17th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 May 2016, Budapest, Hungary

MODIFICATION AND VALIDATION OF A METHOD FOR ESTIMATING THE LOCATION OF A POINT STATIONARY SOURCE OF PASSIVE NON-REACTIVE POLLUTANT IN AN URBAN ENVIRONMENT

George C. Efthimiou[1], Spyros Andronopoulos[1], Alexandros Venetsanos[1], Ivan V. Kovalets[2], Konstantinos Kakosimos[3] and Christos D. Argyropoulos[3]

[1]Environmental Research Laboratory, INRASTES, NCSR Demokritos, Patriarchou Grigoriou & Neapoleos Str., 15310, Aghia Paraskevi, Greece
[2]Department of Environmental Modelling, Institute of Mathematical Machine and System Problems, National Academy of Sciences of Ukraine, Kiev, Ukraine
[3]Chemical Engineering Department & Mary Kay 'O Connor Processes Safety Center, Texas A&M University at Qatar

Abstract: In Kovalets et al. (2011), a method was developed to estimate the location and rate of an unknown point stationary source of passive atmospheric pollutant in a complex urban geometry. The algorithm was implemented in the ADREA-HF Computational Fluid Dynamics code and was applied in a complex urban geometry (the MUST wind tunnel experiment). This approach has also been evaluated for the Michelstadt experiment (Tsiouri et al., 2014). In the present paper, a major change in the data assimilation code includes the implementation of a two-step approach: At first only the source coordinates are analysed using a correlation function of measured and calculated concentrations. In the second step the source rate is identified by minimizing a quadratic cost function. The validation of the new algorithm is performed for the source location by simulating a wind tunnel experiment on atmospheric dispersion among buildings of a real urban environment. Good results of source location estimation are obtained. Special attention is given on the grid influence regarding the obtained results.

Key words: *CFD; Data assimilation; Atmospheric dispersion; Inverse problem; Source identification*

INTRODUCTION

The characterization of an unknown atmospheric pollutant's source following a release is a special case of inverse atmospheric dispersion problem. Such kind of inverse problems are to be solved in a variety of application areas such as emergency response and indoor air quality (Kovalets et al., 2011; Koracin et al., 2011, Matsuo et al., 2015, Sharan et al., 2012, Singh et al., 2013).

In the urban or industrial scale, few studies have applied Computational Fluid Dynamics (CFD) combined with different source estimation techniques (Chow et al., 2008; Keats et al., 2007; Libre et al., 2012; Bady et al., 2009). Kovalets et al. (2011) developed an effective variational algorithm of source inversion combined with an urban-scale CFD model. The performance of the algorithm was evaluated against measurements obtained in MUST wind tunnel experiment and with the Michelstadt wind tunnel dataset (Tsiouri et al., 2014). The medium performance of the algorithm in the second case created second thoughts about the calculation of the source location and rate as well as the effect of various numerical parameters (e.g. grid resolution and numerical schemes). The purpose of the present study is to modify the cost function of the source inversion algorithm of Kovalets et al. (2011) for the non-simultaneous calculation of the source location and rate. The effects of the grid resolution used for the numerical simulations on the obtained results are also investigated. The evaluation and demonstration of the proposed methodology is performed on a new experiment called CUTE, the description of which is given below.

MODIFICATION OF THE COST FUNCTION OF THE SOURCE INVERSION ALGORITHM

The use of the Source Inversion (SI) algorithm (Kovalets et al., 2011) for the Michelstadt experiment (Tsiouri et al., 2014) produced unsatisfactory results regarding the distance between the true and the estimated source location and the true to estimated source rate ratio. The former was in the order to 300 m, while the latter was in the order of 10. The cause of the discrepancy between the results and the measurements was the 'overfitting' effect. According to this effect, the calculation errors which are introduced by the wrong source location and led to significant underestimation of the concentration were compensated by the overestimated source rate. Thus, the resulting quadratic cost function reached minimum for the wrong combined solution (source location and source rate). This problem is typical for non-linear least squares fitting. In context of data assimilation this problem is especially important when the number of measurements is insufficiently small. For example, in Kovalets et al. (2009) the cost function was minimized with respect to both wind field and source rate. It was found that for too coarse resolution of the monitoring network the Normalized Mean Squared Error of the concentration field (closely related to the value of cost function) decreased in the data assimilation process while the error of the wind field (being part of the solution vector) increased (Kovalets et al. (2009), p. 3519, par. 1).

A proposed solution to the above mentioned problem is the separation of SI algorithm into the following two steps: Step 1: The source coordinates are analysed; Step 2: Only source rate is analysed.

The aforementioned idea is not new and was also expressed by the authors of the paper Brown and Robins (2010) in private discussion with Ivan Kovalets. They developed a method which allows the source coordinates-only estimation, while unknown source rate was not considered. However, it is not clear how they deal with the issue of unknown source rate in calculating posterior probabilities of the concentrations. Furthermore, it is known that Bayesian inference is under certain assumptions equivalent to least squares approach of source term estimation (see Tarantola, 2005, and more details can be found in Kovalets et al., 2013, Sec. 2.1). However from the implementation point of view the methodology in Brown and Robins (2010) is very different from that in Kovalets et al. (2011).

In the present study, we propose a modified version of the SI algorithm (Kovalets et al., 2011) by using, instead of cost function J defined by formula (7) (in Kovalets et al., 2011), a correlation coefficient of measured and calculated concentrations:

$$J = \frac{\left\langle \left(c^c - \left\langle c^c \right\rangle\right)\left(c^o - \left\langle c^o \right\rangle\right) \right\rangle}{\sqrt{\left\langle \left(c^c - \left\langle c^c \right\rangle\right)^2 \right\rangle}\sqrt{\left\langle \left(c^o - \left\langle c^o \right\rangle\right)^2 \right\rangle}} \to \min \tag{1}$$

where $\langle \; \rangle$ denotes arithmetic averaging over all measurements while minimum is sought over all possible source locations. Other notation in (1) as well as everywhere below (until otherwise is stated) is similar to that in Kovalets et al. (2011).

Equation (1) is minimized with respect to source coordinates only, while arbitrary source rate, q^s, is used for the minimization procedure. The justification for this exclusion of source rate from the control vector is that J does not depend on q^s. Indeed, let us consider two calculated concentration fields c_1^c and c_2^c obtained with different source rates: $q_1^s \, and \, q_2^s$. Since equation of concentration transport is linear with respect to q^s, then the following relationship holds:

$$c_2^c = c_1^c \cdot \left(\frac{q_2^s}{q_1^s} \right) = c_1^c \cdot \alpha \tag{2}$$

From equation (2), it is obvious that the calculated values of J_1 and J_2 for c_1^c and c_2^c will be the same. Thus, J does not depend on source rate provided that it is constant. For non-constant source rates, $q_1^s(t)$ and $q_2^s(t)$, the values of J_1 and J_2 generally speaking will be different. Only if $q_1^s(t)$ and $q_2^s(t)$ are similar, i.e. if $q_2^s(t) / q_1^s(t) = \alpha = const$, then again J_1 and J_2 will be the same.

Therefore, the usage of correlation coefficient as a cost function (equation (1)) solves the problem of separate identification of source coordinates and source rate. An arbitrary value of 'first-guess' source rate q_0^s could be used for calculating the correlation coefficient since it will not influence the solution. However, the drawback of the present approach is that prior information can not be used by applying regularization terms.

Therefore, in step 1 of the present approach, the source coordinates are identified using the algorithm of Kovalets et al. (2011) but with the cost function defined above (equation 1) instead of the cost function of Kovalets et al. (2011). In this algorithm, the source location is assumed to coincide with one of the grid nodes. The location of the corresponding grid node k^s for which minimum of the correlation coefficient (equation (1)) is reached is considered as the analysed source location.

As a second step, the source rate can be identified by minimizing the quadratic cost function with respect to q^s:

$$J = \sum_{n=1}^{K} \left(c_n^c - c_n^o \right)^2 \rightarrow \min \tag{3}$$

By using a Source Receptor Function (SRF), similarly applied as in Kovalets et al. (2011) for the node k^s (being the solution found in Step 1),equation (3) can be expressed as (cf. formula (15) from Kovalets et al., 2011):

$$J = \sum_{n=1}^{K} \left(q^s c_{n,k^s}^* - c_n^o \right)^2 \xrightarrow[q^s]{} \min \tag{4}$$

The solution of problem (4) can obviously be obtained analytically by equating to zero the derivative of J with respect to q^s:

$$\frac{\partial J}{\partial q^s} = \sum_{n=1}^{K} \left(q^s c_{n,k^s}^* - c_n^o \right)^2 = \sum_{n=1}^{K} 2 c_{n,k^s}^* \left(q^s c_{n,k^s}^* - c_n^o \right) = 0 \tag{5}$$

The solution of (5) is obviously:

$$q^s = \frac{\sum_{n=1}^{K} c_{n,k^s}^* c_n^o}{\sum_{n=1}^{K} \left(c_{n,k^s}^* \right)^2} \tag{6}$$

THE COMPLEX URBAN TERRAIN EXPERIMENT (CUTE)

The CUTE (Complex Urban Test Experiment) data set includes results from field and wind tunnel measurements. The data set is dedicated to test Emergency Response Tools/Atmospheric Dispersion Models predicting dispersion processes in urban areas. The second part of the CUTE dataset (CUTE cases 2 to 4) consists of wind tunnel data measured in the model of the European city centre, where the CUTE field test was carried out.

The test was done in the downtown area of a typical Central European city. The area of interest is densely built-up with building heights between 25 m and 35 m. There is no significant elevation and no major urban greenery in the modelled area.

The CUTE wind tunnel dataset consists of concentration data only. Concentration time series of tracer gas from continuous releases were measured at pedestrian level. Concentration measurements are available for three cases. Each case represents a different source location. However, only one case (case 3) was used during the evaluation of the model.

The chosen source is located between houses near the river on the opposite side of the harbour. Case 3 was selected because here the source is at a different location than the release during the field test, representing a different concentration field and therefore a different test case for model evaluation. For case 4, the release was at the downwind side of river, resulting in fewer measurement points and short travel times of released material. More information about the experiment can be found in http://www.elizas.eu/images/Documents/Model%20Evaluation%20Case%20Studies_web.pdf.

THE SIMULATIONS OF THE INVERSE PROBLEM

The computational domain for simulating the developing flow and dispersion within and above the three-dimensional obstacle array extended in the streamwise direction from x = −540 m to 2105.02 m, with the upstream wall of the first obstacle at x = 0 m. The simulations were conducted with a domain height of $6H_{max}$ (H_{max} is the maximum building height equal to 108 m), which was sufficiently deep to ensure that the flow changes near the surface (within and above the obstacles) were not being moulded (or influenced) by the boundary conditions imposed at the top of the computational domain. In the spanwise (or y) direction, the computational domain spanned −540 m ≤ y ≤ 1755.02 m. The vertical x-z centre plane at y ≈ 300 m contained the ground-level source.

Sensitivity analysis of the solution to the discretization of the computational domain is performed in the present paper. A medium and a fine grid have been selected. The number of the control volumes for the medium grid is 891,132 while for the fine grid is 1,372,800. The grid was equidistant between the buildings and increases logarithmically outside the urban area. For the medium grid the minimum cell size in horizontal direction is 12.52 m while for the fine grid it is 9.969 m. Both grids have the same vertical resolution with a minimum cell size close to the ground equal to 1.0 m.

The boundary conditions of the 3D domain are the same for the forward and the inverse problems. In order to solve the adjoint equation (8) in Kovalets et al. (2011) the inverse problem uses the flow results of the forward problem i.e. the mean velocities *u*, *v*, *w*, the turbulent kinetic energy *k* and the dissipation rate ε.

The non-stationary adjoint equation has been integrated in backward direction through the same time interval as the forward problem (1000 s) to achieve established distributions of adjoint variables. The time step was kept constant and equal to 1 s.

THE RESULTS OF THE INVERSE PROBLEM

Based on the solution of the adjoint equation, the adjoint variables have been pre-calculated for all sensors in the whole computational domain and then stored in binary files.

The performance of the modified algorithm of source estimation was evaluated by two parameters: horizontal $\left(r_H = \sqrt{\left(x^s - x_t^s\right)^2 + \left(y^s - y_t^s\right)^2} \right)$ and vertical $\left(r_V = \left| z^s - z_t^s \right| \right)$ distances between the estimated and the true source locations. The results are presented in Table 1. The medium grid presents a rather high discrepancy from the true source location. On the other hand the fine grid predicted perfectly the source. These results indicate that even for a small number of sensors (32 in this experiment) the grid plays a major role for the performance of the methodology.

Table 1. Horizontal and vertical distances of the estimated source location from the true source location.

	r_H (m)	r_V (m)
Medium grid	43.33	14.74
Fine grid	0	0

CONCLUSIONS

In the present work an existing method for estimating the location and rate of a stationary point source of passive nonreactive pollutant in an urban environment was modified and evaluated. A major change in the data assimilation code included the implementation of a two-step approach: At first only the source coordinates were analysed using a correlation function of measured and calculated concentrations. In the second step, the source rate was identified by minimizing a quadratic cost function. The validation of the new algorithm was performed for the source location by simulating a wind tunnel experiment on atmospheric dispersion among buildings of a real urban environment. Good results of source location estimation have been achieved when all available measurements (32) were used to solve the inverse problem. It was found that the grid resolution plays an important role for the inverse problem.

ACKNOWLEDGMENTS

This publication was made possible by a NPRP award [NPRP 7-674-2-252] from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

Bady M., S. Kato and H. Huang, 2009: Identification of pollution sources in urban areas using reverse simulation with reversed time marching method. *Journal of Asian Architecture and Building Engineering*, **8**, 275-282.

Brown G. and P. Robins, 2010: Source term estimation for rapid hazard assessment. HARMO13, Paris, France, 924-928.

Chow F.K., B. Kosovic´ and S. Chan, 2008: Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations. *Journal of applied meteorology and climatology*, **47**, 1553-1572.

Keats A., E. Yee and F.-S. Lien, 2007: Bayesian inference for source determination with applications to a complex urban environment. *Atmospheric Environment*, **41**, 465–479.

Koracin D., R. Vellore, D.H. Lowenthal, J.G. Watson, J. Koracin, T. McCord, D.W. DuBois, L.-W.A. Chen, N. Kumar, E.M. Knipping, N.J.M. Wheeler, K. Craig and S. Reid, 2011: Regional source identification using lagrangian stochastic particle dispersion and HYSPLIT backward-trajectory models. *Journal of the Air & Waste Management Association*, **61**, 660-672.

Kovalets I.V., S. Andronopoulos, A. Venetsanos and J.G. Bartzis, 2011: Identification of strength and location of stationary point source of atmospheric pollutant in urban conditions using computational fluid dynamics model. *Mathematics and Computers in Simulation*, **82**, 244-257.

Kovalets I., R. Hofman, P. Seibert and S. Andronopoulos, 2013: Review of existing approaches of source inversion to be integrated in real-time nuclear emergency response system JRODOS. Report of the EU FP7 PREPARE Project No. PREPARE(WP4)-(14)-01.

Kovalets I.V., V. Tsiouri, S. Andronopoulos and J.G. Bartzis, 2009: Improvement of source and wind field input of atmospheric dispersion model by assimilation of concentration measurements: Method and applications in idealized settings. *Applied Mathematical Modelling*, **33**, 3511-3521.

Libre J-M., S. Guérin, A. Castellari, A. Tripathi, M. Leguellec, T. Mailliard and C. Souprayen, 2012: Source determination in congested environment through Bayesian inference. *Int. J. Environment and Pollution*, **48**, 174–184.

Matsuo T., A. Kondo, H. Shimadera, T. Kyuno and Y. Inoue, 2015: Estimation of indoor contamination source location by using variational continuous assimilation method. *Build. Simul.*, **8**, 443–452.

Sharan M., S.K. Singh and J.P. Issartel, 2012: Least square data assimilation for identification of the point source emissions. *Pure Appl. Geophys.*, **169**, 483–497.

Singh S.K.,·M. Sharan and J.-P. Issartel, 2013: Inverse modelling for identification of multiple-point releases from atmospheric concentration measurements. *Boundary-Layer Meteorol*, **146**, 277–295.

Tarantola A., 2005: Inverse problem theory and methods for parameter estimation. SIAM, Philadelphia, USA.

Tsiouri V., I. Kovalets, K.E. Kakosimos, S. Andronopoulos and J.G. Bartzis, 2014: Evaluation of advanced emergency response methodologies to estimate the unknown source characteristics of the hazardous material within urban environments. HARMO16, Varna, Bulgaria.