

An optimized inverse modelling method for determining the location and strength of a point source releasing airborne material in urban environment

George C. Efthimiou^{a1}, Ivan V. Kovalets^b, Alexandros Venetsanos^a, Spyros Andronopoulos^a, Christos D. Argyropoulos^c, Konstantinos Kakosimos^c

^a Environmental Research Laboratory, INRASTES, NCSR Demokritos, Patriarchou Grigoriou & Neapoleos Str., 15310, Aghia Paraskevi, Greece

^b Department of Environmental Modelling, Institute of Mathematical Machine and System Problems, National Academy of Sciences of Ukraine, Kiev, Ukraine

^c Department of Chemical Engineering & Mary Kay 'O Connor Processes Safety Center, Texas A&M University at Qatar, Education City, Doha, Qatar, PO Box 23874

Abstract

An improved inverse modelling method to estimate the location and the emission rate of an unknown point stationary source of passive atmospheric pollutant in a complex urban geometry is incorporated in the Computational Fluid Dynamics code ADREA-HF and presented in this paper. The key improvement in relation to the previous version of the method lies in a two-step segregated 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 by simulating the MUST wind tunnel experiment. A grid-independent flow field solution is firstly attained by applying successive refinements of the computational mesh and the final wind flow is validated against the measurements quantitatively and qualitatively. The old and new versions of the source term estimation method are tested on a coarse and a fine mesh. The new method appeared to be more robust, giving satisfactory estimations of source location and emission rate on both grids. The performance of the old version of the method varied between failure and success and appeared to be sensitive to the selection of model error magnitude that needs to be inserted in its quadratic cost function. The performance of the method depends also on the number and the placement of sensors constituting the measurement network. Of significant interest for the practical application of the method in urban settings is the number of concentration sensors required to obtain a “satisfactory” determination of the source. The probability of obtaining a satisfactory solution – according to specified criteria – by the new method has been assessed as function of the number of sensors that constitute the measurement network.

Keywords: ADREA-HF; CFD; Inverse modelling; Source term estimation; Urban environment

1. Introduction

¹ Corresponding author. Tel: +30 2106503405. E-mail address: gefthimiou@ipta.demokritos.gr (G. C. Efthimiou).

The characterization of an unknown atmospheric pollutant's source (i.e., assessment of its location and emission rate) following a release (i.e., using the available observations) is a special case of inverse atmospheric dispersion problem. The inverse dispersion problems are solved by combining in some optimal way results of computational dispersion models and existing measurements. Such kind of inverse problems are to be solved in a variety of application areas such as emergency response (e.g. Kovalets et al., 2011; Sharan et al., 2012; Singh et al., 2013; Ben Salem et al., 2016) pollution control decisions (Koracin et al., 2011), indoor air quality (Matsuo et al., 2015) and monitor of nuclear testing (Pudykiewicz, 1998). A review of source term estimation methods for atmospheric dispersion events is given in Hutchinson et al. (2016).

In the urban or industrial plant spatial scale, where complex and turbulent meteorological conditions arise due to the interaction of the air flow with the buildings, there are few researchers that have combined Computational Fluid Dynamics (CFD) with source estimation techniques (Mons et al., 2017; Xue et al., 2017; Kumar et al., 2016; Kumar et al., 2015; Libre et al., 2012; Kovalets et al., 2011; Bady et al., 2009; Chow et al., 2008; Keats et al., 2007). In Mons et al. (2017) a non-intrusive ensemble-based variational data assimilation scheme enhanced by proper orthogonal decomposition was deployed in conjunction with a Large Eddy Simulation (LES) solver. Their methodology was proved to be able to accurately identify both pollutant source and wind characteristics with a computational cost of $O(10)$ LES simulations. Xue et al. (2017) used a two-step deterministic approach for identifying the point source with a constant emission rate in built-up urban areas. The analytic form of the marginal posterior probability density function of the source location was derived to estimate the source location. The emission rate was then estimated using the conditional posterior distribution. Their method was tested using data obtained in two wind tunnel scenarios of contaminant dispersion in typical urban geometries represented by block arrays. CFD modelling and the adjoint equations were used to calculate the building-resolving source-receptor relationship required in the identification. In Kumar et al. (2016) a source inversion methodology was based on the renormalization inversion theory coupled with a building resolving CFD (i.e. fluidyn-PANACHE) modelling approach which estimates the release height along with the projected location on the ground surface and the intensity of an unknown continuous point source in an urban area. In Kumar et al. (2015) a methodology was described combining a renormalization inversion technique with a building-resolving CFD approach for source retrieval in the geometrically complex urban regions. The renormalization inversion approach was based on an adjoint source-receptor relationship. Libre et al. (2012) CFD (i.e. Fluidyn-PANEP) was used to model the dispersion while the Bayesian inference algorithm enabled the estimation of the source term in real time. The modelling platform based on these developments produces real time 3D concentration fields owing to precalculated wind fields patterns selected from a database on the basis of real time meteorological data acquisition. Bady et al. (2009) examined the accuracy of the inverse CFD (Star-CD) modelling in a simple laminar flow, but focused on the source location. The pollutant was emitted for variable wind conditions and the wind flow around a single building was investigated for two different sources. In both cases the numerical concentration results of the forward problem were used as concentration measurements for the inverse problem. By using the peak concentration as indicator, they could identify the source of the pollution. Chow et al. (2008) used Bayesian inference via stochastic sampling algorithms with a high-resolution CFD code (FEM3MP) in order to reconstruct an atmospheric release event and to determine the plume source and release rate based on point measurements of concentration. The CFD model has been used as a 'black-box', i.e., for

each set of input parameters the model was run in forward mode to obtain the corresponding set of output parameters. In Keats et al. (2007) a similar probabilistic formulation of inverse problem was adopted and posterior distribution function was evaluated with Monte-Carlo sampling. However instead of using a CFD model (i.e. urbanSTREAM) as a 'black box' tool, a 'source-receptor' function was constructed using the adjoint transport equation, as per Marchuk (1982) and Marchuk (1996).

At this point it should be noted that there are limitations in some cases of inverse modelling. According to Dhall et al. (2006) there is a typical problem for non-linear least squares fitting due to the ill-posed minimization problem and the non-convex cost function. This problem is called 'overfitting' effect. According to this effect, the calculation errors which are introduced by a wrong source location estimation are compensated by error in estimated source rate. Thus, the resulting quadratic cost function reaches minimum for the wrong combined solution (source location and source rate). In context of data assimilation, this problem is especially important when the number of measurements is 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. This 'overfitting' effect was also observed in the work of Tsiouri et al. (2014) in which the Source Inversion (SI) algorithm produced unsatisfactory results regarding the distance between the true and the estimated source location and the true to estimated source rate ratio.

This study presents an integrated and innovative approach, to eliminate the 'overfitting' effect, based on the following improvement: a non-simultaneous determination of the source location and rate is proposed, based on a two-step segregated approach that combines a correlation coefficient and a cost function, instead of using a single cost function as was done by Kovalets et al. (2011). The correlation coefficient of measured and calculated concentrations is used to estimate the source location, while the quadratic cost function is directly solved to calculate the emission rate. In the following part of the paper, the new version of the method is described in detail and then both old and new versions are inter-compared by applying them on the MUST wind-tunnel experimental dataset. This experiment has been selected because of its high quality data and its extensive use in other similar works. The source term estimation method is based on the calculated flow field, so a grid independent solution for the wind flow field is firstly obtained by applying successive refinements of the computational mesh. However, the old and new source term estimation methods are tested on both coarse and fine grids, to assess their dependence on the grid resolution, in combination to other parameters. In the last part of the paper the dependence of the new method's performance on the number of measuring sensors is examined.

2. Methodology

In general, solving the inverse atmospheric dispersion problem consists of estimating the posterior joint probability distribution of the unknown model parameters (usually the source spatial coordinates and emission rate), the measurements and the prior estimations of the unknown model parameters. However this estimation is computationally expensive. Therefore in many practical applications, e.g., in emergency response, the use of variational formulation is preferred, in which

only the location of the maximum of the posterior probability is estimated. If model errors, measurement errors and prior estimation errors follow Gaussian probability distributions, then, locating the maximum of the posterior joint probability distribution coincides with minimizing a cost function, such as the one proposed by Kovalets et al. (2011). In that work, as well as in the present, the unknown pollutant source parameters, that are sought for, are the emission rate q^s and the coordinates x^s, y^s, z^s . The cost function that was minimised by Kovalets et al. (2011) to calculate these unknown parameters was formulated as follows:

$$J = \frac{(x_0^s - x^s)^2}{\sigma_H^2} + \frac{(y_0^s - y^s)^2}{\sigma_H^2} + \frac{(z_0^s - z^s)^2}{\sigma_V^2} + \frac{1}{\sigma_O^2 + \sigma_M^2} \cdot \sum_{n=1}^K (c_n^c - c_n^o)^2 \quad (1)$$

The first three terms in the r.h.s of Eq. 1 are regularization terms that contain prior estimations of the source location (x_0^s, y_0^s, z_0^s) and parameters of probability distributions characterizing errors of prior estimations of source location (assumed to be Gaussian), namely mean squared deviations of horizontal (σ_H^2) and vertical (σ_V^2) coordinates. Regularization terms are added to address the problems of ill-posedness of the inverse problem and non-convexity of the cost function. The last term in Eq. 1 accounts for the mean squared difference between calculated (c_n^c) and measured (c_n^o) concentration values at the measurement point n , while σ_O^2 and σ_M^2 are mean squared errors of measurements and model results respectively. K is the total number of measurement points.

Both source emission rate and coordinates are estimated through minimization of Eq. (1) in the work of Kovalets et al. (2011). However, as already noted in the Introduction, in such inverse problem solution, an “overfitting” effect may occur, in which the quadratic cost function reaches a minimum for the wrong combined solution (large error in source location compensated by small error in source rate or vice versa). Therefore, to avoid this situation, in the present work the estimation of the source location is separated from the estimation of source rate. So, instead of using Eq. (1) the present method is split in two parts, as follows.

In the first part, the following correlation-based cost function of measured and calculated concentrations (i.e. correlation coefficient taken with negative sign) is minimised to calculate the source coordinates:

$$J = - \frac{\langle (c^c - \langle c^c \rangle)(c^o - \langle c^o \rangle) \rangle}{\sqrt{\langle (c^c - \langle c^c \rangle)^2 \rangle} \sqrt{\langle (c^o - \langle c^o \rangle)^2 \rangle}} \quad (2a)$$

where $\langle \rangle$ denotes arithmetic averaging over all sensors positions.

To minimize Eq. (2a), a direct method is applied here. A direct method requires calculating the values of the function J , therefore requires the model-calculated values of concentration c^c at every sensor location for every set of source parameters. Concentrations c^c for every possible source location and emission rate need to be calculated very fast for the direct method to be effective. A way to achieve this is through the use of “Source-receptor functions” (SRFs) which

describe the sensitivity of concentration at a receptor to the parameters of the emitting source. The SRFs for each receptor and for each possible source location can be pre-calculated and stored (of course for a given flow field) and they can be used when needed to calculate c^c . SRFs can be calculated in two ways: (a) by “forward”-in-time solution of the concentration transport equation, with the emission located at each possible source location, calculating the concentration at each receptor due to each possible source (this approach is preferable if the number of potential sources is smaller than the number of receptors), and (b) by “backward”-in-time or solution of the adjoint concentration transport equation, with the emission located at each receptor, calculating the adjoint concentration at each possible source location (this approach is preferable if the number of potential sources is higher than the number of receptors). In the present case, the emission source can be located anywhere in the computational domain, which means a theoretically infinite number of potential source locations. Therefore adjoint runs of the dispersion model are carried out in this paper to calculate SRFs. An Eulerian approach is selected for the formulation of the adjoint concentration equation, in which receptor points are considered as sources, and the flow field is reversed. The values of “adjoint” concentrations are calculated on a numerical (i.e., discrete) grid. Each of these values represents the sensitivity of the sensor to a source located at that grid node. So, the possible source locations considered in this approach are primarily the computational grid nodes. Additional potential source locations could be considered by computing (e.g., by interpolation) the values of the adjoint variable at points lying between the grid nodes.

Using the adjoint concentrations, Eq. (2a) is written as follows:

$$J = - \frac{\langle (q^s c^* - \langle q^s c^* \rangle) (c^o - \langle c^o \rangle) \rangle}{\sqrt{\langle (q^s c^* - \langle q^s c^* \rangle)^2 \rangle} \sqrt{\langle (c^o - \langle c^o \rangle)^2 \rangle}} \quad (2b)$$

In Eq. (2b), the modelled concentration c^c at a sensor location is expressed through the SRF, as the product of the adjoint concentration c^* - obtained by the solution of the adjoint dispersion equation for that sensor – with an arbitrary source rate q^s . The use of an arbitrary source rate is possible because J does not depend on q^s for a stationary source of a passive and non-reactive tracer. 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 (for a passive and non-reactive tracer), then the following relationship holds:

$$c_2^c = c_1^c \cdot \left(\frac{q_2^s}{q_1^s} \right) = c_1^c \cdot \alpha \quad (3)$$

From Eq. 3, 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 will in general 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-based cost function (Eq. 2b) enables the separate identification of source coordinates from the source rate, which will be calculated in the 2nd part. An arbitrary value of ‘first-guess’ source rate q_0^s is used for calculating the correlation coefficient since it does not influence the solution. However, the drawback of the present approach is that prior information cannot be used through regularization terms, as were the first three terms in the r.h.s of Eq. (1) – used by Kovalets et al. (2011). This is a drawback because inverse problems are frequently ill-posed, i.e., the solution can often be sensitive to input data, unless they are properly regularized by using prior information about source properties. This is a subject for future investigation.

In the second part of the methodology, the source rate is identified by minimizing – with respect to q^s – a quadratic cost function of the following generic form:

$$J = \sum_{n=1}^K (c_n^c - c_n^o)^2 \rightarrow \min \quad (4)$$

Having identified in the 1st part the grid node k^s , where the source is located, Eq. (4) is expressed for the node k^s through the use of SRFs, similarly to Eq. (2) (cf. formula (15) from Kovalets et al., 2011):

$$J_{k^s} = \sum_{n=1}^K (q^s c_{n,k^s}^* - c_n^o)^2 \xrightarrow{q^s} \min \quad (5)$$

where c_{n,k^s}^* is the adjoint variable for the node k^s .

The solution of problem (5) can be obtained analytically by equating to zero the derivative of J_{k^s} with respect to q^s :

$$\frac{\partial J_{k^s}}{\partial q^s} = \frac{\partial}{\partial q^s} \sum_{n=1}^K (q^s c_{n,k^s}^* - c_n^o)^2 = \sum_{n=1}^K 2c_{n,k^s}^* (q^s c_{n,k^s}^* - c_n^o) = 0 \quad (6)$$

The solution of (6) is:

$$q^s = \frac{\sum_{n=1}^K c_{n,k^s}^* c_n^o}{\sum_{n=1}^K (c_{n,k^s}^*)^2} \quad (7)$$

The source emission rate is calculated from Eq. (7). Note that positivity of c_{nk}^* (which is the solution of parabolic adjoint equation) guarantees positiveness of the estimated emission rate.

In brief the computational steps of the algorithm are the following:

1. Solution of the hydrodynamic problem for the specific geometry and boundary conditions: the wind flow field is resolved using the CFD code ADREA-HF. The 3-dimensional, steady-state fields of the 3 wind velocity components and the turbulence-related quantities (e.g.,

the turbulence kinetic energy and the turbulence kinetic energy dissipation rate in case of using a $k-\epsilon$ turbulence closure (Argyropoulos and Markatos, 2015) are calculated.

2. Calculation of Source Receptor Function (SRF) for each sensor:
 - a. Since concentration is calculated by the CFD model at the nodes of the computational grid, while the locations of concentration-measuring sensors cannot, in the general case, coincide with the nodes of the computational grid (this is especially true for a CFD code that uses a structured mesh of parallelepiped cells), model-concentrations need to be spatially interpolated to the sensors positions, for the latter to be used as source terms in the adjoint dispersion equations. Therefore, the coefficients of linear interpolation from the computational grid to each sensor's location are calculated.
 - b. The adjoint concentration equation is solved for each sensor using the hydrodynamic fields obtained in the 1st step (inversed) and with its source term defined by the linear interpolation operator from step 2a. The values of the resolved adjoint variable at each grid node obtained for a sensor represent the values of the SRF for that sensor if the source were located at the particular grid node.
 - c. The resulting gridded values of SRF for each sensor are stored.
3. Finding the minimum of the cost function: the cost function (2b) is evaluated at each grid node, i.e., considering that the source is located in the particular grid node. The node at which the smallest value of the cost function is reached (i.e. correlation is maximized) coincides with the source location.
4. Having calculated the SRFs in step 2 and identified the grid node of the source in step 3, the source rate is calculated by Eq. (7).

It is reminded that the above modifications (Eqs. 2 and 4 instead of Eq. 1) have been introduced in the algorithm in order to eliminate the 'overfitting' effect. It is expected to obtain better performance in the prediction of the source location and emission rate.

A point needed to be addressed is that practical application of the described method could be complicated by the fact that in reality even if release rate is stationary measured concentrations will not be stationary due to turbulence and influence of large eddies existing in atmospheric boundary layer. Therefore it is recommended to use time integrated measurements in source inversion procedure.

For further details on the theoretical background and the adjoint equation solved, the reader is referred to Kovalets et al. (2011). The source inversion algorithm is integrated within the ADREA-HF CFD model (Venetsanos et al., 2010) which was very extensively validated against laboratory and field experiments on atmospheric dispersion in urban atmospheric environment (e.g. Efthimiou et al., 2017a,b,c, Efthimiou et al., 2016, Efthimiou et al., 2015, Efthimiou et al., 2014; Efthimiou et al., 2011a,b, Efthimiou and Bartzis, 2011 and many others).

3. Results and Discussion

3.1. The MUST wind tunnel experiment

The methodology has been validated against data of the MUST wind tunnel experiment (Bezpalcova and Harms, 2005) which have been scaled up for the conditions of the corresponding field experiment (Yee and Biltoft, 2004). Hence the CFD ADREA-HF code (Efthimiou et al., 2016) has been setup for the simulation in the field scale and all the experimental and computational parameters below are given in the field scale.

In MUST experiment the obstacles were arranged in 12 rows, each consisting of 10 obstacles. The obstacles were nearly identical and had average length, width and height $12.2\text{ m} \times 2.42\text{ m} \times 2.54\text{ m}$ respectively. The contaminant's concentration has been measured by a 256-detectors array arranged along obstacle rows in the part of the domain covered by the plume (Fig. 1). All detectors were placed at the same height equal to 1.28 m. Non-zero concentrations have been measured by nearly all (244 out of 256) detectors. Here it is noted that only the measurement data that were available to the authors (through data base of COST Action 732) are taken into account.

The wind flow was characterized by neutral stratification, wind speed at the roof level $U_{ref} = 8\text{ m/s}$, (field scale) and wind direction -45° (Fig. 1) in the experimental coordinate system.

The contaminant originated from a point source located at the ground level (Fig. 1). The volume flow rate of gas at the source was: $\approx 3.3 \times 10^{-6}\text{ m}^3\text{s}^{-1}$.

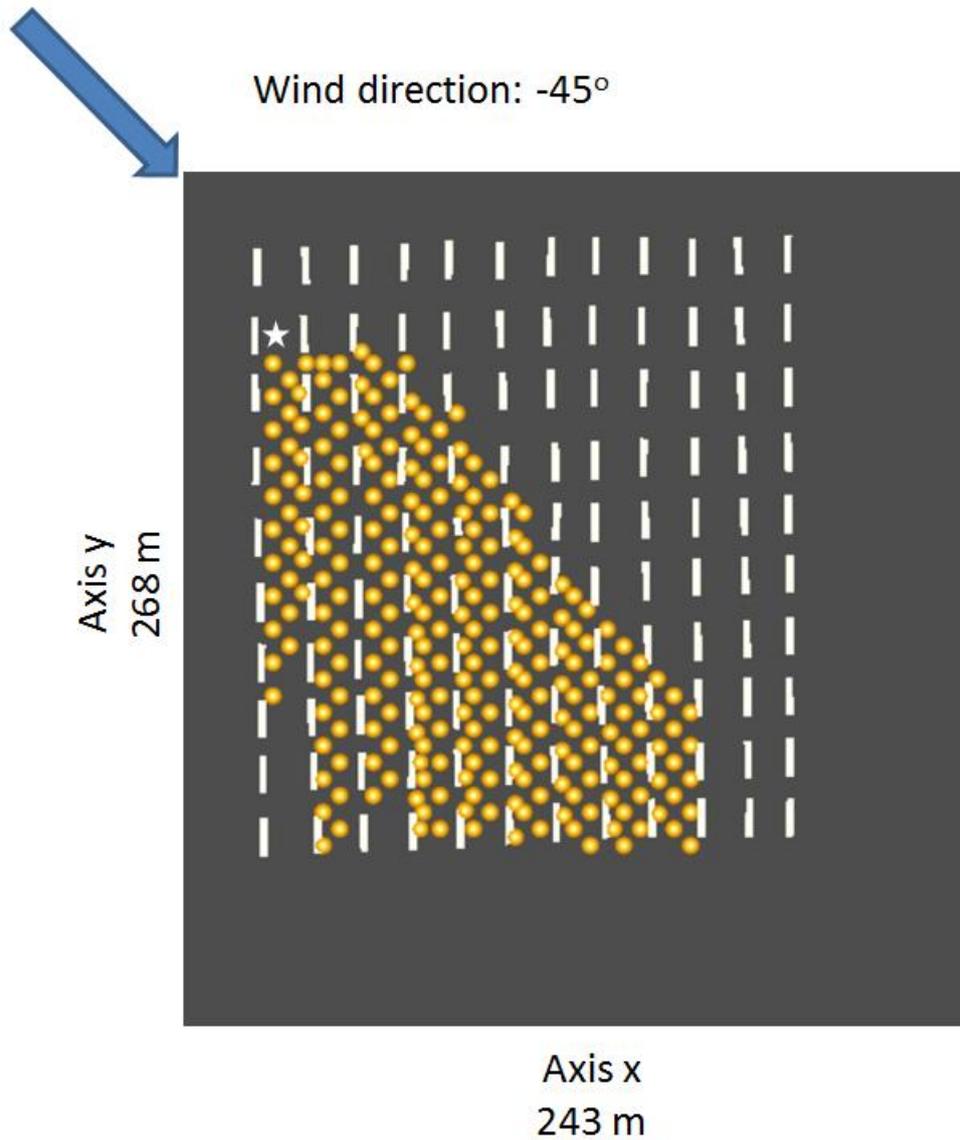


Fig. 1. The computational domain of the MUST case. The sensors are presented with yellow circles. The source is depicted with a star. Small offsets of the containers are observed (see Biltoft, 2001).

3.2. The wind field computational simulations

The wind flow problem was solved in a 3D rectangular computational domain with the x - and y -axes as presented in Fig. 1 and the z -axis in the vertical direction. The domain dimensions are presented in the first column of Table 1.

Table 1. Computational domain dimensions and details of the three numerical grids that have been tested

Domain dimensions $x/y/z$ (m)	Grid characterization	Total number of cells	Number of cells in each axis			Minimum/maximum cell sizes (m)		
			x	y	z	x	y	z

243/268/21	“Coarse”	209,950	85	95	26	2.5/7.5	2.5/7.5	0.2/2.06
	“Medium”	820,742	164	185	26	1.25/5.8	1.25/5.8	0.2/2.06
	“Fine”	2,940,742	312	353	26	0.625/5.3	0.625/5.3	0.2/2.06

The domain extends horizontally by 17.55 m (6.9 H , $H=2.54$ m being the buildings height) upwind of the first buildings and by 52.65 m (20.7 H) downwind of the last buildings. The vertical dimension of the domain is about 8.3 H . The above dimensions conform to the recommendations of COST Action 732 (Franke et al., 2007).

One of the crucial factors affecting the quality of the flow field solution is the computational mesh resolution. To ensure grid-independency of the obtained flow field solution, three spatial discretizations have been tested. Details are listed in Table 1. The horizontal resolution is increased going from the “coarse” to “fine” grid, while the vertical resolution is identical for the three grids. The grid refinement ratio for a 3D mesh is defined as the cubic root of the ratio between the numbers of grid elements in the two meshes. This parameter should be greater than 1.3 to allow the discretization error to be separated from the other sources of error (Celik, 2004). As can be seen from Table 1 the refinement ratios are $r_{21} = 1.57$ and $r_{32} = 1.53$, where subscript 1 indicates the coarse grid, 2 the medium grid and 3 the fine grid.

The boundary conditions for the hydrodynamic variables at each boundary plane or surface of the domain are presented in Table 2.

Table 2. Boundary conditions for the hydrodynamic variables (u, v, w : velocity components in the x -, y -, z -axis respectively, k : turbulence kinetic energy, ε : turbulence kinetic energy dissipation rate)

Plane	Boundary condition
-x	Inlet: fixed vertical profiles for $u, v, k, \varepsilon, w=0$
+x	Outlet: $\partial\varphi/\partial x = 0, \varphi=u, v, w, k, \varepsilon$
-y	Outlet: $\partial\varphi/\partial y = 0, \varphi=u, v, w, k, \varepsilon$
+y	Inlet: fixed vertical profiles for $u, v, k, \varepsilon, w=0$
Ground	Standard wall functions, roughness length = 0.0165m
Top	Fixed values for u, v, k, ε, w is calculated from the cell mass balance
Building walls	Standard wall functions, roughness length = 1×10^{-5} m

The fixed vertical profiles at the inlet planes (-x and +y) for u, v, k and ε have been constructed based on the experimental corresponding profiles, while for w a zero value has been set. The outlet boundary conditions at planes +x and -y are justified, as these planes are located at a sufficient distance (see above) downwind of the last buildings. For turbulence modelling, the standard k - ε model (Launder and Spalding, 1974) was used to be in agreement with Kovalets et al. (2011).

The problem was treated as a transient case. The unsteady Reynolds-Averaged Navier-Stokes equations for total mass, the three components of momentum, the turbulence kinetic energy and its dissipation rate were solved. At the end of the solution, steady-state values were produced. The total calculation time (1200s) was set about equal to the travel time from the source point to the outlet of the computational domain of a hypothetical particle moving with the lowest speed observed in the experiment. This time was sufficient to reach a steady-state situation. The time step was automatically adapted according to the desired Courant–Friedrichs–Lewy (CFL) number whose maximum value was set to 2 based on the authors’ experience for numerical stability (better convergence behaviour of the solver). Nevertheless, as an implicit scheme is adopted and a steady-state solution is sought for, the magnitude of the final time step does not play here an important role. A first order scheme was used for the time derivative.

For the calculation of pressure, the linear system that arises from the discretized equations was solved using the preconditioned BiCGstab method. A block-ILU[0] type of preconditioner was used (additive Swartz preconditioner, Saad, 2003).

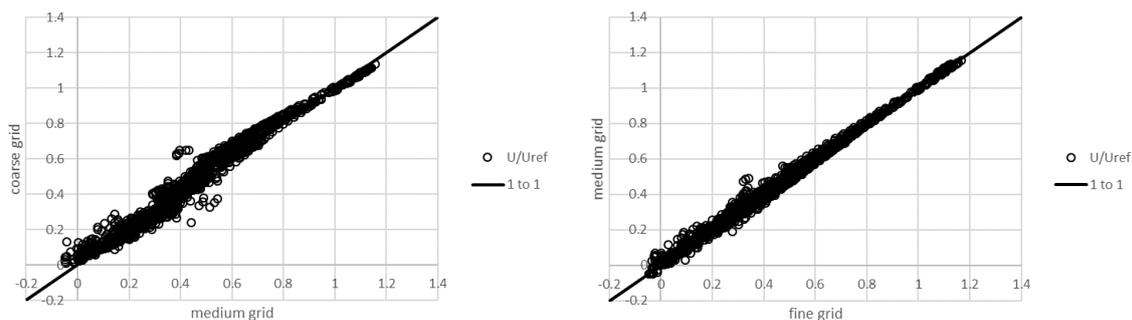
The 1st order upwind scheme has been used for approximation of convective terms. The second order central-difference scheme has been used for approximation of diffusion terms. It should be noticed that, since a grid independent numerical solution has been attained, the selection of the numerical schemes for the spatial derivatives does not have significant effects on the results.

The computations were run in parallel using four cores of a processor. ADREA-HF is parallelized for shared memory architectures using the OpenMP protocol.

3.3. Results of grid sensitivity study

The flow field was computed by running the CFD model on the three meshes with increasing spatial resolution, as shown in Table 1. The calculated velocity components were inter-compared to evaluate the grid independency of the solution. A qualitative evaluation was carried out using scatter plots. The measurement positions of the experiment have been selected as the set of points where the velocity components were compared. There were 1753, 1219 and 534 measurement locations for the u , v , and w wind velocity components respectively.

Scatter plots comparing the calculated velocity components between the three mesh resolutions (coarse to medium and medium to fine) are presented in Fig. 2. The scatter plots reveal a very close prediction between the medium and the fine meshes, indicating that grid independency is attained towards the fine grid resolution. Therefore, the fine grid was chosen for the following solution of the inverse problem.



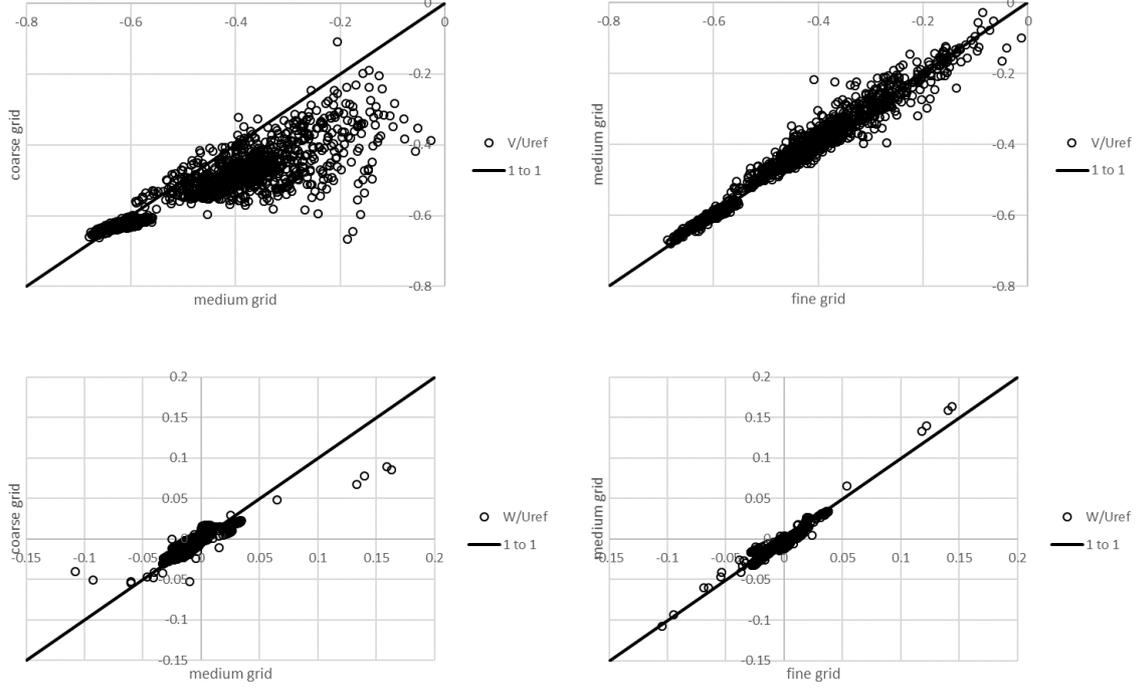


Fig. 2. Comparison of calculated wind velocity components u , v , w between the three mesh resolutions (coarse to medium and medium to fine).

3.4. Validation of the hydrodynamic problem solution

It is anticipated that the accuracy of the inverse source term estimation depends on the accuracy of the hydrodynamic (flow field) problem solution. The quality of the flow field simulation has been validated for the fine grid resolution (Table 1) according to the conclusions of the grid independency study. For the validation the calculated non-dimensionalised velocities U/U_{ref} , V/U_{ref} and W/U_{ref} (where U_{ref} is a reference velocity equal to 8 ms^{-1} for the experiment being simulated) have been compared with the corresponding experiment values.

The comparison has been performed using the Hit Rate (HR). This metric was first proposed by COST Action 732 (Schatzmann et al., 2010). In principle, the Hit Rate specifies the fraction of model results that differ within an allowed range D or R from the comparison data. D is the allowed range of relative error and R is the allowed range of absolute error. Although the broad quantitative application of the hit rate poses some questions, its use as a qualitative metric presents some advantages over other similar metrics (e.g. fractional bias, normalized mean square error) because of its simplicity and applicability to variables that can take both negative and positive values (Kakosimos and Assael, 2013). The formal definition of HR is the following:

$$HR = \frac{N}{n} = \frac{1}{n} \sum_{i=1}^n N_i \text{ with } N_i = \begin{cases} 1 & \text{if } \left| \frac{P_i - O_i}{O_i} \right| \leq D \text{ or } |P_i - O_i| \leq R \\ 0 & \text{else} \end{cases} \quad (8)$$

where n is the total number of compared couples of values, P_i is the model result and O_i is the observation (normalised velocities). In other words a Hit is added if one of the following conditions is fulfilled:

1. $|P_i - O_i|$ is smaller than the allowed absolute deviation R , or
2. $|(P_i - O_i)/O_i|$ is smaller than the allowed fractional deviation D

Based on COST Action 732 (Schatzmann et al., 2010) the proposed values of R for the MUST wind tunnel experiment and for each velocity component are: for U/U_{ref} $R = 0.008$, for V/U_{ref} and W/U_{ref} $R = 0.007$. According to the VDI guideline (VDI, 2005), that structured the application of HR metric, the allowed relative difference D equals 0.25.

The Hit Rates obtained for the calculated wind velocity components on the fine grid resolution are presented in Table 3, separately for four groups of sensors, according to their height above ground: a) sensors placed above twice the mean building height ($z > 2H$), b) sensors placed between the mean building height and twice the mean building height ($H < z \leq 2H$), c) sensors placed between half the mean building height and the mean building height ($0.5 H < z \leq H$), and d) sensors placed up to half the mean buildings height ($z \leq 0.5 H$). The mean building height H is 2.55m. As can be seen in Table 3, no V -component measuring sensors were placed during the experiment above $2H$. It is noticed that the model presents a very good performance above the buildings ($z > H$), where the Hit Rate has almost the ideal value ($=1.0$) for the velocities U/U_{ref} and V/U_{ref} . Below the building roofs the Hit Rate gradually decreases for all velocity components. U/U_{ref} presents in total better agreement with the experimental data than V/U_{ref} and W/U_{ref} . Very close to the ground the velocities V/U_{ref} and W/U_{ref} present the lowest Hit Rates.

Table 3. Hit Rates of calculated velocity components on the fine grid resolution. The sensors have been grouped according to their height above ground. The mean building height $H = 2.55\text{m}$.

Group of sensors	Hit rate		
	U/U_{ref}	V/U_{ref}	W/U_{ref}
$2H < z$	1.0	-	0.31
$H < z \leq 2H$	0.99	1.0	0.18
$0.5H < z \leq H$	0.75	0.43	0.10
$z \leq 0.5H$	0.48	0.14	0.13
Total	0.75	0.43	0.18

Besides the above quantitative metrics, scatter plots of non-dimensionalised calculated versus observed velocities were used as qualitative validation. The scatter plot of velocity components U/U_{ref} and V/U_{ref} for the fine grid is presented in Fig. 3. The same groups of sensors have been used as shown in Table 3. It can be observed that there is good agreement with the experiment above the average height of the roofs while below the average height of the roofs the model tends to underpredict the experimental values, while the scatter of the points increases. The underestimation becomes higher as the sensors approach the ground. Nevertheless the majority of the points lies within the factor-of-2 area, delineated by the dashed lines.

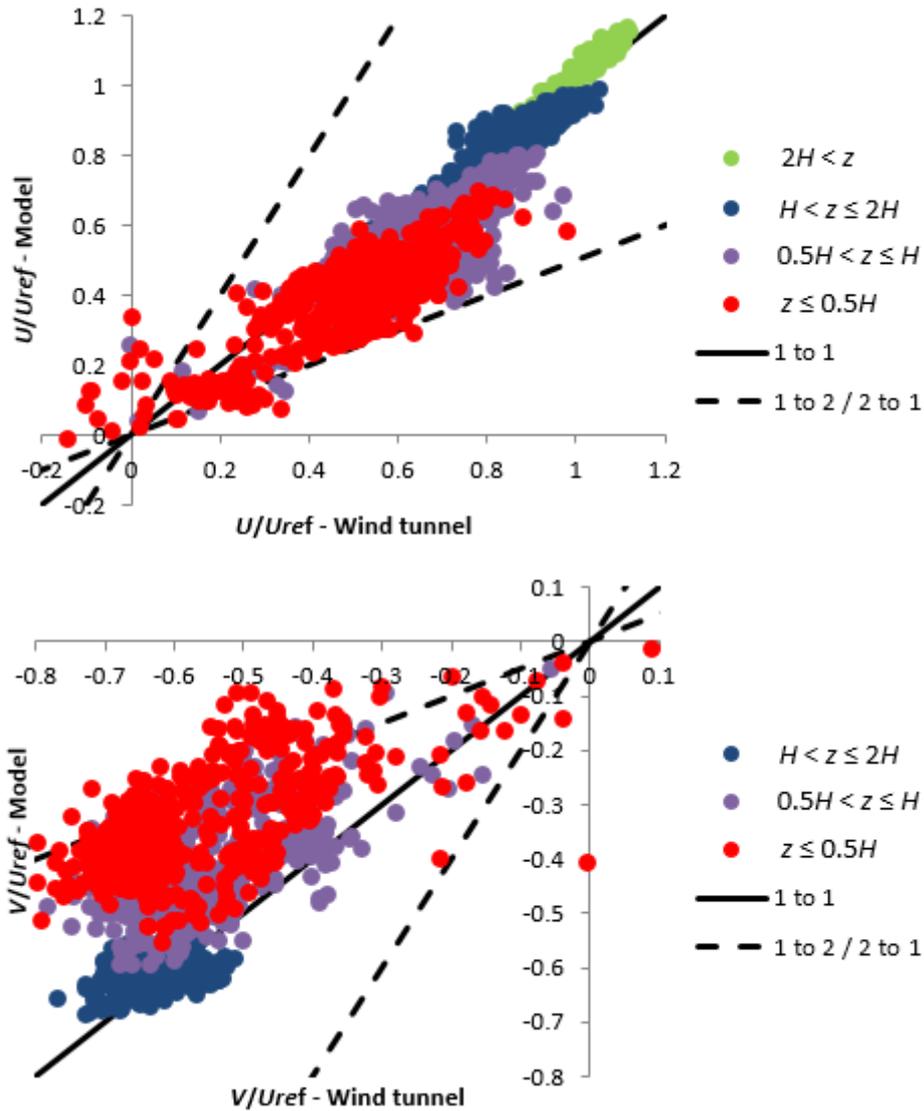


Figure 3. Scatter plot of calculated versus observed velocities U/U_{ref} and V/U_{ref} at experimental sensors locations; sensors are classified according to their height above ground; buildings height $H = 2.55\text{m}$

The observed discrepancies between calculated and experimental values of wind velocities inside the street canyons can be attributed to unsteady (low frequency) flow features that have also been reported in past relevant studies (Hertwig et al., 2012) and which cannot be simulated by RANS models as the one used here. Nevertheless, as can be seen in the following sections, these discrepancies do not have a major effect on the performance of the inverse problem.

3.5. The inverse source term estimation computations

Having established a grid-independent solution for the flow field, the adjoint concentration equation was solved on the fine grid resolution for each sensor as source term. The linear interpolation coefficients between the computational grid and the sensors locations are first calculated, as mentioned at the end of Section 2. The flow and turbulence related variables needed for the formulation of the adjoint equations are taken from the solution of the hydrodynamic problem. The

boundary conditions of the adjoint variable c^* for each boundary plane or surface of the domain are presented in Table 4.

Table 4. Boundary conditions for the adjoint variable c^*

Plane	Boundary condition
-x	$c^*=0$
+x	$c^*=0$
-y	$c^*=0$
+y	$c^*=0$
Ground	$\partial c^*/\partial z = 0$
Top	$\partial c^*/\partial z = 0$
Building walls	$\partial c^*/\partial s = 0, s=x, y, z$

The total calculation time was set equal to the one used for the hydrodynamic computations, i.e., 1200s, which was adequate for obtaining a steady-state solution. For the discretization of the convective term in the adjoint equation the upwind scheme was used. The time step was kept constant and equal to 1s.

3.6. Evaluation of the SRF

The calculated source-receptor functions are approximate, and therefore they need to be evaluated. For this purpose, the forward-in-time dispersion problem has been solved with the source located at the true location of the experiment and the true emission rate q_t^s . As a result, concentrations have been calculated at the sensors positions. These concentrations (denoted as c^F) have been compared with the concentrations resulting from the calculated SRFs, i.e, from the calculated value of the adjoint variable at the true source location for the corresponding sensors $c^{SRF} = q_t^s c^*$. As statistical index for this comparison, the relative Mean Absolute Error (MAE) $MAE = \frac{\langle |c^F - c^{SRF}| \rangle}{\langle c^F \rangle}$ has been used, where triangle brackets denote arithmetic averaging. The ideal value is MAE = 0. The value obtained here was MAE = 0.1 In order to clarify the significance of such error in SRF, the same statistical parameter has been calculated on the basis of comparison of results of the forward run with the measurements which gave MAE = 0.89. Therefore, it is clear that the error introduced by SRF is significantly less (by a factor of about 9) than the model error as compared to measurements, and it therefore could be considered as negligibly small.

In Figure 4 the scatter plot comparing concentrations calculated by the SRFs with concentrations obtained by forward computations at the sensors locations is presented. It is seen that all points lie very close to the 1:1 line.

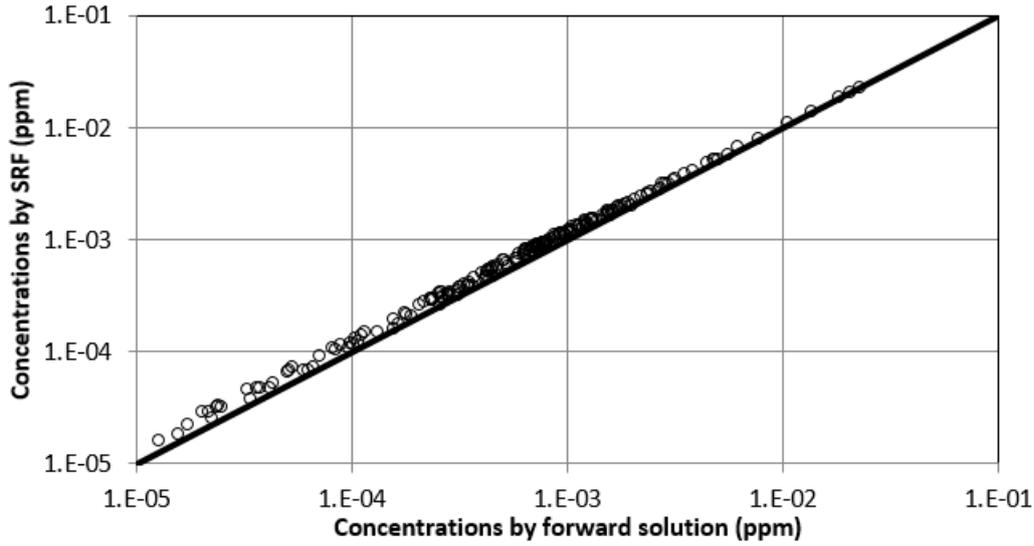


Figure 4. Scatter plot comparing concentrations calculated by SRF with concentrations obtained by forward computation at the sensors locations.

3.7. Source location and emission rate estimation - validation of results

Having calculated the SRFs, the next step is to calculate the source location and emission rate. In this paper, the methodology of Kovalets et al. (2011), denoted hereafter for economy reasons as “old” and realized through minimization of the single equation (1), has been compared with the “new” two-step method, realized through minimization of equation (2b) and solution of equation (7). For both methods the effect of the computational grid resolution on the results is examined. Also for both methods all available concentration sensors have been used.

The old method provides the option to use prior information (first guess) regarding the source location (x_0^s , y_0^s , z_0^s , σ_H , σ_V in equation (1)). If no prior information is used, then the first three terms of equation (1) (regularization terms) are neglected. In this paper both options of the old method have been tested (i.e., with and without prior information). If prior information was used, then the selected values were $x_0^s = 80$ m, $y_0^s = 255$ m, $z_0^s = 0$ m, $\sigma_H = 170$ m and $\sigma_V = 2.5$ m following Kovalets et al. (2011). It has been assumed that the measurement error σ_O would be negligible in comparison to the model error σ_M . Different values of the model error have been tested, to assess the effect of this parameter.

Regarding the new method, the correlation-based cost function (2b) has been calculated at all grid points and its minimum was found. The node with the smallest value of the cost function was the calculated source location. Then the source rate has been calculated by Eq. (7).

In order to quantify the error in locating the source, the horizontal $r_H = \sqrt{(x^s - x_t^s)^2 + (y^s - y_t^s)^2}$ and vertical $r_V = |z^s - z_t^s|$ distances of the estimated source location (x^s , y^s , z^s) from the true source location (x_t^s , y_t^s , z_t^s) = (95.74, 246.52, 0.0), have been used. The distances of the first-guess source

location are $r_{H,0} = 17.88$ m and $r_{V,0} = 0.0$ m. As already noted, the predicted source (x^s, y^s, z^s) is located at the centre of the computational cell with the minimum value of the cost function (2). Therefore the minimum achievable values of r_H and r_V are those of the distance of the centre of the computational cell where the source is actually located from the coordinates of the source. For the coarse grid $r_{H,\min} = 0.53$ m while for the fine grid $r_{H,\min} = 0.187$ m. For both grids $r_{V,\min} = 0.1$ m.

Concerning the source rate the relative source rate ratio $\delta q = \max\left[\left(q^s/q_t^s\right), \left(q_t^s/q^s\right)\right]$ has been calculated which is always greater than unity for both underestimated and overestimated source rates.

The distances between calculated and actual source location and the emission rate ratios that have been obtained with the new and old methods for the coarse and fine grids are presented in Table 5. For the coarse grid (upper part of Table 5), the new and old (with and without-prior information) methods present the same – very satisfactory – performance, as long as the model error in the old method is assumed to be less than or equal to 10^{-8} . As the assumed model error σ_M is increased in the old method (implying that less and less confidence is attributed to the model results) the solution converges to the first guess and the corresponding $r_{H,0}$ (see above).

On the fine grid (lower part of Table 5), the new method presents again a very satisfactory performance, especially regarding the horizontal distance between estimated and true source locations. On the other hand, the old method fails to predict both the location and the emission rate of the source if no prior information is given, or if prior information is given and the model error is assumed to be small (10^{-9}). These results indicate that the correlation-based cost function used in the new method, having less degrees of freedom, is more robust computationally than the quadratic cost function used in the old method. So on the fine grid, that introduces more degrees of freedom, i.e., more possible source locations, the old method necessitates the use of prior information. It is apparent however that in this case, the old method is sensitive to the value of the model error σ_M . When the model error σ_M increases by one order of magnitude (from 10^{-9} to 10^{-8}), the old method is successful in locating the source and estimating the source rate. But again when the model error is increased from 10^{-7} to 10^{-6} , the old method converges to the first guess. These results indicate that for the old method to perform satisfactorily, the model error should be guessed correctly within 1 order of magnitude. In practice however, it is a difficult task to assess the model error in advance (i.e. when the true solution of the inverse problem is unknown). Some methods which allow for solution of such task exist (e.g. Wahba and Wendelberger, 1980). Their application is outside of the scope of this work.

On the other hand it should be noted that, not including prior information is a drawback of the new method as well. Despite the new method giving good results on the two settings studied here, it cannot be concluded that it will be successful in other settings. A possible way to include prior information in the new method is to restrict the boundaries of the domain where the source can be found. This is a subject of a future study.

Table 5. Horizontal and vertical distances between estimated and true source locations, and relative source rate ratios obtained by the 2 methods on the coarse and fine grids

Results of the Coarse grid			
Method	r_H (m)	r_V (m)	δq (-)
New	2.99	0.1	1.42
Old (no prior information)	2.99	0.1	1.42
Old (prior information, $\sigma_M=10^{-9}$)	2.99	0.1	1.42
Old (prior information, $\sigma_M=10^{-8}$)	2.99	0.1	1.42
Old (prior information, $\sigma_M=10^{-7}$)	5.44	0.1	1.06
Old (prior information, $\sigma_M=10^{-6}$)	18.48	0.1	1.11
Old (prior information, $\sigma_M=10^{-5}$)	18.48	0.1	1.11
Results of the Fine grid			
Method	r_H (m)	r_V (m)	δq (-)
New	1.20	0.1	1.69
Old (no prior information)	235.06	14.65	1.83×10^{16}
Old (prior information, $\sigma_M=10^{-9}$)	235.06	14.65	1.83×10^{16}
Old (prior information, $\sigma_M=10^{-8}$)	0.20	0.1	1.62
Old (prior information, $\sigma_M=10^{-7}$)	2.99	0.1	1.50
Old (prior information, $\sigma_M=10^{-6}$)	17.58	0.1	1.15
Old (prior information, $\sigma_M=10^{-5}$)	17.27	0.1	1.18

3.8. Sensitivity analysis for the number of sensors

The performance of the source estimation method (regarding both source location and emission rate) depends on the number and placement of sensors that provide the concentration measurements to be assimilated. This is an important aspect, in particular for the practical applicability of the method in urban settings. Therefore, in this chapter the following question will be answered: What is the least number of sensors required to obtain a satisfactory solution – according to specified criteria – at a given probability level?

Since the number of measurement points in the simulated experiment is quite large ($N_o = 244$) this gives the opportunity to calculate the above mentioned dependence on the number and distribution of sensors. For this purpose, a given number of sensors $K < N_o$ is fixed (several different values of K are examined) and the source term is estimated for an increasing number of network configurations consisting of K sensors randomly sampled out of the N_o available sensors. The total number of such configurations is equal to binomial coefficient $C_K^{N_o}$. The solution is considered as 'good' if it satisfies the following threshold criteria: distance deviation $r_H \leq 15$ m, $r_V \leq 2.5$ m and emission rate ratio $\delta q \leq 4$ following Kovalets et al. (2011). Generally it is very difficult to define what good solution of source inversion problem is. The reason for choosing such criteria for δq is explained in Kovalets et al. (2011) by referring to other studies (Platt and DeRiggi, 2010) in which the reported δq reached a value up to 10. The criteria for r_H and r_V were selected on the basis of typical dimensions of buildings. However there are also other justifications for such criteria. First of all typical length scale of the territory covered by sensors is about 200 m. Hence we consider solution to be 'good' if the achieved error in position of source is by one order of magnitude less than the original uncertainty in source position. On the other hand if we divide the total simulation area by the number of measurements we obtain that there is approximately 1 sensor/16 m². Therefore 15 m is roughly the length scale of

'good' solution because better resolution is not likely to be achieved with the given spatial density of sensors.

The total number of possible network configurations is extremely large for $K > 4$, making it impossible to examine all of them. Figure 5 presents the number of possible combinations for different values of K sensors selected out of 244.

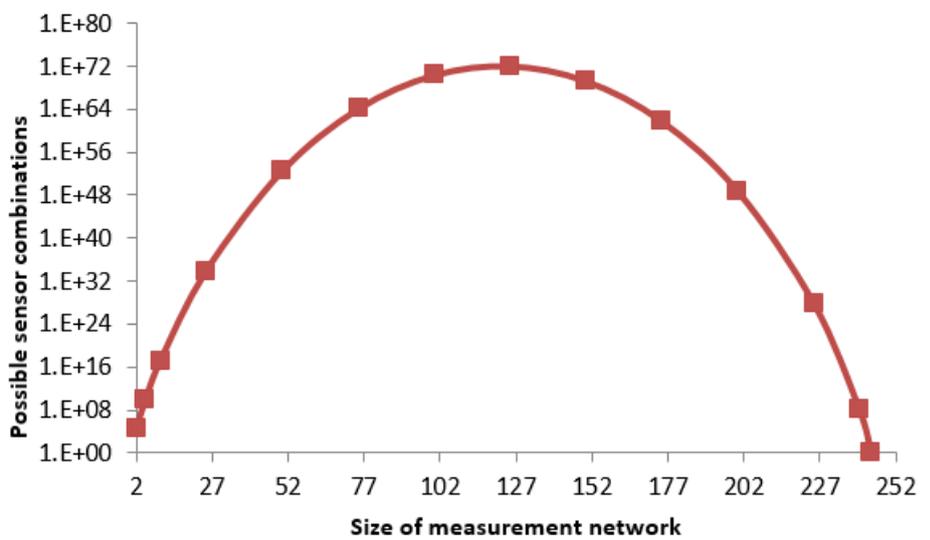


Figure 5. Number of possible network configurations for different sizes of networks selected out of 244 sensors.

Therefore, a random sampling algorithm is used (which is described in details in Appendix A of Kovalets et al., 2011) to generate repeatedly measurement networks of a given size K , out of the N_0 stations and 2000 measurement networks have been examined for each K . The inverse problem is solved for each of these generated networks. The ratio of the number N_{good} of networks giving 'good' results according to the above criteria to the total number N_{tot} of so far processed networks, is found to converge quickly to a stationary value which is the probability of achieving good solution for the given size K of measurement network.

Fig. 6, shows how the ratio N_{good}/N_{tot} converges to asymptotic values for $K = 150, 100, 50$ and 25 ($N_0 = 244$) as the number of processed measurement configurations (N_{tot}) increases.

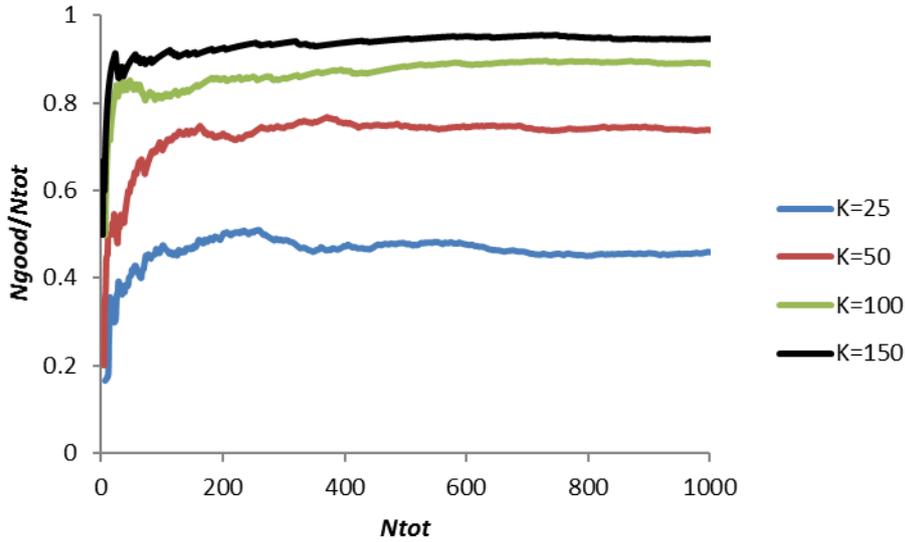


Figure 6. Convergence of the relative number of measurement networks giving ‘good’ solutions of the inverse problem as the total number of examined network configurations increases, for measurement networks of sizes: $K = 25, 50, 100, 150$.

Fig. 7 shows the probability of achieving good solution of the inverse problem – obtained through the use of the random sampling algorithm – , i.e., the asymptotic values of Figure 5, as function of the size of the measurement network, varying from $K = 25$ to $K = N_o = 244$. It can be seen that the probability of achieving good solution monotonically increases with increasing number of sensors used for the solution of the inverse problem. The probability of obtaining good solution equal to 90% or higher is achieved with measurement networks consisting of 114 sensors or more. Note that with the old method as reported by Kovalets et al. (2011) achieving result of the same quality required at least 150 sensors. In practice number of stationary monitoring stations is typically much less than the required value. Lack of stationary stations could be complemented with mobile measurements which require special algorithms for their assimilation (Hutchinson et al., 2017).

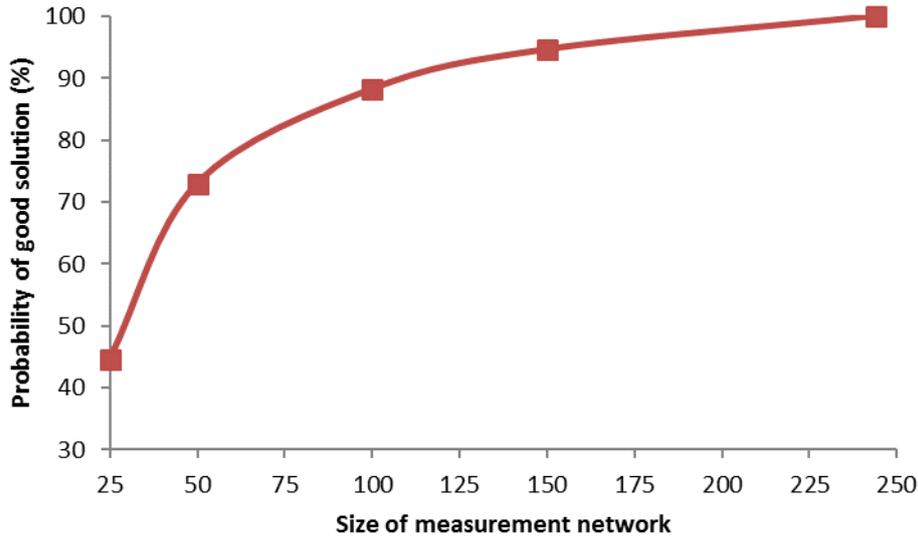
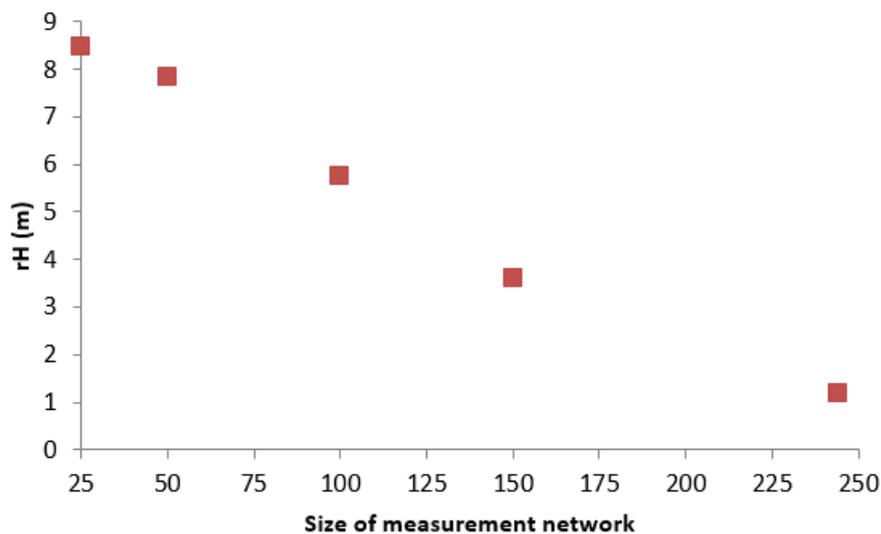


Fig. 7. Dependence of probability of achieving good solution of inverse source estimation problem on the number of measurement sensors used.

Figure 8 presents the average errors in estimation of source characteristics obtained over the 1000 networks of sensors examined for each network size of 25, 50, 100 and 150 sensors. In the plots, the errors obtained with the full network of 244 are also included (values of “fine grid” in Table 5). The average error in horizontal distance (Figure 8a) decreases monotonically with increasing size of measurement network. The average error in vertical distance (Figure 8b) remains approximately constant for measurement networks consisting of 25 to 100 sensors and decreases for larger networks. The average error in emission rate (Figure 8c) decreases with increasing size of measurement network, but with a small rate.



(a)

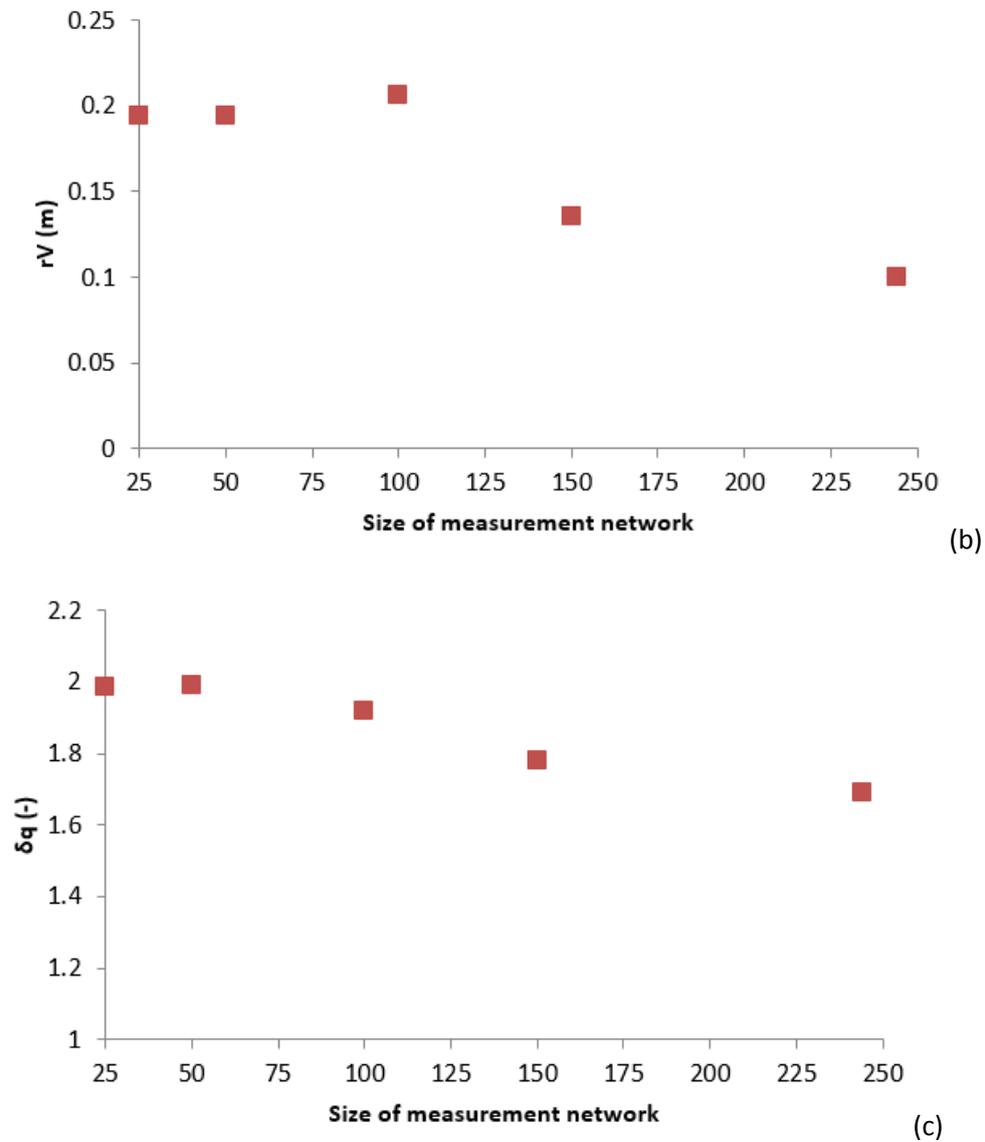


Figure 8: Average errors obtained for the different sizes of sensor networks that have been examined: (a) error in horizontal distance; (b) error in vertical distance; (c) error in emission rate

4. Conclusions

Substantial modifications to the inverse source term estimation method of Kovalets et al. (2011) have been introduced and tested in this paper. The modified approach consists of two-steps, separating the estimation of the location and emission rate of the source:

1. The source coordinates are analysed using a correlation function of measured and calculated concentrations.
2. The source emission rate is calculated by minimizing a quadratic cost function.

Our motivation for introducing the modifications was that in some cases of inverse modelling there is a typical problem called ‘overfitting’ effect according to which, the calculation errors which are introduced by a wrong source location estimation are compensated by error in estimated source rate. Thus, the resulting quadratic cost function reaches minimum for the wrong combined solution (source location and source rate).

The validation of the new algorithm was performed by estimating the source location and emission rate of the MUST wind tunnel experiment, regarding atmospheric dispersion of a passive tracer from a continuous point source among buildings of urban-like environments. The computational grid resolution was examined first in order to achieve a grid independent solution of the flow field. The old and new / modified algorithms for inverse source term estimation have been inter-compared on both coarse and fine computational grids. The new method presented a more robust behaviour, and has resulted in a satisfactory determination of the source location (in the vertical and horizontal directions) and emission rate on both grids. The performance of the old method proved to be very sensitive to the selection of the value of the model error that enters in the cost function, giving from very poor to very good results. Therefore a correct assessment of the model error, within 1 order of magnitude, is important for the old method to be successful, especially for fine spatial grid resolutions with a large number of degrees of freedom in the quadratic cost function.

However, not including prior information can be a drawback of the new method as well, in different settings than those studied in this paper. A possible way to include prior information in the new method would be to restrict the boundaries of the domain where the source could be located. This will be subject of a future study.

It is noted that the above results have been achieved when all available measurements were used to solve the inverse problem. However, the performance of the method depends also on the number and the placement of sensors that provide the concentration data to be assimilated. Of significant interest for the practical application of the method in urban settings is the number of concentration sensors required to obtain a “satisfactory” determination of the source. In order to assess the dependence of the method’s performance on the size of the measurement network, calculations have been performed for the MUST experiment in which different networks, consisting of a given number of randomly selected sensors out of all available sensors, were used for source estimation. A level of 90% or higher probability of achieving good solution was reached for measurement networks consisting of at least 114 sensors.

Future work includes further validation of the present methodology using other experiments with more complex geometric configurations (e.g., the experiments used by Efthimiou et al., 2016) and the extension of the method for time-varying releases (e.g., puff releases).

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

Argyropoulos C.D., Markatos N.C., 2015. Recent advances on the numerical modelling of turbulent flows. *Applied Mathematical Modelling*, 39, 2, 693-732.

Bady M., Kato S. Huang H., 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.

Ben Salem N., Salizzoni P., Soulhac L., 2016. Estimating accidental pollutant releases in the built environment from turbulent concentration signals, *Atmospheric Environment*, doi: 10.1016/j.atmosenv.2016.10.050.

Bezpalcova K., Harms F., 2005. EWTL Data Report/Part I: Summarized Test Description Mock Urban Setting Test, Technical Report, Environmental Wind Tunnel Lab., Center for Marine and Atmospheric Research, University of Hamburg.

Biltoft C.A., 2001. Customer report for mock urban setting test, DPG document no, WDTC-FR-01-121. West Desert Test Center, US Army Dugway Proving Ground, Dugway.

Celik I.B., 2014. Procedure for estimation and reporting of discretization error in CFD applications. In: Statement on the control of numerical accuracy of the journal of fluids engineering (editorial policy). ASME Fluids Engineering Division.

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

Dhall J.M., Lewis S., Lakshmivarahan, S. D., 2006. *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge University Press, 655 p.

Efthimiou, G.C., Andronopoulos, S., Bartzis, J.G., 2017a. Prediction of dosage-based parameters from the puff dispersion of airborne materials in urban environments using the CFD-RANS methodology. *Meteorology and Atmospheric Physics*, <https://doi.org/10.1007/s00703-017-0506-0>

Efthimiou G.C., Andronopoulos S., Tavares R., Bartzis J.G. 2017b. CFD-RANS prediction of the dispersion of a hazardous airborne material released during a real accident in an industrial environment, *Journal of Loss Prevention in the Process Industries*, 46, 23-36.

Efthimiou G.C., Andronopoulos S., Bartzis J.G., Berbekar E., Harms F., Leitl B. 2017c. CFD-RANS prediction of individual exposure from continuous release of hazardous airborne materials in complex urban environments, *Journal of Turbulence*, 18, 2, 115-137.

George C. Efthimiou, Eva Berbekar, Frank Harms, John G. Bartzis, Bernd Leitl, 2015. Prediction of the peak concentration and concentration distribution of a continuous point source release in a semi-idealized urban canopy using CFD-RANS modelling, *Atmospheric Environment*, 100, 48-56.

George C. Efthimiou, John G. Bartzis, 2014. Atmospheric dispersion and individual exposure of hazardous materials. Validation and intercomparison studies, *Int. J. Environment and Pollution*, 55, 76-85.

- G.C. Efthimiou, J. G. Bartzis, 2011. Atmospheric dispersion and individual exposure of hazardous materials. *Journal of Hazardous Materials*, 188, 375-383.
- G.C. Efthimiou, J.G. Bartzis, N. Koutsourakis, 2011a. Modelling concentration fluctuations and individual exposure in complex urban environments. *Journal of Wind Engineering and Industrial Aerodynamics*, 99, 349-356.
- G. Efthimiou, J.G. Bartzis, S. Andronopoulos, A. Sfetsos, 2011b. Air dispersion modelling for individual exposure studies. *International Journal of Environment and Pollution*, Vol. 47, pp. 302 - 316.
- Efthimiou G.C., Andronopoulos S., Tolia I., Venetsanos A., 2016. Prediction of the upper tail of concentration probability distributions of a continuous point source release in urban environments. *Environ Fluid Mech*, 16, 5, 899–921, DOI 10.1007/s10652-016-9455-2.
- Franke J., Hellsten A., Schlünzen H., Carissimo B. (Editors) 2007. Best practice guideline for the CFD simulation of flows in the urban environment. COST Action 732 Quality assurance and improvement of microscale meteorological models, ISBN: 3-00-018312-4, University of Hamburg.
- Hertwig D., Efthimiou G.C., Bartzis J.G., Leitl B., 2012. CFD-RANS model validation of turbulent flow in a semi-idealized urban canopy, *Journal of Wind Engineering and Industrial Aerodynamics*, 111, 61–72.
- Hutchinson M., Oh H., Chen W.-H., 2017. A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors, *Information Fusion*, 36, 130-148, doi: 10.1016/j.inffus.2016.11.010.
- Kakosimos K.E., Assael M.J., 2013. Application of Detached Eddy Simulation to neighbourhood scale gases atmospheric dispersion modelling. *Journal of Hazardous materials*, 261, 653-668.
- Keats A., Yee E. Lien F.-S., 2007. Bayesian inference for source determination with applications to a complex urban environment. *Atmospheric Environment*, 41, 465–479.
- Koracin D., Vellore R., Lowenthal D.H., Watson J.G., Koracin J., McCord T., DuBois D.W., Chen L.-W.A., Kumar N., Knipping E.M., Wheeler N.J.M., Craig K., Reid S., 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., Andronopoulos S., Venetsanos A., Bartzis J.G., 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.V., Tsiouri V., Andronopoulos S., Bartzis J.G., 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.
- Kumar P., Feiz A.-A., Singh S.K., Ngae P., Turbelin G., 2015. Reconstruction of an atmospheric tracer source in an urban-like environment. *Journal of Geophysical Research-Atmospheres*, 120, 24, 12589-12604.

- Kumar P., Singh S.K., Feiz A.-A., Ngae P., 2016. An urban scale inverse modelling for retrieving unknown elevated emissions with building-resolving simulations. *Atmospheric Environment*, 140, 135-146.
- Launder B.E., Spalding D.B., 1974. The numerical computation of turbulent flows. *Comp. Meth. Appl. Mech. Eng.*, 3, 269–289.
- Libre J.-M., Guérin S., Castellari A., Tripathi A., Leguellec M., Mailliard T., Souprayen C., 2012. Source determination in congested environment through Bayesian inference. *Int. J. Environment and Pollution*, 48, 174–184.
- Marchuk G., 1982. *Mathematical Modelling in the Environmental Problems*, Nauka, Moscow.
- Marchuk G., 1996. *Adjoint Equations and Analysis of Complex Systems*, Kluwer Academic Publishers, Dordrecht, Netherlands.
- Matsuo T., Kondo A., Shimadera H., Kyuno T., Inoue Y., 2015. Estimation of indoor contamination source location by using variational continuous assimilation method. *Build. Simul.*, 8, 443–452.
- Mons V., Margheri L., Chassaing J.-C., Sagaut P., 2017. Data assimilation-based reconstruction of urban pollutant release characteristics. *Journal of Wind Engineering and Industrial Aerodynamics*, 169, 232-250.
- N. Platt, D. DeRiggi, Comparative investigation of source term estimation algorithms using fusion field trial 2007 data—linear regression analysis, in: *Proc. of the 13th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, Paris, France, 1–4 June, 2010, pp. 901–905, Full presentation available at <http://www.harmo.org>.
- Pudykiewicz J., 1998. Application of adjoint tracer transport equations for evaluating source parameters. *Atmos. Environ.*, 32, 3039–3050.
- Saad Y., 2003. *Iterative methods for sparse linear systems*. 2nd edn. SIAM, Philadelphia.
- Schatzmann M., Olesen H., Franke J., (Editors) 2010. *Model evaluation case studies: Approaches and results*. COST Action 732 *Quality assurance and improvement of microscale meteorological models*, ISBN: 3-00-018312-4. University of Hamburg.
- Sharan M., Singh S.K., Issartel J.P., 2012. Least square data assimilation for identification of the point source emissions. *Pure Appl. Geophys.*, 169, 483–497.
- Singh S.K., Sharan M., Issartel J.P., 2013. Inverse modelling for identification of multiple-point releases from atmospheric concentration measurements. *Boundary-Layer Meteorol*, 146, 277–295.
- Tsiouri V., Kovalets I., Kakosimos K.E., Andronopoulos S., Bartzis J.G., 2014. Evaluation of advanced emergency response methodologies to estimate the unknown source characteristics of the hazardous material within urban environments. 16th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 8-11 September 2014, Varna, Bulgaria.

VDI Guideline 3783/9, 2005. Environmental meteorology—Prognostic microscale wind field models—Evaluation for flow around buildings and obstacles. Beuth Verlag, Berlin.

Venetsanos, A., Papanikolaou, E., Bartzis, J., 2010. The ADREA-HF CFD code for consequence assessment of hydrogen applications. *Int. J. Hydrogen Energy*. 35, 3908–3918.

Wahba G., Wendelberger J. Some new mathematical methods for variational objective analyses using splines and cross-validation. *Monthly Weather Review* 108 (1980) 1122-1143.

Xue F., Li X., Zhang W., 2017. Bayesian identification of a single tracer source in an urban-like environment using a deterministic approach. *Atmospheric Environment*, doi: 10.1016/j.atmosenv.2017.05.046.

Yee E., Biltoft C., 2004. Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles, *Bound.-Layer Meteorol.*, 111, 363–415.