Project title:

Development of sensor-based Citizens' Observatory Community for improving quality of life in cities

Acronym: **CITI-SENSE** Grant Agreement No: **308524**

EU FP7- ENV-2012

Collaborative project

Deliverable D 6.3

Data fusion of crowdsourced air quality observations and dispersion model data for urban-scale air quality mapping

Work Package 6

Date: 30. 09. 2015

Version: 1.0

Leading Beneficiary:	Norwegian Institute for Air Research (NILU)
Author(s):	P. Schneider (NILU), N. Castell (NILU), W. Lahoz (NILU), I. Vallejo (NILU)
Dissemination level:	RE (Restricted)

Versioning and contribution history

Version	Date issued	Description	Contributors
0.1	15. 09. 2015	Initial draft	P. Schneider (NILU), N. Castell (NILU), W. Lahoz (NILU) , I. Vallejo (NILU)
0.2	21. 09. 2015	Feedback from William Lahoz included	W. Lahoz (NILU), P. Schneider (NILU)
0.3	23. 09. 2015	Reviewer comments by Milos Davidovic addressed	M. Davidovic (VINCA), P. Schneider (NILU)
0.4	29. 09. 2015	Reviewer comments by David Broday addressed	D. Broday (Technion), P. Schneider (NILU)
0.5	30. 09. 2015	Comments by Sonja Grossberndt addressed	S. Grossberndt (NILU), P. Schneider (NILU)
1.0	30. 09. 2015	Final version	P. Schneider (NILU)

Peer review summary

Internal review 1			
Reviewer	David Broday (Technion)		
Received for review	15. 09. 2015	Date of review	29. 09. 2015

Internal review 2			
Reviewer	Milos Davidovic (Vinca)		
Received for review	16. 09. 2015	Date of review	23. 09. 2015

Executive Summary

The low-cost microsensors deployed within the CITI-SENSE project have significant potential for improving high-resolution mapping of air quality in the urban environment. However, one major shortcoming of the data obtained by such sensors is that there are significant data gaps in both space and time. To overcome this issue, we present a data fusion method based on geostatistics that allows for merging crowdsourced observations of air quality with the information from an urban-scale air quality model. The performance of the methodology is evaluated using a simulated dataset of NO_2 over the city of Oslo, Norway. First results indicate that the method is capable of producing a concentration field that replicates the spatial patterns of a simulated true concentration field. Cross-validation of the fused concentration field carried out against a second set of simulated observations shows that the technique provides a robust way of bias-correcting the model information and that the prediction errors are on the order of the same magnitude as the perturbation applied to the simulated values. The achievable accuracy of the data fusion method is highly dependent on the number of observations and their spatial distribution.

This page intentionally left blank.

Contents

This page intentionally left blank.

1 Introduction

With an ever-increasing amount of environmental observations available through crowdsourcing, 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 and in particular in many large urban agglomerations (Schneider et al., 2015), with significant impacts on societal health and economy (World Health Organization, 2013; Guerreiro et al., 2014). 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 deployed reference stations is generally not able to adequately capture the spatial variability in air pollutants.

Recent technological advances related to sensor technology have resulted in comparatively low-cost and small devices for measuring air quality. Using 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 the crowdsourced data with spatially continuous data from a model.

We present a data fusion technique for combining near real-time crowdsourced observations of urban air quality with output from an urban-scale air pollution dispersion model that allows for providing highly detailed, up-to-date maps of urban air quality. Data fusion is conceptually similar to data assimilation (Kalnay, 2003; Lahoz and Schneider, 2014). Data fusion 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.

The manuscript is organized as follows: Section 2 gives an overview about previous work on crowdsourcing for air quality as well as data assimilation and data fusion studies related to air quality mapping. Section 3 describes the datasets and the model used and provides details on the data fusion methodology used.

Section 4 provides the results and discusses their implications. Section 5 summarizes the study, provides conclusions and presents an outlook for future work.

2 Background

In the following section we provide a short overview of existing crowdsourcing and citizen science efforts and describe previous work on data assimilation as well as data fusion techniques and report on some previous research on using these methods for applications in urban-scale air quality mapping.

2.1 Crowdsourcing for urban air quality

The traditional way of making observations of air quality in the urban environment involves reference air quality monitoring stations placed in strategic location throughout a city. While providing highly accurate measurements of a large variety of air pollutants, such stations tend to be highly complex, require significant maintenance, and are thus very expensive to set up and operate. As a result, the network of air quality monitoring stations has always been extremely sparse even in highly developed countries, and is close to non-existent in developing countries, which often tend to have severe issues with air pollution. The low spatial density of air quality monitoring stations limits their usefulness for continuously mapping urban air quality as small-scale spatial patterns cannot be properly resolved.

In recent years, technological progress has resulted in a new class of low-cost sensor for measuring various air pollutants such as NO_2, NO, CO, O_3, black carbon, and particulate matter (PM_{10}, $PM_{2.5}$) (Spinelle et al., 2015; Kumar et al., 2015). Integrated within a platform for communication through the mobile network or directly linking to a user-carried smartphone, such devices are generally inexpensive, small, and lightweight. While their accuracy is at this point not sufficient to replacement observations from calibrated reference air quality instrumentation, their other characteristics mean that they can be deployed in large numbers throughout a region, for example within a citizen science or crowdsourcing framework. Such a high-density network of low-cost devices has the potential to provide spatial and temporal information about urban air quality at spatial scales and resolutions that were not possible in the past with conventional monitoring systems.

Several studies have investigated the potential use of this new generation of air quality monitoring devices, particularly for mobile measurements using portable sensors. Hasenfratz et al. (2012a) presented a system for using small and portable devices for monitoring ozone concentration using smartphones and show how to perform an on-the-fly calibration of low-cost gas sensors based on co-located

measurements (Hasenfratz et al., 2012b). The potential of crowdsourced observations of air quality to be used in conjunction with statistical land-use regression models was also demonstrated for mapping urban-scale air quality (Hasenfratz et al., 2014). Mead et al. (2013) deployed a network of low-cost, electrochemical gas sensors in Cambridge, United Kingdom, and found that even though these devices were designed for sensing at parts-per-million mixing ratios, they can actually be used at the parts-per-billion level given suitable configuration and operation. They conclude that sparse, static, traditional monitoring stations are not capable of fully characterizing the urban environment and that low-cost microsensors are able to provide a much more complete assessment of the small-scale spatial and temporal variability of urban air quality.

Peters (2013) describes the use of bicycle-based mobile measurements for measuring PM_{10} and ultrafine particles along a street network. Similarly, Peters et al. (2014) demonstrate the bicycle-based measurement of black carbon and ultrafine particles for studying the micro-variability of street-level air pollution and the exposure of bicyclists. Castell et al. (2014) show how mobile technologies and low-cost sensors can be applied for environmental monitoring and in particular for measuring air quality at the street level. Other studies use novel technologies such as smartphones and personal air pollution sensors to examine the diurnal variability in personal air pollution levels and the relationship between modeled and measured exposure to air pollutants in different microenvironments (Nieuwenhuijsen et al., 2015). Van den Bossche et al. (2015) used bicycle-based mobile monitoring to map the black carbon in an urban environment and found that it can provide insight into the spatial variability within the urban environment, given a sufficient number of passes along each road segment in order to reduce short-term temporal variability. Moltchanov et al. (2015) deployed a network of metal-oxide wireless sensors and found that in general the individual network nodes exhibited high inter-node consistency and sensitivity to their respective local microenvironments. However, sensor-specific temporal variation of the calibration parameters was observed, which was corrected using observations from a nearby air quality monitoring station.

2.2 Data assimilation and data fusion for urban scale mapping of air quality

Data assimilation in general is a way of combining model information with observations in a mathematically objective way (by providing the best linear unbiased estimate). As such it provides the possibility to create self-consistent and realistic representations of the Earth system (Kalnay, 2003; Lahoz and Schneider, 2014). In this process, value is added to both the observations and the model: The spatial and temporal gaps in the observations are interpolated in a meaningful way, while the model is constrained by the observations. Data assimilation is carried out using many different techniques. Some data assimilation methods can actively interact with the model. These include variational

methods such as 2D-VAR, 3D-VAR, and 4D-VAR (Lahoz and Schneider, 2014), and sequential methods such as the Kalman Filter (KF) (Kalman, 1960), the Extended Kalman Filter (Bouttier and Courtier, 1999) and the Ensemble Kalman Filter (EnKF) (Evensen, 2003). Lahoz and Schneider (2014) provide a comprehensive overview of these types of data assimilation methods. Other data assimilation methods, which we refer to here as *data fusion* techniques, are based on a wide variety different statistical techniques. The goal of all these methods is to combine the various data sources and to provide an optimal estimate of the spatial distribution of the parameter in question (Denby et al., 2008). While spatial interpolation techniques have been used in the past for data fusion purposes in various fields, not many studies exist that apply such methods for urban-scale air quality mapping.

Van De Kassteele et al. (2006) performed statistical mapping of annual surface PM_{10} concentrations over Western Europe by fusing observations from air quality monitoring stations with both dispersion model output of the LOng Term Ozone Simulation-EURopean Operational Smog model (LOTOS-EUROS) (Schaap et al., 2008) and satellite data of aerosol optical thickness acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS). Similarly, van de Kassteele et al. (2009) used kriging with external drift to combine NO_x (NO_x = NO + NO_2) observations with the output from a dispersion model. They found that a reduction in the density of the reference monitoring network for air quality can be compensated by geostatistical methods. Tilloy et al. (2013) used a Kalman filter approach to assimilate air quality observations from 9 fixed stations into an urban air quality model. Applying a model covariance function that heavily relies on the shape of the street network, they find that the root mean squared error of the assimilated fields is reduced by 30-50 % over the model fields when station density is high. Silibello et al. (2014) compared two data assimilation approaches, namely the successive corrections methods and optimal interpolation, to combine air quality observations with the output of a regional air quality model. Johansson et al. (2015) describe an infrastructure for performing data fusion of meteorological and air quality data for local and regional domains and find that hourly local NO_2 concentrations can be estimated more accurately with data fusion techniques than with conventional extrapolation methods.

Denby et al. (2008) assessed the performance of two data assimilation methods used together with the LOTOS-EUROS model (Schaap et al., 2008) for assessing PM_{10} exceedances on the European scale. The compared methods included the EnKF (Evensen, 2003) and spatial interpolation based on a combination of residual kriging and linear regression. They found that assimilating the observations based on the spatial interpolation technique provided significantly superior results over those obtained from EnKF for their specific case study, with the root mean squared error (RMSE) of the daily mean concentrations of PM_{10} at pre-selected validation stations being 9.2 $\mu g\ m^{-3}$ for the spatial interpolation technique and 13.5 $\mu g\ m^{-3}$ for EnKF. The poor performance of EnKF in this is case was primarily due to model bias which the used bias correction scheme was not able to properly account for. The residual kriging technique was found

to be much less sensitive to model bias. Based on these results in this paper we also use a combination of residual kriging and linear regression as our data fusion method.

To our knowledge, no previous studies have applied data assimilation and data fusion techniques for combining data from an urban-scale dispersion model with crowdsourced observations of air quality.

3 Methodology

3.1 The EPISODE dispersion model

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. Furthermore, EPISODE provides schemes for advection, turbulence, deposition, and chemistry. 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.

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 emission are known most reliably. Figure 1 shows an example of EPISODE output for NO_2 over the city of Oslo, Norway, at the original spatial resolution of 1 km. However, as 1000 m horizontal spatial resolution is not sufficient for use in detailed urban mapping we apply a downscaling procedure using a dense network of receptor points and EPISODE's capability to calculate the concentration at a point based on the raster-based areal emission and vector-based line source emissions.

3.2 Downscaling EPISODE data

The 1000 m × 1000 m horizontal spatial resolution at which EPISODE is generally run is not sufficient for providing the detailed spatial patterns that are required when using the data for fusion with crowdsourced observations. For this reason we use a downscaling procedure to obtain high-resolution concen-

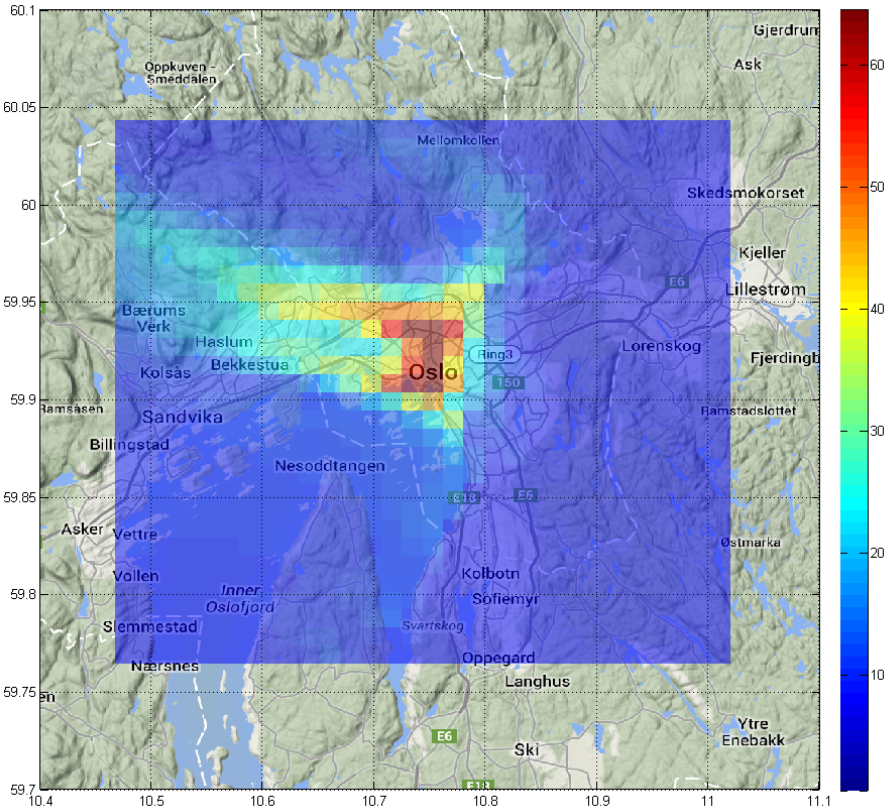

Figure 1 – Example of hourly output of the surface layer from the EPISODE dispersion model at original spatial resolution (1000 m × 1000 m with a 20 m layer height) for concentrations of NO_2, given in units of μg m^{-3}. Shown is the area of greater Oslo, Norway on a typical moderately polluted winter day.

trations 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 modelling 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 vector-based emissions from line sources. The resulting high-density set of concentrations is then interpolated to the desired output resolution using geostatistical techniques (ordinary kriging).

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 lies in the vertical representativity of the model output. 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 the two datasets can therefore not be directly compared with each other, the lower vertical representativity of the downscaled map has the advantage of being a more realistic estimate of what people are exposed to at the street level.

Figure 2 shows an example result of the downscaling methodology for the 2011 annual average concentrations of NO_2 in the area of greater Oslo, Norway. The top left panel shows the original gridded output of EPISODE. The downtown area of Oslo slightly north of the 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. This method of downscaling does not use any other information besides what is available in the original 1 km × 1 km resolution concentration field generated by the EPISODE model. It is therefore not able to provide any additional information about fine-scale spatial patterns. 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 Figure 2 shows the downscaled maps after spatial interpolation of the receptor point values using ordinary kriging.

Using EPISODE and the described downscaling methodology, high-resolution annual average concentration fields were derived , which were used as a basemap or *climatology* (long-term mean) for the data fusion process. Figure 3 shows examples of basemaps for NO_2, PM_{10}, and $PM_{2.5}$ derived from the model and the corresponding experimental and fitted semivariograms. 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). While the maps for NO_2 and PM_{10} show similar spatial patterns with quite rapid gradients, which are primarily linked to line-source emissions from road transport, the map for $PM_{2.5}$ shows overall smoother patterns indicating a stronger link to areal emissions such as domestic heating and woodburning.

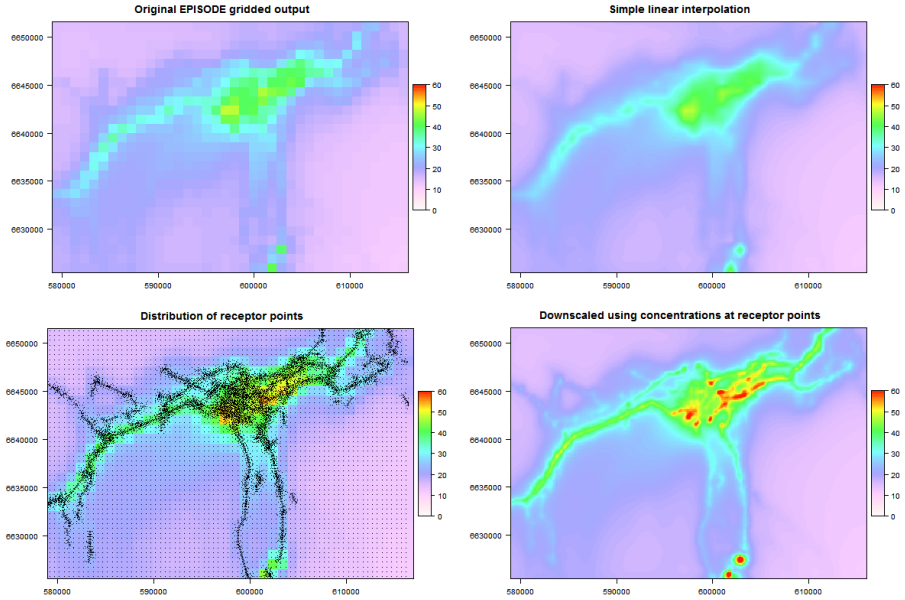

Figure 2 – Downscaling methodology for EPISODE. The panels show the 2011 annual mean surface NO_2 concentration in the area of Oslo, Norway, in units of $\mu g\ m^{-3}$. Top left panel: Original gridded output from the EPISODE model at 1000 m horizontal spatial resolution. Top right panel: The original 1000 m concentration grid regridded to 100 m horizontal spatial resolution using simple linear interpolation. 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 the receptor points. Note that the receptor points are located at 2 m above the surface, whereas the original gridded concentration field represents the vertical average over the lowermost 20 m of the atmosphere. This causes the downscaled concentrations to be higher than the original gridded concentration field. The coordinates are given in units of meters in the UTM32N projection using the WGS84 datum.

This is also clear in the respective semivariograms for the three maps (Figure 3). 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 the sill value) of 9493 m with a nugget effect (a non-zero semivariance at the origin typically representing varaibility at distance smaller than the sampling distance) of 0.86. The semivariogram for PM_{10} was fitted with an exponential model with 12.7 sill and a range of 7553 m with a nugget effect of 0.42. Finally, the semivariogram for $PM_{2.5}$ was fitted with a spherical model of 2.24 sill and a range of 11868 m. It is clear from both the maps in Figure 3 and the empirical as well as fitted semivariograms that both NO_2 and PM_{10} exhibit a similar behavior in terms of spatial gradients, whereas the model output for $PM_{2.5}$ indicates less strong spatial gradients.

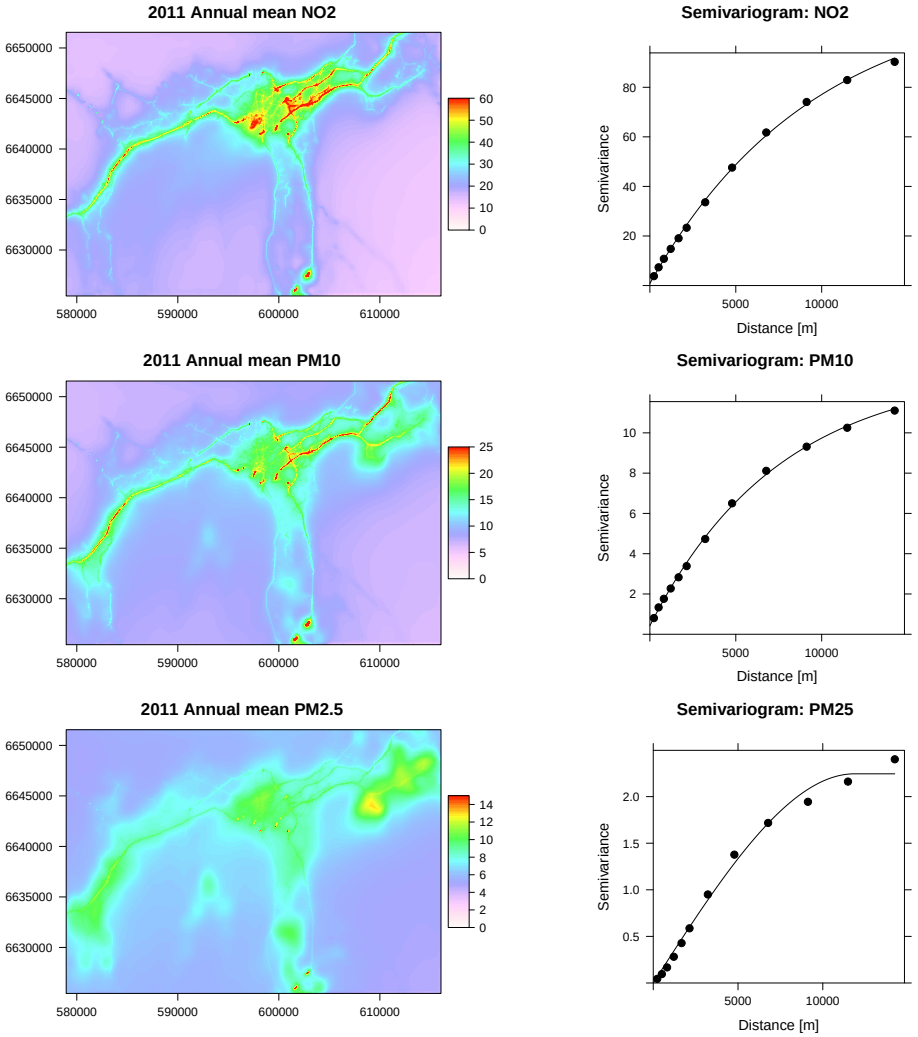

Figure 3 – Downscaled model-derived basemaps (left column) for NO_2, PM_{10}, and $PM_{2.5}$ as produced by the EPISODE model and the corresponding semivariograms (right column) for Oslo, Norway. All concentrations given in μg m^{-3}. The coordinates are given in units of meters in the UTM32N projection using the WGS84 datum.

3.3 Simulated observations

In order to test the system against a known reference, observations were simulated from a modeled concentration field, which was considered as the true state of the atmospheric composition for this evaluation. The hourly concentration field for 8 January 2013 at 08:00 CEST was used here for this purpose,

representing typical winter conditions in Oslo. Observations were simulated first retrieving the modeled concentration at a given set of locations. These were then separately perturbed with 1) a simulated bias (Gaussian with zero mean and a standard deviation of 5 μg m^{-3}) and 2) a simulated random error (Gaussian with zero mean and a standard deviation of 10 μg m^{-3}). These standard deviations have been chosen to roughly approximate the typical error characteristics of real low-cost microsensors resulting from comparison exercises against reference air quality monitoring stations.

3.4 Data fusion methodology

The data fusion methodology applied here is based on geostatistical principles (Isaaks and Srivastava, 1989; Cressie, 1993; Goovaerts, 1997; Kitanidis, 1997; Wackernagel, 2003; Webster and Oliver, 2007; Sarma, 2009; Chilès and Delfiner, 2012). It uses universal kriging to combine observations with model data by predicting the concentrations at unknown location by simultaneously interpolating the observations and using the model data to provide information about the spatial patterns.

In contrast to ordinary kriging, 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 where no observations are available, and a kriging component that provides the small-scale random variation.

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

$$\hat{Y}(s_0) = c + a_1 \cdot x_1(s_0) + a_2 \cdot x_2(s_0) + \ldots + a_p \cdot x_p(s_0) + \varepsilon(s_0) \qquad (1)$$

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

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1(s_0) & \cdots & x_p(s_0) \\ 1 & \vdots & & \vdots \\ 1 & x_1(s_n) & \cdots & x_p(s_n) \end{bmatrix} \begin{bmatrix} c \\ a_1 \\ \vdots \\ a_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \boldsymbol{Xa} + \boldsymbol{\varepsilon} \qquad (2)$$

where $\boldsymbol{Y}$ indicates the estimated values at all prediction locations, $\boldsymbol{X}$ represents the values of the predictor variables at all locations, $\boldsymbol{a}$ is the vector of regression coefficients, $\boldsymbol{\varepsilon}$ indicates the vector of residuals error that is estimated using kriging with the know semivariogram model, n is the number of prediction locations and p is the number of predictor variables.

In practice, the spatial trend or drift ε of the mean is estimated here using a single predictor variable, which is the annual average concentration map provided by the EPISODE chemical dispersion model (see Figure 3). 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.

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). The variogram model types were kept the same as those derived for the model-derived basemaps (Figure 3), while the respective range of the models was allowed to vary by up to 30% around the values derived for the basemaps. The nugget and sill parameters were allowed to vary freely.

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. Denby et al. (2008) showed that the theoretical back-transformed expectation value of a concentration C is given as

$$E[C] = \exp\left(\mu + \frac{\sigma^2}{2}\right) \tag{3}$$

where μ and σ represent the mean and standard deviation of the log-normal-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

$$\hat{Z}(s_0) = \exp\left[\hat{Y}(s_0) + \frac{\sigma^2(s_0)}{2}\right] \tag{4}$$

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

$$\mathrm{var}[C] = \left[\exp(\sigma^2) - 1\right] \cdot \exp\left[2\mu + \sigma^2\right] \tag{5}$$

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

$$\delta(s_0) = \sqrt{\exp\left[\sigma^2(s_0) - 1\right] \cdot \exp\left[2 \cdot \hat{Y}(s_0) + \sigma^2(s_0)\right]} \tag{6}$$

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).

3.5 Quantification of mapping accuracy

Three methods for estimating the uncertainty and quantifying the mapping accuracy have been used in this study.

Kriging uncertainty All kriging variants including the one used here provide a spatial estimate of kriging uncertainty. This essentially represents the uncertainty associated with the spatial interpolation process itself and should not be interpreted as an overall uncertainty of the resulting prediction map. Figure 5 (bottom left panel) shows an example of such a map.

Leave-one-out cross-validation One of the most common ways of assessing the mapping accuracy is to compare the the estimated concentrations against measured values. Ideally, if an independent set of accurate observations exists, it can provide the basis for a validation. However, in many cases no such dataset exists. In that case the original set of observations can be split into a dataset used for prediction and a dataset used for validation. This is called cross-validation (Cressie, 1993; Goovaerts, 1997) and is for example implemented in K-fold cross validation, where the observation dataset is randomly split into K equally size parts (commonly $K = 5$ or $K = 10$) and one of those parts is used solely for validation. This process is then repeated K times such that each part is used for validation. If the total number of observations is quite low, as it is the case in this study, using such relatively large fractions of the data solely for validation would lower the number of observations used for prediction by an unacceptable magnitude (and thus significantly decrease the prediction accuracy). In this

case, a technique called leave-one-out cross-validation, which is a special case of K-fold cross validation where $K = n$ with n being the number of observations (Hengl, 2009), can be used. This technique excludes only one sample at a time from the prediction process in order to be used solely for validation purposes, and thus preserves the vast majority of observations for the actual mapping. Here we therefore use leave-one-out cross-validation in order to estimate the mapping accuracy against the original observations.

The following statistical metrics have been used to evaluate the results from the cross validation: The *Mean Prediction Error* (MPE) which is the same as the average bias, is calculated as

$$\mathrm{MPE} = \frac{1}{N}\sum_{i=1}^{N}\left[Y(s_i) - \hat{Y}(s_i)\right] \tag{7}$$

where N is the number of observations, $Y(s_i)$ is the observed concentration at point s_i, $\hat{Y}(s_i)$ is the estimated concentration at point s_i. An MPE closer to zero indicates a better estimation. In addition, the *standard deviation of the differences* s was calculated as

$$s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(Y(s_i) - \hat{Y}(s_i)\right) - \left(\overline{Y(s_i) - \hat{Y}(s_i)}\right)\right]^2} \tag{8}$$

where $\overline{Y(s_i) - \hat{Y}(s_i)}$ indicates the mean of the differences between observations and estimates. The *Root Mean Squared Error* (RMSE) is calculated as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[Y(s_i) - \hat{Y}(s_i)\right]^2} \tag{9}$$

A smaller RMSE value indicates a better estimation. Finally, the *Mean Absolute Error* was calculated as

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|Y(s_i) - \hat{Y}(s_i)\right| \tag{10}$$

A smaller MAE indicates a better estimation. In addition to these statistics, it can also be helpful to fit a linear regression model to the pairs of predictions and observations. The resulting regression statistics such as intercept, slope, and the coefficient of determination (R^2) can provide additional information on the accuracy of the fused map with respect to the original observations.

Figure 4 shows a typical scatterplot visualizing sample results of a leave-one-out cross-validation exercise for data fusion of simulated NO_2 values. In general the data points appear to follow the 1:1 reference line reasonably well. However, it can be seen that relatively low concentrations between 40 μg m^{-3} and 55 μg

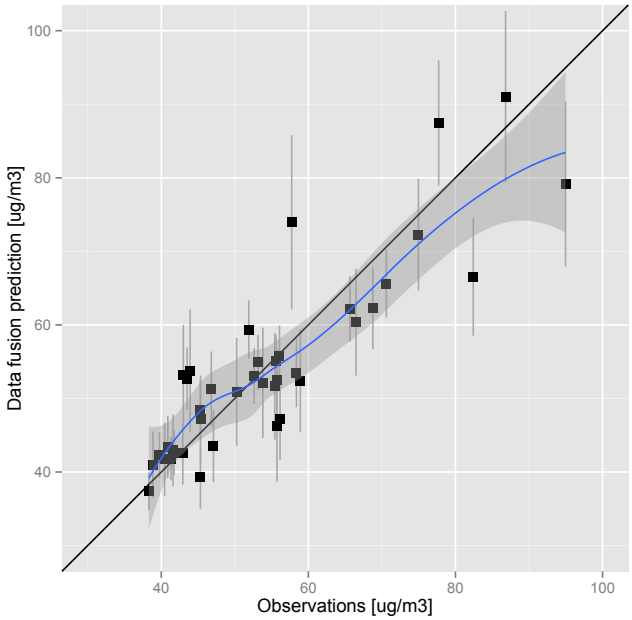

Figure 4 – Typical scatterplot of observations against concentrations estimated within the data fusion process, given in μg m^{-3}. In this case the observations were simulated from a "true" concentration field by adding random noise to the true concentrations. The gray error bars indicate the prediction uncertainty related to the spatial interpolation process. The black line shows the 1:1 reference line. The blue line indicates a robust locally weighted regression fit (LOESS) (Cleveland, 1979; Cleveland and Grosse, 1991) to the data and the surrounding dark gray area shows the corresponding 95% confidence interval of the Loess fit.

m^{-3} are estimated much more accurately than concentrations of greater than 60 μg m^{-3}. The spatial predictions of higher concentrations in this case tend to slightly underestimate the true values in general, as can be seen by the Loess fit (Cleveland, 1979; Cleveland and Grosse, 1991) being slightly below the 1:1 line for observations exceeding 60 μg m^{-3}. In addition there are two positive outliers, whose confidence interval do not intersect with the confidence interval of the Loess fit. Such overestimates likely are due to the fitted semivariogram (and particularly its range) not being able to replicate very sharp local gradients near the street network.

Ensemble data fusion The previous two methods of quantifying the mapping accuracy do not take into account the uncertainty in the observations. This can be accomplished by perturbing the observations with error characteristics that have been measured in laboratory and field validation experiments. This creates and ensemble of observations. Subsequently, the data fusion algorithm can be

applied to the entire ensemble. The ensemble mean and its standard deviation then can be seen as the most likely concentration field and its corresponding uncertainty, given the uncertainty in the observations.

4 Results and Discussion

In the following we describe the results from applying the data fusion technique with urban-scale model data and crowdsourced observations. We show some examples and discuss the implications.

4.1 Data fusion with simulated observations

Figure 5 shows an example of applying the data fusion methodology to a simulated example for the city of Oslo, Norway. The top left panel shows a "true" concentration field which is supposed to be recreated using incomplete data. The truth field here represents the modeled concentration of NO_2 in the Oslo area at 08:00 CEST on 8 January 2013. The top center panel shows the two datasets that are available for the data fusion. This includes in the background the model proxy dataset, in this case the 2014 annual average concentration field of NO_2 while the points represent observations of NO_2 which were simulated from the "truth" field using a random perturbation of 10 $\mu g\ m^{-3}$. The locations of the simulated observations are the same as used in a real-world deployment of static sensor nodes in the CITI-SENSE project (`www.citi-sense.eu`). Note that the color scale used is the same for both datasets, so the observations indicate significantly higher concentrations than the model-based proxy dataset would predict.

The top right panel of Figure 5 shows the result of a fusion of the two datasets shown in the top center panel, following the methodology described in Section 3.4. It can be observed that that the spatial patterns in general are quite well replicated. Even more importantly, the overall levels appear quite similar to those in the truth field. One area that is not picked up well by the data fusion process is the relatively high concentrations in the southwestern corner of the truth field. This is because there were no observations available in this area.

The bottom left panel of Figure 5 shows the uncertainty associated with the universal kriging process and the locations of the simulated observations (as points). It can be seen that the interpolation error is quite low in areas where many observations are available. However 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

Figure 5 – 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.

give an indication of the overall true uncertainty since the measurement error in the observations is not considered at this point.

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 Figure 5). 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. 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.

For evaluating the success of the data fusion algorithm it can further be helpful to compute a difference image between the truth field and the fused result. Such a map is shown in the bottom right panel of Figure 5. 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).

In order to evaluate 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 selected randomly throughout the mapping domain was selected. At these "validation stations" the corresponding values of model basemap, true concentration field and fused concentration field were extracted.

Figure 6 shows an example where the NO_2 concentrations of the basemap were significantly higher than those of the simulated true concentration field. The truth dataset used here represents the modeled concentration field on 1 July 2013 at 08:00 CEST. The figure shows that the data fusion process is able to correct the bias resulting from the basemap and to provide concentration values at the 20 locations that are much closer to the true concentration fields in most cases. It should be noted that in a small number of cases, for example for stations 1 and 15, the fused values are higher than the model-estimated values. This can occur when the nearest observation site had a quite high measurement error and thus negatively affected the concentration field in its surroundings. Overall, the statistics show that the data fusion was able to improve the initial estimate by the model quite significantly. While the RMSE between model and true values was 18.4 $\mu g\ m^{-3}$, the RMSE between fused estimate and true values was reduced to 9.2 $\mu g\ m^{-3}$.

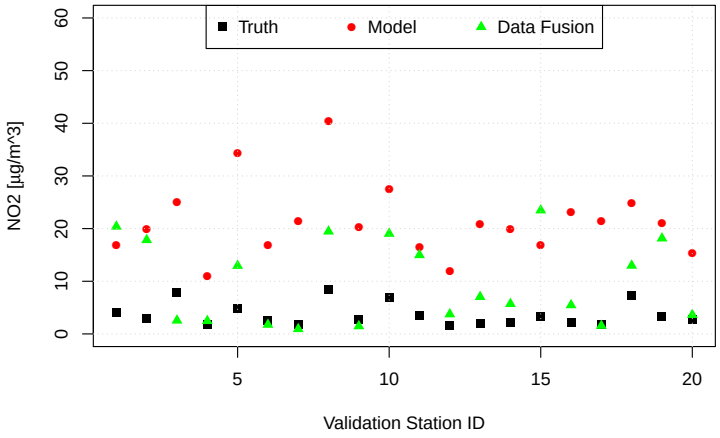

Figure 6 – Comparison of NO_2 concentration for truth map, basemap, and fused results, extracted at 20 validation stations which were randomly placed throughout the study area. The figure shows the results for a summer day (1 July 2013 08:00 CEST) with quite low overall NO_2 concentrations, for which the annual average basemap significantly overestimates the true concentrations. The data fusion algorithm corrects the basemap such that the fused results are able to quite closely replicate the true state of the atmosphere.

Similarly, Figure 7 shows an example where the overall levels of the basemap are significantly lower than those of the simulated true concentration field. Once again 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 concentration 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 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 sam-

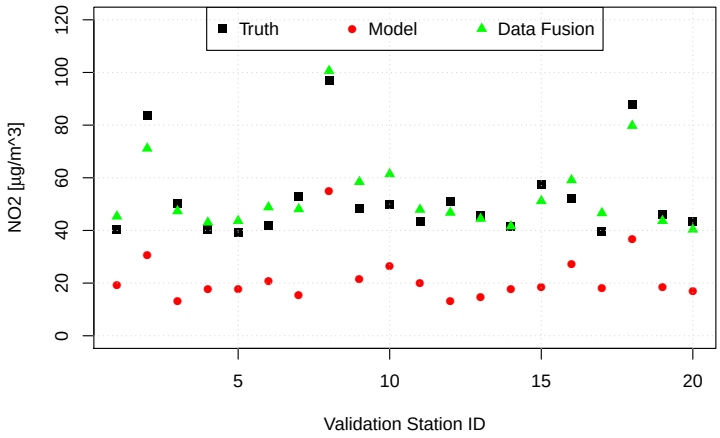

Figure 7 – Same as Figure 6 but for a winter day (8 January 2013 08:00 CET) with generally quite high NO_2 concentrations. Here the long-term average concentrations given by the basemap significantly underestimate the true concentrations, while the data fusion process is able to replicate the true values quite closely.

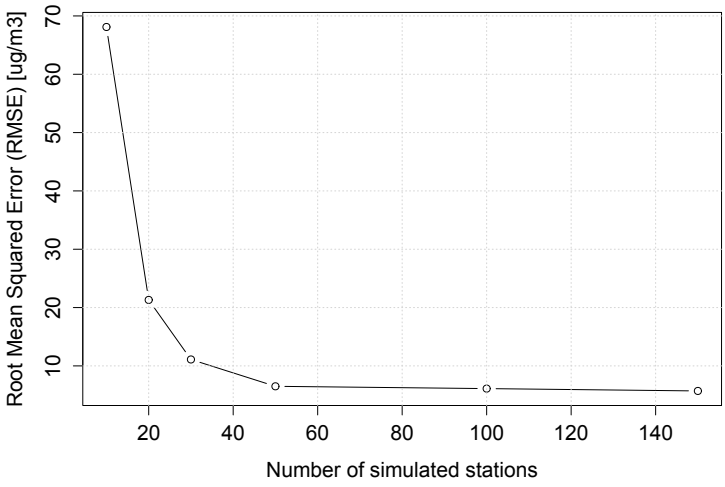

Figure 8 – Relation between the mapping accuracy as measured by the RMSE of the leave-one-out cross validation and the number of simulated stations throughout the entire mapping domain.

pled 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). Figure 8 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 $\mu g\ m^{-3}$ at 10 simulated observations to around 5 $\mu g\ m^{-3}$ at 50 simulated observations. From this point on any further decreases in RMSE are not significant. The step from 10 to 20 observations already reduces the corresponding RMSE by over 300 %. RMSEs of 10 $\mu g\ m^{-3}$ and less were achieved for more than 30 observations. However, it should be noted that the relationship shown in Figure 8 varies with each new simulated dataset and should thus be seen only as indicative.

4.2 Data fusion with real observations

At the time of the writing of this deliverable, the planned observation network of Geotech sensors in the various CITI-SENSE locations was not active yet, so no real-world data was available for testing the mapping approach. However, as soon as the main deployment phase of CITI-SENSE will start, the data fusion methodology will be applied to the near-real time observations made by the Geotech sensors. A system has been developed to fully automatize the process. First, the last hour's set of city-wide Geotech observations is retrieved for all CITI-SENSE locations from the Snowflake server. Subsequently, for each hour, each CITI-SENSE location, and each species (NO_2, PM_{10}, $PM_{2.5}$) the data fusion code is run using the corresponding Geotech observations and a basemap depicting annual average concentrations. These basemaps were producted for Oslo using the EPISODE dispersion model and for the other CITI-SENSE locations using a land-use regression model (Wang et al., 2014).

Finally, resulting concentration maps are converted to maps depicting the Common Air Quality Index (CAQI) (Van Den Elshout et al., 2014). An example is shown in Figure 9, where the bottom left panel indicates the CAQI grid values (ranging from 0 to 100) calculated for a fused map of NO_2 on 8 January 2013 in Oslo. It can be seen that the general spatial patterns of the concentration field as well as the spatial gradients remain visible even after the oversion to the index. The bottom right panel of Figure 9 shows the CAQI classes representing air quality (ranging from *very low* to *very high*). Due to the coarse nature of the classes the spatial gradients have now disappeared, however the general spatial patterns of the concentration field can still be observed.

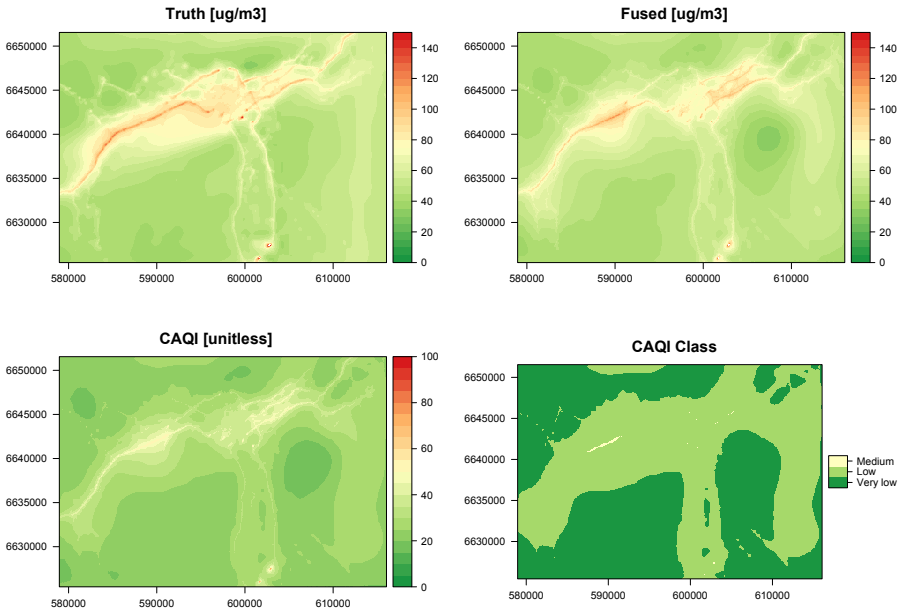

Figure 9 – "Truth" map for a winter day in Oslo (top left panel), the result of applying the data fusion algorithm to the modeled basemap and simulated observation to replicate the true state of the atmosphere (top right panel), the grid values of the Common Air Quality Index (CAQI) (Van Den Elshout et al., 2014) derived from the fused concentration map (bottom left panel), and the CAQI classes derived from the CAQI grid values (bottom right panel).

5 Conclusions

A methodology is presented to combine observations from a large number of crowdsourced air quality monitoring devices at static locations with 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.

For evaluation purposes, the methodology has been tested using simulated datasets for which a "true" concentration field was known. Point-based observations were then sampled from this "truth" field and a random error component was added. The outcome of the data fusion process has then been evaluated against the original "true" concentration field both for the entire mapping domain and at simulated validation stations. The results indicate that the concentration field provided by the data fusion technique is able to quite well replicate the original

concentration field in terms of both spatial patterns and absolute values. While there is a strong dependence of the achievable mapping accuracy on the total number of available stations providing air quality observations, the mapping accuracy for, e.g. NO_2, was found to reach RMSE values of less than 5 $\mu g\ m^{-3}$ when a total number of 50 or more simulated stations were used throughout the mapping domain.

We show that data assimilation and data fusion of crowdsourced air quality observations with model information offers a novel way of generating spatially detailed maps of air quality in the urban environment. Testing of the methodology has so far been limited to simulated observations, however the data fusion technique will be applied to real-world data collected by the Geotech sensors as soon as they are deployed throughout the various location as part of the full deployment phase of CITI-SENSE. Additional future work will focus on improving the characterization of the spatial representativity and the uncertainty of the crowdsourced observations.

Acknowledgments

Funding for this work has been provided by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 308524.

This page intentionally left blank.

References

Bouttier, F. and Courtier, P. (1999). Data assimilation concepts and methods March 1999. Technical report, European Centre for Medium-Range Weather Forecasts.

Castell, N., Kobernus, M., Liu, H.-Y., Schneider, P., Lahoz, W., Berre, A. J., and Noll, J. (2014). Mobile technologies and services for environmental monitoring: The Citi-Sense-MOB approach. *Urban Climate*, (DOI:10.1016/j.uclim.2014.08.002).

Chilès, J.-P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons.

Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.

Cleveland, W. S. and Grosse, E. (1991). Computational methods for local regression. *Statistics and Computing*, 1(1):47–62.

Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley-Interscience, New York.

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

Denby, B., Schaap, M., Segers, A., Builtjes, P., and Horálek, J. (2008). Comparison of two data assimilation methods for assessing PM10 exceedances on the European scale. *Atmospheric Environment*, 42(30):7122–7134.

Denby, B., Sundvor, I., Cassiani, M., de Smet, P., de Leeuw, F., and Horálek, J. (2010). Spatial mapping of ozone and SO2 trends in Europe. *The Science of the total environment*, 408(20):4795–806.

Denby, B. R., Sundvor, I., Schneider, P., and Thanh, D. V. (2014). Air quality maps of NO2 and PM10 for the region including Stavanger Sandnes Randaberg and Sola. Technical Report TR 1/2014, NILU - Norwegian Institute for Air Research, Kjeller, Norway.

Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press, New York.

Guerreiro, C. B., Foltescu, V., and de Leeuw, F. (2014). Air quality status and trends in Europe. *Atmospheric Environment*, 98:376–384.

Hand, E. (2010). Citizen Science - People power. *Nature*, 466(August):685–687.

Hasenfratz, D., Saukh, O., Sturzenegger, S., and Thiele, L. (2012a). Participatory Air Pollution Monitoring Using Smartphones. In *Proceedings of the 2nd International Workshop on Mobile Sensing, April 16-20, 2012, Beijing, China*, pages 1–5.

Hasenfratz, D., Saukh, O., and Thiele, L. (2012b). On-the-fly calibration of low-cost gas sensors. In Picco, G. P. and Heinzelman, W., editors, *Wireless Sensor Networks - Proceedings of the 9th European Conference on Wireless Sensor Networks, EWSN 2012, Trento, Italy, February 15-17, 2012.*, Lecture Notes in Computer Science, pages 228–244. Springer Berlin Heidelberg.

Hasenfratz, D., Saukh, O., Walser, C., Hueglin, C., Fierz, M., Arn, T., Beutel, J., and Thiele, L. (2014). Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive and Mobile Computing*, 16(Part B):268–285.

Hengl, T. (2009). *A practical guide to geostatistical mapping*. Number EUR 22904 EN. Office for Official Publications of the European Communities, Amsterdam.

Hengl, T., Heuvelink, G. B. M., and Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computers and Geosciences*, 33(10):1301–1315.

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

Horálek, J., Smet, P. D., Kurfürst, P., Leeuw, F. D., and Benešová, N. (2013). European air quality maps of PM and ozone for 2011 and their uncertainty. Technical Report 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 Magazine*, 14(06).

Isaaks, E. H. and Srivastava, R. M. (1989). *Applied geostatistics*. Oxford University Press, New York.

Johansson, L., Epitropou, V., Karatzas, K., Karppinen, A., Wanner, L., Vrochidis, S., Bassoukos, A., Kukkonen, J., and Kompatsiaris, I. (2015). Environmental Modelling & Software Fusion of meteorological and air quality data extracted from the web for personalized environmental information services. *Environmental Modelling and Software*, 64:143–155.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(Series D):35–45.

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 Press.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L., and Britter, R. (2015). The rise of low-cost sensing for managing air pollution in cities. *Environment International*, 75:199–205.

Lahoz, W. A. and Schneider, P. (2014). Data assimilation: making sense of Earth Observation. *Frontiers in Environmental Science*, 2(16):1–28.

Mead, M., Popoola, O., Stewart, G., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J., McLeod, M., Hodgson, T., Dicks, J., Lewis, a., Cohen, J., Baron, R., Saffell, J., and Jones, R. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, 70:186–203.

Moltchanov, S., Levy, I., Etzion, Y., Lerner, U., Broday, D. M., and Fishbain, B. (2015). On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Science of The Total Environment*, 502:537–547.

Nieuwenhuijsen, M. J., Donaire-Gonzalez, D., Rivas, I., de Castro, M., Cirach, M., Hoek, G., Seto, E., Jerrett, M., and Sunyer, J. (2015). Variability in and agreement between modelled and personal continuously measured black carbon levels using novel smartphone and sensor technologies. *Environmental Science & Technology*, 49:2977–2982.

Peters, J. (2013). Monitoring PM10 and Ultrafine Particles in Urban Environments Using Mobile Measurements. *Aerosol and Air Quality Research*, pages 509–522.

Peters, J., Van den Bossche, J., Reggente, M., Van Poppel, M., De Baets, B., and Theunis, J. (2014). Cyclist exposure to UFP and BC on urban routes in Antwerp, Belgium. *Atmospheric Environment*, 92:31–43.

Petersen, W. B. (1980). User's guide for HIWAY-2 - A highway air pollution model. Technical Report EPA-600/8-80-018, United States Environmental Protection Agency.

Sarma, D. D. (2009). *Geostatistics with Applications in Earth Sciences*. Springer Science & Business Media, Dordrecht, The Netherlands.

Schaap, M., Timmermans, R. M., Roemer, M., Boersen, G., Builtjes, P. J., Sauter, F. J., Velders, G. J., and Beck, J. P. (2008). The LOTOS EUROS model: description, validation and latest developments. *International Journal of Environment and Pollution*, 32(2):270.

Schneider, P., Lahoz, W. A., and van der A, R. (2015). Recent satellite-based trends of tropospheric nitrogen dioxide over large urban agglomerations worldwide. *Atmospheric Chemistry and Physics*, 15:1205–1220.

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

Silibello, C., Bolignano, A., Sozzi, R., and Gariazzo, C. (2014). Application of a chemical transport model and optimized data assimilation methods to improve air quality assessment. *Air Quality, Atmosphere & Health*, (doi:10.1007/s11869-014-0235-1).

Slørdal, L. H., Walker, S.-E., and Solberg, S. (2003). The Urban Air Dispersion Model EPISODE applied in AirQUIS 2003 - Technical Description. Technical report, NILU - Norwegian Institute for Air Research, Kjeller, Norway.

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., and Bonavitacola, F. (2015). Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical*, 215:249–257.

Tilloy, A., Mallet, V., Poulet, D., Pesin, C., and Brocheton, F. (2013). BLUE-based NO2 data assimilation at urban scale. *Journal of Geophysical Research: Atmospheres*, 118(4):2031–2040.

Van De Kassteele, J., Koelemeijer, R. B. a., Dekkers, a. L. M., Schaap, M., Homan, C. D., and Stein, A. (2006). Statistical mapping of PM10 concentrations over Western Europe using secondary information from dispersion modeling and MODIS satellite observations. *Stochastic Environmental Research and Risk Assessment*, 21(2):183–194.

van de Kassteele, J., Stein, A., Dekkers, A. L. M., and Velders, G. J. M. (2009). External drift kriging of NOx concentrations with dispersion model output in a reduced air quality monitoring network. *Environmental and Ecological Statistics*, 16(3):321–339.

Van den Bossche, J., Peters, J., Verwaeren, J., Botteldooren, D., Theunis, J., and De Baets, B. (2015). Mapping spatial variation in urban air quality using mobile measurements: development and validation of a methodology based on an extensive dataset. *Atmospheric Environment*, 105:148–161.

Van Den Elshout, S., Léger, K., and Heich, H. (2014). CAQI common air quality index - update with PM2.5 and sensitivity analysis. *Science of the Total Environment*, 488-489:461–468.

Wackernagel, H. (2003). *Multivariate Geostatistics*. Springer Science & Business Media.

Wang, M., Beelen, R., Bellander, T., Birk, M., Cesaroni, G., Cirach, M., Cyrys, J., de Hoogh, K., Declercq, C., Dimakopoulou, K., Eeftens, M., Eriksen, K. T., Forastiere, F., Galassi, C., Grivas, G., Heinrich, J., Hoffmann, B., Ineichen, A., Korek, M., Lanki, T., Lindley, S., Modig, L., Mölter, A., Nafstad, P., Nieuwenhuijsen, M. J., Nystad, W., Olsson, D., Raaschou-Nielsen, O., Ragettli, M., Ranzi,

A., Stempfelet, M., Sugiri, D., Tsai, M. Y., Udvardy, O., Varró, M. J., Vienneau, D., Weinmayr, G., Wolf, K., Yli-Tuomi, T., Hoek, G., and Brunekreef, B. (2014). Performance of multi-city land use regression models for nitrogen dioxide and fine particles. *Environmental Health Perspectives*, 122(8):843–849.

Webster, R. and Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. John Wiley & Sons.

World Health Organization (2013). Review of evidence on health aspects of air pollution – REVIHAAP Project. Technical report, World Health Organization, Copenhagen, Denmark.

