Elsevier Editorial System(tm) for Mathematics and Computers in Simulation
Manuscript Draft

Manuscript Number: MATCOM-D-10-00505R2

Title: Identification of strength and location of stationary point source of atmospheric pollutant in urban conditions using computational fluid dynamics model

Article Type: Research Paper

Keywords: Inverse problem, CFD model, data assimilation, urban atmospheric dispersion, hazardous pollutant

Corresponding Author: Dr Ivan Vasilievich Kovalets, Ph.D.

Corresponding Author's Institution: Institute of Mathematical Machines and Systems Problems NAS of Ukraine

First Author: Ivan V Kovalets, Ph.D

Order of Authors: Ivan V Kovalets, Ph.D; Spyros Andronopoulos, Ph.D.; Alexander G Venetsanos, Ph.D.; John G Bartzis, Prof.

Abstract: In this paper a method is presented which allows for estimation of the location and rate of an unknown point stationary source of passive atmospheric pollutant in a complex urban geometry. The variational formulation is used in which the cost function characterizing difference of the calculated and measured concentrations is minimized with respect to source coordinates and rate. The minimization problem is solved by a direct algorithm using a source-receptor function which is calculated by solving adjoint equation. The algorithm has been implemented in ADREA-HF computational fluid dynamics code and has been applied in complex urban geometry. Validation of the algorithm has been performed by simulation of a wind tunnel experiment on atmospheric dispersion among an array of rectangular obstacles. Good results of source parameters estimation have been achieved. The measuring sensors to be used in source estimation had been randomly selected out of 244 available sensors by a specially designed random sampling algorithm which allowed for estimation of the probability of `good' source term estimation with the given number of sensors. In case of `perfect model' (when synthetic measurements were used for source estimation) good results were achieved with 90% or higher probability for arbitrary measurement networks consisting of 15 or more sensors even though prior estimation of source location was not used in this case. Several tests had been performed with the use of real measurements which differed by prior estimation of source location. In all cases 90% or higher probability of obtaining good solution was reached only for measurement networks consisting of 150 or more sensors. Hence further improvement of the source estimation algorithm can be achieved first of all by improving the performance of forward model.

Cover Letter

Dear Professor R. Beauwens,

I would like to submit the revised article entitled Identification of strength and location of stationary point source of atmospheric pollutant in urban conditions using computational fluid dynamics model" by I.V, Kovalets, S. Andronopoulos, A.G. Venetsanos and J.G. Bartzis (Ref. No.: MATCOM-D-10-00505R1).

The authors are most grateful to Editor and referees for their constructive criticism. We have considered their remarks, and the paper has been revised accordingly.

The responses to referee's comments are attached.

Sincerely yours,

Ivan Kovalets

Identification of strength and location of stationary point source of atmospheric pollutant in urban conditions using computational fluid dynamics model

Ivan V. Kovalets[a,*], Spyros Andronopoulos[b], Alexander G. Venetsanos[b], John G. Bartzis[c]

[a] *Institute of Mathematical Machines and Systems Problems NAS of Ukraine, prosp. Glushkova 42, 03187, Kiev, Ukraine*
[b] *Environmental Research Laboratory, Institute of Nuclear Technology and Radiation Protection, NCSR 'DEMOKRITOS', 15310, Ag. Paraskevi, Attikis, Greece*
[c] *Department of Mechanical Engineering, University of Western Macedonia, Bakola and Sialvera str., 50100, Kozani, Greece*

Abstract

In this paper a method is presented which allows for estimation of the location and rate of an unknown point stationary source of passive atmospheric pollutant in a complex urban geometry. The variational formulation is used in which the cost function characterizing difference of the calculated and measured concentrations is minimized with respect to source coordinates and rate. The minimization problem is solved by a direct algorithm using a source-receptor function which is calculated by solving adjoint equation. The algorithm has been implemented in ADREA-HF computational fluid dynamics code and has been applied in complex urban geometry. Validation of the algorithm has been performed by simulation of a wind tunnel experiment on atmospheric dispersion among an array of rectangular obstacles. Good results of source parameters estimation have been achieved. The measuring sensors to be used in source estimation had been randomly selected out of 244 available sensors by a specially designed random sampling algorithm which allowed for estimation of the probability of 'good' source term estimation with the given number of sensors. In case of 'perfect model' (when synthetic measurements were used for source estimation) good results were achieved with 90% or higher probability for arbitrary measurement networks consisting of 15 or more sensors even though prior estimation of source location was not used in this case. Several tests had been performed with the use of real measurements which differed by prior estimation of source location. In all cases 90% or higher probability of obtaining good solution was reached only for measurement networks consisting of 150 or more sensors. Hence further improvement of the source estimation algorithm can be achieved first of all by improving the performance of forward model.

Keywords: inverse problem, CFD model, data assimilation, urban atmospheric dispersion, hazardous pollutant

1. Introduction

The problem of identification of unknown characteristics of atmospheric pollutant's source emerging following some release is a particular case of inverse problem of atmospheric dispersion. Such kind of inverse problems are to be solved in variety of application areas. For instance, verification of the Comprehensive Nuclear Test Ban Treaty requires solution of global to regional scale inverse atmospheric dispersion problem with instantaneous source of unknown location and released mass [21]; another example is the source term

☆Abbreviations: SRF - Source Receptor Function, NMSE - Normalized Mean Squared Error, FB - Fractional Bias, MAE - Mean Absolute Error.

*Corresponding author. Tel.: +38-044-5261438, fax: +38-044-5266187
Email addresses: ik@env.com.ua (Ivan V. Kovalets), sandron@ipta.demokritos.gr (Spyros Andronopoulos), venets@ipta.demokritos.gr (Alexander G. Venetsanos), bartzis@uowm.gr (John G. Bartzis)

estimations following accidental releases from nuclear power plants which require the solution of regional to local-scale atmospheric dispersion problems with known source location but highly varying in time source rate [5],[12],[23]. Other examples of inverse atmospheric dispersion problems are reviewed in [8]. Recently the increased urbanization led to increased vulnerability of population living in urban areas to hazardous airborne pollutants. Therefore solution of inverse urban-scale atmospheric dispersion problem for identification of source with unknown location and source rate is of great importance in case of deliberate releases [22] as well as in many cases of accidental releases of hazardous matter (such as the case described in [7]).

In general statement of the inverse problem joint probability distributions of the unknown model parameters, measurements and prior estimations are considered [24]. In variational formulation of the inverse problem the maximum of the posterior probability distribution of the model parameters is found by minimizing the specified cost function [13]. Most commonly the variational statement of the inverse problem leads to linear (for instance, when only source rate is adjusted) or nonlinear regression problem. However there are two main difficulties in using the variational approach. The first is that the inverse problems are usually ill-posed, i.e. the solution of the inverse problem could be not stable with respect to the small changes in the input data or the minimum of the cost function could be non-unique. The second difficulty is that the cost function could be non-convex and thus even if unique well defined minimum exist it could be difficult to find it. That circumstance could be particulary important in context of inverse atmospheric dispersion problems (e.g., [12]). One possible way to overcome those difficulties is to use prior information in the form of regularization terms in the cost function. This method requires not only prior (first guess) estimations of the unknown parameters but also approximate values of the errors of the prior estimations which enter as weighting factors in regularization terms of the cost function. If those errors are overestimated the effect of regularization could be small. If the errors are underestimated regularization could lead to erroneous results.

Additional difficulties in source inversion of atmospheric pollutant arise in the case of urban-scale atmospheric dispersion problem. The most important difficulty is that the atmospheric dispersion in urban surface layer is governed by highly complex and turbulent meteorological conditions which are the consequences of interactions of the air flow with buildings [9]. Numerical solution of fluid dynamics equations resolving the necessary scales of atmospheric motions with computational fluid dynamics (CFD) models is highly challenging and computationally expensive task. Recently a few studies have been performed which combined advantages of CFD models describing urban meteorology with different source estimation techniques. In [6] the inverse atmospheric dispersion problem has been solved following the general approach of [24], where conditional posterior probability distribution of source parameters with the given particular model and available observations is evaluated. In that work a 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 [10] a similar probabilistic formulation of inverse problem has been adopted and posterior distribution function has been evaluated with Monte-Carlo sampling, however instead of using CFD model as a 'black box' tool, a 'source-receptor' function has been constructed using the adjoint transport equation and the approach previously developed in [14],[15].

While estimation of posterior density distribution is the most general and comprehensive solution of inverse problem, it still requires significant computational resources; while in practical applications especially in real-time it could be sufficient to use variational formulation to evaluate the location of maximum of that distribution. Under conditions of Gaussian probability distributions of model errors, measurement errors and errors of prior estimations the location of maximum of the posterior probability distribution coincides with the solution of special minimization problem, which is described below. Hence the objective of the present work is to develop an effective variational algorithm of source inversion which could be used in combination with urban-scale CFD model, and to evaluate it's performance against measurements obtained in wind tunnel experiment [4].

The structure of the paper is the following. In chapter 2 the inverse problem is presented as continuous problem of cost function minimization. In chapter 3 the source-receptor function is introduced. With the help of SRF the continuous minimization problem is transformed to discrete minimization problem and the direct minimization algorithm is presented. The details of the algorithm implementation in the ADREA-HF CFD code are also discussed. The results of solution of the inverse problem are presented in chapter 4. The presented results were obtained using the algorithm of random selection of measurement network, which is

described in Annex. Finally the conclusions are formulated in the last section.

2. Problem formulation

We consider the advection-diffusion equation of a passive conservative contaminant originating from a point source:

$$\frac{\partial c}{\partial t} + u_i \frac{\partial c}{\partial x_i} - \frac{\partial}{\partial x_i} D \frac{\partial c}{\partial x_i} = q^s \delta_\varepsilon (x - x^s) \delta_\varepsilon (y - y^s) \delta_\varepsilon (z - z^s) = f^s (x, y, z), \tag{1}$$

where c is Reynolds averaged concentration, D is coefficient of turbulent diffusion, u_i are Reynolds-averaged and fixed in time velocity components in Cartesian coordinate system with coordinates $x_i = (x, y, z)$, $i = 1, 2, 3$. Right-hand side f^s in (1) describes stationary point source located at (x^s, y^s, z^s) and having source rate q^s[kg/s]; function of scalar argument $\delta_\varepsilon(.)$ in (1) is stepwise function:

$$\delta_\varepsilon (t) = \begin{cases} 1/\varepsilon, & |t| \le \varepsilon \\ 0, & |t| > \varepsilon \end{cases} . \tag{2}$$

Note that δ_ε converges to Dirac delta function when $\varepsilon \to 0$, while equation (1) with delta function in r.h.s has unique solution [27]. Therefore the value of ε could always be considered to be small enough so that solution of (1) negligibly differs from the solution corresponding to point source (with Dirac delta function in r.h.s). Thus ε is formal parameter and it's actual value is of no importance. The only reason for using that parameter is because scalar product is not defined in space of generalized functions and thus presentation of adjoint formalism will be not strict if Dirac delta function was used. From the definition of δ_ε obviously follows that: $\int\limits_\Omega f^s \cdot d\Omega = q^s$ where integration is performed over the spatial domain Ω in which equation (1) is considered and $d\Omega$ $[m^3]$ is infinitesimal volume.

The solution of equation (1) is considered in spatio-temporal domain $G = [0, T] \times \Omega$, where T is integration time. Hence (1) is complemented with initial conditions: $c(x, y, z, 0) = 0$ and with boundary conditions corresponding to zero-fluxes through the boundaries:

$$\partial c / \partial \bar{n} = 0, \quad (x, y, z) \in \partial\Omega. \tag{3}$$

Here $\partial\Omega$ is boundary of spatial domain, and $\bar{n}$ is normal vector to it. In this paper we will be interested in solution of (1)-(3) at $T \to \infty$: $c(x, y, z) = c(x, y, z, T)$. Thus from now, unless otherwise is stated, we will use letter $c = c(x, y, z)$ in order to mention solution of stationary problem:

$$Lc = u_i \frac{\partial c}{\partial x_i} - \frac{\partial}{\partial x_i} D \frac{\partial c}{\partial x_i} = q^s \delta_\varepsilon (x - x^s) \delta_\varepsilon (y - y^s) \delta_\varepsilon (z - z^s) = f^s (x, y, z) \tag{4}$$

satisfying the same boundary conditions (3). In (4) operator L has been introduced which will be used below. Velocity and turbulent kinetic energy fields on which D depends are pre-calculated by the CFD model and concentration equation (1)-(3) is then solved until a sufficiently large T when concentration field could be considered established and converged to solution of (4).

Now assume that measurements are performed in simulation domain Ω and that they indicate non-zero stationary values of concentration c_n^o at set of measurement points with coordinates $\bar{r}_n^o = (x_n^o, y_n^o, z_n^o)$: $1 \le n \le K$, where K is total number of measurement points. The measured concentrations are combined in vector $\bar{c}^o$.

Consider the concentration c_n^c calculated with equation (4) at measurement point n. The calculated concentration field $c(x, y, z)$ is related to c_n^c with the following functional:

$$c_n^c = \int_\Omega c \cdot p_n \cdot d\Omega = (c, p_n), \tag{5}$$

Equation (5) is also referenced below as 'measurement equation'. In order to represent measurement in a given point Dirac delta function should be used in (5): $p_n = \delta(\bar{r} - \bar{r}_n^0)$. But, in order to be able to use

scalar product, p_n (the so-called 'probing function') should be integrable with square. Therefore we define it as:

$$p_n(x,y,z) = \delta_\varepsilon(\bar{r} - \bar{r}_n^o) = \begin{cases} 1/\varepsilon^3, & |\bar{r} - \bar{r}_n^o| \le \varepsilon \\ 0, & |\bar{r} - \bar{r}_n^o| > \varepsilon \end{cases}, \tag{6}$$

where function of vector argument $\delta_\varepsilon(.)$ converges to Dirac delta function when $\varepsilon \to 0$ and parameter ε is considered to be small enough to represent measurement in given point. In other words the value of ε is so small that the difference between c_n^c calculated with the use of p_n defined by Dirac delta function and with the use of equation (6) is negligibly small. Thus the actual value of ε is of no importance.

Then the problem of estimating the source emission rate and location could be posed as the problem of finding such values of source parameters q^s, x^s, y^s, z^s (combined in a vector of control parameters $\bar{\psi} = (q^s, x^s, y^s, z^s)^T$) that satisfy the least squares criteria and minimize the following cost function:

$$J = \frac{(x_0^s - x^s)^2}{\sigma_H^2} + \frac{(y_0^s - y^s)^2}{\sigma_H^2} + \frac{(z_0^s - z^s)^2}{\sigma_V^2} + \frac{1}{\sigma_O^2 + \sigma_M^2} \cdot \sum_{n=1}^{K} (c_n^c - c_n^o)^2 \to \min. \tag{7}$$

The first three terms in the r.h.s of (7) 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. Note that we do not use prior estimations of the source rate since in practice it is very difficult to make reasonable prior estimations of that parameter. The last term in (7) accounts for the mean squared difference between calculated and measured values while σ_O^2 and σ_M^2 are mean squared errors of measurements and model respectively.

It could be shown (see e.g., [24]) that such statement of the inverse problem corresponds to finding maximum of posterior probability distribution of control vector $\bar{\psi}$ given the measurements $\bar{c}^o$, provided that probability distributions of all errors, i.e. errors of prior estimations of horizontal and vertical coordinates of the source, measurements and model predictions are Gaussian with constant variances σ_H^2, σ_V^2, σ_O^2, σ_M^2 respectively.

3. Method of solution

There are different methods of numerical solution of minimization problem (7). Most generally they could be divided in two broad categories: iterative and direct methods. Great advantage could be gained by iterative descent methods that use gradient of cost function with respect to control parameters in case when cost function is convex or at least quasi-convex [18]. However cost function (7) is non-convex with respect to source coordinates and descent algorithm can converge to local minimum. Thus, in present work we use a special kind of direct method for solution of minimization problem. The effective use of direct method requires very fast calculation of c_n^c for every measurement point n and every set of source parameters. The 'source-receptor' function (SRF) which is described below could be used for that purpose.

3.1. Source-receptor function

Source-receptor function is a function relating concentration at given measurement point to source coordinates and source rate. As it was shown in works [14],[15] the most convenient way of finding source-receptor function is the use of adjoint equations.

Adjoint operator L^* to operator L from (4) is defined with Lagrange duality relationship: $(L\varphi, \gamma) = (\varphi, L^*\gamma)$ which should be valid for the arbitrary functions φ and γ defined and integrable with squares on Ω and satisfying boundary conditions (3). Here notation of scalar product is used in the same sense as in (5). The following relationship for adjoint operator: $L^* = -u_i \partial/\partial x_i - \partial/\partial x_i (D \partial/\partial x_i)$ could be easily verified following the same procedure as in [21].

Now let define the adjoint variable c_n^* being the solution of the following adjoint equation:

$$L^* c_n^* = -u_i \frac{\partial c_n^*}{\partial x_i} - \frac{\partial}{\partial x_i} D \frac{\partial c_n^*}{\partial x_i} = p_n, \tag{8}$$

Where p_n in r.h.s is defined by (6) and c_n^* satisfies boundary condition (3). Following from the definition of adjoint operator the solution of (8) satisfies condition: $(L^* c_n^*, c) = (Lc, c_n^*)$ and hence:

$$(p_n, c) = (f^s, c_n^*)\,. \tag{9}$$

Then following (9) and (5) solution of (4) at measurement point n could be calculated as:

$$c_n^c = q^s \cdot c_n^*\,(x^s, y^s, z^s)\,, \tag{10}$$

where the entire expression in the r.h.s of (10) is the SRF.

Hence, in order to define SRF, adjoint equation (8) is to be solved once per each measurement point. Note that adding term $-\partial c_n^*/\partial t$ to the left-hand side of equation (8) yields in the non-stationary equation which is adjoint to the non-stationary forward equation (1). Solution of (8) is obtained by solving the non-stationary adjoint equation from sufficiently large T to $t = 0$.

3.2. Minimization algorithm

For the numerical solution of the inverse problem, forward equation (4) and adjoint equation (8) are discretized on computational grid (particular details of the discretization method are mentioned in the next section). The discretized analogs of the continuous functions $c\,(x, y, z)$ and $c_n^*\,(x, y, z)$ as well as that of source term function $f^s\,(x, y, z)$ are the corresponding vectors $\bar{c}$,$\bar{c}_n^*$,$\bar{f}^s$ defined on computational grid and hence having the size equal to the number of grid cells N: $\bar{c}, \bar{c}_n^*, \bar{f}^s \in R^N$.

The source term (right-hand side of model equation (4)) is approximated with the use of the nearest neighbor interpolation, i.e., a nonzero source term is set only at the node k^s nearest to the assumed source location. Hence the elements of source term vector $\bar{f}^s$ are defined by relationship: $f_k^s = q^s \delta_{k,k^s}$, where δ_{k,k^s} is the Kronecker's delta. Thus, in contrast to continuous formulation in numerical formulation of the atmospheric dispersion problem the set of possible source locations is restricted by the number of grid nodes. That circumstance allows to use direct minimization algorithm as described below.

The approximation of equation (4) could then be written in general form:

$$\underline{\underline{L}}\bar{c} = \bar{f}^s = q^s \delta_{k,k^s}, \tag{11}$$

where matrix $\underline{\underline{L}}$ arises from numerical approximation of operator L in (4).

The correspondence between measured concentration and gridded concentration field is established by weighted linear interpolation:

$$c_n^c = (\bar{p}_n)^T\, \bar{c}\,(t) = (\bar{p}_n, \bar{c})\,. \tag{12}$$

Equation (12) is the discretized analog of measurement equation (5), vector $\bar{p}_n \in R^N$ is vector of weights of interpolation of the gridded concentration field to the position of $n-$th measurement point and, analogously to (5), r.h.s of (12) is scalar product of vectors in Euclidian space R^N. Thus approximation of adjoint equation (8) could be written as:

$$\underline{\underline{L}}^* \bar{c}_n^* = \bar{p}_n, \tag{13}$$

where matrix $\underline{\underline{L}}^*$ arises from numerical approximation of L^* in (8). The discrete source receptor function is then defined by relationship:

$$c_n^c \approx q^s \cdot c_{n,k^s}^*. \tag{14}$$

Note that in order to obtain strict equality in (14) matrix $\underline{\underline{L}}^*$ in (13) should be taken as transpose to matrix $\underline{\underline{L}}$ from (11). However constructing matrix $\underline{\underline{L}}^*$ as approximation of continuous adjoint equation (8) has important implementation advantages. In particular, the last approach is straightforwardly applicable

for the case of higher-order iterative schemes of approximation of advection-diffusion operator L (in that case numerical approximation is represented by nonlinear operator $\tilde{L}(\bar{c})$ instead of matrix $\underline{\underline{L}}$). As it will be shown below with the used approach of constructing adjoint operator $\underline{\underline{L}}^*$ equality (14) holds with sufficient accuracy.

At this step the original continuous minimization problem could be replaced by the following discrete minimization problem: to find such a pair (q^s, k^s) that minimizes cost function (7). Solution of this discrete minimization problem proceeds as follows. Substituting relationship (14) for discrete source-receptor function in (7) one can obtain the following relationship for the value of cost function J_k when source with intensity q is located in $k-$th grid node:

$$J_k(q) = \frac{(x_0^s - x_k)^2}{\sigma_H^2} + \frac{(y_0^s - y_k)^2}{\sigma_H^2} + \frac{(z_0^s - z_k)^2}{\sigma_V^2} + \frac{1}{\sigma_O^2 + \sigma_M^2} \cdot \sum_{n=1}^{N_o} \left(c_{n,k}^* q - c_n^o\right)^2. \tag{15}$$

Firstly at each grid node k the functional $J_k(q)$ is minimized with respect to q and for each k the corresponding minimum value of the cost function $\tilde{J}_k$ is calculated. Then the solution of discrete minimization problem is obtained by selecting the grid node k^s where minimum among all $\tilde{J}_k$ is achieved. This procedure can be written as follows:

$$\begin{array}{l} for\, each\, k:\ J_k(q) \to \min,\ when\ q = q_k^s;\ \tilde{J}_k = J_k(q_k^s). \\ \tilde{J}_k \to \min,\ when\ k = k^s \end{array}. \tag{16}$$

The corresponding pair $(q_{k^s}^s, k^s) = (q^s, k^s)$ is the solution of discrete minimization problem, and the vector $(q^s, x_{k^s}, y_{k^s}, z_{k^s})$ is considered as approximation to the solution of original continuous minimization problem.

3.3. Details of implementation in ADREA-HF CFD code

The ADREA-HF CFD code [26] is a three dimensional transient fully compressible flow and dispersion CFD solver, able to treat highly complex geometries using the porosity formulation on Cartesian grids. Through many years it has been extensively used in different atmospheric dispersion studies and has been validated on the basis of numerous field experiments on atmospheric dispersion [26],[16],[17],[1]. The mixture continuity equation, the three Cartesian mixture momentum equations, the mixture energy equation (enthalpy or internal energy) and the species mass conservation equations are numerically solved in ADREA-HF. Turbulence is modeled with the $k-\varepsilon$ model. The comprehensive description of the governing equations, physical options and boundary conditions relevant to this study is given in [2] while recently developed extensions are described in [26].

The details of numerical algorithms used in ADREA-HF model are described in [3]. The control volume approach is used for the discretization of the conservation equations on Cartesian grids. For the numerical calculations, the flow domain is divided into rectangular cells. The cells are used as control volumes for the integration of partial differential equations involving scalar variables such as concentration and other. For the vector parameters such as the velocity components, additional control volumes are introduced based on the staggered grid concept [19].

Intersection of the geometry with the grid is treated in two complementary ways. As for explicitly resolved geometry the rectangular cells are incomplete in cases when they are crossed by irregular surfaces. Boundary surfaces arising from intersection of the control volumes with the geometry are fully described by set of parameters (such as orientation, surface area, distance from the center of grid node and other), which are calculated by DELTA code [25]. The appropriate boundary conditions are set at those boundary surfaces (e.g. condition (3) for concentration). Intersection of the subgrid geometry with the domain is treated with the porosity formulation. Resulting grid cells can be fully active (porosity 1), fully blocked by solid parts (porosity 0) or partially active (porosity between 0 and 1). All the necessary for numerical approximation parameters such as cell's porosities, volumes, areas of open surfaces and other are also calculated by DELTA code.

The following numerical approximations were used in the present study. The fully implicit numerical schemes were used for numerical stability. Time derivatives were approximated with first order in time

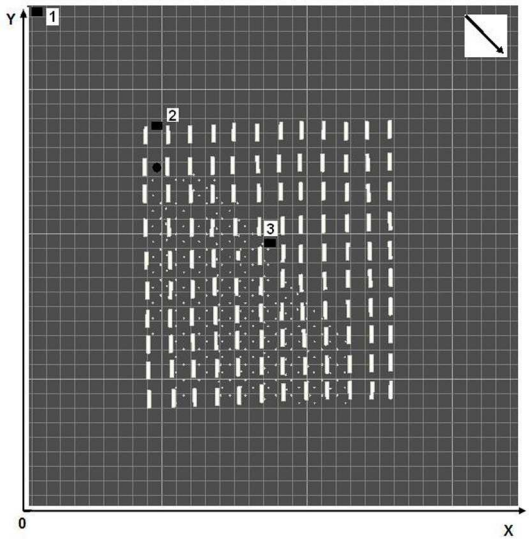

Figure 1: Computational domain of ADREA-HF. Positions of buildings are shown together with true source location (black circle), different prior estimations of source locations (black squares) and sensors (white dots) in computational domain of ADREA-HF. Black arrow indicates wind direction.

backward scheme. 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. The approximation of pressure gradients in momentum equations is of second order since in the case of staggered grid the pressure nodes are located at the faces of the velocity control volumes. The special version of ADREA/SIMPLER approach (see also [19],[11]) is used to iteratively update pressure and velocity fields. The resulting pressure equation is solved with Bi-CGStab method using various types of pre-conditioners described in [11].

As far as inverse problem is formulated above only for the passive contaminant (which doesn't affect the wind flow) the wind field is pre-calculated, stored and then used for solving atmospheric dispersion problem. The adjoint equation (8) is solved by integrating non-stationary adjoint equation from sufficiently large T to $t = 0$ when distribution of c_n^* is fully established. The right part of adjoint equations – weights $\bar{p}_n$ from 'measurement' equation (12) are defined by bilinear interpolation algorithms.

The adjoint equation (8) is solved separately for each sensor which defines the right part p_n and the obtained solution c_n^* is stored in binary file. Hence for a given measurement network with K sensors required number of integrations of equation (8) is K.

When the adjoint variables c_n^* defining source-receptor function (14) have been calculated for all K sensors they are used in minimization algorithm (16). As tests have shown, the minimization algorithm itself is nearly immediate procedure as compared to the time spent for the solution of adjoint equations (8). Since solutions of equations (8) for different sensors are independent between each other, the proposed algorithm is suitable for parallelization, and the potential improvement in calculation time by parallelization is by a factor K, if such number of processors is available.

4. Results of calculations

The described methodology has been validated against data of the wind tunnel experiment [4] which has been scaled for the conditions of the field MUST experiment [28]. Hence ADREA-HF 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 [4] the atmospheric dispersion among rows of rectangular obstacles was investigated. As it is shown in Figure 1 the obstacles were arranged in 12 rows consisting of 10 obstacles. The obstacles were nearly identical and had average length, width and height: 12.2x2.42x2.54 m. The contaminant's concentration has been measured by an array consisting of 256 detectors arranged along obstacle rows in the part of the domain covered by plume (Figure 1). All detectors were arranged at the same height equal to 1.28 m. Therefore non-zero concentrations have been measured by nearly all (244 out of 256) detectors (shown at Figure 1). Here we should note that only those measurement data are described that were available to authors (through data base of COST 732 project).

The wind flow was characterized by neutral stratification, wind speed at the roof level: $U_{ref} = 8\,m/s$, and wind direction: -45 deg. in experimental coordinate system (corresponding to 45 deg. in the coordinate system shown in Figure 1). The contaminant originated from a point source (shown by black circle in Figure 1) located at the ground level. The volume flow rate of gas at the source was: $\approx 3.3 \cdot 10^{-6}\,m^3/s$.

ADREA-HF was setup for the conditions of the experiment [4] with the following parameters. The computational domain which is shown at Figure 1 covered all building blocks present in the experiment. The center of the grid coincided with the center of the part of experimental domain covered by building blocks. Size of the computational domain has been set to 340x340x20 m., with horizontal grid size varying from $dx = dy = 5.5\,m$ inside the part of the domain covered by building blocks and growing in the outer domain with a rate $dx_{i+1}/dx_i = 1.2$. The vertical grid size was set to $dz = 0.2\,m$ near the ground and then increasing with a rate $dz_{i+1}/dz_i = 1.1$. The resulting number of grid nodes had been 52x52x25 in x,y,z directions respectively.

The wind field has been pre-calculated and the stationary atmospheric dispersion problem has been solved by integrating concentration equation (1) from $t = 0$ to $t = T = 200\,s$ which was enough for achieving fully established concentration distribution. The non-stationary adjoint equation has been integrated in backward direction through the same time interval to achieve established distributions of adjoint variables.

First of all results of forward run by ADREA-HF have been evaluated against measurements. On the basis of all available measurements the normalized mean squared error (NMSE) and fractional bias (FB) have been calculated according to standard definitions: $NMSE = \left\langle (C_m - C_o)^2 \right\rangle \Big/ \left(\langle C_m \rangle \langle C_o \rangle \right)$, $FB = 2\,\langle C_m - C_o \rangle \,/\, \langle C_m + C_o \rangle$, where triangle brackets denote arithmetic averaging, indices "m" and "o" stands for model and observations respectively. The calculated values: $NMSE = 1.94$, $FB = 0.56$ indicated satisfactory quality of the forward runs taking into account the problem complexity.

Based on the solution of the adjoint equation, the adjoint variables has been pre-calculated for all sensors in the whole computational domain and then stored in binary files. In the described configuration it required 2 minutes of computational time per one sensor on PC Intel Core-2 2.93 GHz. As it was noted in subsection 3.2 the calculated source-receptor function (14) is approximate, and therefore the accuracy of the calculated SRF has been evaluated by comparing results of the forward run with the results given by relationship (14). The relative mean absolute error (MAE) of results obtained with the use of SRF has been evaluated: $MAE = \langle |(C_f - C_{SRF})\,/C_f| \rangle$, where indices "f" and "SRF" stand for the results obtained with the forward model and with the use of SRF respectively. The obtained value was MAE=0.07. In order to clarify the significance of such error in SRF calculations the same statistical parameter has been calculated on the basis of comparison of results of the forward runs with measurements which gave MAE=2.65. From these results it is clear that the error introduced by SRF is significantly less (by a factor of about 40) than the model error as compared to measurements, and it could be considered as negligibly small.

The performance of the developed algorithm of source estimation was evaluated by three parameters: 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 from the true source location (x_t^s, y_t^s, z_t^s) and the relative source rate ratio: $\delta q = \max\left[(q^s/q_t^s), (q_t^s/q^s) \right]$ where index "t" stays for the true source rate. Thus δq is always greater than unity for both underestimated and overestimated source rates.

When data of all available measurement sensors ($N_o = 244$) were used for reconstruction of source parameters the following values were obtained (without use of any prior estimations of source location): $r_H = 11\,m$, $r_V = 0.8m$, $\delta q = 3.1$ which is quite good result as compared to results of other studies such as [20]. For example among the results reported in [20] the value of δq reached up to the value of 10. Note, that detec-

tors shown at Figure 1 almost totally covered the contaminated part of the domain within obstacle arrays and the measured concentration values spanned the range from $1.8 \cdot 10^{-8} kg/m^3$ to $2.6 \cdot 10^{-12} kg/m^3$. Since very low and zero values are not used in data assimilation procedure the described experiment could be representative of more general situation when the available measurements totally cover the contaminated territory.

In practice however the actual number of measurements is very limited, and hence the natural question arises: what results could be expected from the measurement network with less sensors and how robust they are (how they may change depending on sensors locations)? That question could be reformulated as follows. For the given number of sensors and fixed 'satisfactory' criteria with what probability good results could be achieved?

Since the number of measurement points in the simulated experiment is quite large ($N_o = 244$), one could fix a given number of sensors $K < N_o$ and try to estimate source term with all possible network configurations consisting of K sensors out of N_o available sensors. The total number of such configurations is equal to binomial coefficient: $C_K^{N_o}$. One can consider that solution is 'good' if it satisfies the threshold criteria: $r_H \leq r_H^{thresh}$, $r_V \leq r_V^{thresh}$, $\delta q \leq \delta q^{thresh}$. Then when $C_K^{N_o}$ is large the ratio P of the number of good solutions to the total number of configuration networks $P = N_{good}/C_K^{N_o}$ could be interpreted as probability of finding good solution for the measurement network of a given size K.

The difficulty in following the above approach arises from the fact that $C_K^{N_o}$ could be extremely large. For instance, if $K = 2$ the total number of measurement networks $C_2^{244} = 29646$ is not large and all possible configurations could be directly processed. However, with $K > 4$, the total number of networks is completely infeasible for direct processing (e.g. in case of $K = 100$, this number is: $C_{100}^{244} \sim 10^{70}$).

In order to overcome this difficulty a random sampling algorithm has been developed in which the measurement networks are chosen randomly and the ratio of the number of 'good' networks (N_{good}) to the total number of the processed networks (N_{tot}) converges quickly to the stationary value: $N_{good}/N_{tot} = P$ and hence to the probability of achieving good solution for the given size of measurement network. The details of the random sampling algorithm are described in the Appendix.

The performance of the random sampling algorithm is illustrated in Figure 2-a), which shows the dependence of N_{good}/N_{tot} on the number of processed measurement configurations for the case of $K = 2$, $N_o = 244$. Results obtained with direct and random sampling algorithms are presented. The following criteria for the 'good' solution have been used: $r_H \leq 15m$, $r_V \leq 2.5m$, $\delta q \leq 4$ which is quite reasonable because from the one hand it roughly corresponds to the average quality of solutions reported in literature (average value of δq as follows from the results reported in [20] is around: $\delta q \approx 4$) and from the other – it corresponds to the case of finding source location within the region having horizontal and vertical scale comparable to the scale of single building. As it is seen from this figure, the random sampling algorithm quickly converges to the stationary value of N_{good}/N_{tot} which coincides with the corresponding value finally obtained with the full processing of all possible measurement configurations. Subsequently random sampling algorithm was applied for different values of K and it was stopped when stationary value of N_{good}/N_{tot} had been achieved (as it is demonstrated by the Figure 2-b) for the case of $K = 100$).

With the approach described above the dependencies of the probability of achieving good solution on the size of measurement network had been calculated in several tests listed in Table 1. In the Test 1 'synthetic' measurements were used (i.e., the values calculated with the forward model and true source parameters at the locations of measurement points were treated as 'measurements'). In this test prior estimation of source location was not used. The Test 1 could be treated as the case of 'perfect model' because error of the model as compared to synthetic measurements was very small (when exact source parameters were used). Though this error was not exactly zero because the use of SRF introduces some small error as it was discussed above. In the Test 2 real measurements were used for source inversion and prior estimation of source location again was not used.

All other tests (No 3-7) differed by prior estimations of source location and by the value of the assumed error σ_H in horizontal coordinates of source estimation. Thus the sensitivity of the source inversion algorithm to the use of prior estimations of source location and to the values of regularization parameters had been investigated. Different prior estimations of source location are shown at Figure 1. True source coordinates in

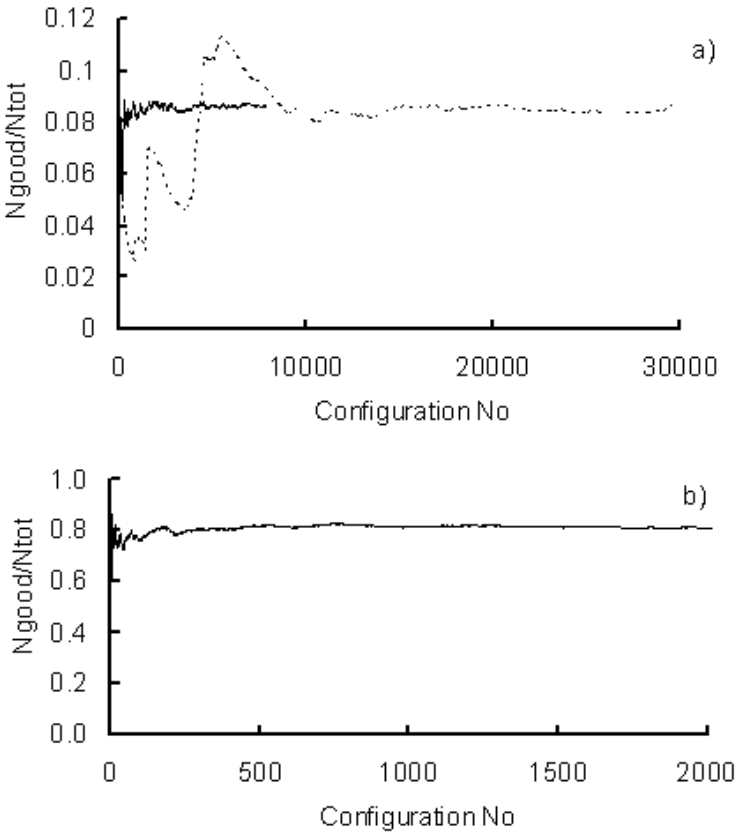

Figure 2: Relative number of 'good' solutions of the inverse problem obtained for the case of total number of available sensors $N_o = 244$ and different measurement networks of the sizes: $K = 2$ (a) and $K = 100$ (b). Solid lines show results obtained with random sampling algorithm. Dashed line shows results obtained with the full processing of all possible measurement configurations (calculated and shown only for the case of $K = 2$). Prior estimations of source location were not used in presented cases.

Table 1:
Summary of computational tests.

Test No	If true or synthetic measurements were used (T/S)	Prior estimation of source location (No, corresponding to Figure 1.)	σ_H, m
1	S	-	-
2	T	-	-
3	T	1	340
4	T	3	340
5	T	2	340
6	T	1	170
7	T	2	170

model coordinate system were: $(x_t^s, y_t^s, z_t^s) = (95m, 246m, 0m)$ while source coordinates of prior estimations shown at Figure 1 and listed in Table 1 were: (0 m, 340 m, 0 m) (estimation No. 1), (80 m, 250 m, 0 m) (estimation No. 2) and (170 m, 170 m, 0 m) (estimation No 3). Thus estimations No. 1 and 3 were located quite far from the true source (at the distances 133 m and 106 m upwind and downwind from the source respectively). Estimation No. 2 was closest to the true source location ($r_H = 15.6\,m$) representing the situation when user possesses some additional information about source location. Estimations No. 1 and No. 2 represented usual practice when prior estimation is taken upwind to the existing measurements of pollutant and close to the centerline of the observed plume. The distances of the estimations No. 1 and No. 3 from the true source location had been set approximately equal to the scale of the problem (the along-wind size of the plume as represented by measurements was about 120 m).

The values of error parameters entering cost functon (7) has been assigned in the following way. In tests No. 3-5 the value of σ_H has been equal to the size of the computational domain: $\sigma_H = 340m$. Such assumption seems to be quite realistic in practice. In tests No. 6-7 the value of σ_H had been reduced by the factor of 2 ($\sigma_H = 170m$). The reduced value of σ_H refereed to the correct assumption that prior estimation is not too far from the true source location and thus convergence of source inversion algorithm to the true solution had to be enhanced. The value of the assumed error σ_V in vertical position of the prior estimation of source location has been the same in all tests from No. 3 to No. 7 and equal to building height: $\sigma_V = 2.5m$. This assumption clearly reflects assumption that the source is located somewhere below roof tops. It was assumed that error in model prediction σ_M is much more than the measurement error σ_O. This is very realistic assumption (see e.g. [10]). Therefore $\sigma_M^2 + \sigma_O^2 \approx \sigma_M^2$. The value of σ_M had been assumed to be approximately equal to the average measured (volume) concentration: $\sigma_M = 10^{-9} \approx \langle C_m \rangle$. Though actually this value of σ_M is somewhat underestimated, but it is still realistic while in practice one cannot know it exactly in advance.

Figure 3 shows the probability of achieving good solution of the inverse problem calculated with the use of random sampling algorithm for different sizes of measurement networks varying from $K = 1$ to $K = N_o = 244$ and in different tests listed in Table 1. As it could be seen from the figure in all cases the probability of achieving good solution monotonically increases with increasing number of sensors used for the solution of the inverse problem. In case of the 'perfect model' (Test 1) good results are achieved much more quickly than in all other cases. The probability of obtaining good solution equal to 90% or higher in the Test 1 is achieved with measurement networks consisting out of 15 sensors or more.

Results of the tests No. 2-7 with real measurements are quite different for small number of measurements used in source inversion (i.e. when $K < 25$). This is easily seen from the Figure 3-b) in which x - coordinate is plotted in logarithmic scale. As it could be seen from that figure use of prior estimations lead to increase in probability of achieving good solution. This effect is significant in tests No. 4 and 7 in which the prior estimation of source location was closest to the true source location. For example probability of achieving

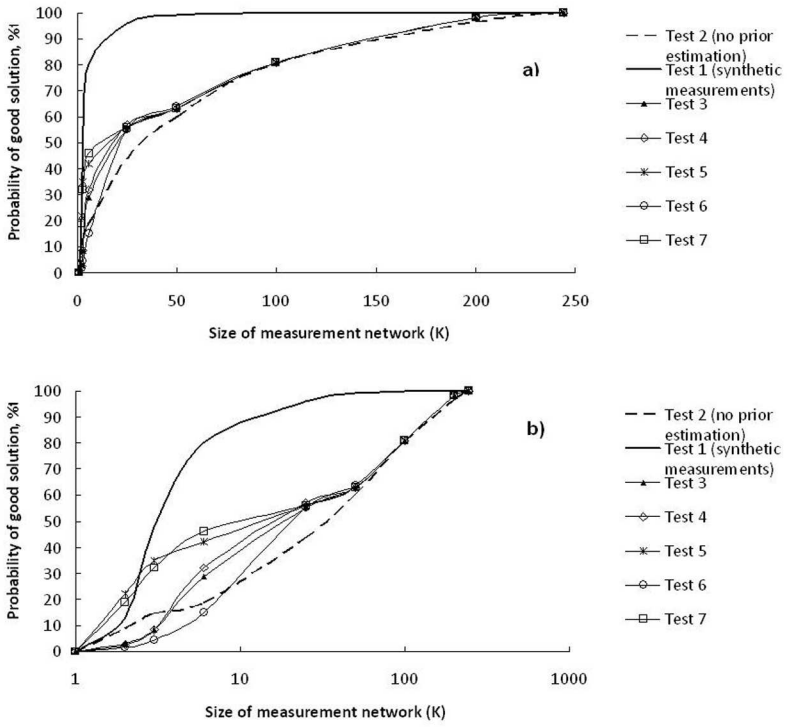

Figure 3: Dependence of probability of achieving good solution of inverse problem on the number of sensors K used in source estimation. Results of different tests listed in Table 1 are shown. Figure b) is the same as a) but horizontal axis is in logarithmic scale.

good solution with the use of only 6 measurements in those tests reached 40% while in the Test 2 (without the use of prior estimation) it was only 18%. However the same value for the case of 'perfect' model (Test 1) was 80% though prior information was not used in Test 1. Therefore one could come to the conclusion that improving of the quality of model could be of more benefit for source inversion than the use of prior information.

However in all tests (No. 2-7) when real measurements were used with increasing K ($K > 50$) probabilities of achieving good solution were the same. The 90% or higher probability of achieving good solution was reached only for measurement networks consisting of 150 or more sensors. This is 10 times more than the corresponding value obtained in the case of 'perfect model' (Test 1), which required only 15 measurements to achieve the same probability. This result also supports the above conclusion that the quality of model is of primary importance for obtaining good results of source inversion.

5. Conclusions

In the presented work a method for estimating the location and rate of a point stationary source of passive non-reactive pollutant in an urban environment has been developed. The method is based on variational formulation in which the cost function characterizing the difference between the calculated and measured concentrations is minimized with respect to source coordinates and rate. The minimization problem is solved with a direct algorithm based on the use of source-receptor function which allows for quick estimation of concentration in a given point for arbitrary source parameters. The computational time is spent mainly for the calculation of the SRF which requires integration of as many adjoint equations as the number of sensors N_o used in the source term estimation. Since the adjoint equations solutions for different

sensors are independent the algorithm has high potential for parallelization. The developed source inversion algorithm has been implemented in the CFD code ADREA-HF and it is able to run in complex urban geometry. Validation of the algorithm has been performed on the basis of simulation of the wind tunnel experiment [4] on atmospheric dispersion among an array of rectangular obstacles scaled for the conditions of the MUST field experiment [28]. For the particular computational configuration of ADREA-HF used to simulate the experiment, SRF calculation required 2 minutes of computational time per one sensor on a 2.93 GHz Intel Core-2 PC. Good results of source parameters estimation have been achieved when all available measurements ($N_o = 244$) were used to solve the inverse problem.

In order to test the performance of the algorithm depending on the size of the measurement network, calculations have been performed in which different randomly selected sensors out of all available sensors were used for source estimation. For this purpose a special random sampling algorithm has been developed which allowed for calculating the probability of achieving good solution for the given number of sensors used in source estimation. In case of the 'perfect model' (i.e. when 'synthetic measurements' as calculated by the forward model and true source term were used for source estimation) good results of the solution of the inverse problem were achieved with 90% or higher probability for arbitrary measurement configurations consisting of 15 or more sensors. When real measurements were used for source estimation and the number of measurements was small ($K < 25$) the results obtained with the use of prior estimations of source location were generally better than the results without the use of prior estimations. However with increasing the number of measurements that difference disappeared. In all cases in which real measurements were used the 90% or higher probability of achieving good solution was reached only for measurement networks consisting of 150 or more sensors. Hence further improvement of the source inversion algorithm can be achieved first of all by improving the performance of the forward model.

6. Acknowledgements

The present work has been supported by the European Commission Seventh Framework Grant agreement no.: 229773 "Enhancing the Research Potential of the NCSR 'Demokritos' Environmental Research Laboratory in the European, National and Regional Research Areas" (PERL). The authors are grateful to George Efthimiou and Nektarios Koutsourakis for preparing ADREA-HF to run on experimental data. The first author is also grateful to Stella Giannissi for assistance in using ADREA-HF user interface.

Appendix A. Algorithm of random selection of measurement network

Here we present the algorithm of random selection of the K measurement points (sensors) out of N_o available. Let us consider the multi-dimensional probability $P(\xi_1, ..., \xi_{N_o})$ that the sensors 1, ..., N_o are being selected or rejected. Random variable $\xi_i = 1$ if i-th sensor is selected and $\xi_i = 0$ if it is rejected. If K and only K different sensors are to be selected the considered probability is non-zero if and only if the number of nonzero values among all $\xi_1, ..., \xi_{N_o}$ is equal to K. In all other cases (when more or less than K sensors are selected and hence the number of nonzero values among $\xi_1, ..., \xi_{N_o}$ is more or less than K) the above probability is zero. We are interested in case when all nonzero probabilities are equal and hence the probabilities of choosing any measurement network out of K sensors are the same.

The basis of the random sampling algorithm is the well-known fact [24] that the considered probability $P(\xi_1, ..., \xi_{N_o})$ could be reproduced as multiplication of conditional probabilities:

$$P(\xi_1, ..., \xi_{N_o}) = P(\xi_1) \cdot P(\xi_2|\,\xi_1) \cdot ... \cdot P(\xi_{N_o}|\,\xi_1, .., \xi_{N_o-1})\,. \tag{A.1}$$

In r.h.s of (A.1) $P(\xi_1)$ is the probability that the first sensor is being selected or rejected; $P(\xi_2|\,\xi_1)$ is conditional probability that the second sensor is being selected or rejected given the information about whether the first sensor had been already selected; $P(\xi_{N_o}|\,\xi_1, .., \xi_{N_o-1})$ is conditional probability that the last sensor is being selected or rejected given the information about selection of each out of the previous sensors. If the probability of being selected is the same for all sensors, then for the first sensor:

$$P(\xi_1) = 1 - Q(\xi_1) = 1 - \left(C_K^{N_o-1} \Big/ C_K^{N_o} \right) = K/N_o, \tag{A.2}$$

where $Q(\xi_1) = C_K^{N_o-1}/C_K^{N_o}$ is the probability that the first sensor was rejected which is equal to the ratio of the total number of configurations of K sensors not including the first sensor to the total number of configurations of K sensors. The well known properties of binomial coefficients were used in deriving the right hand side in (A.2).

Analogously the conditional probability of selecting the second sensor given the value of ξ_1 is obtained by relationship:

$$P(\xi_2|\,\xi_1) = 1 - Q(\xi_2|\,\xi_1) = 1 - \left(C_{K-\xi_1}^{N_o-2} \Big/ C_{K-\xi_1}^{N_o-1} \right) = (K-\xi_1)/(N_o-1), \tag{A.3}$$

where $Q(\xi_2|\,\xi_1)$ is conditional probability of the second sensor to be rejected given the information about whether the first sensor was selected or rejected. Relationship (A.3) reflects the fact that after the decision has been made concerning the selection of the first sensor, the total number of sensors available for the selection becomes $N_o - 1$ while the number of sensors to be selected is: $K - \xi_1$ (i.e. it is equal to $K-1$ if the first sensor was selected and it is equal to K if the first sensor was rejected).

Analogously to (A.3) and (A.2) for arbitrary i the following relationship could be verified:

$$P(\xi_i|\,\xi_1,..,\xi_{i-1}) = 1 - \left(C_{K-\sum_{l=1}^{i-1}\xi_l}^{N_o-i} \Big/ C_{K-\sum_{l=1}^{i-1}\xi_l}^{N_o-i+1} \right) = \left(K - \sum_{l=1}^{i-1} \xi_l \right) \Big/ (N_o - i + 1). \tag{A.4}$$

Note that the right part in relationship (A.4) is valid when: $N_o - i \geq K - \sum_{l=1}^{i-1}\xi_l$. Otherwise according to standard definitions of binomial coefficient $C_{K-\sum_{l=1}^{i-1}\xi_l}^{N_o-i} = 0$ and the corresponding probability $P(\xi_i|\,\xi_{i-1},..,\xi_1) = 1$. Thus if the total number of sensors that were not processed before the selection of the i-th sensor $(N_o - i + 1)$ is equal to the number of sensors $(K - \sum_{l=1}^{i-1}\xi_l)$ that remains to be chosen, then all non-processed sensors are to be selected.

The random-sampling algorithm directly follows from the above. The sensors are processed consequently from $i = 1$ to $i = N_o$ and the probability of selecting the i-th sensor is given by relationship (A.4). From the above it is also straightforward that this procedure guarantees that K and only K sensors will be selected and that the probabilities of selecting all configurations of K sensors with the proposed algorithm are equal.

References

[1] S. Andronopoulos, J. Bartzis, J. Wurtz, D. Asimakopoulos, Modelling the effects of obstacles on the dispersion of denser-than-air gases, J. Hazard. Mater. 37 (1994) 327–352.

[2] J. Bartzis, ADREA-HF: a three-dimensional finite volume code for vapour cloud dispersion in complex terrain,, Technical Report EUR 13580 EN, National Centre for Scientific Research DEMOKRITOS, Athens, Greece, 1991.

[3] J. Bartzis, A. Venetsanos, N. Catsaros, S. Andronpoulos, D. Vlachogiannis, The numerical scheme of the ADREA codes, Technical Report RODOS(WG2)-TN(99)-09, National Centre for Scientific Research DEMOKRITOS, Athens, Greece, 1999. Available at www.rodos.fzk.de.

[4] K. Bezpalcova, F. Harms, 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, 2005.

[5] M. Bocquet, High-resolution reconstruction of a tracer dispersion event: Application to ETEX, Q.J.R. Meteorol. Soc. 133 (2007) 1013–1026.

[6] F. Chow, B. Kosovic, S. Chan, Source inversion for contaminant plume dispersionin urban environments using building-resolving simulations, J. Appl. Meteorol. Clim. 47 (2008) 1533–1572.

[7] G. Cosemans, P. Berghams, R. Ampe, F. Fierens, P. Vanderstraeten, J. Janssens, K.V. den Bruel, T. Tran, The 2008 elemental mercury vapour pollution accident in the brussels capital region: two approaches towards source identification, in: Proc. of the 13-th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes - 1-4 June 2010, Paris, France, pp. 280–284.

[8] I. Enting, Inverse Problems in Atmospheric Constituent Transport, Cambridge University Press, Cambridge, UK, 2002.

[9] B. Fisher, J. Kukkonen, M. Schatzmann, Meteorology applied to urban air pollution problems: COST 715, Int. J. Environ. Pollut. 16 (2010) 560–670.

[10] A. Keats, E. Yee, F. Lien, Bayesian inference for source determination with applications to a complex urban environment, Atmos. Environ. 41 (2007) 465–479.

[11] I. Kovalets, S. Andronopoulos, A. Venetsanos, J. Bartzis, Optimization of the numerical algorithms of the ADREA-I mesoscale prognostic meteorological model for real-time applications, Environ. Model. Softw. 23 (2008) 96–108.
[12] I. Kovalets, V. Tsiouri, S. Andronopoulos, J. Bartzis, Improvement of source and wind field input of atmospheric dispersion model by assimilation of concentration measurements: Method and applications in idealized settings, Appl. Math. Model. 33 (2009) 3511–3521.
[13] S. Lewis, S. Lakshmivarahan, J. Dhall, Dynamic Data Assimilation: A Least Squares Approach, Cambridge University Press, Cambridge, UK, 2006.
[14] G. Marchuk, Mathematical Modelling in the Environmental Problems, Nauka, Moscow, 1982.
[15] G. Marchuk, Adjoint equations and analysis of complex systems, Kluwer Academic Publishers, Dodrecht, Netherlands, 1996.
[16] I. Mavroidis, S. Andronopoulos, J. Bartzis, R. Griffiths, Atmospheric dispersion in the presence of a three-dimensional cubical obstacle: Modelling of mean concentration and concentration fluctuations, Atmos. Environ. 41 (2007) 2740–2756.
[17] P. Neofytou, M. Haakana, A. Venetsanos, A. Kousa, J. Bartzis, J. Kukkonen, Computational fluid dynamics modelling of the pollution dispersion and comparison with measurements in a street canyon in helsinki, Environ. Model. Softw. 13 (2008) 439–448.
[18] J. Orthega, W. Rheinboldt, Iterative solution of nonlinear equations in several variables, Academic Press, New York, 1970.
[19] S. Patankar, Numerical Heat Transfer and Fluid Flow, Hemisphere Publishing Corporation, New York, 1980.
[20] N. Platt, D. DeRiggi, Comparative investigation of source term estimation algorithms using fusion field trial 2007 data - linear regression analysis, in: Proc. of the 13-th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes - 1-4 June 2010, Paris, France, pp. 901–905. Full presentation available at http://www.harmo.org.
[21] J. Pudykiewicz, Application of adjoint tracer transport equations for evaluating source parameters, Atmos. Environ. 32 (1998) 3039–3050.
[22] R. Serafin, E. Barron, H. Bluestein, S. Clifford, L. Duncan, M. Lemone, D. Neff, W. Odom, G. Pfeffer, K. Turekian, T. Warner, J. Wyngaard, L. Geller, V. Turekian, D. Gustafson, J. Demuth, Tracking and Predicting the Atmospheric Dispersion of Hazardous Material Releases, The National Academic Press, Washington, D.C, USA, 2003.
[23] N. Talerko, Reconstruction of 131i radioactive contamination in ukraine caused by the chernobyl accident using atmospheric transport modeling, J. Environ. Radioact. 84 (2005) 343–362.
[24] A. Tarantola, Inverse problem theory and methods for parameter estimation, SIAM, Philadelphia, USA, 2005.
[25] A. Venetsanos, N. Catsaros, J. Wurtz, J. Bartzis, The DELTA-B code. A computer code for the simulation of the geometry of three-dimensional buildings. Code structure and users manual, Technical Report EUR 16326 EN, National Centre for Scientific Research DEMOKRITOS, Athens, Greece, 1995.
[26] A. Venetsanos, E. Papanikolaou, J. Bartzis, The ADREA-HF CFD code for consequence assessment of hydrogen applications, Int. J. Hydrog. Energy 35 (2010) 3908–3918.
[27] V. Vladimirov, Methods of the Theory of Generalized Functions, Taylor and Francis, New York, USA, 2002.
[28] E. Yee, C. Biltoft, Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles, Bound.-Layer Meteorol. 111 (2004) 363–415.

Response to Referee 1.

Referee: "Observe that the authors use prior information and exact one as the terms in the cost function (7) and (15). It is easy to see that in this way the cost function (7) considers true values of the source location, which are not known. Such kind calculation strategies are sometimes called as "inverse crime".
"

Response: No, exact values of source location and source rate are not used in cost function (7) and (15). This misunderstanding could arise due to somewhat misleading formulation in paragraph explaining formula (7) where it was written:
"...and mean squared errors in horizontal distance (σ_H^2) and vertical distance (σ_V^2) from the prior estimation of source location to the true one".

This was corrected to:
"...and parameters of probability distributions characterizing errors in prior estimations of source location (assumed to be Gaussian), namely mean squared deviations of horizontal (σ_H^2) and vertical (σ_V^2) coordinates"

Additionally in paragraph 8 of the section 4, "Results of calculations" the special notation for the true source location was introduced: $\left(x_t^s, y_t^s, z_t^s\right)$ in order to avoid notation mixing between estimated source parameters and the true source parameters.

Referee: "Moreover, it is not clear how the prior coordinates of the source location are estimated"

Response: Reasoning for the choices of prior estimations in source location had been added in paragraph 16 of section 4 "Results of calculations":

"*Estimation No. 2 was closest to the true source location ($r_H = 15.6m$) representing the situation when user possesses some additional information about source location. Estimations No. 1 and No. 2 represented usual practice when prior estimation is taken upwind to the existing measurements of pollutant and close to the centerline of the observed plume. The distances of the estimations No. 1 and No. 3 from the true source location had been set approximately equal to the scale of the problem (the along-wind size of the plume as represented by measurements was about 120 m).*"

Referee: ". In the section 4 in the numerical experiments, the authors consider only 3 estimations of the prior source location. Thus, it is not clear which number of estimations of the prior source location should be considered in general situation and whether the number of estimations could improve accuracy or not."

Response: This question is out of scope of the present paper and it was not studied. The aim of selecting different choices of prior estimations of source location was to study the sensitivity of the algorithm to the different choices of prior source location but not to study the dependance of the quality of solution to the number of prior source estimations. The corresponding explanation had been added in par. 16 of section 4:

"Thus the sensitivity of the source inversion algorithm to the use of prior estimations of source location and to the values of regularization parameters had been investigated."

Referee: "Also in the section 4, on the Figure 1 it is shown that the sensors are concentrated only along the wind direction, but not over the whole domain. In this case, this numerical experiment cannot be claimed as 'general' "

Response: Actually measurements covered most of the domain where nonzero concentrations had been observed and only such measurements had to be used in source inversion. Thus, the described experiment could be representative of more general situation when the available measurements totally cover the contaminated territory. The corresponding explanation had been added to the end of par. 9 of section 4:

"*Note, that detectors shown at Figure 1 almost totally covered the contaminated part of the domain within obstacle arrays and the measured concentration values spanned the range from* $1.8 \cdot 10^{-8}$ *to* $2.6 \cdot 10^{-12}$ *kg/m*^{3}. *Since very low and zero values are not used in data assimilation procedure the described experiment could be representative of more general situation when the available measurements totally cover the contaminated territory.*"

Referee: "Additionally, the choice of σ_H, σ_V is made by 'hand' and not generalized"

Response: Yes, it was out of scope of the present paper to generalize the choice of regularization parameters. Though for this particular application the choice of parameters σ_H, σ_V was already explained in par. 17 of section 4.

Referee: "Overall, after corrections the language of the article became more fluent and transparent. But still there are some proofreading mistakes. It is also recommended, to include in the Introduction the paragraph about the structure of the paper."

Response: The last paragraph had been added in Introduction explaining the structure of the paper. The language had been improved.