EVALUATION OF DATA ASSIMILATION METHOD AT THE URBAN SCALE WITH THE SIRANE MODEL

Chi Vuong Nguyen and Lionel Soulhac

Laboratoire de Mécanique des Fluides et d'Acoustique, Université de Lyon, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Claude Bernard Lyon I, Ecully - France

Abstract: Ambient air quality modelling is widely used by local and national authorities to evaluate population exposure, to locate concentration thresholds exceedances, to investigate the relationship between air pollution and health effects or to quantify the impact of urban and traffic planning. Compared to the monitoring stations, numerical models are less accurate but their spatial resolution is better, in particular when using urban air quality models, which can describe the pollution with a resolution of a few meters. In order to reduce errors in modelling, data assimilation techniques can be used to combine measurements and simulations to produce a better description of the concentration field. The aim of this work is to evaluate a data assimilation method at urban scale, using the Best Linear Unbiased Estimator (BLUE) approach combined with the SIRANE urban air quality model (Soulhac et al., 2012, 2011). The difficulty in BLUE method is based on the estimation of the background error covariance matrix (B). In this work, B is modelled with three different models. This study compares estimates of the NO_2 concentrations on the agglomeration of Lyon for the year 2008, obtained with the BLUE method implemented with these three B models.

Key words: *Dispersion modelling, urban scale, data assimilation, background error covariance model*

Introduction

Ambient air quality modelling is widely used by local and national authorities to evaluate population exposure, to locate concentration thresholds exceedances, to investigate the relationships between air pollution and health effects or to quantify the impact of urban and traffic planning. Compared to the monitoring stations, numerical models are less accurate but their spatial resolution is better, in particular when using urban air quality models, which can describe the pollution with a resolution of a few meters. In order to reduce errors in modelling, data assimilation techniques, widely used at the regional scale (e.g. Wang et al., 2011), can be used to combine measurements and simulations to produce a better description of the concentration field. As far as we are aware, only Tilloy et al. (2013), and Air4EU (2007) have applied these methods with an atmospheric dispersion model at urban scale. The objective of this study is to evaluate the performances of the data assimilation method, called Best Linear Unbiased Estimator (BLUE), with the SIRANE urban air quality model.

Description of the SIRANE model

SIRANE is an operational model which represents the pollutants dispersion at the urban scale (Soulhac et al., 2012, 2011). This model is based on the street network concept proposed by Soulhac (2000). To estimate the mean concentration in the urban canopy, the model takes into account the main flow phenomena in the urban area: advection along the street axis induced by the parallel component of the wind; turbulent diffusion across the interface between the street and the external atmosphere; exchanges at the street intersections. The roughness sub-layer just above the urban canopy is neglected and the flow in the external atmosphere is modelled as a boundary layer flow on a rough surface. This external flow is supposed to be horizontally uniform. In the external atmosphere, the pollutants dispersion is represented by means of a Gaussian plume model with the standard deviations parametrized by the Monin-Obukhov similarity theory. The SIRANE model estimates the hourly mean concentrations assuming stationary conditions at each time step. The input data are the urban geometry of the streets network, the meteorological data, the point, line, and distributed emissions and the background concentration coming from the exterior of the domain.

Data assimilation

In this study, the data assimilation method used is the so called Best Linear Unbiased Estimator (BLUE). In the BLUE framework, the best estimation, called analysis, is a linear combination of an a priori information, called background and provided by a numerical model, and observation data. The analysis is determined with the equation (1) where y is the observation state vector of size m, x^b is the background state vector, x^a is the analysis state vector, H is the observation operator which maps from the background to the observation space, and K is the Kalman gain matrix.

$$x^a = x^b + K\left(y - Hx^b\right) \tag{1}$$

The Kalman gain matrix which minimizes the analysis error variance is given by the equation (2) where B is the background error covariance matrix of size n × n and R is the observation error covariance matrix of size m × m.

$$K = BH^T\left(HBH^T + R\right)^{-1} \tag{2}$$

The role of the B matrix is fundamental since it spreads the correction to points where there is no observation.

Modelling of the observation error covariance matrix (R)

European Directive relative to air quality stipulates that the maximal incertitude for the measurements must be of 15 % for the specie NO_2. Moreover, Tilloy et al. (2013) indicates that the observation errors are dependent of the measured concentrations. We assume in this study that the probability distribution of the observation error is Gaussian and that 95% of the errors are inferior to 15% of the mean measured concentrations. Consequently, in this study, the diagonal R matrix is estimated with the equation (3).

$$R = I_m\left(\frac{0.15\bar{y}}{1.96}\right)^2 \tag{3}$$

Modelling of the background error covariance matrix (B)

Basically, the background error covariance matrix is modelled with a spatial approach. This approach assume that closer (spatially) two points are, more correlated the associated background error are. The first model (M1) applied in this study is the one proposed by Tilloy et al. (2013). This model takes into account the variable d_{ij} which is the shorter distance along the street network between the points s_i and s_j and the variable P_i which is the projection of the point s_i on the closer road (equation (4)). The parameters v_0, L_d, L_p, and α refer respectively to a characteristic variance, a characteristic distance along the street network, a characteristic projection distance and a scaling coefficient.

$$B_{ij} = v_0 \exp\left(-\frac{d_{ij}}{L_d}\right)\exp\left(-\frac{\left|P_i - P_j\right|}{L_p + \alpha \min\left(P_i, P_j\right)}\right) \tag{4}$$

The second model used (M2) in inspired by the one proposed by Blond et al. (2003). The second model take into account the variable ρ_{ij} which is the correlation coefficient between the modelled concentrations at the points s_i and s_j and the variable var_i which is the variance of the modelled concentrations at the point s_i (equations (5) and (6)). The parameters ρ_0, L_p, and β refer respectively to a characteristic correlation coefficient, a characteristic correlation distance and a scaling coefficient. The underlying assumption of this approach is that more correlated the modelled concentration at two points are, more correlated the associated background error are. With this model, the non-diagonal terms of the B matrix are modelled with the equation (5) whereas the diagonal terms are modelled with the equation (6).

$$B_{ij(i\neq j)} = \rho_0 \exp\left(\frac{\rho_{ij} - 1}{L_\rho}\right)\sqrt{B_{ii}B_{jj}} \tag{5}$$

$$B_{ii} = \beta\, var_i \tag{6}$$

Finally, the third model proposed (M3) is a model which combines the two first models (equations (6) and (7)). In this last case, the diagonal terms are also calculated with the equation (6).

$$B_{ij(i\neq j)} = \rho_0 \exp\left(-\frac{d_{ij}}{L_d}\right)\exp\left(-\frac{\left|P_i - P_j\right|}{L_p + \alpha \min\left(P_i, P_j\right)}\right)\exp\left(\frac{\rho_{ij} - 1}{L_\rho}\right)\sqrt{B_{ii}B_{jj}} \tag{7}$$

Results of the data assimilation

This study case consists on estimating the NO_2 hourly mean concentrations in 2008 on the Lyon city. The background state is provided by the SIRANE urban air quality model and the observations are provided by 16 monitoring stations distributed on the city (Figure 1). In this study, the Kalman gain is stationary and is determined by means of the three background error covariance model and the observation error covariance model. The three background error covariance models are parametrized by realizing the χ^2 diagnostic (Tilloy et al., 2013). This diagnostic aims to check the consistency between the available data and the B and R model. The parameters determined with the χ^2 diagnostic are indicated in the Table 1.

Table 1. Background error covariance model parameters

Background error covariance model	Parameters	Values
BLUE (M1)	v_0, L_d, L_p, α	(575 $\mu g^2.m^{-6}$, 12 km, 1 m, 5)
BLUE (M2)	ρ_0, L_ρ, β	(0.50, 0.30, 0.80)
BLUE (M3)	ρ_0, L_d, L_p, α, L_ρ, β	(0.80 ,24 km, 1 m, 10, 0.60, 0.85)

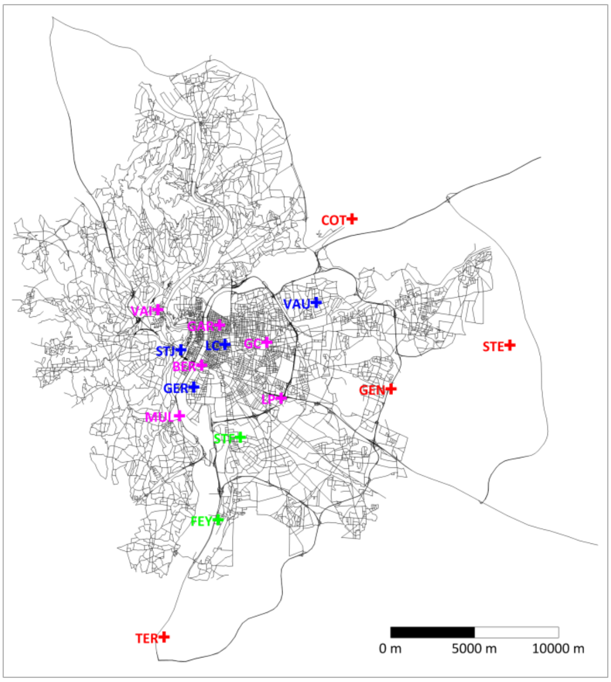

Figure 1. Map of the studied fields and of the street network used in the SIRANE model (red: background station, green: industrial station, blue: urban station, pink: traffic station)

The method used to assess the quality of the data assimilation is the leave-one-out cross validation. This validation consists on estimating the concentrations on a point where observations are available using all the observation available except those relative to the point. This operation is repeated for all the points where observations are available and the results are finally compared to the measurements. This validation aim to assess the performance of the data assimilation where there is no observations available. The statistical indices used to assess the quality of the simulation and of the data assimilation compared to the measurements data are the bias Bias, the relative error RE, the error variance RMSE, and the correlation coefficient Corr (Table 2, where the subscript m and p refer respectively to measured data and predicted data by the SIRANE model or by the data assimilation method).

Table 2. List of the statistical indices used for the comparison of SIRANE and data assimilation performances

Bias	ER	RMSE	Corr
$\overline{C_m - C_p}$	$\overline{\left(\frac{\lvert C_m - C_p \rvert}{C_m}\right)}$	$\sqrt{\overline{\left(C_m - C_p\right)^2}}$	$\frac{\overline{\left(C_m - \overline{C_m}\right)\left(C_p - \overline{C_p}\right)}}{\sigma_m \sigma_p}$

Table 3. Global performances of the SIRANE model and the method BLUE (the green bold values indicate an improvement in respect to the SIRANE model)

Method	Bias [µg.m^{-3}]	ER	RMSE [µg.m^{-3}]	Corr
SIRANE	3.51	0.48	22.31	0.73
BLUE (M1)	**1.41**	**0.38**	**17.68**	**0.83**
BLUE (M2)	**0.73**	**0.39**	**18.11**	**0.82**
BLUE (M3)	**1.04**	**0.36**	**17.22**	**0.84**

Table 4. Performances of the method BLUE (M3) (the values in the brackets are relative to the SIRANE model and the green bold values indicate an improvement in respect to the SIRANE model)

Station	μ_{Mes} [µg.m^{-3}]	Bias [µg.m^{-3}]	ER	RMSE [µg.m^{-3}]	Corr
COT	23.44	1.55 (1.47)	**0.36 (0.46)**	**9.54 (11.31)**	**0.86 (0.80)**
GEN	33.80	2.20 (1.79)	**0.35 (0.51)**	**12.14 (15.48)**	**0.83 (0.71)**
STE	17.96	-2.77 (-2.06)	0.40 (0.36)	5.03 (4.38)	0.96 (0.97)
TER	30.00	5.17 (4.64)	**0.47 (0.55)**	**14.45 (16.12)**	**0.72 (0.63)**
BER	52.22	-6.51 (-0.84)	**0.31 (0.39)**	**17.28 (22.14)**	**0.83 (0.63)**
GAR	73.24	**10.03 (16.41)**	**0.26 (0.35)**	**25.93 (33.45)**	**0.81 (0.69)**
GC	46.91	**3.29 (6.65)**	**0.37 (0.48)**	**18.96 (24.99)**	**0.80 (0.64)**
LP	50.70	-1.46 (0.22)	**0.31 (0.48)**	**17.07 (25.64)**	**0.84 (0.63)**
MUL	78.86	**-0.70 (10.45)**	0.46 (0.44)	**35.59 (39.70)**	**0.65 (0.60)**
VAI	59.30	**14.91 (18.99)**	**0.31 (0.38)**	**23.98 (29.83)**	**0.78 (0.66)**
FEY	33.56	**0.22 (2.01)**	**0.38 (0.54)**	**12.20 (17.64)**	**0.81 (0.59)**
STF	35.94	**-0.07 (2.54)**	**0.42 (0.49)**	**12.47 (18.73)**	**0.89 (0.70)**
GER	38.34	**0.92 (2.31)**	**0.23 (0.45)**	**9.62 (18.55)**	**0.92 (0.66)**
LC	38.01	-8.98 (-4.79)	**0.39 (0.58)**	**13.34 (19.78)**	**0.92 (0.65)**
STJ	37.31	**-1.07 (-2.72)**	**0.33 (0.55)**	**12.59 (19.46)**	**0.86 (0.66)**
VAU	26.95	**-0.12 (-0.95)**	**0.38 (0.72)**	**8.60 (13.64)**	**0.91 (0.75)**

The global performances of the BLUE method are indicated in the Table 3. The results show that the BLUE method, with the three background error covariance models, improves statistically the estimations of the NO_2 hourly concentration compared to the SIRANE model. We note that the global performances of the BLUE with the 3 BECM models are basically similar even if the model M3 lead to slightly better results. The Table 4 indicates the local results obtained with the model M3. Locally, the results show an improvement of the statistical indices RE, RMSE, and Corr for almost all the stations with the implementation of the BLUE method (M3). However, the data assimilation performances are less satisfying for the bias. The bias is sometimes improved but also sometimes much worst. Let remember that the term “best estimation” is here synonymous of minimization of the estimation error variance, in other words minimization of the RMSE. Moreover the BLUE method assumes that the background and the observations are unbiased and by consequent do not aim to correct the bias. This can partly explain the unsatisfactory results for the bias.

Conclusion

The performances of the BLUE method at the urban scale have been assessed with the SIRANE urban air quality model. In this study, three background error covariance models are applied. The first model is the one proposed by Tilloy et al. (2013) which uses the spatial approach and the second model is inspired by the one of Blond et al. (2003) which model the B matrix in function of the correlation coefficient relative to the modelled concentrations. The third model is a combination of the two first ones. The results of the leave-one-out cross validation show that the BLUE method, with these three models, improves the global estimations of the NO_2 hourly concentrations considering the statistical indices bias, RE, RMSE, and Corr with slightly better improvement with the third model. However, the bias is sometimes much worse locally. In this study, one of the limits is the stationary aspect of the B matrix. This means that the variations of B due to atmospheric stability, to the changes of meteorological conditions and of the emissions are not taken into account.

Others approach to model the B matrix as ensemble methods can be considered. Likewise, complementary studies relative to the characterization of the background error can lead to a better modelling of the B matrix. Let note that the data assimilation method need available observation data whose are moreover spatially heterogeneous. So, it is also necessary to consider other approaches to improve the estimations of urban air quality model as coupling several dispersion models with different spatial scale.

Blond, N., Bel, L., Vautard, R., 2003. Three-dimensional ozone data analysis with an air quality model over the Paris area. J. Geophys. Res. Atmospheres 108, ACH 10–1 – ACH 10–17.

Denby, B., 2007. Basic data assimilation:application to the urban scale (No. 6th Framework Programme : Deliverable D7.1.6). Air4EU.

Gandin, L.S., 1963. Objective analysis of meteorological fields. Israel Program for Scientific Translations;1965; [available from the U.S. Dept. of Commerce, Clearinghouse for Federal Scientific and Technical Information, Springfield, Va.].

Soulhac, L., Salizzoni, P., Cierco, F.-X., Perkins, R., 2011. The model SIRANE for atmospheric urban pollutant dispersion; part I, presentation of the model. Atmos. Environ. 45, 7379–7395. doi:10.1016/j.atmosenv.2011.07.008

Soulhac, L., Salizzoni, P., Mejean, P., Didier, D., Rios, I., 2012. The model SIRANE for atmospheric urban pollutant dispersion; PART II, validation of the model on a real case study. Atmos. Environ. 49, 320–337. doi:10.1016/j.atmosenv.2011.11.031

Tilloy, A., Mallet, V., Poulet, D., Pesin, C., Brocheton, F., 2013. BLUE-based NO 2 data assimilation at urban scale. J. Geophys. Res. Atmospheres 118, 2031–2040. doi:10.1002/jgrd.50233

Wang, X., Mallet, V., Berroir, J.-P., Herlin, I., 2011. Assimilation of OMI NO2 retrievals into a regional chemistry-transport model for improving air quality forecasts over Europe. Atmos. Environ. 45, 485–492. doi:10.1016/j.atmosenv.2010.09.028

WHO, 2013. Outdoor air pollution a leading environmental cause of cancer deaths [WWW Document]. URL http://www.euro.who.int/en/health-topics/environment-and-health/urban-health/news/news/2013/10/outdoor-air-pollution-a-leading-environmental-cause-of-cancer-deaths (accessed 3.12.15).