Data fusion for air quality mapping using sensor data: feasibility and added-value through an application in Nantes

Alicia Gressent[1], Laure Malherbe[2*] and Augustin Colette[1]

[1] *alicia.gressent@ineris.fr, augustin.colette@ineris.fr*
[2] *laure.malherbe@ineris.fr*
*Corresponding author

Abstract. *The recent technological developments and the increased interest for public information lead to a fast-growing use of microsensors for air quality monitoring. Measurement campaigns are conducted to assess the potential of these low-cost instruments by deploying fixed sensors (e.g. on top of buildings, street lights or reference stations) and/or mobile sensors (e.g. on top of cars, bikes, or carried by citizens). These experiments allow to measure pollutant concentrations at high resolution in space and time. The large amount of collected information offers new opportunities of developments in air quality modelling and mapping. This work aims to take the best of these sensors despite the related measurement uncertainty to produce urban air pollution maps at fine spatial and temporal resolution. A geostatistical methodology (data fusion) is presented, which uses sensor observations as well as dispersion model outputs. It is applied to PM_{10} data in the French city of Nantes. It involves new challenges such as the consideration of the quick change of the sensor location if it is mobile, the temporal variability of the measurements, the analysis of numerous and heterogeneous data, the spatial representativeness of the measurements and the measurement uncertainties. Also, efforts still need to be done on the sampling design to ensure appropriate spatial coverage of the considered domain and get more accurate estimates.*

Keywords. *Air Quality Mapping; Microsensor; Data Fusion; PM_{10}*

1 Introduction

Air quality monitoring is conventionally based on a network of stations which allows a continuous report of pollutant concentrations. The related measurement uncertainty is constrained by the European existing legislation [1, 2] ensuring observation accuracy. Nevertheless, the installation and maintenance of such a network are expensive and so the number of stations in each region is limited. The use of numerical modelling on various scales (regional, urban, local) has thus increased during the last 15 years to supplement station observations and support air quality assessment.
In parallel, the technological progress allowed the development of miniaturized and low-cost instruments to measure pollutant concentrations [3]. Many projects of crowdsourcing and citizen science are emerging. In addition, field measurement campaigns are conducted to assess the potential of these low-cost devices by deploying fixed sensors (on top of buildings, street lights, reference stations) and/or mobile sensors (on top of cars, bikes, or carried by citizens) offering higher spatial coverage than

reference stations. Because microsensors suffer from metrological weaknesses, a calibration is generally applied to the raw data [4, 5].

The large amount of collected information offers new opportunities of development in air quality modelling and mapping at urban scale that are the scope of recent studies. Statistical methodologies are broadly used to derive air quality maps from sensor data, in particular the Land Use Regression models (LUR), but they generally do not take spatial dependence into account. Geostatistical approaches have been less frequently applied to such type of data but provide significant advantages to combine sensor measurements and auxiliary information such as dispersion model outputs [6].

In this paper, data collected from fixed and mobile micro-sensors are used together with urban-scale modelling data to map PM_{10} in the city of Nantes (France).

2 PM_{10} data

PM_{10} sensor data were provided by AtmoTrack (https://atmotrack.fr), a French company created in 2015 in Nantes. PM_{10} data measured at reference monitoring stations (quarter-hourly mean concentrations) and simulation data (ADMS-Urban model) on the city of Nantes were provided by the French air quality monitoring association Air Pays de la Loire (http://www.airpl.org/).

2.1 Sampling routes and frequency of measurements

In November 2018, PM_{10} sensor measurements were collected in Nantes. During this sampling period, the company deployed 16 fixed sensors including 3 sensors at the Victor Hugo station (reference station for traffic typology) and 3 other sensors at the Bouteillerie station (reference station for urban background typology). In addition, 19 mobile sensors were installed on-board of driving school cars to measure PM_{10} concentrations over numerous routes each day of the sampling period. The vehicles routes ensure a satisfactory spatial coverage over the entire urban area even if they are totally dependent on the driving school car itineraries and on the lesson time (only daytime).

2.2 Measurement accuracy

Considering the measurement uncertainty, the three available datasets (data from the reference stations, the fixed sensors and the mobile sensors) can be related to three monitoring networks of respectively low (up to 25%), medium (up to 50%) and high (up to 125%) uncertainty (Figure 1).

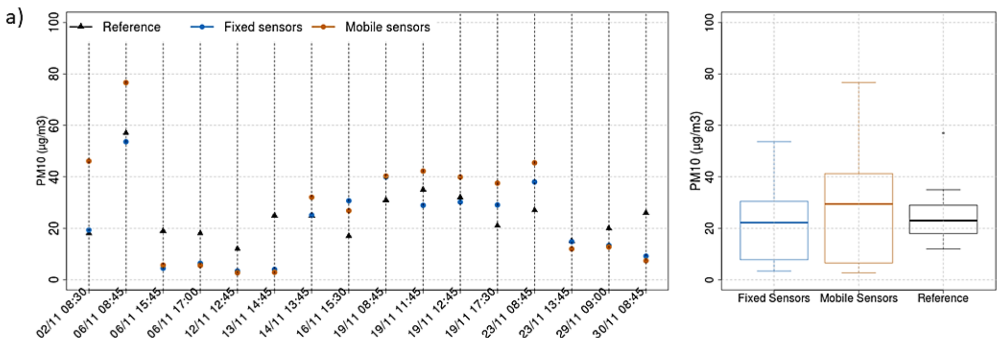

Figure 1: Comparison of the three networks (fixed sensors in blue, mobile sensors in orange and reference station in black) at Victor Hugo station for November 2018. Example of PM_{10}.

Microsensors offer a unique spatial and temporal coverage of pollutant concentrations. However, the accuracy of the measurements and their meaning, in case of mobile sensors, are real challenges to include them in air quality maps. In the following sections, a methodology of data fusion is detailed and a first test using fixed and mobile sensor data in Nantes (France) is presented.

3 Data fusion

Kriging [7] involves deriving linear combination of the data which ensures minimal estimation variance under a non-bias condition. Its strength is to give an information about the uncertainty of the estimated map.
Among the kriging methods, the universal or external drift kriging makes it possible to consider auxiliary information to increase the estimation accuracy. The main hypothesis is that the global mean is not constant through the domain and relies on explanatory variables, entailing an additional condition on the kriging weights. This approach has long been applied to air quality mapping [8, 9, 10, 11] and was used in this work to perform data fusion between:

- the hourly average concentrations measured by the fixed and mobile microsensors (after bias correction) as main variable;
- the 2016 annual average concentrations of the pollutant simulated by the ADMS-Urban dispersion model (https://www.cerc.co.uk/environmental-software/ADMS-Urban-model.html) as drift of the mean.

In addition, the measurement uncertainty of the sensors was taken into account by defining the variance of measurement errors (hereafter VME) as an input of the calculation.

4 Results

4.1 Estimation of PM_{10} concentration fields

Data fusion was performed for 27/11/2018, the day for which the amount of data was the largest. At every measurement position, the hourly mean of the observations is calculated, and external drift kriging is applied. The mobile and fixed sensor observations at 5pm and the annual modelled concentration field are presented for PM_{10} in figure 2a). Figure 2b) presents the VME for the same sampling routes. In this case, the measurement uncertainty is set to 25%, i.e. to the maximum uncertainty of the reference station observations. The uncertainty definition is totally arbitrary here and could be considered between 25% to 125%. Note that the fixed station measurements are not included in this estimation because they were used to correct and prepare the sensor data before kriging.

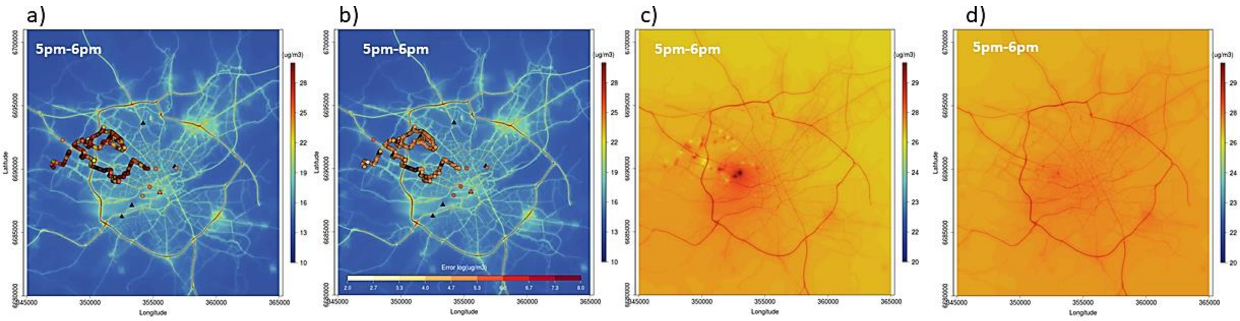

Figure 2: Data fusion of the sensor data on 27/11/2018 at 5pm: the 2016 annual average concentrations simulated by ADMS-Urban and the hourly-averaged sensor data (a), the variance of the measurement errors (b), the fused map with 25% uncertainty on measurements (c), and the fused map with 75% uncertainty on measurements (d).

As shown by the fused maps (Figure 2c and 2d), the modelled annual average allows to define the general patterns of the pollutant fields. Then the sensor observations which are associated with higher concentrations (by a factor of two) increase the concentration levels in the estimation domain, with some PM_{10} hotspots where data were collected (Figure 2c). When data fusion is performed with higher VME (75%, figure 2d), the hotspots are not represented anymore and the local effects of the sensor data is minimized.

5 Conclusion

The recent technological developments for miniaturizing the instruments that measure outdoor ambient air offer new possibilities for air quality modelling and mapping. The new portable and low-cost devices could provide observations of pollutants with higher spatial coverage than the reference monitoring networks. As long as several challenges can be dealt with (measurement uncertainty, representativeness of the sampling…), they could help to produce more accurate pollution maps. In this work, we investigate the potential added value of these data for air quality mapping by applying a data fusion technique. The dataset refers to PM pollution in the French city of Nantes. Hourly averaged sensor data and the annual mean concentration field simulated by ADMS-Urban model are combined by external drift kriging to estimate hourly PM_{10} concentrations, taking the variance of the measurement error into account. Those calculations were performed for one day but the next step of this work will be to consider each hour of the whole sampling period (November 2018). Further investigations will be carried out to estimate the influence of the amount of data, their position and their related uncertainties on the interpolation results. In addition, several ways of improvement have been identified such as the consideration of the spatial anisotropy in kriging and the application of spatiotemporal kriging. Besides geostatistical methods, machine learning techniques will be tested allowing to learn about historical data to improve the current estimate.

Acknowledgments

We thank AtmoTrack company for providing their sensor data and Air Pays de La Loire for providing the reference station observations and the results from their simulations with the ADMS-Urban model.

References

[1]. Directive 2008/50/EC of the European Parliament and the Council of 21 May 2008 on ambient air quality and cleaner air for Europe.

[2]. Directive 2004/107/EC of the European Parliament and the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air.

[3]. Kumar, P., et al. (2015). The rise of low-cost sensing for managing air pollution cities. Environment International, 75, 199-205.

[4]. Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., 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, http://dx.doi.org/10.1016/j.snb.2015.03.031.

[5]. Maag, B., Zhou Z., Thiele L. (2015). A survey on sensor calibration in air pollution monitoring deployments. IEEE Internet of Things journal, 5 (6), 4857 – 4870.

[6]. Schneider, P., Castell, N., Vogt, M., Dauge, F. R. and Lahoz, W. A., (2017). Mapping urban air quality in near real-time using observations from low cost sensors and model information. Environment International, 106, 234-247.

[7]. Chilès, J.-P., Delfiner, P.P. (2012). Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sons, 726 pp.

[8]. Wackernagel H., Lajaunie C., Blond N., Roth C., and Vautard R. (2004). Geostatistical risk mapping with chemical transport model output and ozone station data. Ecological Modelling, 179, 177–185.

[9]. Malherbe L., Ung A., Colette A., Debry E. (2011). Formulation and quantification of uncertainties in air quality mapping. ETC/ACM Technical Paper 2011/9.

[10]. Lahoz, W.A., Schneider, P.P. (2014). Data assimilation: making sense of earth observation, Frontiers in Environmental Science, 2 (16), 1–28, http://dx.doi.org/10.3389/fenvs.2014.00016.

[11]. Beauchamp M., de Fouquet Ch., Malherbe L. (2017). Dealing with non-stationarity through explanatory variables in kriging-based air quality maps. Spatial Statistics, 22 (1), 18-46.