CrossMark

Mapping urban air quality in near real-time using observations from low-cost sensors and model information

Philipp Schneider[*], Nuria Castell, Matthias Vogt, Franck R. Dauge, William A. Lahoz, Alena Bartonova

NILU - Norwegian Institute for Air Research, PO Box 100, Kjeller 2027, Norway

ABSTRACT

The recent emergence of low-cost microsensors measuring various air pollutants has significant potential for carrying out high-resolution mapping of air quality in the urban environment. However, the data obtained by such sensors are generally less reliable than that from standard equipment and they are subject to significant data gaps in both space and time. In order to overcome this issue, we present here a data fusion method based on geostatistics that allows for merging observations of air quality from a network of low-cost sensors with spatial information from an urban-scale air quality model. The performance of the methodology is evaluated for nitrogen dioxide in Oslo, Norway, using both simulated datasets and real-world measurements from a low-cost sensor network for January 2016. The results indicate that the method is capable of producing realistic hourly concentration fields of urban nitrogen dioxide that inherit the spatial patterns from the model and adjust the prior values using the information from the sensor network. The accuracy of the data fusion method is dependent on various factors including the total number of observations, their spatial distribution, their uncertainty (both in terms of systematic biases and random errors), as well as the ability of the model to provide realistic spatial patterns of urban air pollution. A validation against official data from air quality monitoring stations equipped with reference instrumentation indicates that the data fusion method is capable of reproducing city-wide averaged official values with an R^2 of 0.89 and a root mean squared error of 14.3 µg m^{-3}. It is further capable of reproducing the typical daily cycles of nitrogen dioxide. Overall, the results indicate that the method provides a robust way of extracting useful information from uncertain sensor data using only a time-invariant model dataset and the knowledge contained within an entire sensor network.

1. Introduction

With an ever-increasing amount of environmental observations available through methods such as crowdsourcing, citizen science, and participatory sensing, one of the major emerging challenges is how to best make sense of the vast amount of collected observations and how to provide citizens and other end-users with a relevant value-added product. Air pollution is a major environmental concern in many areas worldwide, with significant impacts on societal health and economy (World Health Organization, 2016; Guerreiro et al., 2014). Poor air quality is of particularly significant concern for many large urban agglomerations (Baklanov et al., 2016; Schneider et al., 2015). However, detailed observation-based urban-scale air quality maps are very scarce as the traditional highly accurate observation network is very costly and the resulting low number of air quality monitoring stations with reference equipment is generally not able to adequately capture the

small-scale spatial variability of air pollutants in the urban environment.

Recent technological advances related to sensor technology have resulted in comparatively low-cost and small devices for measuring air quality (Castell et al., 2014; Borrego et al., 2016; Spinelle et al., 2015; Kumar et al., 2015; Mead et al., 2013; Aleixandre and Gerboles, 2012; Piedrahita et al., 2014; Snyder et al., 2013; De Nazelle et al., 2013). Applying various elements from Citizen Science (Hand, 2010; Serrano Sanz et al., 2014) and crowdsourcing (Howe, 2006), a high-density network of such low-cost air quality sensors has significant potential for improving spatial mapping in general and in urban areas in particular. However, most datasets of observations made within a crowdsourcing framework contain substantial data gaps and the observations are generally highly irregular point measurements, which are only representative of a relatively small area. This poses a significant challenge in using such observations for mapping applications. One way to

overcome these issues is to combine such crowdsourced information with model data, which has complete spatial coverage.

We present a geostatistical data fusion technique for combining near real-time observations of urban air quality from low-cost sensors platforms with output from an urban-scale air pollution dispersion model, with the objective of providing highly detailed, up-to-date maps of urban air quality. Data fusion is conceptually similar to data assimilation (Kalnay, 2003; Lahoz et al., 2010; Lahoz and Schneider, 2014). It describes a set of techniques for merging two or more datasets and thus generating a product of higher overall quality. Data fusion techniques, as a subset of data assimilation (Lahoz and Schneider, 2014), allow for combining observations with model data in a mathematically objective way (through the best linear unbiased estimate) and therefore provide a means of adding value to both the observations and the model. The gaps in the observations are filled and the model is constrained by the observations. The model further provides detailed spatial patterns in areas where no observations are available. As such, data fusion of observations from high-density low-cost sensor networks together with models can contribute to significantly improving urban-scale air quality mapping.

As the use of low-cost microsensors for air quality applications became possible only relatively recently, not many studies have been carried out for using this information for mapping urban-scale air quality. While there are already numerous studies using such sensors for general monitoring and personal exposure assessment (Peters et al., 2013; Nieuwenhuijsen et al., 2015; Castell et al., 2014; Piedrahita et al., 2014; Steinle et al., 2013; Snyder et al., 2013; De Nazelle et al., 2013), the number of studies using such sensor devices specifically for mapping urban air quality are quite limited. Those that are relevant include primarily those investigating the use of mobile air quality sensors for generating longer-term average maps along the street network and areas in which the mobile measurements are representative (Van den Bossche et al., 2015; Mueller et al., 2016; Peters et al., 2014) or for the urban area as a whole (Hasenfratz et al., 2015). Other studies have used a network of fixed passive samplers for creating longer-term average maps of urban air quality, for example using generalized additive models (Mueller et al., 2015) or applying land-use regression techniques (e.g. Beelen et al., 2013). Even though they used observations from official air quality monitoring stations and not low-cost sensor data, Tilloy et al. (2013) showed that data assimilation of air quality observations from 9 fixed sites into an urban air quality model is feasible and can account for up to 50% reduction in root mean squared error in areas of high station density.

To our knowledge, no previous studies have applied geostatistical data fusion techniques for combining near real-time data from a network of fixed low-cost microsensor with data from an urban-scale dispersion model.

2. Data and methodology

2.1. Low-cost sensor observations

We deployed a network of AQMesh platforms for monitoring air quality in Oslo, Norway. AQMesh units (provided by Environmental Instruments Ltd, UK, www.aqmesh.com) are battery driven stationary platforms which measure the four gaseous components carbon monoxide (CO), nitrogen oxide (NO), nitrogen dioxide (NO_2), ozone (O_3) and particle count. The AQMesh platform also measures air temperature, relative humidity and atmospheric pressure. The data is post-processed by the manufacturer with the aim to correct cross-interferences as well as the effect of temperature and relative humidity. The version of the AQMesh platform that was used here is v3.5. This version includes an O_3-filtered NO_2 sensor from Alphasense, which is designed to reject O_3 and hence eliminate cross-sensitivity issues. CO, NO and O_3 are measured by electrochemical sensors from the Alphasense Series B. While AQMesh units can be configured to deliver 15-min averaged

data, we used here the standard averaging period of 1 h to reduce random noise. An integrated GPRS modem in each unit allows data transfer to the AQMesh database server. The data were then downloaded from a dedicated web-site.

Testing of the sensor platforms was carried out as follows. For the period from 13th April 2015 to 24th June 2015, a total number of 24 AQMesh platforms were co-located at an air quality reference monitoring station at Kirkeveien street, Oslo, Norway. The Kirkeveien station (10.7245 ° E, 59.9323 ° N) is located in a street with busy traffic and is equipped with CEN approved gas analysers for CO, O_3 and nitrogen oxides (NO_x). CO is measured using non-dispersive infrared spectroscopy (EN14626), NO_x is measured using chemiluminescence (EN14211) and O_3 is measured using UV photometry (EN14625). While we focus here on NO_2 as an example for mapping applications, the performance of the sensors for related gases such as NO and O_3 was also evaluated.

The following metrics were used for comparing sensor platform data at time t (M_t) with observations from the reference instrumentation (O_t), where n represents the total number of observations, and σ_M and σ_O represent the respective standard deviations for sensors observations and reference:

- Mean Bias (MB)

$$MB = \frac{1}{n} \sum_{t=1}^{n} (M_t - O_t) \tag{1}$$

- Mean Gross Error (MGE)

$$MGE = \frac{1}{n} \sum_{t=1}^{n} |M_t - O_t| \tag{2}$$

- Normalized Mean Bias (NMB)

$$NMB = \frac{\sum_{t=1}^{n} (M_t - O_t)}{\sum_{t=1}^{n} O_t} \tag{3}$$

- Normalized Mean Gross Error (NMGE)

$$NMGE = \frac{\sum_{t=1}^{n} |M_t - O_t|}{\sum_{t=1}^{n} O_t} \tag{4}$$

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (M_t - O_t)^2}{n}} \tag{5}$$

- Correlation Coefficient (r)

$$r = \frac{1}{n-1} \sum_{t=1}^{n} \left(\frac{M_t - \overline{M}}{\sigma_M} \right) \left(\frac{O_t - \overline{O}}{\sigma_O} \right) \tag{6}$$

Table 1 shows the results of the co-location of the 24 AQMesh nodes, for CO, NO_2, and O_3. The results indicate that the mean bias can be significant for some of the pollutants. For example, for NO_2 the bias can reach 75 parts per billion (ppb). The bias varies from sensor to sensor. For example, for O_3 the bias varies in the range between −29 ppb and 41 ppb. NO sensor measurements show a good agreement with the reference instrumentation, with an average correlation of 0.86. All the NO sensors have a correlation above 0.6. This is not the case for NO_2, where 19 out of the 24 pods have a correlation below 0.6.

The co-location results show that even for the same sensor type and platform version the performance can be very different from sensor to

Table 1
Results of a co-location of 24 AQMesh units at the Kirkeveien reference air quality monitoring station carried out between 13th April 2015 and 24th June 2015. Only data for sensors relevant for NO_2 are shown here. MB indicates the mean bias, MGE the mean gross error, NMB the normalized mean bias, NMGE the normalized mean gross error, RMSE the root mean squared error, and r is the Pearson correlation coefficient. Units for MB, MGE, NMB, NMGE, and RMSE are given in ppb. More information about these results and additional plots can be found in Castell et al. (2017).

Species	Metric	MB	MGE	NMB	NMGE	RMSE	r
NO	Average	−0.54	12.48	0	0.64	16.35	0.86
	Max	12.75	22.01	0.73	1.34	30.94	0.98
	Min	−15.05	4.84	−0.71	0.25	6.97	0.6
NO_2	Average	13.3	26.23	0.98	1.79	30.27	0.49
	Max	74.66	74.69	5.42	5.42	81.6	0.72
	Min	−22.73	12.56	−1.31	0.85	15.52	0.21
O_3	Average	6.76	19.87	0.62	1.64	22.2	0.54
	Max	40.71	40.96	3.52	3.53	44.27	0.81
	Min	−28.66	9.6	−1.9	0.79	11.77	0.09

sensor. To reduce the bias and errors, we applied a linear regression employing the calibration data from the co-location. The slope and offset were calculated for each of the 24 AQMesh platforms. In addition, the calibration was carried out separately for each individual gas sensor within each platform in order to achieve the best possible performance for the various species. This process reduced the average RMSE from 30 ppb to 9 ppb for NO_2. It should be noted that the sensor evaluation was carried out in the spring when NO_2 concentrations are moderate. When pollution levels are higher, the sensors exhibit a much better signal-to-noise ratio and have a lower uncertainty than indicated by Table 1. For this reason, NO_2 observations during the month of January 2016, when pollution levels were relatively high, were selected for demonstrating the mapping methodology. More information about the sensor performance as measured in Oslo can be found in Castell et al. (2017).

The 24 AQMesh units were deployed throughout Oslo by installing them together with kindergarten teachers and students on the premises of various kindergartens throughout Oslo. The kindergarten teachers and the parents of the children have a vested interest in planning the children's outdoor activities according to the local air pollution levels and were therefore very interested in the measurement results. Fig. 1 shows a map of the Oslo study site and the location of the 24 AQMesh pods located on the kindergarten premises.

2.2. The EPISODE dispersion model

We used the EPISODE dispersion model as a source of model information on air quality for our study site. EPISODE is a 3-D Eulerian/Lagrangian dispersion model that provides urban- and regional-scale air quality forecasts of atmospheric pollutants. The model, which is described in detail in Slørdal et al. (2003), is a Eulerian grid model with embedded subgrid models for computing the various pollutant concentrations that result from area-, point-, and line-based emission sources. Applying finite difference numerical methods, EPISODE integrates forward in time and solves the time-dependent advection and diffusion equation on a three-dimensional grid. EPISODE provides schemes for advection, turbulence, deposition, and chemistry, although the latter was not activated here for reasons of computational performance. The EPISODE model as used for this study calculates NO_2 concentrations using the photochemical steady state assumption. This assumption works well under conditions where net ozone formation is not occurring or is limited. Such cases include situations close to the NO_x emission sources where net ozone can be discounted, i.e., in urban areas with ubiquitous NO_x emission sources. This assumption works especially well in urban areas in higher latitude conditions where lower temperatures and lower actinic flux reduce the role played by net ozone formation. EPISODE contains a sub-grid line source model based on a

standard integrated Gaussian model (Petersen, 1980), which computes the concentration levels of non-reactive pollutants from road traffic over distances up to hundreds of meters downwind. Most commonly, EPISODE is used for modeling airborne species such as NO_2, NO_x, PM_{10}, $PM_{2.5}$, CO, and sO_2. Validation studies have shown good correspondence between modeled and measured concentrations of NO_2, PM_{10}, and $PM_{2.5}$ (Oftedal et al., 2009).

The emissions used for the dispersion modeling with EPISODE represent all major emissions sectors including traffic, industry, wood burning, and shipping. The traffic volume (average daily traffic) was estimated using a traffic model. The vehicle fleet distribution (including heavy duty vehicles) and the basic emission factors were obtained from the Norwegian public roads administration. Industrial emissions from point sources were acquired from the Norwegian Environment Agency. Shipping emissions were obtained from the Norwegian Coastal Administration. Other emissions, such as agriculture, offroad, and air transportation were obtained from Statistics Norway or local authorities. As background concentrations for running EPISODE we used results from the regional model ensemble of the Copernicus Atmosphere Monitoring Service (CAMS), comprising a total of seven models (Marécal et al., 2015). The EPISODE model reads these concentrations hourly around the domain of the local-scale simulation.

The EPISODE model is run at hourly time steps for an entire year. The annual average concentration field is then computed by averaging over all hours. While EPISODE can be run at horizontal spatial resolutions down to 100 m, it is most typically run at 1000 m horizontal spatial resolution as this is the spatial scale at which areal emissions are known most reliably. However, this resolution is not sufficient for providing the detailed spatial patterns that are required when using the data for fusion with observations from a network of low-cost sensors. For this reason we use a downscaling procedure to obtain high-resolution concentrations fields of 100 m × 100 m horizontal spatial resolution or even lower (Denby et al., 2014). The downscaling procedure used here exploits the fact that line source emissions such as roads in the model are given at essentially "infinite" spatial resolution as they are stored as vector-based line data. As line sources are a major source of pollutant emissions affecting air quality in urban areas, the model is able to provide information at much higher spatial resolution than what would be expected based on the gridded input data alone. The downscaling is performed following Denby et al. (2014) by distributing a high-density network of receptor points within the modeling domain. These points are distributed both at regular sampling intervals throughout the domain and, in addition, at a substantially higher density along roads and other line sources. The density decreases exponentially with distance from the line source. The model then calculates the concentration at each receptor point taking into account raster-based areal emissions, and Gaussian dispersion from vector-based emissions from line and points sources (Slørdal et al., 2003). The resulting high-density set of concentrations is then interpolated to the desired output resolution using geostatistical techniques (ordinary kriging) (Goovaerts, 1997; Chilès and Delfiner, 2012). It should be noted that the downscaled map is not directly comparable in terms of absolute concentrations to the original gridded concentration fields at coarse resolution. The reason for this is the different vertical representativity of the model output, i.e. the vertical extent for which the given concentrations are valid. While the original 1000 m spatial resolution grids represent the lowermost model layer which ranges from the surface to a height of 20 m, the receptor points are located at 2 m height above the surface. As such, the receptor points generally exhibit significantly higher concentrations than the average over the lowermost 20 m. While a direct comparison between the datasets is therefore not possible, the fact that the downscaled map is representative at a height of 2 m above ground has the advantage of being a more realistic estimate of street-level air pollution.

Fig. 2 shows an example result of the downscaling methodology for the 2011 annual average concentrations of NO_2 in the area of greater
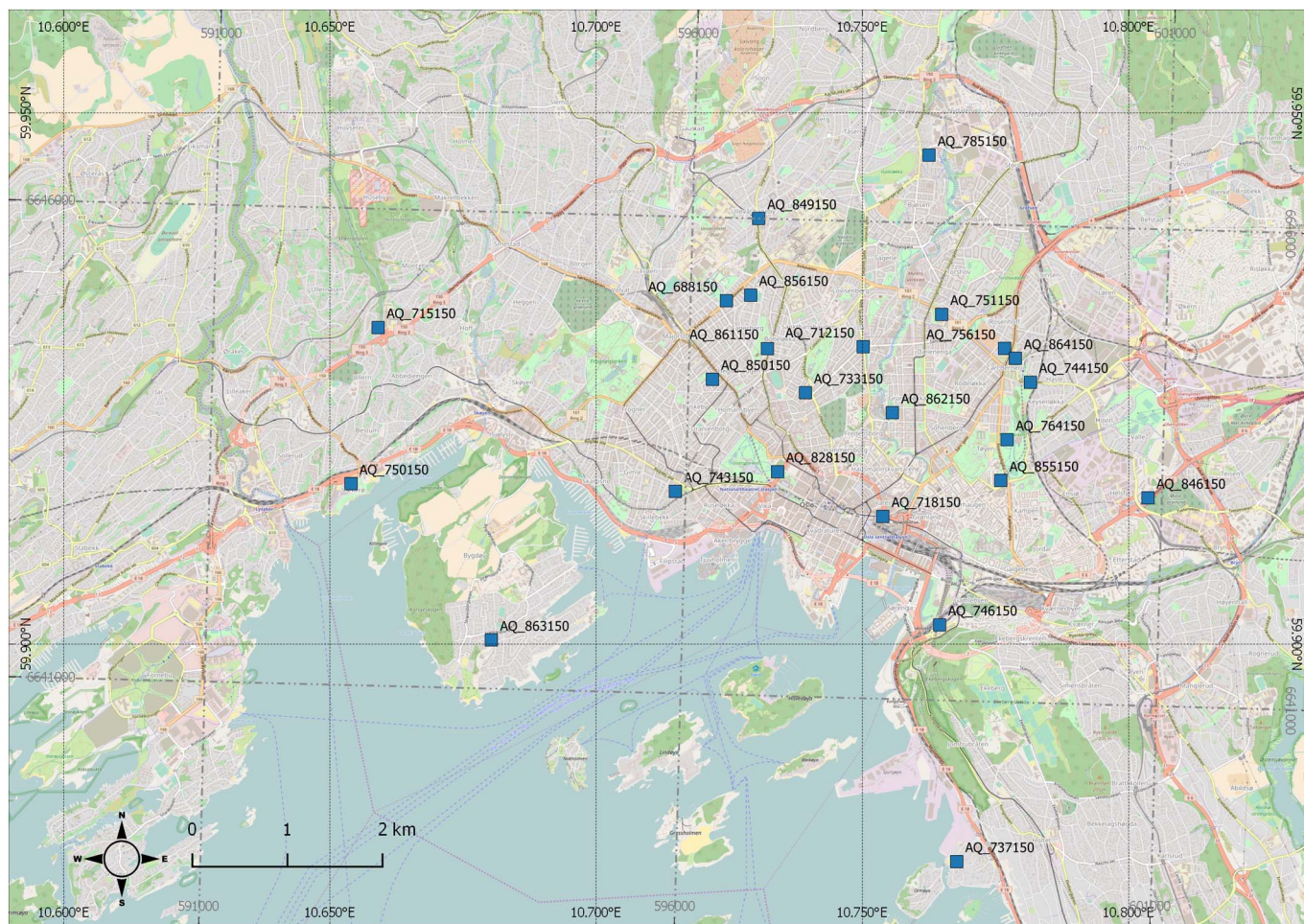
Fig. 1. Overview of the greater Oslo area with the deployment locations of the 24 AQMesh pods marked in blue. Background map data provided by the OpenStreetMap project.

Oslo, Norway. The top left panel shows the original gridded output of EPISODE. The downtown area of Oslo slightly north of the map center exhibits the highest values. A simple bilinear interpolation from 1 km × 1 km to 100 m × 100 m is shown in the top right panel for reference. In contrast, our downscaling method uses a dense network of receptor points (bottom left panel), which were distributed on both sides along major road links in increasing distance intervals up to a distance of 400 m. Outside of those areas the receptor points were distributed on a regular grid pattern of 500 m distance. The bottom right panel of Fig. 2 shows the downscaled map after spatial interpolation of the receptor point concentration values. Using EPISODE and downscaling methodology described above, high-resolution annual average concentration fields were derived, which were used as a *basemap* or *climatology* (long-term mean) for the data fusion process. Fig. 3 shows an example basemap for NO_2 derived from the model and the corresponding experimental and fitted semivariogram. A semivariogram is a function describing the characteristics of the spatial correlation between observations and is frequently used in geostatistics (Goovaerts, 1997; Wackernagel, 2003; Chilès and Delfiner, 2012). The basemap for NO_2 shows spatial patterns with quite steep gradients, which are primarily linked to line-source emissions from road transport. The corresponding semivariogram (Fig. 3 right panel) reflects the spatial gradients of the annual average map. The semivariogram for NO_2 was fitted by an exponential model with 117.0 sill (the semivariance value at which the variogram levels off) and a range (lag distance at which the semivariogram reaches 63% of the sill value for exponential models) of 9493 m with a nugget effect (a non-zero semivariance at the origin typically representing variability at distance smaller than the sampling distance) of 0.86. This model was selected from a choice of six

different models by an automated process based on the smallest residual sum of square with the sample semivariogram.

2.3. Data fusion methodology

The data fusion methodology applied here is based on geostatistical principles (Isaaks and Srivastava, 1989; Cressie, 1993; Isaaks and Srivastava, 1989; Cressie, 1993; Goovaerts, 1997; Kitanidis, 1997; Wackernagel, 2003; Webster and Oliver, 2007; Sarma, 2009; Chilès and Delfiner, 2012). One of the fundamental techniques in geostatistics is kriging, which allows for interpolating between values by modeling them with a Gaussian process that is subject to prior covariances. Our method uses the more advanced universal kriging technique (Goovaerts, 1997) to combine observations with model data. It does this by predicting the concentrations at unknown locations through simultaneously interpolating the observations and using the model data as proxy information to provide information about the spatial patterns (or trend). In contrast to ordinary kriging (Goovaerts, 1997; Chilès and Delfiner, 2012), universal kriging allows for the overall mean to be non-constant throughout the domain and to be a function of one or more explanatory variables. Universal kriging is similar to kriging with external drift and mathematically equivalent to regression kriging (Hengl et al., 2007) or residual kriging (Denby et al., 2010; Horálek et al., 2013) but can perform the linear regression against auxiliary variables and the spatial interpolation of the corresponding residuals in a single step. Universal kriging assumes a non-stationary mean and in addition the presence of local spatial variation. As such the parameter in question is modeled by a deterministic regression component that provides the large-scale spatial variation and provides spatial patterns in areas
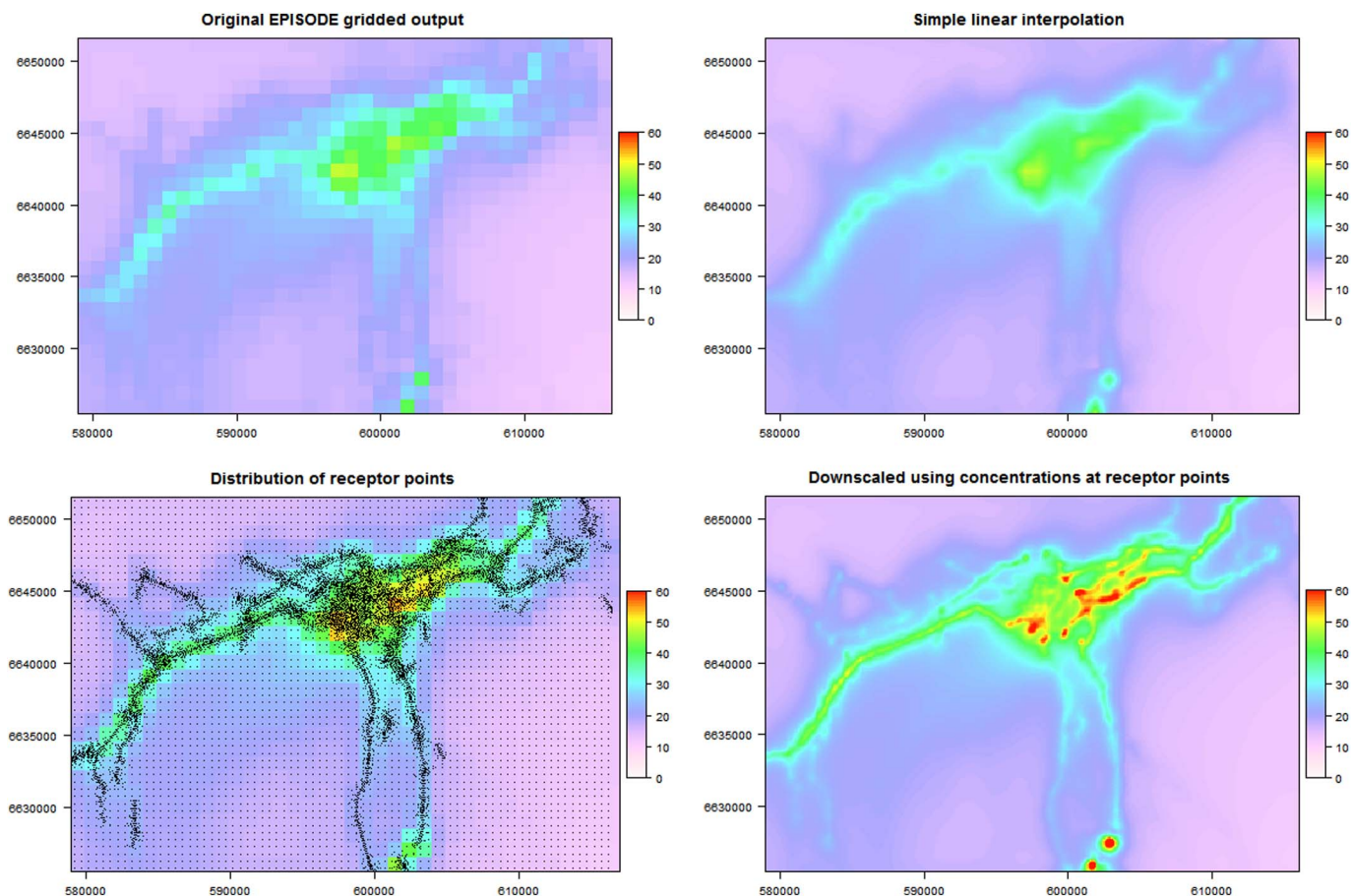
Fig. 2. Downscaling method for EPISODE, shown for the 2011 annual mean surface NO$_2$ field for Oslo, Norway, in units of μg m^{-3}. Top left panel: Original gridded output from the EPISODE model at 1000 m horizontal spatial resolution. Top right panel: The original field regridded to 100 m horizontal spatial resolution using simple linear interpolation as a reference. Bottom left panel: Original gridded EPISODE concentration with locations of receptor points overlaid in black. Bottom right panel: Downscaled concentration field at 100 m horizontal spatial resolution derived through spatial interpolation of receptor point concentrations. Coordinates in UTM32N/WGS84.
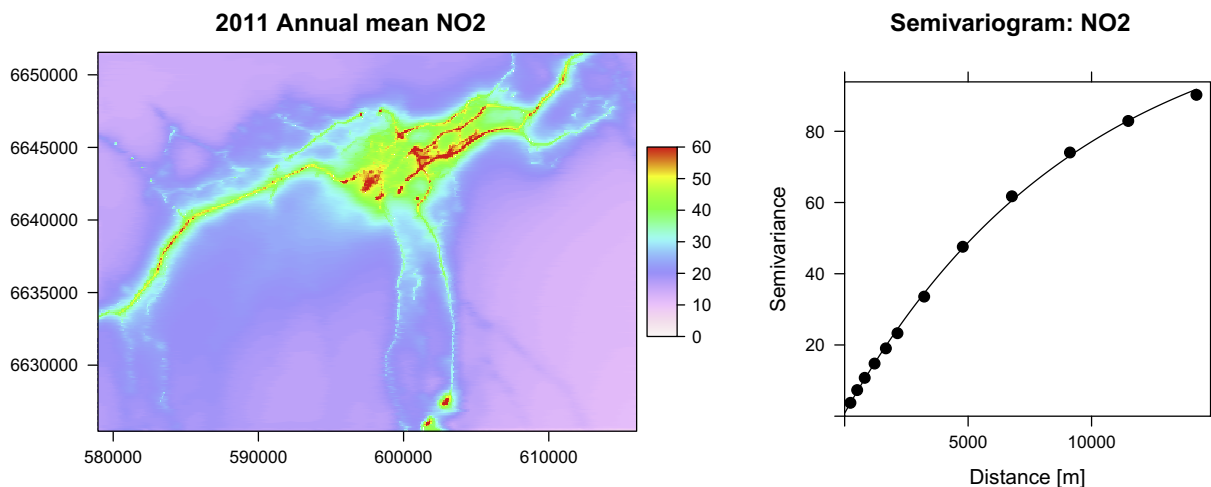

Fig. 3. Downscaled model-derived basemap (left panel) for NO$_2$ as produced by the EPISODE model and the corresponding semivariogram (right panel) for Oslo, Norway. The semivariogram was calculated over a total of 96,831 grid points. The concentrations are given in μg m^{-3}. Coordinates in UTM32N/WGS84.

where no observations are available, and a kriging component that provides the small-scale random variation. Kriging in general and universal kriging in particular is also conceptually similar to data assimilation methods such as Optimal Interpolation (OI), which is often used in meteorology, as well as the Kalman Filter (Bertino et al., 2003).

In general, the estimated concentration $\widehat{Y}(s_0)$ at point s_0 is computed as

$$\widehat{Y}(s_0) = c + a_1 \cdot x_1(s_0) + a_2 \cdot x_2(s_0) + \ldots + a_p \cdot x_p(s_0) + \varepsilon(s_0) \qquad (7)$$

where c is a constant, a_1, a_2, etc are regression coefficients, x_1, x_2, …, x_p are the values of the p predictor variables of the regression component, and ε is a stationary random process with a given semivariogram. In matrix notation we get

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1(s_0) & \cdots & x_p(s_0) \\ 1 & \vdots & & \vdots \\ 1 & x_1(s_n) & \cdots & x_p(s_n) \end{bmatrix} \begin{bmatrix} c \\ a_1 \\ \vdots \\ a_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} = Xa + \varepsilon$$

(8)

where Y indicates the estimated values at all prediction locations, X represents the values of the predictor variables at all locations, a is the vector of regression coefficients, ε indicates the vector of residuals error that is estimated using kriging with the known semivariogram model, n is the number of prediction locations and p is the number of predictor variables.

In practice, the spatial trend or drift of the mean is estimated here using a single predictor variable, which is the annual average concentration map provided by the EPISODE air pollution dispersion model (see Fig. 3), with the intercept set to zero. The observations are provided by the air quality sensors deployed throughout the environment. As such, the system takes the overall spatial patterns of the concentration field from the annual average map, which acts as a *climatology* (essentially a long-term mean), and adjusts this field based on the observations. The annual average concentration map is assumed to be a reasonable and simple-to-generate first-guess of the typical spatial patterns of air quality in the urban environment. It should be noted here that, while the sensors primarily drive the temporal effects, they also capture fine-scale spatial variability that is not adequately represented by the dispersion model in general and the annual average concentration map in particular.

In addition to an urban-scale dispersion model as it was used here, other approaches such as land use regression (Hoek et al., 2008; Beelen et al., 2013) can be used to generate a basemap. As the modeled concentration field remains constant through time in our method, the model does not provide any information on the temporal behavior of the pollutants. This temporal information is extracted entirely from the observations made by the low-cost sensor network. Our method can be readily extended to use basemaps with shorter averaging periods, such as seasonal or monthly means, or even hourly model information. However, this generally requires an air pollution dispersion model to be run operationally. Since our method was designed to work also in locations that do not run such a model, and since we want the results to be primarily driven by the low-cost sensor observations themselves rather than by the model output, we focus here on the performance achievable by using annual average concentration maps.

Before the actual data fusion takes place, both the modeled and observed concentrations are first transformed into log-space using the natural logarithm. This approach follows previous work such as that carried out by Denby et al. (2008), De Smet et al. (2010), and Horálek et al. (2014) and is done because the frequency distribution of observed and modeled concentrations most often resembles the lognormal distribution. A log-transformation therefore is able to convert these distributions into an approximately Gaussian distribution, which is what is assumed for universal kriging. Taking the lognormal distribution of the concentrations into account has further been shown to provide superior mapping accuracy (Denby et al., 2008; Horálek et al., 2013). The theoretical semivariogram required for calculating the covariances in the kriging process was fitted automatically to the empirical semivariogram for each new set of observations (generally at hourly intervals) based on the smallest residual sum of square with the sample semivariogram.

After universal kriging is carried out in log-space, the resulting concentration field and the corresponding mapping uncertainty have to be back-transformed from log-space into the space of the original observations. Denby et al. (2008) showed that the theoretical back-transformed expectation value of a concentration C is given as

$$E[C] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

(9)

where μ and σ represent the mean and standard deviation of the lognormal-transformed data, respectively. In practice the concentration

values resulting from the data fusion process are thus back-transformed by exponentiation with the kriging error as

$$\hat{Z}(s_0) = \exp\left[\hat{Y}(s_0) + \frac{\sigma^2(s_0)}{2}\right]$$

(10)

where $\hat{Z}(s_0)$ is the estimated back-transformed concentration value at point s_0, $\hat{Y}(s_0)$ is the concentration at point s_0 resulting from the data fusion process, and $\sigma(s_0)$ is the kriging standard deviation at point s_0 (De Smet et al., 2010). The theoretical back-transformed variance of the log-normal distribution is computed as

$$\text{var}[C] = [\exp(\sigma^2) - 1] \cdot \exp[2\mu + \sigma^2]$$

(11)

where μ and σ represent the mean and standard deviation of the log-normal-transformed data, respectively (Denby et al., 2008). Thus the back-transformed standard deviation (uncertainty) $\delta(s_0)$ at point s_0 of the fused map can be calculated in practice as

$$\delta(s_0) = \sqrt{\exp[\sigma^2(s_0) - 1] \cdot \exp[2 \cdot \hat{Y}(s_0) + \sigma^2(s_0)]}$$

(12)

where $\sigma(s_0)$ is the kriging standard deviation at point s_0, and $\hat{Y}(s_0)$ represents the concentration at point s_0 resulting from the data fusion process (Denby et al., 2008; De Smet et al., 2010).

3. Results

In the following we describe the results from applying the data fusion technique with both simulated data and measured real-world observations from a low-cost sensor network deployed in Oslo.

3.1. Data fusion with simulated observations

To asses the performance of the method we first tested the methodology using a simulated set of observations sampled from a concentration field, which was assumed to represent the true state of the atmosphere and was used as a reference to judge the outcome of the method against. Fig. 4 shows an example illustrating our approach with simulated data. The top left panel shows an assumed "true" concentration field, which is supposed to be reproduced as closely as possible using incomplete observations. Obviously the true state of the atmosphere is not known in practice, but it is useful for testing the algorithm. The truth field here represents the modeled concentration of NO_2 for a single hour in the Oslo area at 08:00 CEST on 8 January 2013. This hour was selected to represent typical pollution conditions during rush hour on a winter day.

The top center panel shows the two datasets that are available for the data fusion. This includes in the background the modeled proxy dataset (the basemap), which is in this case the 2011 annual average concentration field of NO_2, while the points represent observations of NO_2 which were simulated from the "truth" field using a perturbation with a Gaussian random error of $5 \, \mu g \, m^{-3}$. This value is lower than the error generally expected from observations obtained by a low-cost sensor network, however it helps to illustrate the performance of the mapping technique under ideal conditions. The performance of the method under real-world conditions will be demonstrated in the next section. The locations of the simulated observations are the same as used in the real-world deployment of static AQMesh sensor nodes (see Section 2.1). Note that the color scale used in the top center panel is the same for both datasets, so the simulated observations seen there indicate significantly higher concentrations than the model-based annual average proxy dataset would predict.

The top right panel of Fig. 4 shows the result of a fusion of the two datasets from the top center panel, following the methodology described in Section 2.3. It can be observed that that the spatial patterns in general are replicated quite well. Even more importantly, the overall concentration levels are generally similar to those found in the truth field. One area where the data fusion method does not perform well is
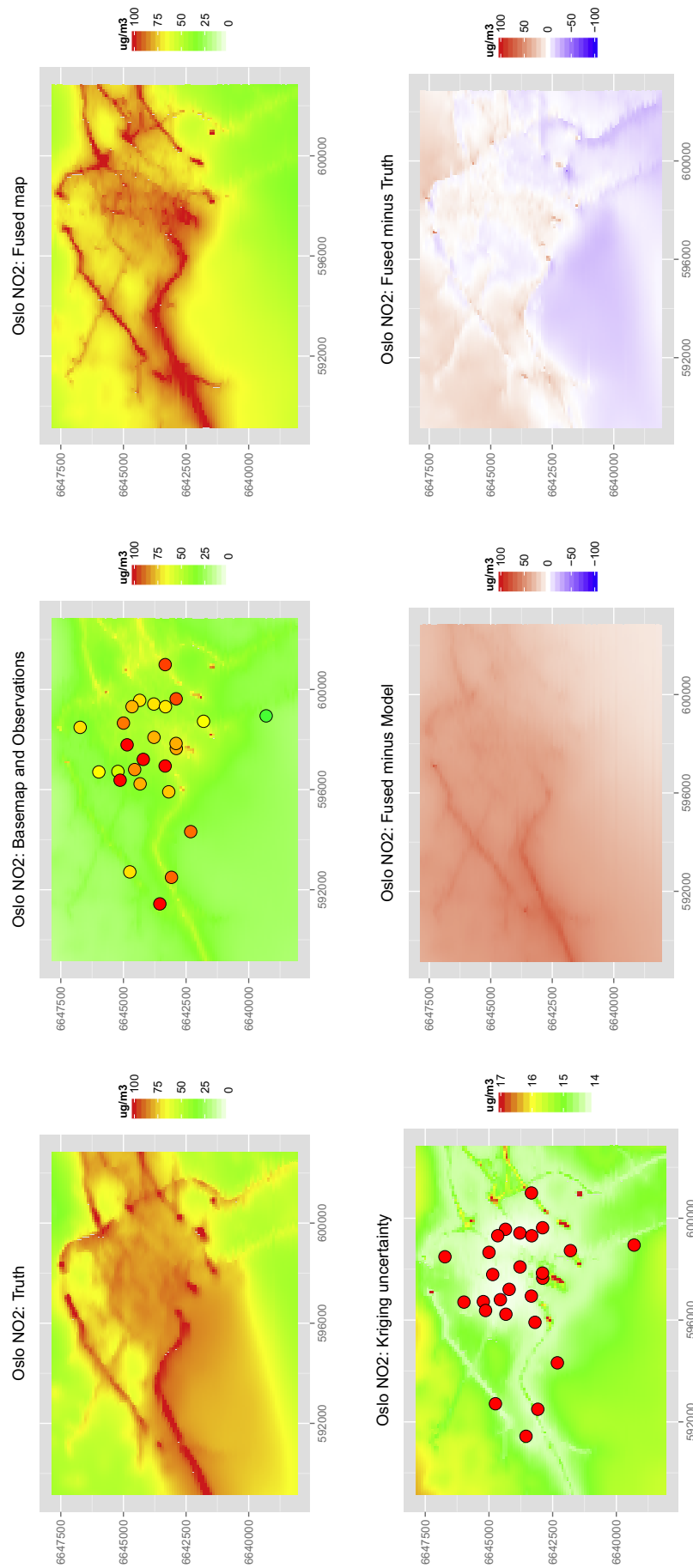
Fig. 4. Example of data fusion with simulated observations. Top left panel: "true" NO$_2$ field (in practice, unknown). Center top panel: model-derived annual average basemap of NO$_2$ and observations simulated from truth field using a random error. Top right panel: map from data fusion algorithm applied to basemap/observations. Bottom left panel: uncertainty associated with data fusion process. Bottom center/right panels: difference between fused map and model and "truth", respectively. All units are given in μg m^{-3}.

the region of relatively high concentrations in the southwestern corner of the truth field. The reason for this is that there were no observations available in this area. The bottom left panel of Fig. 4 shows the uncertainty associated with the universal kriging process and the locations of the simulated observations (as points). The uncertainty is given as the standard deviation of the prediction at each pixel. It can be observed that the interpolation error is relatively low in areas where many observations are available. The uncertainty increases towards the southwest and northwest where no observations are available. The uncertainty map shown here includes both the uncertainty related to the regression component of universal kriging as well as as the uncertainty resulting from the spatial interpolation process. However, it should be noted that this uncertainty does not give an indication of the overall true uncertainty since the measurement error in the observations is not considered. In order to see how the modeled concentration field has been modified during the data fusion process based on the observations, it is helpful to calculate a difference image between the original modeled long-term mean and the fused map (see bottom center panel of Fig. 4). In this example, the difference map indicates that the concentration field has been increased throughout most of the domain, while in the southeastern corner the concentration values have been mostly left as they were predicted by the model. This is related to the simulated observation from the truth field in this area being very close to the value indicated by the annual average concentration map (see top center panel), so no substantial correction was necessary in this region. It can also be observed that the concentrations along some of the larger roads, particularly in the western half of the mapping domain, have been increased more than the surroundings, which is related to the algorithm taking into account that some of road-side observations in this area required larger corrections than background locations.

For evaluating the theoretical performance of the data fusion algorithm it can further be helpful to compute a difference image between the truth field (in practice unknown) and the fused result. Such a map is shown in the bottom right panel of Fig. 4. It can be seen that the differences are quite close to zero in those areas where the majority of the simulated observations were located. In areas outside of the center of the mapping domain the error increases. In the southern half of the mapping domain the differences are primarily negative (indicating that the concentration field had higher values in these areas than the fused map was able to recover), whereas towards the north of the mapping domain, the differences tend to be mostly positive (indicating that the fused map overestimated the true concentration field in this regions).

To evaluate more quantitatively to what extent the data fusion algorithm has been able to combine the model information with observations to replicate the simulated truth field as closely as possible, a set of 20 locations distributed randomly throughout the mapping domain was selected. At these "validation stations" (these are not actual measurements stations but simply randomly selected points at which the true concentration value is known) the corresponding values of model basemap, the true concentration field and the fused concentration field were extracted. The result can be seen in Fig. 5, which shows an example for a winter day (8 January 2013 08:00 CET) where the concentration levels of the long-term average basemap are significantly lower than those of the simulated true concentration field. The data fusion process was able to correct the modeled concentration field using the observations such that the fused values are much closer to true simulated values. The difference between the concentrations from the fused field and the truth varies slightly among the 20 validation stations. This is due to the fact that the validation stations were selected randomly and thus some of them happen to be in areas were no observations were available. The further the validation site is away from observations sites, the higher the difference between the true concentration at the validation site and the estimated concentration will generally be. In this example, the first-guess RMSE between modeled basemap and true observations is 32.2 μg m^{-3}. The data fusion method was able to significantly reduce this RMSE to a value of only 6.2 μg
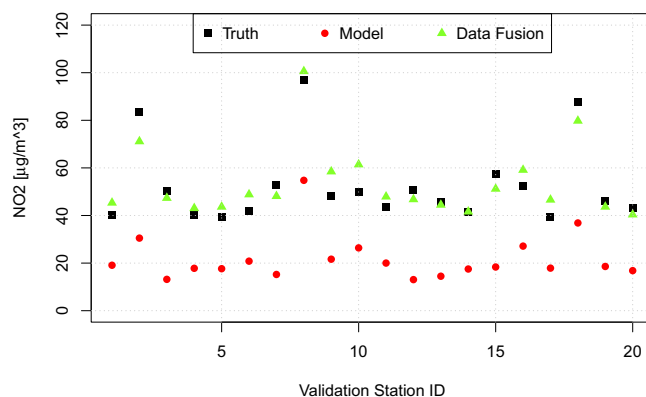

Fig. 5. Comparison of NO$_2$ concentrations for a "truth" map modeled at one specific hour (8 January 2013 08:00 CET) ("Truth"), long-term average basemap ("Model"), and fused results ("Data Fusion"), extracted at 20 validation sites, which were randomly distributed throughout the study domain. The figure shows the results for a winter day with generally quite high NO$_2$ concentrations. The long-term average concentrations given by the basemap significantly underestimate the true concentrations, while the data fusion process is able to reproduce the true values quite closely.

m^{-3}.

The achievable mapping accuracy is dependent on the number of observation sites available throughout the mapping domain. A higher number of sites is able to capture more of the spatial detail, or, when used in combination with the modeled concentration field, is able to adjust the modeled information in more regions than if only few observations are available. In order to test the impact of the number of available observations sites on the mapping accuracy, we sampled a set of 150 observations sites from the "truth"field and ran the data fusion algorithm with an increasing number of sites from this set. The mapping accuracy was subsequently determined using leave-one-out cross-validation (Cressie, 1993). Fig. 6 shows the relationship between the RMSE metric as determined from leave-one-out cross-validation and the number of simulated observations. It can be seen that the RMSE for this particular example decreases significantly from around 70 μg m^{-3} at 10 simulated observations to around 5 μg m^{-3} at 50 simulated observations. From this point on any further decreases in RMSE are not significant as the originally simulated random error is reached. The step from 10 to 20 observations already reduces the corresponding RMSE by over 300%. RMSEs of 10 μg m^{-3} and less were achieved for more than 30 observations. However, it should be noted that the relationship shown in Fig. 6 is given here only for illustration of the general concept since the relationship varies significantly with each new simulated dataset and because the systematic and random errors associated with
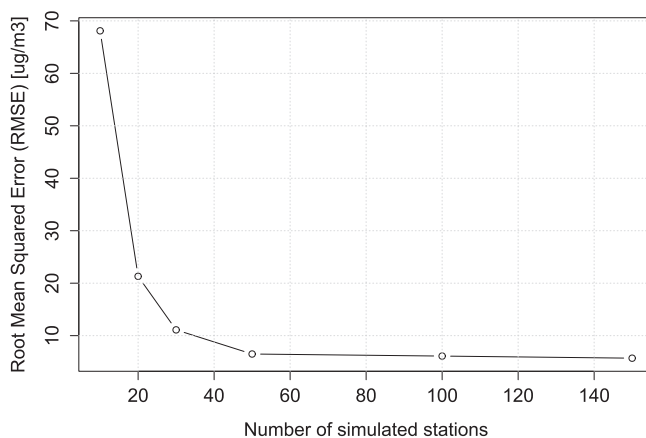

Fig. 6. Relation between the mapping accuracy as measured by the RMSE of the leave-one-out cross validation and the number of simulated stations throughout the entire mapping domain.

real observations from a low-cost sensor network are generally higher than in this simulated example. As such, the shape of the relationship shown in Fig. 6 cannot be transferred to other study sites, where different spatial patterns of air pollution and different models are likely to lead to different results. Nonetheless, geostatistical practice indicates that to calculate somewhat realistic experimental semivariogram models, a minimum number of approximately 20 stations is necessary, with a higher number strongly desirable as this will result in more stable and robust semivariograms. We have not studied the impact of the area of the study site on the mapping accuracy for a given number of sites or the required number of sites for a given area. However, we show in the next section that the method works reasonably well under the right circumstances for a sensor network of 24 sites distributed throughout a moderately sized city such as Oslo, but given this generally low number of observations the method will likely be even more useful when the number of sites increases to 50 or more.

3.2. Application to a network of low-cost sensors

The data fusion methodology described in the earlier sections was applied to real-world observations from a network of AQMesh sensor platforms deployed throughout the city of Oslo, Norway. The devices measured a variety of air pollutants. We focus here solely on NO_2 to explore its potential for mapping urban-scale air quality when combined with model information.

Fig. 7 shows an example of the data fusion methodology for NO_2, in this case for 6 January 2016 at 09:00 UTC. This date was chosen as it represents typical winter conditions with moderately high NO_2 concentrations in Oslo. A day with relatively high pollution levels was chosen as the AQMesh sensors then exhibit a high signal-to-noise ratio and generally perform best. The top left panel of the figure illustrates the two input datasets that are required by the data fusion algorithm: The background map shows the long-term average concentration of NO_2 as modeled by the EPISODE chemical dispersion model, whereas the point markers indicate both the location of the the sensor platforms as well as the magnitude of the NO_2 concentration observed by each device. It can be observed that in this instance the observations are overall significantly higher than the long-term average concentrations. When the data fusion algorithm is applied to these datasets, the resulting concentration field (top right panel) is much more consistent with the observations. The general spatial patterns provided by the modeled long-term average basemap are preserved, however the overall magnitude of the concentration field is adjusted in magnitude to match the observation both globally throughout the entire domain as well as locally in the vicinity of the observations made by the low-cost sensor platforms. As such, the data fusion method is capable of adding value to both the model information as well as the observations. In addition, each fused concentration field is associated with a map of uncertainty (bottom left panel) which illustrates qualitatively how the reliability of the mapping result varied in space and gives quantitative information about each grid cell's mapping uncertainty. Finally, the bottom right panel of Fig. 7 shows the basemap correction, i.e. the
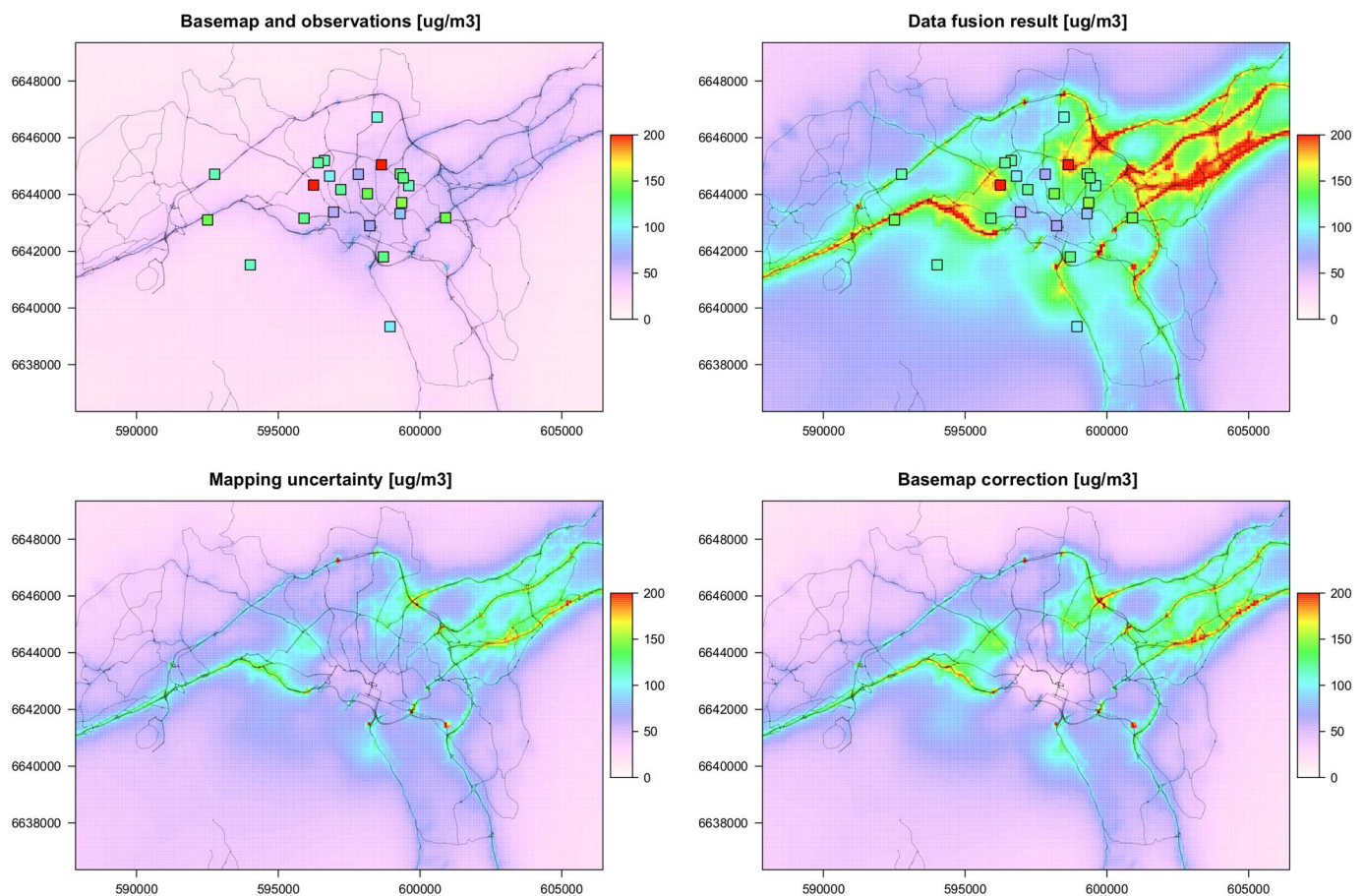

Fig. 7. Example of the data fusion process combining observations from a low-cost sesnor network with a modeled basemap, here shown for NO_2 on 6 January 2016 at 9:00 UTC. The top left panel shows both input datasets required for the methodology, namely the temporally constant model-derived basemap and the observations from the low-cost sensor network. Note that the color scale applies equally to the basemap and the observations, indicating that the observations exhibit significantly higher NO_2 concentrations than the basemap. The top right panel shows the result of the data fusion process as well as the observations from the low-cost sensor platforms for reference, indicating that the fused concentration field is much more in line with the observations. The bottom left panel gives the absolute uncertainty associated with the data fusion, while the bottom right panel shows the correction that the data fusion applied to the basemap to achieve the result. All concentrations shown in units of µg m^{-3}. The faint thin black lines indicate the major roads in Oslo for reference. Coordinates in UTM32N/WGS84.
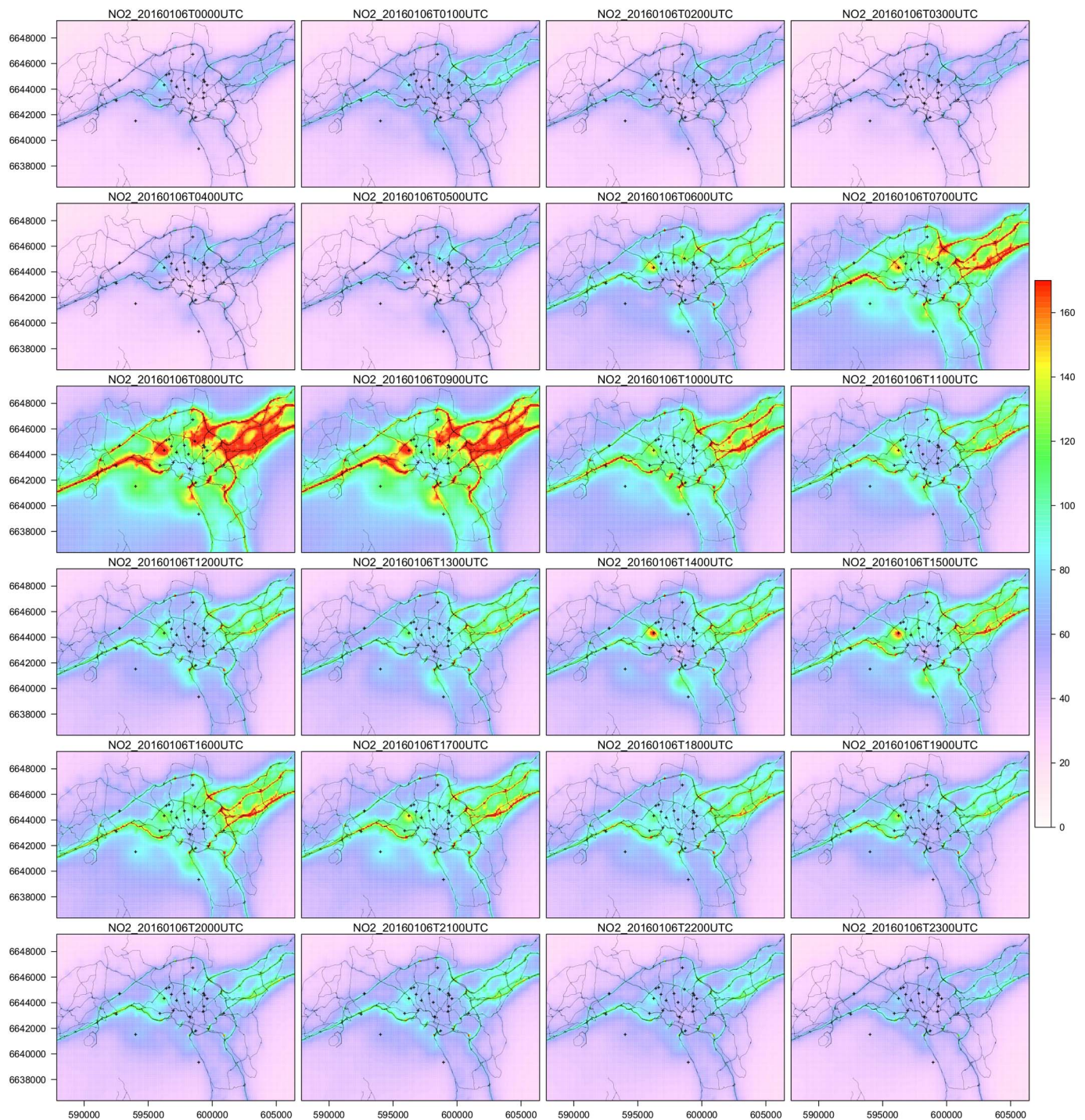
Fig. 8. Example of 24 h of data fusion results in Oslo, combining NO$_2$ measurements from the AQMesh units with a long-term average basemap derived from the EPISODE model, here shown for 6 January 2016. The concentration is shown here in units of µg m^{-3}. Note that the time is given in UTC, whereas the local time is in CET. The faint thin black lines indicate the major roads in Oslo for reference. The black point markers show the location of the deployed sensors. Coordinates in UTM32N/WGS84.

amount by which each grid cell of the basemap (top left panel) had to be adjusted to achieve the data fusion result (top right panel). In this case all correction values are positive as all of the observations were significantly higher than the concentrations given by the basemap, however the correction can also take on negative values in other cases.

The data fusion algorithm was applied for every hour during the study period of January 2016. Fig. 8 shows an example of 24 h of the resulting data fusion maps for 6 January 2016. Qualitatively it can be observed that the method provides realistic spatial patterns of NO$_2$ concentrations throughout the day, with the highest concentrations

located along major road segments with high average daily traffic such as the three ring roads, major feeder roads in and out of the city, as well the large interchanges between such roads segments. The downtown area of Oslo shows lower than expected values, possibly due to a sensor located in that area that was subject to a constant bias (Castell et al., 2017).

One of the important points that can be observed in Fig. 8 is that the typical bi-modal diurnal cycle of NO$_2$ is clearly visible. Following generally very low concentrations throughout the night, the concentrations begin to increase throughout the city at around 6:00 UTC
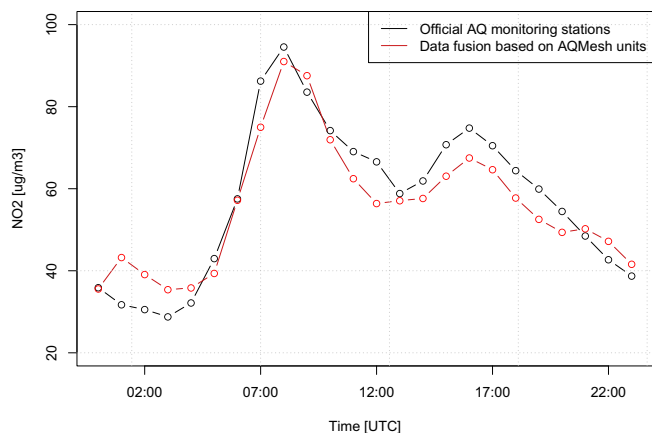
Fig. 9. Comparison of the NO$_2$ concentrations averaged over all air quality monitoring stations in Oslo against the corresponding concentrations of NO$_2$ derived at the same locations from the fused maps for 6 January 2016.

(7:00 local time) and reach a maximum at around 8:00 UTC (9:00 local time) during the peak of the morning rush hour. At this point the fused maps indicate concentrations exceeding 160 µg m^{-3} along the major ring roads and highways carrying the traffic into the city. The concentration in the remainder of the Oslo urban areas varies between approximately 80 µg m^{-3} and 120 µg m^{-3} at this point. It should be noted here that each of the maps used observations from the previous hour, which means that the map labeled with 8:00 UTC used the observations from the low-cost sensors between 7:00 UTC and 8:00 UTC.

Following the morning rush hour the concentrations then decrease slightly and stay more or less constant throughout mid-day. Towards the mid-to late-afternoon the NO$_2$ levels then increase again slightly and reach a peak at about 16:00 UTC (17:00 local time), corresponding to the relatively early evening rush hour as it is typical in Norway. Finally, the overall concentrations decrease again throughout the

evening and around 23:00 UTC reach approximately the same levels as during the previous night of around 60–80 µg m^{-3} along the major roads and between 0 µg m^{-3} and 40 µg m^{-3} in most of the other areas.

To illustrate how the methodology is capable of reproducing the typical daily cycle of NO$_2$, Fig. 9 shows the average fused concentration value extracted at all of Oslo's air quality monitoring stations for each hourly panel of Fig. 8. We can see that the NO$_2$ concentrations follow the typical diurnal cycle with a distinct, high peak during the morning rush hour, followed by lower concentration through mid-day, and a secondary, broader peak during the afternoon/evening rush hour. In addition to the diurnal cycle obtained from the fused maps, Fig. 9 also shows the corresponding average measured concentrations at the official air quality monitoring stations equipped with reference instruments. It can be seen that there is a very good correspondence between the two time series, not only in terms of the overall temporal behavior throughout the day and the time of the rush-hour related concentration peaks but also in terms of the relative magnitude of the concentration peaks. In addition, the two time series also indicate generally similar absolute concentrations ranging between approximately 30 µg m^{-3} and 100 µg m^{-3}. Given the quite large uncertainties associated with the individual sensor pods (see Table 1 as well as Castell et al., 2017 and Borrego et al., 2016), this is a very promising result, indicating that despite the systematic biases and significant random errors of the individual low-cost sensor platforms, it is possible to obtain quite accurate and realistic estimates of NO$_2$ concentrations when an entire network of such devices is used. Valuable signals are hidden within the random noise of the sensors, and while they are not always visible at the individual sensor level, such signals can be extracted from an entire network of sensors with the help of methods such as data fusion or data assimilation, as demonstrated here.

To get a better overview of the performance of the data fusion method over time, Fig. 10 shows an entire month comparing the NO$_2$ data measured by the official air quality monitoring stations with the average time series extracted from the hourly data fusion at the same
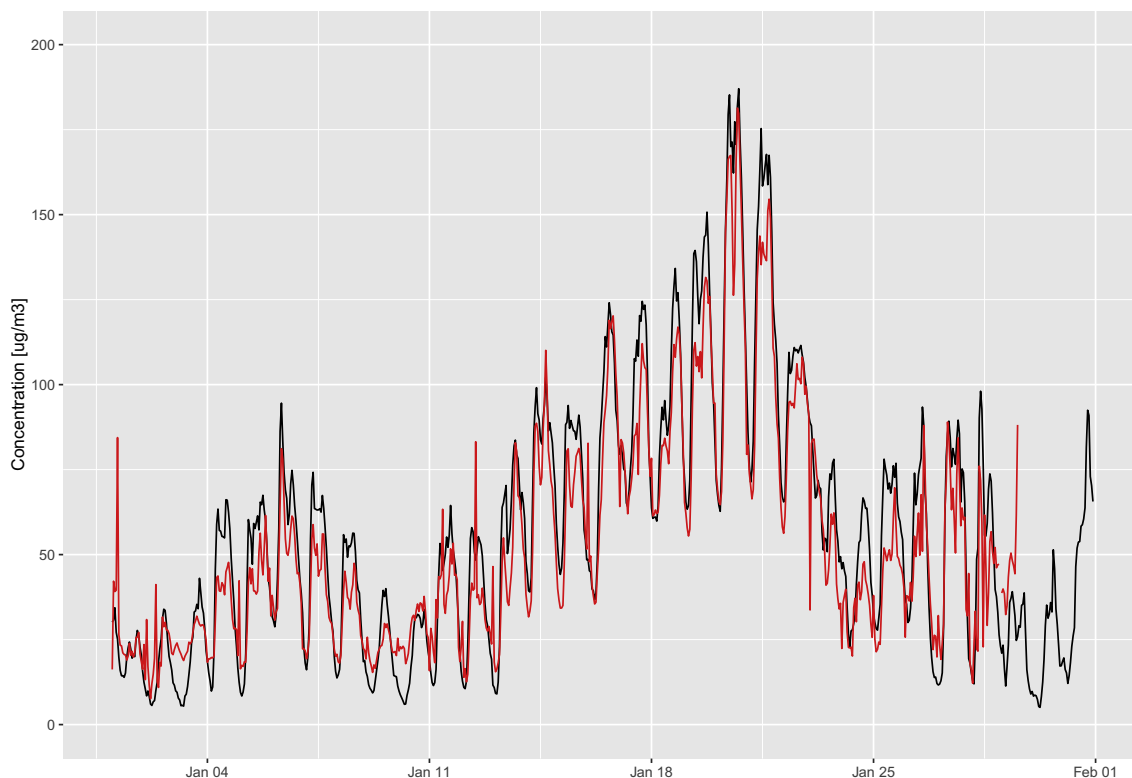

Fig. 10. Time series of observed reference concentrations of NO$_2$ for January 2016 averaged over all stations in Oslo (black line) and the corresponding average NO$_2$ concentration extracted from the data hourly fusion maps at all stations (red line). The data fusion maps at the end of the month were subject to a sensor malfunction and are therefore not shown here.
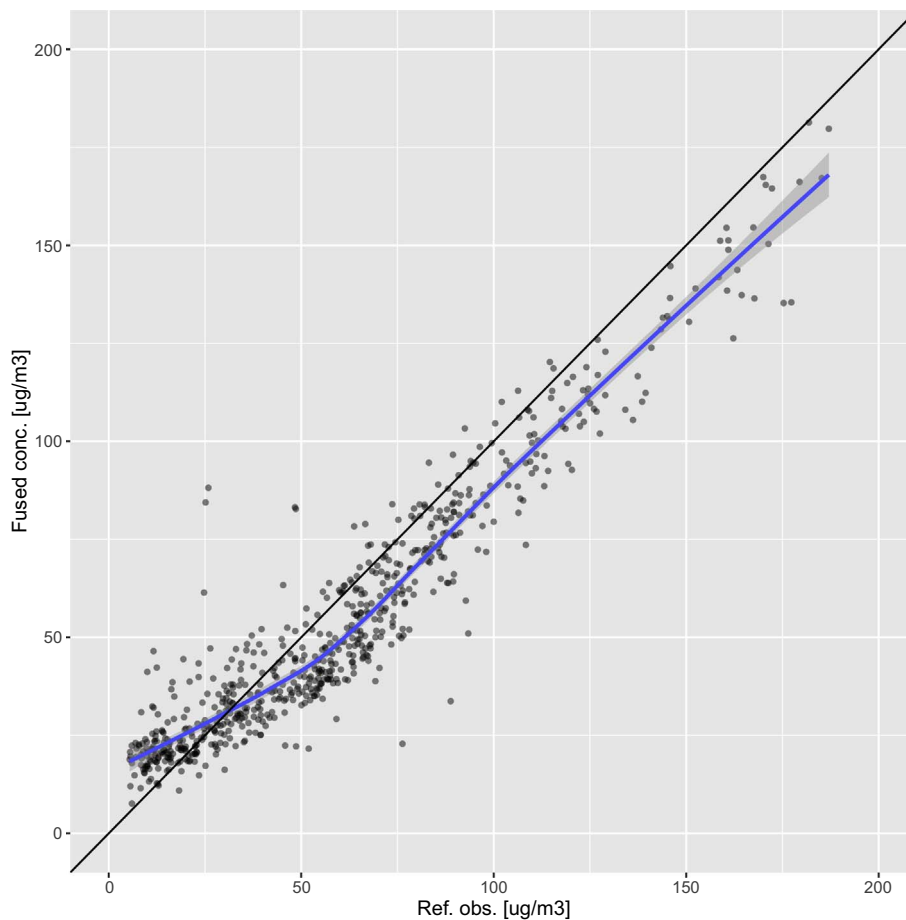
Fig. 11. Scatter plot of observed reference observations of NO_2 averaged over all stations in Oslo against the corresponding average NO_2 concentration extracted from the data hourly fusion maps at all stations. The blue line indicates a Loess fit (Cleveland, 1979) to the data with the gray areas representing the 95% confidence intervals. The black line indicates the 1:1 reference line.

locations. We can observe a very good agreement between the two time series, both in terms of long-term temporal patterns as well as the diurnal cycles throughout the month. The long-term patterns indicate relatively low concentrations of below 30 µg m^{-3} at the beginning of the month (except one outlier in the data fusion results on the first of January). The concentrations increase overall for both time series after the 4th of January and maintain levels of around 50 µg m^{-3} before dropping down again to levels around 25 µg m^{-3} on the 10th of January. Beginning on the 13th of January, the levels of NO_2 increase rapidly throughout the middle of the month until they reach their peak with daily maximum values of 160 µg m^{-3} on the 20th of January. This is followed by a rapid drop in concentrations to levels between 10 µg m^{-3} and 80 µg m^{-3}, where they remain throughout the rest of the month. The data fusion maps computed for the end of the month were affected by a sensor malfunction and are not shown here. In addition to the long-term temporal patterns, the daily extreme values are generally captured quite well by the data fusion method.

Fig. 11 shows a scatter plot of the two time series displayed in Fig. 10. It indicates that there is a very strong relationship between observations made with reference equipment at the official air quality monitoring stations and the corresponding concentrations of NO_2 as predicted at the same location by the data fusion methodology. The data points clearly follow the 1:1 line very closely, although there is a slight underestimation with respect to the reference observations for NO_2 values over 50 µg m^{-3}. For very low concentration of less than 20 µg m^{-3}, the data fusion method slightly overestimates the expected NO_2 concentrations. The relationship has a mean bias of 5 µg m^{-3} with a standard deviation of 13.1 µg m^{-3}, and a root mean square error of 14.3 µg m^{-3}.

4. Discussion

The results shown here indicate that a geostatistical data fusion method is able to combine observations from a network of low-cost sensors and time-invariant model information in such a way that the resulting concentration fields have realistic model-derived spatial patterns and at the same time are subject to observations-based absolute values. Overall, the strength of the relationship ($R^2 = 0.89$) between the data fusion results and the averaged reference data is promising given a) the very high uncertainties, biases, and calibration issues observed at the individual sensor level (Castell et al., 2017; Borrego et al., 2016), b) the fact that only a constant annual average concentration field from a different year (2011) was used as spatially exhaustive model information, and c) the fact that these spatial predictions are all made by the data fusion algorithm at locations where no AQMesh sensor devices had been installed, but that are solely derived from air quality observations made by low-cost sensors in the general vicinity and some basic and time-invariant information about the typical spatial patterns of air pollution given by a model.

Currently, individual low-cost microsensors for air quality generally tend to suffer from relatively high uncertainties, which are driven primarily by significant sensor-to-sensor variability, sensitivity of the sensor to environmental conditions and the resulting calibration drift over time, and the general deterioration of the sensors after deployment (Castell et al., 2017). As such, it is challenging to use the current generation of sensors at the individual sensor level. However, our results indicate that despite the high uncertainties at sensor-level, it is possible to extract useful and realistic information from an entire network of sensors. In addition, we show that data fusion of air quality observations from a network of low-cost sensors with model information offers a way of generating spatially detailed up-to-date maps of air quality in urban areas.

4.1. Limitations

Even though geostatistical data fusion of low-cost sensor data with model information was shown here to provide reasonable results under the right conditions, the method is nonetheless subject to several limitations:

The accuracy of the method is clearly dependent on the accuracy of the available input data. If the spatial patterns of the modeled concentration field do not represent the true average conditions or when the observations from multiple sensors are subject to significant biases, the resulting maps will also be of poor quality. It is therefore important to ensure a realistic modeled concentration field and to individually field-calibrate the sensor nodes before deployment.

The approach shown here is based on geostatistics and is thus dependent on a minimum number of observations distributed throughout the mapping domain. With only 24 data points we are operating at the very minimum of what is feasible for computing a semivariogram to carry out universal kriging. Under some circumstances (e.g. missing data from some sensor nodes, or higher than usual random error) the automated semivariogram fitting failed and no valid output map was produced. A simple analysis using simulated data indicated that mapping accuracy becomes stable with a network of around 50 observation sites. We are planning to study this more in the future with data obtained by a sensor network of 64 AQMesh units that recently was deployed in Oslo.

Finally, the method as demonstrated here only uses an annual average concentration field as spatial *a priori* information. This stems from the fact that we wanted to develop a method that can be used even at sites were no operational air pollution dispersion model is available but at which land-use regression maps can be readily developed (or already exist). However, our method can in principle be used with any *a priori* model information including shorter averaging periods, and initial tests with combining hourly modeled concentration fields with data from the sensor network are currently being carried out.

4.2. Applications and potential

Up-to-date maps of air quality at high spatial resolution such as the ones produced using the data fusion methodology presented here have a wide variety of applications. These include, for example, providing citizens with targeted information about air quality in their individual neighborhood or even street, allowing users to estimate their personal exposure along a previously taken or planned walking or cycling route throughout the city, or even providing custom routing services for computing the least polluted (rather than the shortest) route between two points in the urban environment (Mueller et al., 2016).

Only sensor nodes at static locations were used in this study and our results can only be interpreted within the framework of such measurements. Nonetheless, crowdsourcing of air quality observations often involves observations from mobile platforms (e.g. Van den Bossche et al., 2015; Castell et al., 2014; Peters et al., 2014, 2013). In doing so, for example by equipping pedestrian and cyclists with very small portable devices for measuring air quality (Castell et al., 2014), the citizen science and crowdsourcing aspects can be even further strengthened. However, these types of air quality sensors tend to often exhibit even higher uncertainty than fixed-site ones and, due to their generally very high sampling frequency on the order of seconds, they exhibit significant temporal variability. After substantial post-processing and spatio-temporal averaging, such data can also provide useful information about local air quality. It would be worthwhile to investigate the applicability of data fusion methods such as the one presented here for combining such observations with model information.

Given the rapid development in sensor technology it is reasonable to assume that within the near future, sensors for air quality will significantly improve in accuracy and stability and will even further decrease in cost. While the number of devices deployed within our study was at the very lower limit of what is feasible for geostatistical techniques, more inexpensive and stable sensors in the future will allow for significantly larger and denser sensor networks and thus will increase the value of the presented methodology for urban-scale air quality mapping. Particularly when sensors become truly ubiquitous at some point in the future, it will continue to be necessary to extract useful signals out of the large amount of highly variable observations that they provide, and our method presents one example of how it is possible to make sense of such information.

5. Conclusions

We presented a methodology for combining observations from a network of a moderate to large number of low-cost air quality monitoring devices at fixed locations with long-term average information from a high-resolution urban-scale air quality model. The result of the data fusion process, which is based on geostatistical techniques, is a new value-added map representing the best-guess concentration field at the time at which the observations were made. This concentration field inherits properties from both input datasets, i.e. it inherits the overall spatial patterns shown by the time-invariant modeled long-term average concentration field and at the same it inherits the absolute values provided by the instruments deployed within the sensor network.

The methodology was demonstrated using data collected by a network of 24 low-cost air quality platforms, which were deployed at the premises of kindergartens throughout the urban area of Oslo, Norway. While the instruments measured a wide variety of air pollutants and meteorological parameters, we focused here on NO_2 as one of the primary traffic-related air pollutants. Data fusion maps were created using NO_2 observations from the low-cost sensor network and a time-invariant modeled air quality map for each hour of January 2016. The results indicate that qualitatively the methodology is able to produce realistic high-resolution maps of urban air quality at high temporal resolution. The fused maps provide realistic daily cycles of NO_2 and it should be noted here that the temporal evolution of the maps and the derived time series are entirely driven by the sensor data, whereas the model only provides information on typical spatial patterns. Quantitatively, we found by comparison against observations from air quality monitoring stations equipped with reference instrumentation that the data fusion technique is able to predict the average NO_2 concentration at the stations with a mean bias of $5 \, \mu g \, m^{-3}$, an RMSE value of $14.3 \, \mu g \, m^{-3}$, and a relatively robust correlation with an R^2 value of 0.89. While the mapping methodology was demonstrated here solely for NO_2, it can be readily applied to measurements of PM_{10} and $PM_{2.5}$ to produce up-to-date high-resolution maps of particulate matter pollution in the urban environment. Overall, our results indicate that despite significant uncertainties at the individual sensor level, appropriate processing techniques such as the data fusion method presented here are able to exploit the "swarm knowledge" of the entire sensor network and to extract realistic signals, resulting in the possibility of high-resolution spatio-temporal mapping of urban air quality.

References

Aleixandre, M., Gerboles, M.M., 2012. Review of small commercial sensors for indicative monitoring of ambient gas review of small commercial sensors for indicative monitoring of ambient gas. Chem. Eng. Trans. 30. http://dx.doi.org/10.3303/CET1230029.

Baklanov, A., Molina, L.T., Gauss, M.M., 2016. Megacities, air quality and climate. Atmos. Environ. 126, 235–249. http://dx.doi.org/10.1016/j.atmosenv.2015.11.059.

Beelen, R., et al., 2013, jun, jun. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in europe the ESCAPE project. Atmos. Environ. 72 (2), 10–23. http://dx.doi.org/10.1016/j.atmosenv.2013.02.037.

Bertino, L., Evensen, G., Wackernagel, H.H., 2003. Sequential data assimilation techniques in oceanography. Int. Stat. Rev. 71 (2), 223–241. http://dx.doi.org/10.1111/j.1751-5823.2003.tb00194.x.

Borrego, C., et al., 2016. Assessment of air quality microsensors versus reference methods: the eunetair joint exercise. Atmos. Environ. 147 (2), 246–263. http://dx.doi.org/10.1016/j.atmosenv.2016.09.050.

Castell, N., Dauge, F.R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., Bartonova, A.A., 2017. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environ. Int. 99, 293–302. http://dx.doi.org/10.1016/j.envint.2016.12.007.

Castell, N., Kobernus, M., Liu, H.-Y., Schneider, P., Lahoz, W., Berre, A.J., Noll, J.J., 2014. Mobile technologies and services for environmental monitoring: the citi-sense-MOB approach. Urban Climate 14, 370–382. http://dx.doi.org/10.1016/j.uclim.2014.08.002.

Chilès, J.-P., Delfiner, P.P., 2012. Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sonspp. 726.

Cleveland, W., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74 (368), 829–836.

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley-Interscience, New York, pp. 900.

De Nazelle, A., Seto, E., Donaire-Gonzalez, D., Mendez, M., Matamala, J., Nieuwenhuijsen, M.J., Jerrett, M.M., 2013. Improving estimates of air pollution exposure through ubiquitous sensing technologies. Environ. Pollut. 176, 92–99. http://dx.doi.org/10.1016/j.envpol.2012.12.032.

De Smet, P., Horálek, J., Conková, M., Kurfürst, P., De Leeuw, F., Denby, B.B., 2010. European air quality maps of ozone and PM10 for 2008 and their uncertainty analysis. In: Tech. Rep. ETC/ACC Technical Paper 2010/10. European Topic Centre on Air and Climate Change, Bilthoven, Netherlands.

Denby, B., Schaap, M., Segers, A., Builtjes, P., Horálek, J.J., 2008. Comparison of two data assimilation methods for assessing PM10 exceedances on the European scale. Atmos. Environ. 42 (30), 7122–7134. http://dx.doi.org/10.1016/j.atmosenv.2008.05.058.

Denby, B., Sundvor, I., Cassiani, M., de Smet, P., de Leeuw, F., Horálek, J.J., 2010. Spatial mapping of ozone and SO2 trends in Europe. Sci. Total Environ. 408 (20), 4795–4806. http://dx.doi.org/10.1016/j.scitotenv.2010.06.021.

Denby, B.R., Sundvor, I., Schneider, P., Thanh, D.V.D.V., 2014. Air quality maps of NO2 and PM10 for the region including Stavanger sandnes randaberg and sola. In: Tech. Rep. TR 1/2014. NILU - Norwegian Institute for Air Research, Kjeller, Norway.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, pp. 483.

Guerreiro, C.B., Foltescu, V., de Leeuw, F.F., 2014. Air quality status and trends in Europe. Atmos. Environ. 98, 376–384. http://dx.doi.org/10.1016/j.atmosenv.2014.09.017.

Hand, E., 2010. Citizen Science - People power. Nature 466 (August), 685–687. http://dx.doi.org/10.1038/news.2010.106.

Hasenfratz, D., Saukh, O., Walser, C., Hueglin, C., Fierz, M., Arn, T., Beutel, J., Thiele, L.L., 2015. Deriving high-resolution urban air pollution maps using mobile sensor nodes. Pervasive Mob. Comput. 16 (PB), 268–285. http://dx.doi.org/10.1016/j.pmcj.2014.11.008.

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G.D.G., 2007. About regression-kriging: from equations to case studies. Comput. Geosci. 33 (10), 1301–1315. http://dx.doi.org/10.1016/j.cageo.2007.05.001.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D.D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 42 (33), 7561–7578. http://dx.doi.org/10.1016/j.atmosenv.2008.05.057.

Horálek, J., de Smet, P., Kurfürst, P., Leeuw, F.D., Benešová, N.N., 2014. European air quality maps of PM and ozone for 2010 and their uncertainty. In: Tech. Rep. 2014/4. European Topic Centre on Air Pollution and Climate Change Mitigation.

Horálek, J., Smet, P.D., Kurfürst, P., Leeuw, F.D., Benešová, N.N., 2013. European air quality maps of PM and ozone for 2011 and their uncertainty. In: Tech. Rep. ETC/ACM Technical Paper 2008/8. European Topic Centre on Air Quality and Climate Change Mitigation, Bilthoven, Netherlands.

Howe, J., 2006. The rise of crowdsourcing. Wired Mag. 14 (06). http://dx.doi.org/10.1086/599595.

Isaaks, E.H., Srivastava, R.M.R.M., 1989. Applied Geostatistics. Oxford University Press, New York, pp. 561.

Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, Cambridge, UK.

Kitanidis, P.K., 1997. Introduction to Geostatistics: Applications in Hydrogeology. Cambridge University Presspp. 249.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L., Britter, R.R., 2015. The rise of low-cost sensing for managing air pollution in cities. Environ. Int. 75, 199–205. http://dx.doi.org/10.1016/j.envint.2014.11.019.

Data assimilation. In: Lahoz, W.W., Khattatov, B.B., Menard, R.R. (Eds.), Springer, Berlin, Heidelberg. http://dx.doi.org/10.1007/978-3-540-74703-1.

Lahoz, W.A., Schneider, P.P., 2014. Data assimilation: making sense of earth observation. Front. Environ. Sci. 2 (16), 1–28. http://dx.doi.org/10.3389/fenvs.2014.00016.

Marécal, V., et al., 2015. A regional air quality forecasting system over Europe: the MACC-II daily ensemble production. Geosci. Model Dev. 8 (9), 2777–2813. http://dx.doi.org/10.5194/gmd-8-2777-2015.

Mead, M., et al., 2013. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmos. Environ. 70, 186–203. http://dx.doi.org/10.1016/j.atmosenv.2012.11.060.

Mueller, M.D., Hasenfratz, D., Saukh, O., Fierz, M., Hueglin, C.C., 2016. Statistical modelling of particle number concentration in Zurich at high spatio-temporal resolution utilizing data from a mobile sensor network. Atmos. Environ. 126, 171–181. http://dx.doi.org/10.1016/j.atmosenv.2015.11.033.

Mueller, M.D., Wagner, M., Barmpadimos, I., Hueglin, C.C., 2015. Two-week NO2 maps for the city of Zurich, Switzerland, derived by statistical modelling utilizing data from a routine passive diffusion sampler network. Atmos. Environ. 106 (2), 1–10. http://dx.doi.org/10.1016/j.atmosenv.2015.01.049.

Nieuwenhuijsen, M.J., Donaire-Gonzalez, D., Rivas, I., de Castro, M., Cirach, M., Hoek, G., Seto, E., Jerrett, M., Sunyer, J.J., 2015. Variability in and agreement between modelled and personal continuously measured black carbon levels using novel smartphone and sensor technologies. Environ. Sci. Technol. 49, 2977–2982. http://dx.doi.org/10.1021/es505362x.

Oftedal, B., Walker, S.E., Gram, F., McInnes, H., Nafstad, P.P., 2009. Modelling long-term averages of local ambient air pollution in Oslo, Norway: evaluation of nitrogen dioxide, PM10 and PM2.5. Int. J. Environ. Pollut. 36 (2), 110–126. http://dx.doi.org/10.1504/IJEP.2009.021820.

Peters, J., Theunis, J., van Poppel, M., Berghmans, P.P., 2013. Monitoring PM10 and ultrafine particles in urban environments using mobile measurements. Aerosol Air Qual. Res. 13, 509–522. http://dx.doi.org/10.4209/aaqr.2012.06.0152.

Peters, J., Van den Bossche, J., Reggente, M., Van Poppel, M., De Baets, B., Theunis, J.J., 2014. Cyclist exposure to UFP and BC on urban routes in Antwerp, Belgium. Atmos. Environ. 92, 31–43. http://dx.doi.org/10.1016/j.atmosenv.2014.03.039.

Petersen, W.B., 1980. User's guide for HIWAY-2 - a highway air pollution model. In: Tech. Rep. EPA-600/8-80-018. United States Environmental Protection Agency.

Piedrahita, R., et al., 2014. The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. Atmos. Meas. Tech. 3 (2), 3325–3336. http://dx.doi.org/10.5194/amt-7-3325-2014.

Sarma, D.D., 2009. Geostatistics with Applications in Earth Sciences. Springer Science & Business Media, Dordrecht, The Netherlands, pp. 224.

Schneider, P., Lahoz, W.A., van der A, R.R., 2015. Recent satellite-based trends of tropospheric nitrogen dioxide over large urban agglomerations worldwide. Atmos. Chem. Phys. 15, 1205–1220. http://dx.doi.org/10.5194/acp-15-1205-2015.

Serrano Sanz, F., Holocher-Ertl, T., Kieslinger, B., Sanz Garcia, F., Silva, C.G.C.G., 2014. Socientize - White Paper on Citizen Science for Europe.

Slørdal, L.H., Walker, S.-E., Solberg, S.S., 2003. The urban air dispersion model EPISODE applied in AirQUIS 2003 - technical description. In: Tech. rep. NILU - Norwegian Institute for Air Research, Kjeller, Norway.

Snyder, E.G., et al., 2013. The changing paradigm of air pollution monitoring. Environ. Sci. Technol. 47 (20), 369–377. http://dx.doi.org/10.1021/es4022602. 11.

Spinelle, L., Gerboles, M., Villani, M.G., Aleixandre, M., Bonavitacola, F.F., 2015. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: ozone and nitrogen dioxide. Sens. Actuators B 215, 249–257. http://dx.doi.org/10.1016/j.snb.2015.03.031.

Steinle, S., Reis, S., Eric, C.C., 2013. Science of the total environment quantifying human exposure to air pollution - moving from static monitoring to spatio-temporally resolved personal exposure assessment. Sci. Total Environ. 443, 184–193. http://dx.doi.org/10.1016/j.scitotenv.2012.10.098.

Tilloy, A., Mallet, V., Poulet, D., Pesin, C., Brocheton, F.F., 2013. BLUE-based NO2 data assimilation at urban scale. J. Geophys. Res. Atmos. 118 (4), 2031–2040. http://dx.doi.org/10.1002/jgrd.50233.

Van den Bossche, J., Peters, J., Verwaeren, J., Botteldooren, D., Theunis, J., De Baets, B.B., 2015. Mobile monitoring for mapping spatial variation in urban air quality: development and validation of a methodology based on an extensive dataset. Atmos. Environ. 105, 148–161. http://dx.doi.org/10.1016/j.atmosenv.2015.01.017.

Wackernagel, H., 2003. Multivariate Geostatistics. Springer Berlin Heidelbergpp. 387.

Webster, R., Oliver, M.A.M.A., 2007. Geostatistics for Environmental Scientists. John Wiley & Sonspp. 330.

World Health Organization, 2016. Ambient air pollution: a global assessment of exposure and burden of disease. In: Tech. rep. World Health Organization.