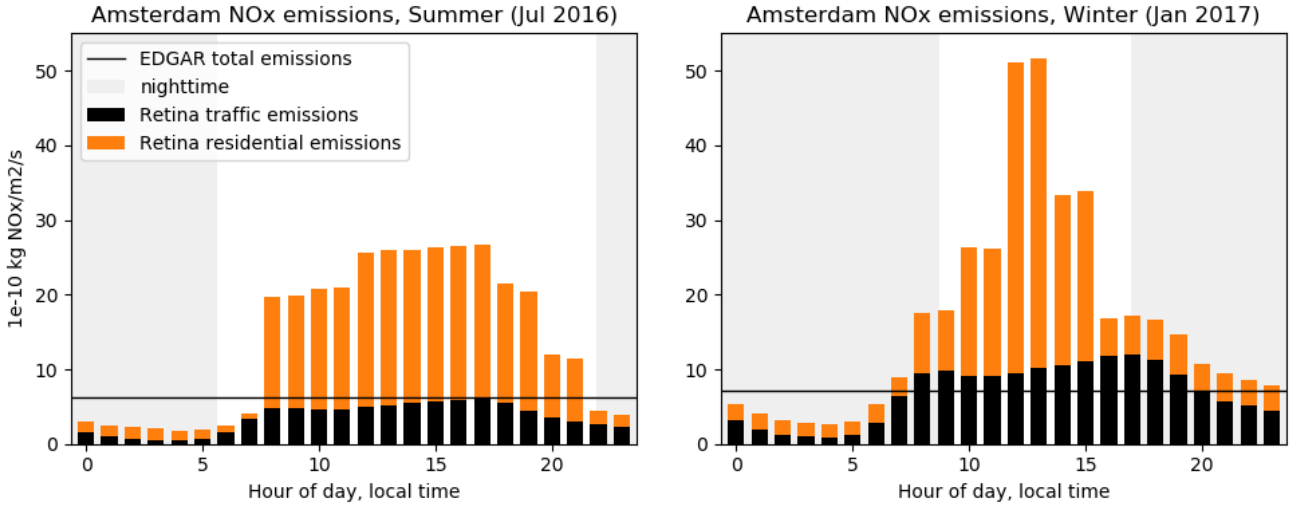


Figure 5: Diurnal emission cycles after calibration of emission factors in different seasons.

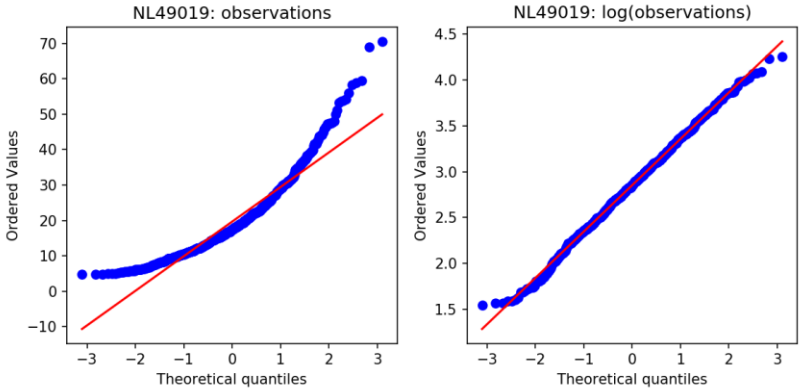
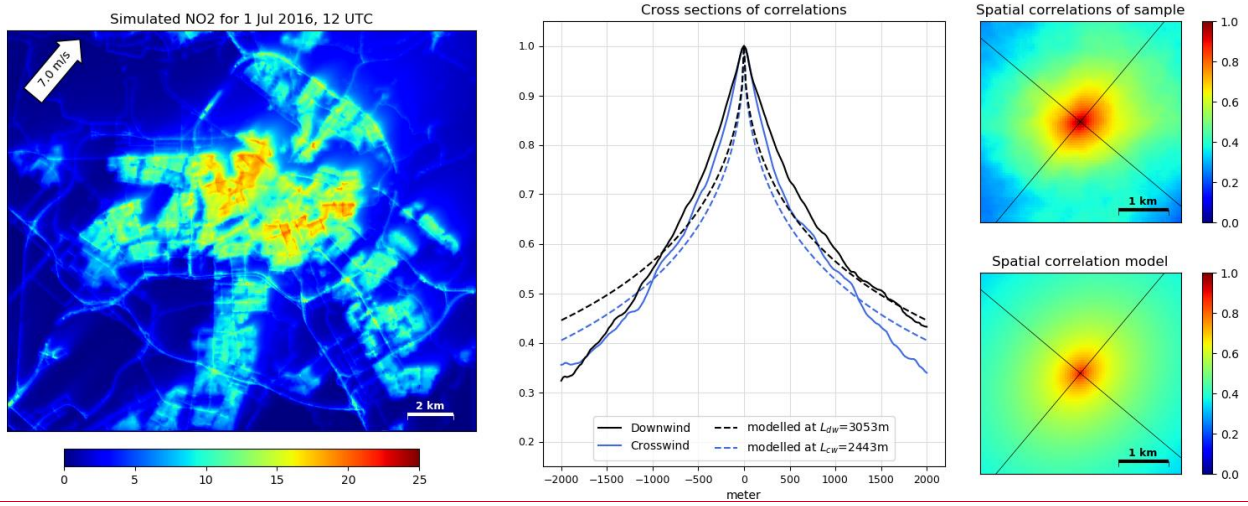


Figure 46: (left) Distribution of the NO_2 observations at reference station Oude Schans in July 2016 compared to a standard normal distribution. (right) The logarithm of the observed values correspond better to a Gaussian distribution, shown by the quantile value pairs being almost on a straight line.



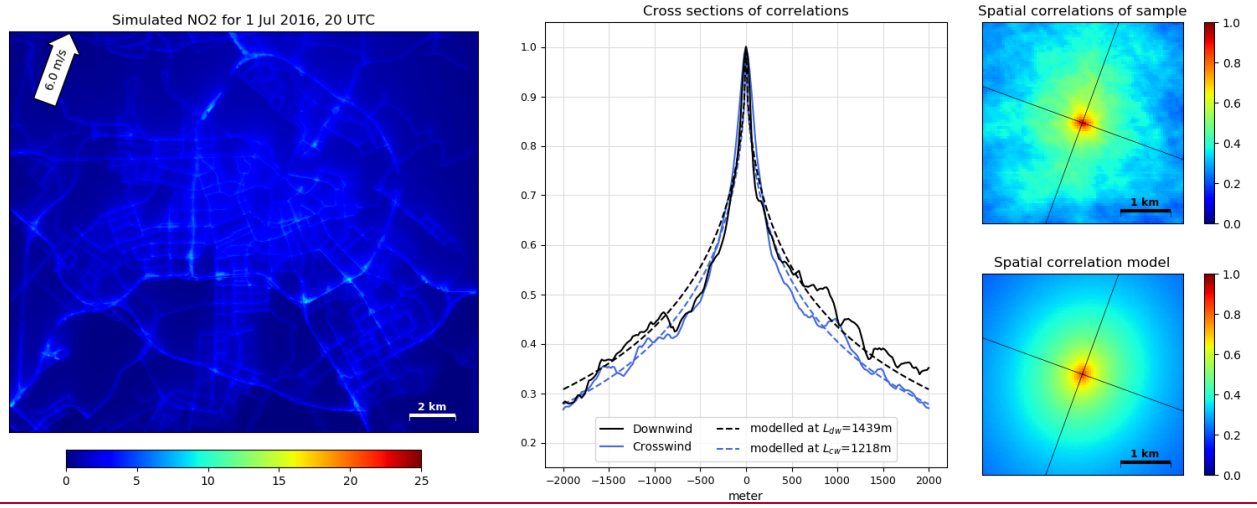
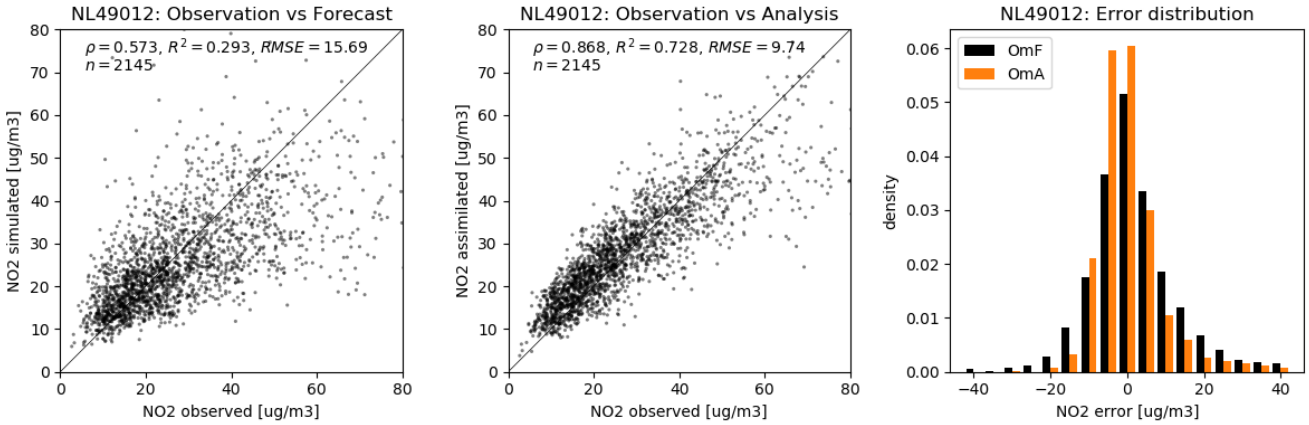


Figure ~~57~~: Left panels show ~~Simulated~~ simulated NO_2 concentration fields at two different hours. The middle panels show the spatial correlations along the downwind and crosswind axes based on a sample of n=1000. The right panels show ~~and~~ the spatial correlations of ~~a~~ the sample ~~(n=1000)~~ and the resulting ~~modeled~~modelled spatial correlation model. ~~Units are in meters.~~



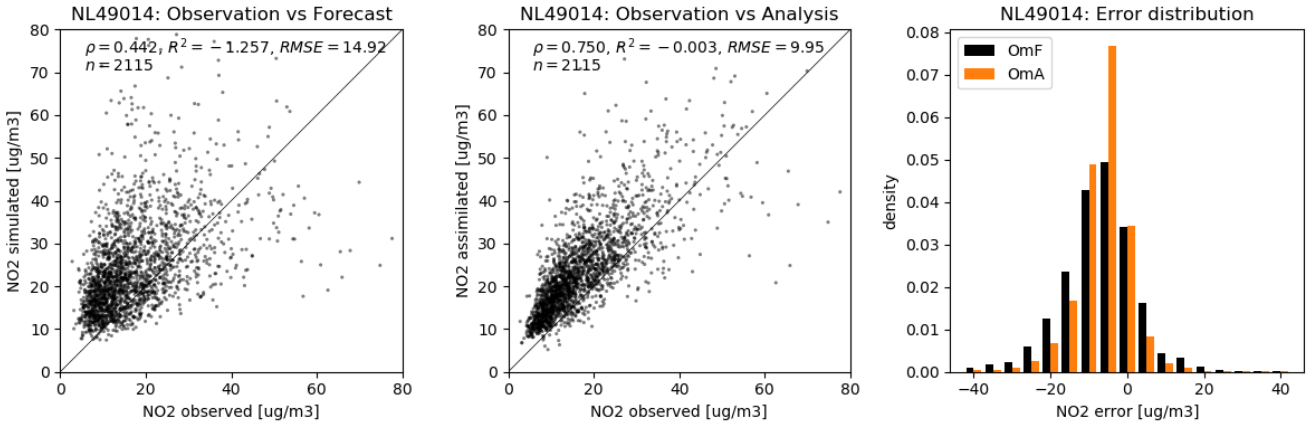


Figure ~~68~~: Validation of hourly time series for the period 1 June to 31 August, 2016, for a well performing street location (above) and a less performing urban background location (below). Statistics of the n data pairs are given in correlation (ρ), coefficient of determination (R^2), and RMSE. The right hand panels compare the error distributions: the observation minus forecast (OmF) against the observation minus analysis (OmA).~~(top) 8-day snapshot of NO_2 time series of observation, simulation, and assimilation at location "Oude Schans". Assimilation is performed with data from reference stations alone (green line), and with additional data from nearby low-cost sensor SD04 (blue line). (bottom) Scatter plots of observation against simulation and assimilation for the June 15 - August 15 2016 period. Statistics of the *n* data pairs are given in correlation (ρ), coefficient of determination (R^2), and RMSE.~~

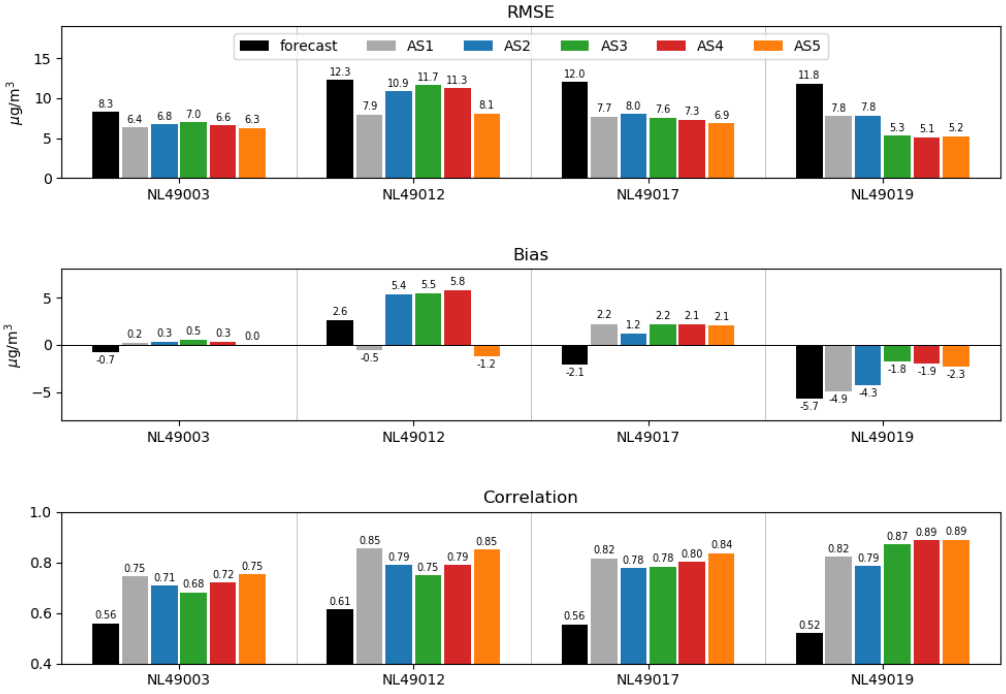
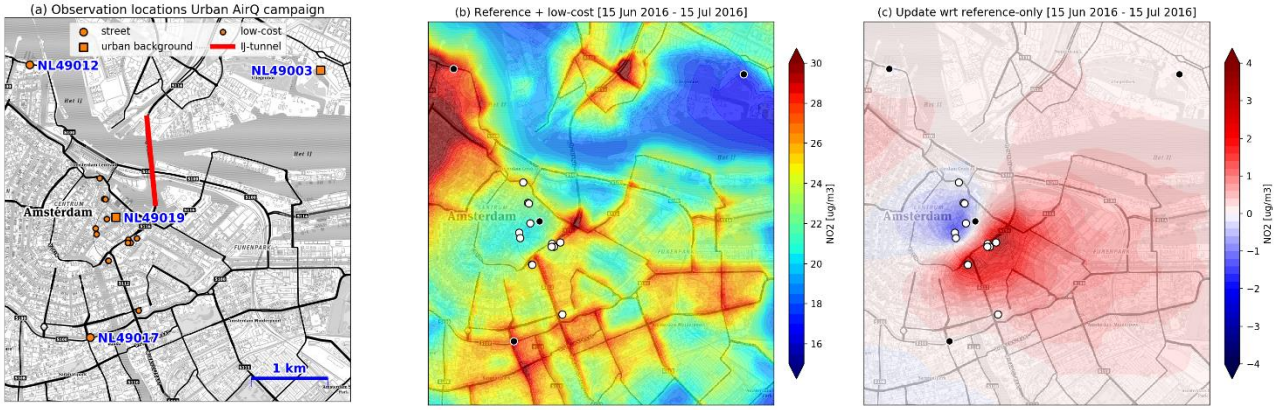


Figure 9: Validation of the model forecast and five different assimilation scenarios at four central reference sites, for the period June 15 to August 15, 2016.



1185

Figure 10~~7~~: (a) Observation sites during the Urban AirQ campaign. (~~a~~b) 30-day average of NO_2 concentrations in the ~~center~~ centre of Amsterdam, after assimilation of both reference measurements (black dots) and low-cost measurements (white dots). (~~b~~c) Changes in spatial pattern when low-cost measurements are included in the analysis. (Basemap source: © Mapbox © OpenStreetMap contributors 2019. Distributed under a Creative Commons BY-SA License.)

1190

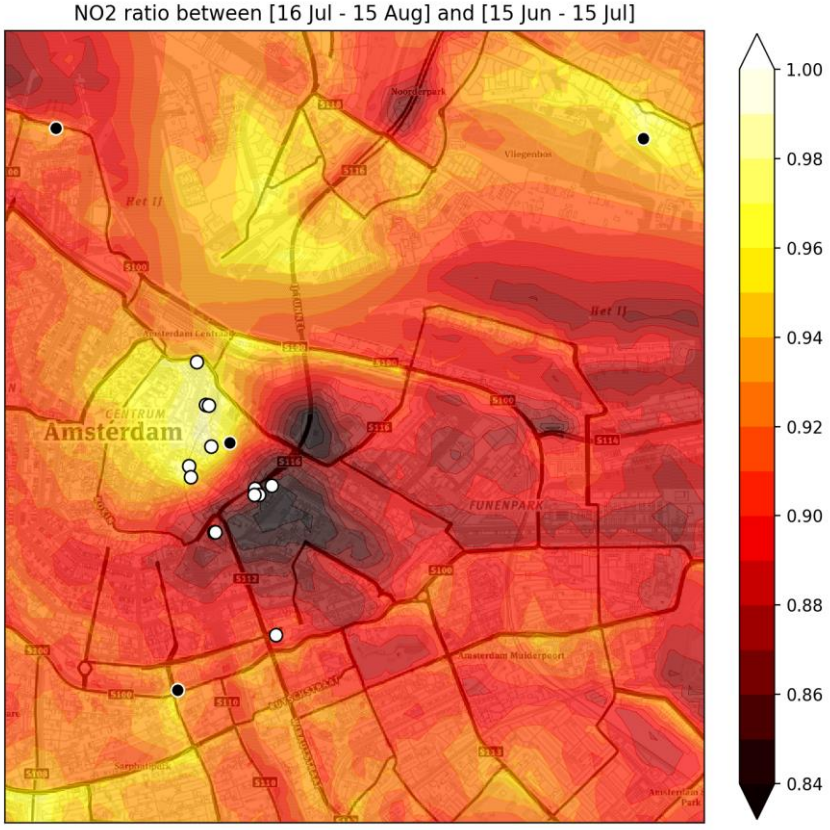


Figure ~~8~~11: Reduction of NO_2 during the holiday period. Largest reduction of concentrations is found in the vicinity of access ways to the IJ-tunnel, which was closed for maintenance. Concentrations in the historic ~~center~~centre remain unchanged. (Basemap source: © Mapbox © OpenStreetMap contributors 2019. Distributed under a Creative Commons BY-SA License.)

1195

Table 1: Overview of AERMOD simulation settings

Road width	20 m
Emission height traffic	0.5 m
Emission height residential	10 m
Initial vertical extension of concentration layer (sigma Z_0)	10 m
Receptor grid	Road following
Receptor height	1.5 m
Urban surface roughness length	1 m
NO_2/NO_X ratio	Ozone Limited Method (OLM), Primary emission ratio 10%
NO_X lifetime	2 h
Other AERMOD modelling options	Optimizing model runtime for sources (FASTALL) Address low wind speed conditions (LOWWIND3) Assuming flat terrain (FLAT)

Table 2: Summary of simulation input data for Amsterdam

Emission	Highway locations	OpenStreetMap (OSM, 2017): street segments labelled motorway and trunk
	Urban road locations	OpenStreetMap (OSM, 2017): street segments labelled primary, secondary, tertiary
	Highway traffic flow	National Data Warehouse for Traffic Information (NDW, 2019): weekly cycle of vehicle counts at 29 selected locations (2016), interpolated to street segments
	Urban traffic flow	Amsterdam municipality (personal communication): weekly cycle of vehicle counts at 24 locations (2016), interpolated to street segments
	Population data	Statistics Netherlands (CBS, 2016): population density (2014) gridded at 100 m resolution
Observation	Background NO_2	Copernicus Atmosphere Monitoring Service (CAMS, 2019): NO_2 analysis from model ensemble; minimum value found in 3x3 grid around domain ~~center~~centre
	Background O_3	Copernicus Atmosphere Monitoring Service (CAMS, 2019): O_3 analysis from model ensemble; mean value found in 3x3 grid around domain ~~center~~centre
Meteorology	Meteorology (surface)	Integrated Surface Database (ISD, 2019): hourly observations from Schiphol Airport weather station
	Meteorology (upper air)	Integrated Global Radiosonde Archive (IGRA, 2019): daily radio sounding at De Bilt (NL)

Table 3: Validation results at reference locations, June 1-August 31, 2016

ID	name	type	n[1)]	mean obs.	CAMS ensemble			Model forecast			Assimilated observations			
					RMSE[2)]	bias	corr	RMSE[2)]	bias	corr	RMSE[2)]	bias	corr	dist[3)]
NL49002	Amsterdam - Haarlemmerweg	street	2145	42.2	31.4	-25.6	0.49	22.6	-14.3	0.55	18.6	-14.5	0.83	0.99
NL49007	Amsterdam - Einsteinweg	street	2145	38.1	29.2	-21.4	0.42	19.6	-6.9	0.57	16.5	-6.2	0.72	1.26
NL49012	Amsterdam - Van Diemenstr.	street	2145	29.1	20.2	-12.5	0.53	15.7	-2.7	0.57	9.7	-0.5	0.87	0.99
NL49017	Amsterdam - Stadhouderskade	street	2140	30.1	17.9	-13.5	0.45	14.3	1.9	0.50	9.0	-2.7	0.78	1.60
NL49020	Amsterdam - Jan van Galenstraat	street	2131	34.8	24.0	-18.2	0.59	16.6	-4.7	0.58	11.1	-5.3	0.86	1.26
NL49003	Amsterdam - Nieuwend. dijk	urban backgr.	2145	16.6	8.6	0.1	0.60	10.5	2.0	0.47	7.5	0.8	0.71	3.28
NL49014	Amsterdam - Vondelpark	urban backgr.	2115	17.3	9.0	-0.7	0.52	14.9	7.9	0.44	9.9	6.5	0.75	1.73
NL49019	Amsterdam - Oude Schans	urban backgr.	2124	20.7	10.3	-4.1	0.59	13.8	5.8	0.50	8.7	4.6	0.81	1.60
NL49021	Amsterdam - Kantershof	urban backgr.	2082	14.9	7.5	1.6	0.65	10.7	5.6	0.56	8.0	4.4	0.73	7.33
NL49022	Amsterdam - Sportp. Ookmeer	urban backgr.	2124	14.3	8.4	2.4	0.65	9.2	3.4	0.66	8.0	3.7	0.80	3.89
NL49565	Oude Meer - Aalsmeerderdijk	rural	2127	17.3	9.1	-0.6	0.57	9.0	-2.4	0.59	8.0	-3.0	0.73	5.94
NL49703	Amsterdam - Spaarnwoude	rural	2125	13.0	8.7	3.7	0.61	8.1	2.1	0.60	7.5	2.4	0.71	4.47
NL49546	Zaanstad - Hemkade	industry	2145	22.9	14.3	-6.2	0.63	15.0	-8.1	0.66	13.0	-8.3	0.83	3.26
NL49704	Zaanstad - Hoogtij	industry	2120	19.6	12.7	-3.0	0.66	13.4	-6.0	0.72	12.1	-6.4	0.84	3.72
NL49561	Badhoevedorp - Sloterweg	undecided	2145	20.5	10.6	-3.9	0.64	10.8	-2.9	0.61	8.9	-4.2	0.79	3.96
Average street locations				**34.9**	**24.5**	**-18.2**	**0.50**	**17.8**	**-5.3**	**0.55**	**13.0**	**-5.8**	**0.81**	

Average urban background locations	**16.8**	**8.8**	**-0.1**	**0.60**	**11.8**	**4.9**	**0.53**	**8.4**	**4.0**	**0.76**	
Average all locations	**23.4**	**14.8**	**-6.8**	**0.57**	**13.6**	**-1.3**	**0.57**	**10.4**	**-1.9**	**0.78**	

1) Number of samples
2) In units $\mu g/m^3$
3) The distance to the nearest observation site, in km