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

ARTICLE INFO

Keywords:
Air quality
Urban air quality
Crowdsourcing
Mapping
Low-cost microsensors
Nitrogen dioxide

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 $\mu g m^{-3}$. 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., 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 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 sensor 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; Steinele 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 (NO2), ozone (O3) 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 O3-filtered NO2 sensor from Alphasense, which is designed to reject O3 and hence eliminate cross-sensitivity issues. CO, NO and O3 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, O3 and nitrogen oxides (NOx). CO is measured using non-dispersive infrared spectroscopy (EN14626), NOx is measured using chemiluminescence (EN14211) and O3 is measured using UV photometry (EN14625). While we focus here on NOx as an example for mapping applications, the performance of the sensors for related gases such as NO and O3 was also evaluated.

The following metrics were used for comparing sensor platform data at time t (Mt) with observations from the reference instrument (Ot), where n represents the total number of observations, and σM and σO represent the respective standard deviations for sensors observations and reference:

- **Mean Bias (MB)**
  \[ \text{MB} = \frac{1}{n} \sum_{t=1}^{n} (M_t - O_t) \]  
  \[ (1) \]

- **Mean Gross Error (MGE)**
  \[ \text{MGE} = \frac{1}{n} \sum_{t=1}^{n} |M_t - O_t| \]  
  \[ (2) \]

- **Normalized Mean Bias (NMB)**
  \[ \text{NMB} = \frac{\sum_{t=1}^{n} (M_t - O_t)}{\sum_{t=1}^{n} O_t} \]  
  \[ (3) \]

- **Normalized Mean Gross Error (NMGE)**
  \[ \text{NMGE} = \frac{\sum_{t=1}^{n} |M_t - O_t|}{\sum_{t=1}^{n} O_t} \]  
  \[ (4) \]

- **Root Mean Squared Error (RMSE)**
  \[ \text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (M_t - O_t)^2}{n}} \]  
  \[ (5) \]

- **Correlation Coefficient (r)**
  \[ r = \frac{\sum_{t=1}^{n} \frac{(M_t - \bar{M})(O_t - \bar{O})}{\sigma_M \sigma_O}}{n-1} \]  
  \[ (6) \]

Table 1 shows the results of the co-location of the 24 AQMesh nodes, for CO, NO2, and O3. The results indicate that the mean bias can be significant for some of the pollutants. For example, for NO2 the bias can reach 75 parts per billion (ppb). The bias varies from sensor to sensor. For example, for O3 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 NO2, 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 sensor.
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات