Least Square Data Assimilation for Identification of the Point Source Emissions

MAITHILI SHARAN,[1] SARVESH KUMAR SINGH,[1] and J. P. ISSARTEL[2]

Abstract—The identification of single and multiple-point emission sources from limited number of atmospheric concentration measurements is addressed using least square data assimilation technique. During the process, a new two-step algorithm is proposed for optimization, free from initialization and filtering singular regions in a natural way. Source intensities are expressed in terms of their locations reducing the degree of freedom of unknowns to be estimated. In addition, a strategy is suggested for reducing the computational time associated with the multiple-point source identification. The methodology is evaluated with the synthetic, pseudo-real and noisy set of measurements for two and three simultaneous point emissions. With the synthetic data, algorithm estimates the source parameters exactly same as the prescribed in all the cases. With the pseudo-real data, two and three point release locations are retrieved with an average error of 17 m and intensities are estimated on an average within a factor of 2. Finally, the advantages and limitations of the proposed methodology are discussed.

Key words: Inverse modeling, multiple-point source identification, least square, data assimilation.

1. Introduction

The rapid urbanization, industrialization and Chemical Biological and Radiological (CBR) releases raise an issue of major concern around the world regarding the contamination of environment, public health and national security. Rapid detection and an early response can reduce the extent of subsequent contamination and associated mortality of the incident. This assessment requires information about number of sources, their locations, emission rates, time, and duration of releases. However, in most of the realistic events, the difficulty increases when distributed contaminant sensor networks detect concentrations over threshold value but have no particular idea about the releases including their origin (LIU and ZHAI, 2007). This necessitates the development of a methodology that can help in identifying potential contaminant sources from limited concentration measurements.

Several researchers have focused on the problem of recovering the parameters of the unknown sources on the basis of available concentration measurements. Notable efforts in this direction are from PUDYKIEWICZ (1998), ROBERTSON and LANGNER (1998), PENENKO et al. (2002), BOCQUET (2005a, b), YEE (2005, 2006), YEE et al. (2006), KEATS et al. (2007a, b), ELBERN et al. (2007), ISSARTEL et al. (2007), SHARAN et al. (2009), etc. In all these studies, the problem of identification of source parameters was primarily restricted to single-point source. The identification of the multiple-point simultaneous releases is a challenging task and assumes significant importance in real world applications. In an extensive study, HAUPT (2005) emphasized that identification of the multiple-point sources becomes difficult when (1) two sources are seen at the same angle from the receptor but at different distances, (2) sources are located far away from the receptors and (3) meteorological conditions are not variable then distinguishing the contributions from different sources become difficult. However, the inverse problem of separating several influences merged into a set of concentration measurements has not been explored completely (ISSARTEL et al. 2011).

In recent years, few studies addressed the inverse problem of multiple-point source identification from finite number of noisy concentration measurements. MATTHES et al. (2005) attempted the identification of source locations using the least square by spatially

[1] Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India. E-mail: mathilis@cas.iitd.ac.in
[2] Centre d'Etudes du Bouchet, 5, rue Lavoisier, BP 3, 91710 Vert le Petit Cedex, France.

distributed electronic noises in an indoor release experiment for industrial storage of toxic chemicals conducted at the Sigma-Aldrich Company (Germany). Later, the identification of the multiple sources in the atmosphere was briefly addressed by YEE (2007), but the approach was based on the assumption that the number of sources is known a priori. Later, YEE (2008) extended the theory for reconstruction of an unknown number of contaminant sources using probabilistic inference in conjunction with Metropolis-coupled reversible-jump Markov chain Monte Carlo (MCMC) method. However, the implementation of the theory was computationally expensive and limited to noisy synthetic data only. Recently, LUSHI and STOCKIE (2010) described an inverse Gaussian plume approach for estimating atmospheric pollutant emissions from four-point sources in a large lead–zinc smelting operation in Trail, British Columbia. The study was performed only for the estimation of emission rates but did not include the identification of locations of point sources. Recently, ISSARTEL *et al*. (2011) proposed a renormalization algorithm for identification of multiple-point sources using the concepts from quantum mechanics and differential geometry.

Least square technique is often used in parametric estimation as well as in the data assimilation problems in various geophysical applications (LEWIS *et al*. 2006). KRYSTA *et al*. (2006) have used a least square approach as a tool in the inverse modelling to estimate the source parameters from the concentration measurements. This method is not fully explored in the source identification as (1) the method becomes computationally expensive in the source-oriented modeling involving forward computation of the advection diffusion equation (RAO, 2007), (2) singularity arises at the point of receptors in receptor-oriented modeling (RAO, 2007) and (3) classical optimization methods may not be adequate as they require a priori knowledge of the initial guess of unknown parameters, which is not feasible in reality. In view of these, an attempt has been made here to overcome some of these inherent problems in the source identification.

In the present study, an approach based on least square technique is described to retrieve the multiple-point simultaneous releases from the atmospheric concentration measurements. As a part of the study, an algorithm is proposed for minimizing the sum of square of residuals. The proposed algorithm is advantageous as (1) it does not require any prior knowledge regarding the location of the sources and (2) takes care of singularities in a natural way. In addition, an approach is coupled with the inversion algorithm to reduce the computational time taken by estimation procedure. The study is evaluated with the pseudo-real data generated from the diffusion experiment conducted at Indian Institute of Technology (IIT) Delhi, India.

2. Methodology

In this study, identification of several simultaneous point sources is explored using a finite set of concentration measurements $\mu_1, \mu_2, \ldots, \mu_n$ within the framework of least square theory (LEWIS *et al*. 2006). It is primarily assumed that the emission is continuously distributed through out the domain and no particular region is taken as a prior release location. The number of simultaneous releases is assumed to be known a priori. The emissions are taken linear with respect to the measurements. Continuous emission is considered from the ground level sources. The methodology begins with the identification of single-point emission for clarity and then it is extended to two and multiple-point simultaneous emissions.

2.1. Least Square Formulation

The least square theory is based on minimizing the error cost function of sum of squares of residuals between the receptor's measured and model predicted concentrations. The initial requirement of the formulation is a source-receptor relationship which describes mapping between source and receptors. Here, we followed a receptor oriented approach to avoid the unnecessary sampling of the source parameters and forward computation of the advection diffusion equation. The sensitivity of the potential source with respect to each sampled measurement is described by introducing the adjoint functions (MARCHUK, 1995) as:

$$\mu_i = q\,a_i \quad \text{for } i = 1, 2, \ldots, n, \qquad (1)$$

where n is the number of receptors, a_i is the adjoint function corresponding to the ith receptor and q is an unknown source strength. Adjoint functions are estimated from adjoint of the dispersion model assuming the unit release at the receptor's location (SHARAN et al, 2009). The adjoint function essentially describes backward transport of the pollutant's concentration from the receptors. Mathematically, if σ_i is the source and L is a linear operator, a direct source-receptor relationship is $L(\sigma_i) = \mu_i$. Using the properties of an inner product ($\langle .,. \rangle$), a fundamental relationship can be described as (ISSARTEL and BAVEREL, 2002), $\langle L(\sigma_i), \mu_i \rangle = \langle \sigma_i, L^*(\mu_i) \rangle = \langle \sigma_i, a_i \rangle$ in which L^* is the adjoint of the linear operator.

Generally, the concentrations measured by receptors will not agree with those predicted by the model owing to noise imposed on the concentration data and turbulence parameters, which by its varying nature is expected to have a complicated structure. For this purpose, the model predicted adjoint functions and the receptor measured concentrations are related as:

$$\mu_i = q\,a_i + \varepsilon_i \quad \text{for } i = 1, 2, \ldots, n, \qquad (2)$$

where ε_i is an additive noise associated with i^{th} concentration measurement.

The least square method for estimation of source parameters (location and intensity) of a single-point source from n observed concentration measurements μ_i's is based on minimizing the sum of square of residuals represented by the function J as (LEWIS et al. 2006):

$$J(\mathbf{x}, q) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [\mu_i - qa_i(\mathbf{x})]^2, \qquad (3)$$

subject to the constraints $q > 0$ and $\mathbf{x}_\ell \leqslant \mathbf{x} \leqslant \mathbf{x}_u$ where $\mathbf{x} = (x, y)$ is a position vector. The vectors $\mathbf{x}_\ell$ and $\mathbf{x}_u$ are, respectively, lower and upper limits of the computational domain containing the monitoring network. Measurement errors are assumed uncorrelated with equal variance (LEWIS et al. 2006).

The conditions for minimization of J lead to a system of non-linear algebraic equations in terms of parameters to be estimated. The complexity grows with increasing number of degrees of freedom or parameters. A number of optimization algorithms such as Newton–Raphson (BEYER, 1964), Steepest–Descent (DEBYE, 1909), Levenberg–Marquardt (LEVENBERG, 1944; MARQUARDT, 1963), conjugate gradient (HESTENES and STIEFEL, 1952) exist in the literature for the estimation of source parameters. These algorithms differ from the use of iterative techniques and on the appropriate choice of the initial values of the unknown parameters with respect to iterations. Since the locations of the sources are not known a priori and it is not feasible to prescribe their initial values, existing optimization methods fail for such estimation problems. In view of this, an alternative algorithm is proposed here.

The algorithm proposed here, is essentially a two step minimization process, in which as a first step, for a fixed location, the function J is minimized with respect to q to obtain its estimate and then fixed location along with estimated $\widehat{q}$ is used to compute the value of function $\widehat{J}$. This process is repeated for all the grid points of the domain and the corresponding values of $\widehat{J}$ are stored. In the second step, a sequential search algorithm is applied to look for the global minimum among the stored values of $\widehat{J}$. The parameters corresponding to the global minimum of $\widehat{J}$ will be the estimation of the source intensity and its location.

Incidentally, this algorithm can be represented mathematically for the single-point source as follows: The function J (Eq. 3) is rewritten in the matrix notation as:

$$J(\mathbf{x}, q) = \frac{1}{2}[\boldsymbol{\mu} - q\mathbf{a}(\mathbf{x})]^{\text{T}}[\boldsymbol{\mu} - q\mathbf{a}(\mathbf{x})], \qquad (4)$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^{\text{T}}$ and $\mathbf{a}(\mathbf{x}) = (a_1(\mathbf{x}), a_2(\mathbf{x}), \ldots, a_n(\mathbf{x}))^{\text{T}}$ are the vector of measurements and adjoint functions, respectively. The superscript 'T' denotes the transpose.

For a fixed $\mathbf{x}$, first order derivative of J with respect to q is given by:

$$\frac{\partial J}{\partial q} = q\mathbf{a}^{\text{T}}(\mathbf{x})\mathbf{a}(\mathbf{x}) - \boldsymbol{\mu}^{\text{T}}\mathbf{a}(\mathbf{x}). \qquad (5)$$

The condition ($\partial J/\partial q = 0$) leads to an estimate $\widehat{q}$ as:

$$\widehat{q} = \frac{\boldsymbol{\mu}^{\text{T}}\mathbf{a}(\mathbf{x})}{\mathbf{a}^{\text{T}}(\mathbf{x})\mathbf{a}(\mathbf{x})}. \qquad (6)$$

Notice that the second order derivative of J, $\partial^2 J/\partial q^2 = \mathbf{a}^T(\mathbf{x})\mathbf{a}(\mathbf{x})$ (square of adjoint function at fixed $\mathbf{x}$) is positive which implies that estimated $\widehat{q}$ (Eq. 6) minimizes the function J for a fixed $\mathbf{x}$. The minimum value of J corresponding to estimate $\widehat{q}$ (Eq. 6) for a fixed $\mathbf{x}$ is obtained from simplifying Eq. (4) as:

$$\widehat{J}\left(\mathbf{x},\widehat{q}\right) = \frac{1}{2}\left[\boldsymbol{\mu}^T\boldsymbol{\mu} - \frac{(\boldsymbol{\mu}^T\mathbf{a}(\mathbf{x}))^2}{\mathbf{a}^T(\mathbf{x})\mathbf{a}(\mathbf{x})}\right], \qquad (7)$$

Notice that, in Eq. (7) the first term on RHS (Right Hand Side) is a square of measurement vector and is independent of $\mathbf{x}$ and the second term on RHS is also a square term with negative sign. Hence, the minimization of $\widehat{J}$ with respect to $\mathbf{x}$ in Eq. (7) is equivalent to the maximization of the function

$$S(\mathbf{x}) = \frac{(\boldsymbol{\mu}^T\mathbf{a}(\mathbf{x}))^2}{\mathbf{a}^T(\mathbf{x})\mathbf{a}(\mathbf{x})}. \qquad (8)$$

The point at which $S(\mathbf{x})$ becomes maximum in the domain will be the estimate of location of the source. Once the location of the source is identified, its intensity is computed from Eq. (6) at the estimated location $\mathbf{x}_e = (x_e, y_e)$.

For single-point emission, the estimate $\widehat{q}$ (Eq. 6) turns out to be an explicit function of $\mathbf{x}$. In addition, $\widehat{J}$, the minimum value of J estimated at $\widehat{q}$ for fixed $\mathbf{x}$, becomes an explicit function of $\mathbf{x}$ which is relatively easier to optimize with respect to $\mathbf{x}$. In view of these facts, the proposed algorithm becomes simpler for single-point emission. Now, we describe the source estimation for two simultaneous releases.

2.2. Two-Point Sources

For two simultaneous point releases, the error function J is written as:

$$J(\mathbf{x}_1, \mathbf{x}_2, \mathbf{q}) = \frac{1}{2}[\boldsymbol{\mu} - q_1\mathbf{a}(\mathbf{x}_1) - q_2\mathbf{a}(\mathbf{x}_2)]^T$$
$$\times [\boldsymbol{\mu} - q_1\mathbf{a}(\mathbf{x}_1) - q_2\mathbf{a}(\mathbf{x}_2)], \qquad (9)$$

in which the components of $\mathbf{q} = (q_1, q_2)^T$ are the intensities of two simultaneous point releases corresponding to the locations $\mathbf{x}_1$ and $\mathbf{x}_2$ respectively. Henceforth, the subscripts '1' and '2' will correspond to first and second source.

For fixed $\mathbf{x}_1$ and $\mathbf{x}_2$, conditions ($\partial J/\partial q_1 = 0$ and $\partial J/\partial q_2 = 0$) for obtaining the critical points lead to a system of equations, written in matrix form, as:

$$\mathbf{Aq} = \mathbf{B}, \qquad (10)$$

where:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T\mathbf{a}_1 & \mathbf{a}_1^T\mathbf{a}_2 \\ \mathbf{a}_2^T\mathbf{a}_1 & \mathbf{a}_2^T\mathbf{a}_2 \end{bmatrix}; \mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} \mathbf{a}_1^T\boldsymbol{\mu} \\ \mathbf{a}_2^T\boldsymbol{\mu} \end{bmatrix} \qquad (11)$$

Here $\mathbf{a}_1 = \mathbf{a}(\mathbf{x}_1)$ and $\mathbf{a}_2 = \mathbf{a}(\mathbf{x}_2)$ are the vectors of adjoint functions evaluated at the points $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. Notice that $\mathbf{A}$ is the Gram matrix of $\mathbf{a}_1$ and $\mathbf{a}_2$ which implies that it is a real symmetric and positive. It is definite, i.e. det $\mathbf{A} \neq 0$, if and only if these vectors are linearly independent. Then, the system of equations can be solved for the estimated value of $\mathbf{q}$ denoted as $\widehat{\mathbf{q}}$

$$\widehat{q}_1 = \frac{\left(\mathbf{a}_1^T\boldsymbol{\mu}\right)\left(\mathbf{a}_2^T\mathbf{a}_2\right) - \left(\mathbf{a}_2^T\boldsymbol{\mu}\right)\left(\mathbf{a}_1^T\mathbf{a}_2\right)}{\det \mathbf{A}} \text{ and}$$
$$\widehat{q}_2 = \frac{\left(\mathbf{a}_2^T\boldsymbol{\mu}\right)\left(\mathbf{a}_1^T\mathbf{a}_1\right) - \left(\mathbf{a}_1^T\boldsymbol{\mu}\right)\left(\mathbf{a}_1^T\mathbf{a}_2\right)}{\det \mathbf{A}}. \qquad (12)$$

The "Hessian matrix" of $J(\mathbf{x}_1, \mathbf{x}_2, \mathbf{q})$ with respect to q_1 and q_2 (square matrix of second order partial derivatives) is simply $\mathbf{A}$. Since this matrix is positive definite implying all its eigenvalues are real and positive, it is non-singular and J attains a local minimum at the critical point $\widehat{\mathbf{q}} = \left(\widehat{q}_1 \ \widehat{q}_2\right)^T$ (Eq. 12) (AYRES, 1962; GOLUB and VANLOAN, 1996). Thus, the pairs of $\mathbf{x}_1$ and $\mathbf{x}_2$ in the domain ensuring the linear independence of $\mathbf{a}_1$ and $\mathbf{a}_2$ are considered.

The minimum value of J is obtained from Eq. (9) for $\widehat{\mathbf{q}}$ and $\mathbf{x}_1$ and $\mathbf{x}_2$ and the resulting value is denoted as $\widehat{J}$. In fact, this $\widehat{J}$ is a function of $\mathbf{x}_1$ and $\mathbf{x}_2$. Since the minimization of $\widehat{J}$ with respect to $\mathbf{x}_1$ and $\mathbf{x}_2$ is not obvious, it is minimized numerically using a sequential algorithm.

Over all algorithm for the source identification is summarized as: (1) choose the pair of $\mathbf{x}_1$ and $\mathbf{x}_2(\mathbf{x}_1 \neq \mathbf{x}_2)$ in the domain such that Hessian matrix is positive definite, (2) estimate $\widehat{\mathbf{q}}$ and $\widehat{J}$ for the pair of $\mathbf{x}_1$ and $\mathbf{x}_2$, (3) store the set $\{\widehat{J}, \widehat{\mathbf{q}}, \mathbf{x}_1, \mathbf{x}_2\}$ of values, (4) repeat the steps (1–3) for all possible pairs of $\mathbf{x}_1$ and $\mathbf{x}_2$ in the domain, (5) employ a sequential algorithm to find the minimum $\widehat{J}$ among its stored

values, (6) values of $\widehat{\mathbf{q}}$, $\mathbf{x}_1$ and $\mathbf{x}_2$ corresponding to the minimum $\widehat{J}$ in (5) will be the desired estimation of source parameters.

Now, we generalize this approach for the retrieval of m-simultaneous releases.

2.3. Multiple-Point Sources

For m unknown simultaneous point sources, the function J is written as:

$$J(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \mathbf{q}) = \frac{1}{2} \left[\boldsymbol{\mu} - \sum_{i=1}^{m} q_i \mathbf{a}(\mathbf{x}_i) \right]^{\mathrm{T}}$$
$$\times \left[\boldsymbol{\mu} - \sum_{i=1}^{m} q_i \mathbf{a}(\mathbf{x}_i) \right] : \ \mathbf{x}_i \neq \mathbf{x}_j,$$
(13)

where q_i is the intensity of the ith source and the condition $\mathbf{x}_i \neq \mathbf{x}_j$ indicates that the locations of the unknown sources are distinct. For a fixed set of $\mathbf{x}_1$, $\mathbf{x}_2,\ldots,\mathbf{x}_m$, the condition $(\partial J/\partial q_i = 0$ for $i = 1, 2,\ldots,m)$ for estimation of critical points will lead to a system of m equations, written in matrix notation as:

$$\mathbf{Aq} = \mathbf{b}, \qquad (14)$$

in which:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a_1^T a_1} & \mathbf{a_1^T a_2} & \cdot & \cdot & \cdot & \mathbf{a_1^T a_m} \\ \mathbf{a_2^T a_1} & \mathbf{a_2^T a_2} & \cdot & \cdot & \cdot & \mathbf{a_2^T a_m} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{a_m^T a_1} & \mathbf{a_m^T a_2} & \cdot & \cdot & \cdot & \mathbf{a_m^T a_m} \end{bmatrix}_{m \times m} ;$$

$$\mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ \cdot \\ \cdot \\ \cdot \\ q_m \end{bmatrix}_{m \times 1} \text{ and } \mathbf{b} = \begin{bmatrix} \mathbf{a_1^T} \boldsymbol{\mu} \\ \mathbf{a_2^T} \boldsymbol{\mu} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a_m^T} \boldsymbol{\mu} \end{bmatrix}_{m \times 1} .$$

Now, the outlines given for the retrieval of two-point emissions in Sect. 2.2 are followed for the identification of m-point sources.

The Hessian matrix generated from second order partial derivatives of J with respect to $q_1, q_2,\ldots, q_m$ for a given set $\mathbf{x}_1, \mathbf{x}_2,\ldots, \mathbf{x}_m$ is found to be the same as

the coefficient matrix $\mathbf{A}$. This matrix is examined for its positive definiteness by showing that all its computed eigenvalues are real and positive. In case any of the eigenvalue failed to be real and positive, another combination of $\mathbf{x}_1, \mathbf{x}_2,\ldots, \mathbf{x}_m$ is chosen. For the chosen set of $\mathbf{x}_1, \mathbf{x}_2,\ldots, \mathbf{x}_m$ the coefficient matrix $\mathbf{A}$ is non-singular and then the system of m equations in m unknowns (Eq. 14) is solved numerically using Gauss elimination method to compute $\widehat{\mathbf{q}}$

Now the set of $\mathbf{x}_1, \mathbf{x}_2,\ldots, \mathbf{x}_m$ along with the computed $\widehat{\mathbf{q}}$ is used to estimate the minimum value of J denoted as $\widehat{J}$ (Fig. 1). These set of values $\{\widehat{J}, \mathbf{x}_1, \mathbf{x}_2,\ldots,\mathbf{x}_m, \widehat{\mathbf{q}}\}$ are stored. This process is repeated for all the possible combination of $\mathbf{x}_1, \mathbf{x}_2,\ldots, \mathbf{x}_m$ in the domain. A sequential algorithm is utilized for searching the minimum of $\widehat{J}$ among all its stored values. The values of $\mathbf{x}_1, \mathbf{x}_2,\ldots, \mathbf{x}_m$ and $\widehat{\mathbf{q}}$ corresponding to the min $(\widehat{J})$ are identified as the source parameters. These outlines are given in a flow diagram (Fig. 1).

3. Diffusion Data

For evaluation of the proposed technique for the source retrieval, a diffusion data is required. The measurements utilized for identification of single as well as multiple-point emissions are described here briefly.

3.1. Measurements for Single-Point Emission

For the identification of single-point emission, data from IIT diffusion experiment conducted at Delhi (28°52′ N, 77°18′ E) for surface release of tracer SF_6 in low-wind conditions is considered. Details of the experiment including atmospheric stability are given in SINGH et al. (1991) and SHARAN et al. (1996). The tracer was released at a height of 1 m above the ground and the receptors were also placed approximately at the same height. For the computations, source and receptors are assumed at ground level.

In all, 14 test runs were conducted. Only seven (runs 1, 6, 7, 8, 11, 12 and 13) corresponding to the unstable steady conditions are chosen for the analysis (SHARAN et al., 2009). Run 2 was ignored due to a relatively large variability in wind direction and the
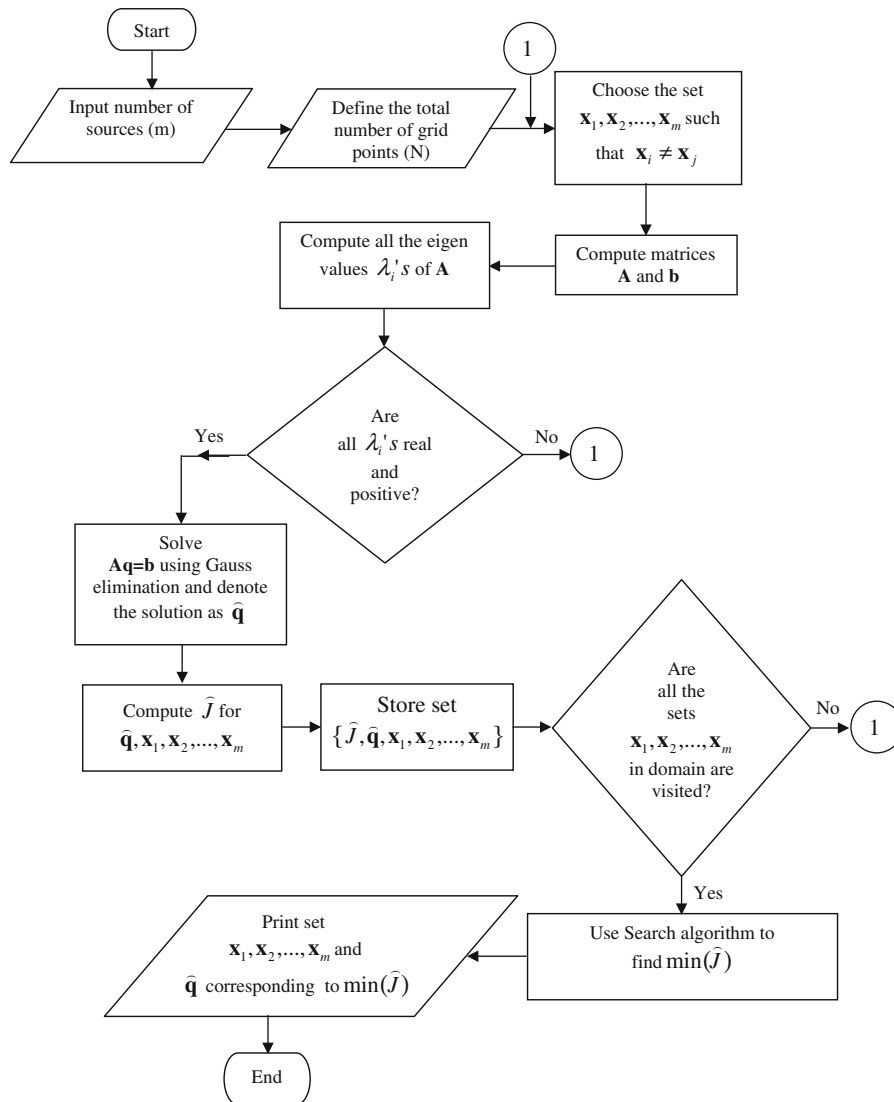
Figure 1
Flow-diagram for the steps of proposed retrieval algorithm

remaining runs because of neutral and stable conditions. In runs 1, 6, 7 and 11, the release point was located at the centre of the circular arcs, whereas in the remaining runs (8, 12 and 13), the release point was shifted 100 m towards northwest. In runs 1, 8, 12 and 13, the release rate was 5,000 µg/s while in the remaining runs (6, 7 and 11), it was 3,000 µg/s. The monitoring network involved 20 receptors on 50, 100 and 150 m, and in some cases on 200 m circular arcs with 45° angular spacing between them (Fig. 2). When the source is shifted from the centre onto the arc at 100 m, the corresponding receptor is moved to the centre. In some of the runs, a few measurements are discarded (SHARAN *et al.*, 1996), thus reducing the effective number of receptors in the monitoring network.

Wind and temperature measurements were obtained at four levels (2, 4, 15 and 30 m) from a 30 m micrometeorological tower. The values of wind speed, wind direction, atmospheric stability and mixing height required in the dispersion model are taken from SHARAN *et al.* (1996).
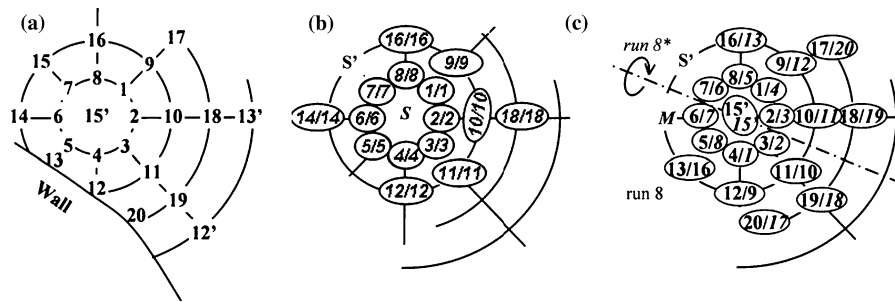
Figure 2
Layout of the site for IIT Delhi tracer diffusion experiment (**a**). The samplers 1–20, 12′, 13′ and 15′ are essentially arranged on circles of radius 50, 100 and 150 m, with regular angular spacing of 45°. In case of shifted source runs, source at the centre S and receptor 15 are interchanged with each other corresponding to positions S′ and 15′ (adopted from SHARAN *et al.* 2009). **b** Representation of pseudo-runs 11–13 and 11–8 obtained by combining the run 11 with central source at S and runs 13 and 8 with shifted source at S′. Within each ellipse the right and left *italic numbers* indicate a sampler respectively in run 11 and 13. **c** Schematic similar to **b** for pseudo-run 8–8* in which run 8 is combined with itself after a mirror symmetry sending the source to M (adopted from ISSARTEL *et al.* 2011)

3.2. Pseudo Measurements for Multiple-point Emissions

Diffusion data is very scarce. Mostly available diffusion data in the literature pertains to the single-point emission. For the evaluation purpose, a pseudo data set is generated from the IIT diffusion experiment by combining the single release with different source locations under similar meteorological conditions.

Notice that, in IIT data, the monitoring network remains same for all the runs only the release location and wind direction vary. Thus, it allows us, with minor modifications, to combine the runs with similar meteorological conditions and different release locations, just as if the two releases had occurred simultaneously. In view of this, four pseudo runs have been prepared by combining the runs with same wind conditions and different location of the source.

Winds are almost same in runs 11 (central source), 8 and 13 (shifted source). Runs 11 (central source), 8 and 13 (shifted source) have almost same wind speed (1.1, 0.9, 1.1 m/s respectively) and direction (125°, 121°, 141°). The pseudo-runs 11–8 and 11–13 are thus prepared by adding the measurements at the receptors common to the combined runs.

A third pseudo-run is prepared based on the fact that wind direction in run eight (shifted source) is close to a symmetry axis of the monitoring network in the direction of 112.5°. Under this mirror symmetry, the source 100 m northwest of the centre of the monitoring network becomes 100 m west of it. Most receptors coincide with the mirror image of another receptor (Fig. 2). The pseudo-run 8–8* is obtained by adding the concentrations observed at the common receptors; the mirror symmetric run is indicated by a star (*).

Finally, a next pseudo-run 11–8–8* is prepared by combining the runs 11, 8 and mirror symmetric run 8* to simulate a triple-point release. For the pseudo run 11–8–8*, wind direction is taken in the direction of 112.5° to follow the mirror symmetry of runs 8 and 8*, and wind speed is averaged between runs 8 and 11.

4. Numerical Computations

The numerical computations are performed for retrieval of both single as well as multiple-point emissions. In both the cases, the computations are performed for (1) the complete set of concentrations observed or pseudo generated and (2) the analogous set of synthetic measurements generated using the dispersion model given in SHARAN *et al.* (1996). A primary step of the inversion algorithm is the generation of adjoint functions. The adjoint functions are generated from the dispersion model (SHARAN *et al.*, 1996) in the backward mode assuming unit release at the receptor's locations and rotating the wind direction by 180° (SHARAN *et al.*, 2009). The implementation of the inversion algorithm requires a discretization of the domain. For this purpose, a square domain of size 995 m × 995 m is chosen and

discretized into 199 × 199 grid points. Each mesh is a square of 5 m × 5 m. Notice that no more than one receptor or source lies in the same grid. The domain and grid discretization remain same for the single as well as multiple-point emissions. The centre of the arcs is at the grid point (100, 100). Depending on the run, source is either at (100, 100) or (86, 114). Note that adjoint function is singular at the position of the receptor. In order to resolve this singularity, the mesh containing the receptor and the neighboring meshes are further subdivided into 99 × 99 grid points and an average value of the adjoint function is computed for the receptor's mesh. The steps for the inversion algorithm for single as well as multiple-point emissions are discussed in Sect. 2. The computations are performed on an Intel® Core™ 2 Duo E8135 @ 2.66 GHz desktop machine.

5. Results and Discussion

Source estimation has been carried out for single as well as two- and three-point emissions using two types of data: (1) synthetic and (2) real. Synthetic data for each run is the concentration at the receptors generated from the dispersion model, described as noise free and minimizes the errors associated with the model. Therefore, it is used to verify the mathematical consistency of the inversion algorithm. However, real data corresponds to concentration measurements sampled during the experiments. Accordingly, the results are presented in the following subsections.

5.1. Single-Point Emission

The single-point source reconstruction is performed with the seven runs (1, 6, 7, 8, 11, 12 and 13) of IIT-Delhi diffusion experiment. With the synthetic data, the location of the source in the runs (1, 6, 7 and 11) having the source at the centre of the monitoring network is retrieved exactly at the grid point (100, 100) whereas in the runs (8, 12 and 13) with shifted release, the retrieved location (86, 114) coincides with that prescribed (Table 1). The intensity in the runs (1, 8, 12 and 13) is retrieved approximately 2,999.8 µg/s in lieu of 3,000 µg/s. Similarly an approximate 4,999.7 µg/s intensity is retrieved in place of prescribed 5,000 µg/s in runs (6, 7 and 11). The exact retrieval of the source parameters with the synthetic data provides a mathematical consistency of the technique used here.

With real data, source is reconstructed in all the runs with an average error of 20.5 m from the original release locations (Table 1). The location is estimated in run six at the grid point (100, 100) with a negligible error and at point (92, 102) in run seven with a maximum error of 41.2 m. The intensity of the source is estimated within a factor two in all the runs (Table 1).

5.2. Errors in the Retrieval

With the real data, the errors in the retrieval are described in terms of the departure of the real concentration measurements from their ideal synthetic values. This departure is traditionally called

Table 1

Reconstruction results for single-point release

Run	1	6	7	8	11	12	13
Experimental release							
Location	(100, 100)	(100, 100)	(100, 100)	(86, 114)	(100, 100)	(86, 114)	(86, 114)
Intensity	5,000	3,000	3,000	5,000	3,000	5,000	5,000
Least square estimate							
Location	(99, 103)	(100, 100)	(92, 102)	(91, 111)	(97, 103)	(82, 117)	(88, 115)
Intensity	9,529	5,967	4,928	4,155	5,279	9,544	3,664
Error							
E_L	15.8	0	41.2	29.2	21.2	25	11.2
θ_e^0	17.3	25.2	20	4.7	13.7	10.1	6.1

The experimental point releases (first and second rows) with least square estimates (third and fourth rows) are indicated in terms of location in grid and intensity in µg/s. On the fifth and sixth rows, inversion error is presented in terms of departure of retrieved location from experimental release location grid coordinate (represented as E_L in meters) and the angular deviation (θ_e^0 in degrees)

measurement errors (COHN, 1997) though it is due to not only instrumental inaccuracy but also to lack of representativity of the dispersion model. For the ideal measurements, the noise free measurement vector always lies along or parallel to the adjoint vector (Eq. 1). However, for a noisy measurement, the equality relation (Eq. 1) will not be true. Thus, the measurement vector will lie along an angle to the adjoint vector. This angular deviation indicates the error incurred in the retrieval of intensity.

As explained in Sect. 2, the location of the source is estimated from the observed noisy measurements μ_r at a location $\mathbf{x}_e$ such that $S(\mathbf{x}_e) = \frac{(\mu_r^T \mathbf{a}(\mathbf{x}_e))^2}{\mathbf{a}^T(\mathbf{x}_e)\mathbf{a}(\mathbf{x}_e)}$ is maximum. This is equivalent to maximize $\frac{S(\mathbf{x}_e)}{\|\mu_r\|^2} = \left[\left(\frac{\mu_r}{\|\mu_r\|}\right)^T \frac{\mathbf{a}(\mathbf{x}_e)}{\|\mathbf{a}(\mathbf{x}_e)\|}\right]^2$, where $\|\mathbf{p}\|^2 = \mathbf{p}^T\mathbf{p}$ for any vector $\mathbf{p}$. Since vectors $\frac{\mu_r}{\|\mu_r\|}$ and $\frac{\mathbf{a}(\mathbf{x}_e)}{\|\mathbf{a}(\mathbf{x}_e)\|}$ both have norm 1, the estimated location $\mathbf{x}_e$ of the release is the one such that the vector $\frac{\mathbf{a}(\mathbf{x}_e)}{\|\mathbf{a}(\mathbf{x}_e)\|}$ minimizes the angular distance $\theta_e = \arccos\left(\left(\frac{\mu_r}{\|\mu_r\|}\right)^T \frac{\mathbf{a}(\mathbf{x}_e)}{\|\mathbf{a}(\mathbf{x}_e)\|}\right)$ from the observations μ_r. Thus, this angular departure θ_e provides an indication of the intensity of the noise contained in the data. If there is no noise in the data, i.e. $\mu_r = \mu = q_0\mathbf{a}(\mathbf{x}_0)$, one would obtain $\mathbf{x}_e = \mathbf{x}_0$ with $\theta_e = 0$. This is verified in the computations with the synthetic data, where the estimates of intensities are almost exactly same to those prescribed.

Similarly, error estimates are computed for real data in all the runs (Table 1) and it is observed that the adjoint vector does not coincide with the measurement vector exactly and depart by an average angular distance of 13.8° in all the runs. A relatively higher value of angular departure θ_e indicates a relatively large deviation in the estimated intensity from that prescribed.

5.3. Two and Three Simultaneous Point Sources

The proposed algorithm is used for the identification of multiple-point sources with the pseudo-real as well as corresponding model generated synthetic data for (1) two unknown sources (runs 11–13, 11–8 and 8–8*) and (2) three unknown sources (run 11–8–8*). With the synthetic data, the grid points [(100, 100), (86, 114)] are estimated as two point release locations in case of runs (11–13 and 11–8) and grid points [(86, 114), (80, 100)] in run 8–8* which are exactly same as prescribed. Similarly, in run 11–8–8*, the points [(100, 100), (86, 114), (80, 100)] are estimated as three source locations exactly similar to those prescribed. In all these four runs, the strength at each location is estimated same as prescribed within a maximum round-off error of 0.02%.

With the pseudo-real data, in average, the releases are retrieved 17 m away from their true locations (Table 2). In two-point sources (runs 11–13, 11–8 and 8–8*), the locations are retrieved at a minimum error of 5 m in run 11–13 and maximum 40 m in run 8–8* from their original release locations. In runs (11–13 and 11–8), the corresponding intensities are retrieved within a factor of 2. However, in run 8–8*, the error in the retrieval is relatively large and increases up to a factor 3.8. Similarly, in three-point sources (run 11–8–8*), all the three source locations

Table 2

Reconstruction results for two and three point simultaneous releases with pseudo-real data

Pseudo run	11–13 SS′		11–8 SS′		8–8* SM		11–8–8* SS′M		
Experimental release									
Location	(100, 100)	(86, 114)	(100, 100)	(86, 114)	(86, 114)	(80, 100)	(100, 100)	(86, 114)	(80, 100)
Intensity	3,000	5,000	3,000	5,000	5,000	5,000	3,000	5,000	5,000
Least square estimate									
Location	(103,104)	(88,116)	(99,102)	(86,115)	(78,114)	(80,101)	(97,101)	(82,115)	(82,100)
Intensity	3,403	4,357	3,970	7,115	7,449	18,914	4,715	9,623	5,159
Error									
E_L	25	14	11	5	40	5	16	21	10

Notations and legends are same as in Table 1

(a)

(b)
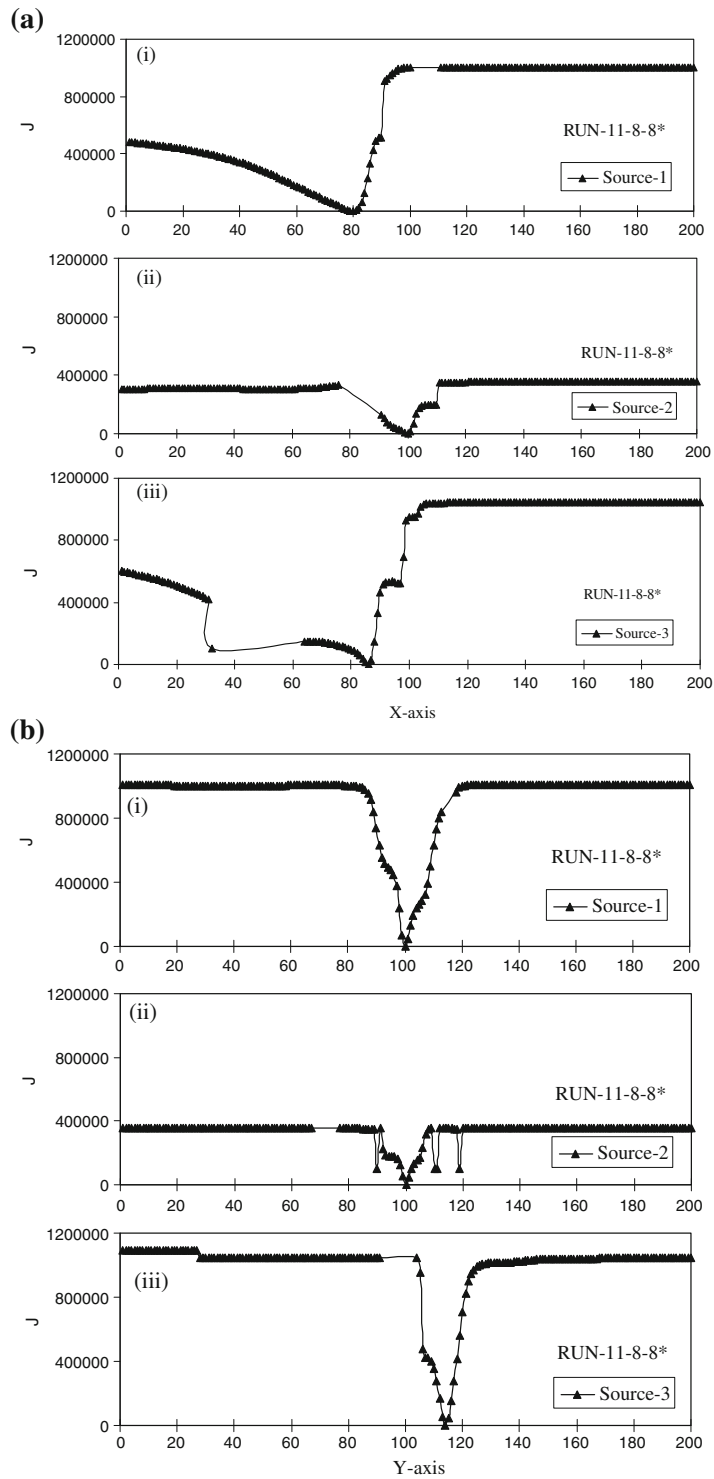
are retrieved within 21 m and intensities are retrieved within a factor of 2. A slight discrepancy may be explained in view of the facts that (1) in average, the real runs contain 19 samplers but only 16 can be combined in the pseudo-runs and (2) the pseudo-runs are not natural runs and the averaging of the winds

◀

Figure 3

a The variation of function J along x-axis in three-point sources (run 11–8–8*) for (1) fixed y-coordinate (grid point 100) of *source-1* and other two fixed sources at grid point (100,100) (*source-2*) and (86, 114) (*source-3*) (2) fixed y-coordinate (100) of source-2 and other two fixed sources at grid point (80,100) (*source-1*) and (86, 114) (*source-3*) and (3) fixed y-coordinate (114) of *source-3* and other two fixed sources at grid point (80,100) (*source-1*) and (100, 100) (*source-2*). **b** The variation of function J along x-axis in three-point sources (run 11–8–8*) for (1) fixed x-coordinate (grid point 80) of *source-1* and other two fixed sources at grid point (100,100) (*source-2*) and (86, 114) (*source-3*), (2) fixed x-coordinate (100) of *source-2* and other two fixed sources at grid point (80,100) (*source-1*) and (86, 114) (*source-3*) and (3) fixed x-coordinate (86) of *source-3* and other two fixed sources at grid point (80,100) (*source-1*) and (100, 100) (*source-2*)

from two real runs is not fully consistent with the addition of the corresponding concentration measurements.

In case of single-point emission, an expression is obtained to quantify the error incurred in the retrieval of intensity in terms of angular departure of the measurement vector from the adjoint vector. However, such an expression is not feasible to quantify the errors in case of more than one-point sources. In addition, it is difficult with the real data to quantify the measurement errors in which representativity term is likely to be dominant.

In the present study, the identification of two-point and three-point sources includes estimation of six and nine unknown parameters respectively. Since intensities are estimated as a function of locations of the sources, the degree of freedom of source identification of the two- and three-point simultaneous releases reduces to four and six respectively. Even in this case, the visualization of all these estimated parameters for two- as well as three-point sources is not possible collectively on a paper or in two-dimension. To overcome this, a representation is drawn in Fig. 3 showing the variation of the function J in case of three-point sources (run 11–8–8*) with respect to each estimated parameters keeping the other source parameters constant. It clearly shows the occurrence of global minimum of J at the estimated parameters.

5.4. Reduction in Computational Time

The general methodology for minimization of function J for a number of simultaneous point emissions is based on visiting all the doublets or

triplets or m-set of grid points in the domain for two-, three- and m-point sources, respectively. This approach is time consuming as the number of pairs $N(N-1)/2 \approx 7.8 \times 10^8$ (N = total number of grid points in the domain) for two-point sources or triplets $N(N-1)(N-2)/6 \approx 10^{13}$ for three-point sources or m-set $N(N-1)\ldots(N-m+1)/(m!) \approx 10^{4m+1}$ for m-point sources involved in the optimization are very large. The computational time taken in the estimation of source parameters for simultaneous three-point releases is approximately 90 h. This is expected to increase further with the increase in the number of sources.

In order to minimize the computational time, an alternative approach is adopted here. In this approach, the source parameters are estimated in two steps: (1) as a first step, a gross estimation of the source parameters is carried out by visiting one point out of five grid points in each direction using the algorithm described in Sect. 2 and then (2) these estimated set of source locations are refined by using the same retrieval algorithm for all the grossly estimated locations in their 11-grid point neighborhood. This reduces roughly the visit of number of pairs by a factor of 5^4, number of triplets by 5^6 and m-set by 5^{2m} for two-, three- and m-sources respectively. This approach reduces the computational time to 2 min for two-point and 12 min for three-point simultaneous sources. However, both the generalized and modified approaches lead to the similar results.

5.5. Noisy Measurements

Generally, the concentration measurements are not available for the multiple-point emissions for evaluating the inversion approach for the retrieval of more than one-point emission sources. In such a situation, researchers (YEE, 2008; LUSHI and STOCKIE, 2010) have generated the concentration measurements for evaluation purposes by adding a proportion of noise in the model generated concentrations. The numerical simulations with such noisy measurements are also important for understanding the effect of possible errors in concentration measurements on emission estimates (LUSHI and STOCKIE, 2010).

The model generated concentration measurements are scaled by a normally distributed random number

chosen from the interval $(1-\alpha, 1 + \alpha)$ for values of $\alpha = 0.1, 0.2$ and 0.3. These values of α correspond to 10, 20 and 30% noise. The error in the location as well as in the intensity is found to increase as the percentage of noise increases in the concentration measurements (Table 3). In all four pseudo-runs, the maximum error in the estimation of source location increases up to 16 m from the original release location and departure in the retrieval of intensities increases up to 30% in comparison to the prescribed with the increase in random noise (10–30%).

6. Advantages and Limitations

The present study is focused for the identification of multiple-point surface releases in steady state conditions using a least square approach. The algorithm is described within a general framework for the retrieval of *m*-point emission sources. This is applied (1) with the real data for single-point source and (2) with pseudo-real corresponding to two and three-point emission sources. The source parameters are retrieved almost exactly with the model generated synthetic data. These parameters are also retrieved with real and pseudo-real data. In the following, we discuss the advantages/limitations with the approach used here for the identification of sources.

6.1. Advantages

The proposed inversion algorithm is based on simple concepts of least square theory. This algorithm is advantageous in comparison to other methods as:

1. Minimization of the sum of the squares of residuals (Eq. 3) leads to a system of non-linear algebraic equations which needs an iterative algorithm to determine an approximate solution. The existing algorithms have limited applicability as they require the initial value of the parameters to begin the iterative process. However, such information is, in fact, not needed in the proposed algorithm as in reality there is no idea from where plume has originated.
2. The set of locations for which system of equations becomes ill-posed or ill-conditioned are discarded

in a natural way and computations are performed only for those set of locations where a local minimum of the function J exists.
3. Source intensities are expressed in terms of their locations reducing the degree of freedom of unknowns to be estimated.
4. The proposed algorithm allows visiting a relatively less number of set of grid points in order to search for the location of the sources, resulting in a significant reduction in the computational time.

6.2. Assumptions/Limitations

The proposed inversion algorithm is subject to the following assumptions: (1) emissions and the concentration measurements are related linearly, (2) the observation error covariance matrix is taken as an identity matrix, while in reality observations may have small order of covariance, (3) the domain is discretized into grids in such a way that no more than one source or receptor lies in the same grid and (4) the dispersion model used here describes perfect relationship between the emission and the corresponding measurements implying the model representativity errors, if any, are negligibly small. However, in reality, models can not be exactly representative of observations in all the atmospheric conditions. The uncertainty in meteorological variables influences the representation of the concentration measurements by a dispersion model which eventually affects the source retrieval. Recently, we are able to retrieve the source reasonably well (SHARAN et al., 2009) with IIT diffusion data in convective conditions in spite of a sparse monitoring network whereas the sources are severely under-estimated with the Idaho low wind diffusion experiment in stable conditions even with a dense monitoring network because of the model representativity errors arising due to large variability in the wind direction (SHARAN et al., 2011). This aspect is being investigated further.

The algorithm is limited in its applicability as (1) the number of sources is known a priori, (2) adjoint functions are not assigned any weights according to the visibility of the region from the monitoring network, however, in a recent study, an improvement

Table 3

Reconstruction results for two and three point simultaneous releases with the noisy synthetic data

Pseudo run	Noise (%)	Source parameters	11–13 SS′		11–8 SS′		8–8* SM		11–8–8* SS′M		
Experimental release	0	Location	(100, 100)	(86, 114)	(100, 100)	(86, 114)	(86, 114)	(80, 100)	(100, 100)	(86, 114)	(80, 100)
		Intensity	3,000	5,000	3,000	5,000	5,000	5,000	3,000	5,000	5,000
Least square estimate	10	Location	(100, 99)	(86, 115)	(102, 99)	(86, 114)	(86, 114)	(81, 100)	(97, 101)	(84, 115)	(80, 100)
		E_L	5	5	11.2	0	0	5	15.8	11.2	0
		Intensity	2,984	5,243	2,439	5,421	5,329	4,351	3,518	6,037	5,198
		E_I	0.53	4.9	18.7	8.4	6.6	13	17.3	20.7	4
	20	Location	(100, 99)	(86, 113)	(102, 99)	(87, 114)	(85, 115)	(79, 100)	(98, 101)	(83, 115)	(80, 100)
		E_L	5	5	11.2	5	7	7	11.2	15.8	0
		Intensity	2,860	4,859	2,857	4,539	5,830	5,043	2,641	6,512	5,448
		E_I	4.7	2.8	4.8	9.2	16.6	0.86	12	30.2	9
	30	Location	(100, 99)	(87, 114)	(103, 99)	(86, 113)	(88, 112)	(81, 100)	(100, 99)	(88, 113)	(80, 100)
		E_L	5	5	15.8	5	14.1	5	5	11.2	0
		Intensity	2,987	4,093	2,168	4,535	5,393	5,195	2,873	4,360	4,633
		E_I	0.43	18.1	27.7	9.3	7.9	3.9	4.2	12.8	7.3

The experimental point releases (first and second rows) are shown in terms of location in grid and intensity in ug/s. The least square estimates are presented with 10%, 20% and 30% noises in third, seventh, and eleventh rows in terms of location and in fifth, ninth and thirteenth rows in terms of intensity. On the fourth, eighth and twelfth rows, error E_L in meter is indicated corresponding to noises

The departure in the retrieved intensity from the prescribed release is shown as E_I (in %)

in this direction is proposed by ISSARTEL *et al.* (2007) by renormalizing the domain with the weights according to available visibility from the direction of monitoring network, (3) computational complexity of the method increases in proportion to the number of unknown sources and number of grids in the computational domain and (4) the quantification of errors individually related to each source in the multiple-point source identification is not feasible at this moment, however, a combined estimate of error is reflected by the value of the function J.

6.3. Data Limitation

The proposed algorithm for the identification of sources is evaluated with the (1) model generated synthetic data, (2) noisy data obtained from synthetic measurements after adding a random noise and (3) real data in single-point emission and pseudo real data for two and three point emission sources. The data from IIT diffusion experiment used here corresponds to primarily the single-point emission in convective conditions. Further, the monitoring network is very sparse in this data set. In general, the diffusion data is very limited even for single-point emission sources. Existing studies on the retrieval of multiple-point emissions utilize the noisy synthetic data. Recently, an attempt has been made in the literature (LUSHI and STOCKIE, 2010) to design an experiment with four-point emission sources. Thus, the technique proposed here needs to be evaluated further with the availability of the concentration measurements not only with the single-point emissions in different atmospheric stability conditions but as well as from the simultaneous releases from more than one-point emission sources.

7. Conclusions

In this paper, we have presented an inversion approach based on least square technique for single- and multiple-point source estimation from limited number of atmospheric concentration measurements. The source estimation method is based on two-step minimization of sum of square of residuals between the receptor measured and the model predicted concentrations. The novelty of our approach stems from its simplicity and advantages in comparison to the other classical optimization methods.

The proposed algorithm has been successfully applied to identify the single as well as two and three simultaneous point emissions from synthetic, noisy and with real or pseudo-real concentration measurements from IIT diffusion experiment. With the synthetic measurements, release locations and intensities are retrieved exactly in all the runs for single as well as two and three simultaneous point releases. For the noisy measurements, it is observed that the retrieval error grows significantly as the noise in the concentration measurement increases. The retrieval of source is presented with noisy measurements obtained by adding 10, 20 and 30% random noise to the model generated concentrations.

In case of single-point release with real data, the source is identified with an average error of 20.5 m from the original source location. The corresponding intensity is retrieved within a factor of two in all the runs. An error estimate for the departure of intensities is also given in terms of angular departure of measurement vector from the corresponding model generated vector.

With the pseudo-real measurements, two and three point release locations are retrieved with an average error of 17 m and intensities are estimated on an average within a factor of two. The incurred errors in the retrieval are correlated with the errors involved in the observations and dispersion model. In addition, an alternative simplified approach is proposed in order to reduce the computational time required in the estimation of source parameters.

REFERENCES

AYRES, F. JR., Schaum's Outline of Theory and Problems of Matrices, (Schaum, New York 1962), 219 pp.

BEYER, W.A. (1964), A Note on Starting the Newton-Raphson Method, Communications of the ACM (CACM) 7, 442 pp.

BOCQUET, M. (2005a), *Reconstruction of an atmospheric tracer source using the principle of maximum entropy-I: Theory*, Q. J. R. Meteorol. Soc. 131, 2191–2208.

BOCQUET, M. (2005b), *Reconstruction of an atmospheric tracer source using the principle of maximum entropy-II: Applications*, Q. J. R. Meteorol. Soc. 131, 2209–2223.

COHN, S.E. (1997), *An introduction to estimation theory*, J. of the Met. Soc. of Japan 75, 1B, 257–288.

DEBYE, P. (1909), *Näherungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index*, Mathematische Annalen *67*(4), 535–558.

ELBERN H., STRUNK A. SCHMIDT H., TALAGRAND O. (2007), *Emission rate and chemical state estimation by 4-dimensional variational inversion*, Atmos. Chem. Phys. *7*, 3749–3769.

GOLUB, G.H. and VANLOAN, C.F., Matrix computations, (Johns Hopkins University Press, Baltimore 1996), 694 pp.

HAUPT, S.E. (2005), A demonstration of coupled receptor/dispersion modelling with a genetic algorithm, Atmos. Environ. *39*, 7181–7189.

HESTENES, M.R. and STIEFEL, E. (1952), *Methods of Conjugate Gradients for Solving Linear Systems*, Journal of Research of the National Bureau of Standards *49*, 409–436.

ISSARTEL, J.-P., BAVEREL, J. (2002), *Adjoint backtracking for the verification of the Comprehensive Nuclear Test Ban Treaty*, Atmospheric Chemistry and Physics Discussions *2*, 2133–2150.

ISSARTEL, J.P., SHARAN, M. and MODANI, M. (2007), *An inversion technique to retrieve the source of a tracer with an application to synthetic satellite measurements*, Proc. R. Soc. A *463*, 2863–2886.

ISSARTEL, J.P., SHARAN, M. and SINGH S.K. (2011), A retrieval technique for multiple point emissions from atmospheric concentration measurements, Submitted for publication.

KEATS, A., YEE, E. and LIEN, F.S. (2007a), *Bayesian inference for source determination with applications to a complex urban environment*, Atmos. Environ. *41*, 465–479.

KEATS, A., YEE, E. and LIEN, F.S. (2007b), *Efficiently characterizing the origin and decay rate of a nonconservative scalar using probability theory*, Ecological Modelling *205*, 437–452.

KRYSTA, M., BOCQUET, M., SPORTISSE, B. and ISNARD, O. (2006), *Data assimilation for short-range dispersion of radionuclides: An application to wind tunnel data*, Atmos. Environ. *40*, 7267–7279.

LEVENBERG, K. (1944), *A Method for the Solution of Certain Non-Linear Problems in Least Squares*, The Quarterly of Applied Mathematics *2*, 164–168.

LEWIS, J.M., LAKSHMIVARAHAN, S. and DHALL S.K., Dynamic Data Assimilation: A Least Square Approach, (Cambridge University Press, 2006), 654 pp.

LIU, X. and ZHAI, Z. (2007), *Inverse modeling methods for indoor airborne pollutant tracking: literature review and fundamentals*, Indoor Air *17*, 419–438.

LUSHI, E. and STOCKIE, J.M. (2010), *An inverse Gaussian plume approach for estimating atmospheric pollutant emissions from multiple point sources*, Atmos. Environ. *44*, 1097–1107.

MARCHUK, G.I., *Adjoint equations and analysis of complex systems*, (Kluver Academic Publishers, Dordrecht, 1995).

MARQUARDT, D. (1963), *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, SIAM Journal on Applied Mathematics *11*, 431–441.

MATTHES, J., GRÖLL, L. and KELLER, H.B. (2005), *Source localization by spatially distributed electronic noses for advection and diffusion*, IEEE Trans. Signal Process *53*, 1711–1719.

PENENKO, V., BAKLANOV, A. and TSVETOVA, E. (2002), *Methods of sensitivity theory and inverse modelling for estimation of source term.* Future Generation Computer Systems *18*, 661–671.

PUDYKIEWICZ, J.A. (1998), *Application of adjoint tracer transport equations for evaluating source parameters*, Atmos. Environ. *32*, 3039–3050.

RAO, K.S. (2007), *Source estimation methods for atmospheric dispersion*, Atmos. Environ. *41*, 6964–6973.

ROBERTSON, L. and LANGNER, J. (1998), *Source function estimate by means of variational data assimilation applied to the ETEX-I tracer experiment*, Atmos. Environ. *32*, 4219–4225.

SHARAN, M., ISSARTEL, J.P., SINGH, S. K. and KUMAR, P. (2009), *An inversion technique for the retrieval of single-point emissions from atmospheric concentration measurements*, Proc. R. Soc. A *465*, 2069–2088.

SHARAN, M., ISSARTEL, J.P. and SINGH, S. K. (2011), A point-source reconstruction from concentration measurements in low wind stable conditions, Submitted for publication.

SHARAN, M., SINGH, M.P., YADAV, A.K., AGGARWAL, P. and NIGAM, S. (1996), *A mathematical model for the dispersion of pollutants in low wind conditions*, Atmos. Environ. *30*, 1209–1220.

SINGH, M.P., AGARWAL, P., NIGAM, S. and GULATI, A. (1991), Tracer experiments—a report. Technical report, Centre for Atmospheric Science, IIT Delhi.

YEE, E. (2005), Probabilistic inference: an application to the inverse problem of source function estimation, The Technical Cooperation Program (TTCP) Chemical and Biological Defence (CBD) Group Technical Panel 9 (TP-9) Annual Meeting, Defence Science and Technology Organization, Melbourne, Australia.

YEE, E. (2006), A Bayesian approach for reconstruction of the characteristics of a localized pollutant source from a small number of concentration measurements obtained by spatially distributed "electronic noses", Russian-Canadian Workshop on Modeling of Atmospheric Dispersion of Weapon Agents, Karpov Institute of Physical Chemistry, Moscow, Russia.

YEE, E. (2007), Bayesian probabilistic approach for inverse source determination from limited and noisy chemical or biological sensor concentration measurements, Chemical and Biological Sensing VIII (Augustus W. Fountain III, ed), Proc of SPIE *6554*, 12 pp.

YEE, E. (2008), *Theory for Reconstruction of an Unknown Number of Contaminant Sources using Probabilistic Inference*, Boundary-Layer Meteorology *127*, 359–394.

YEE, E., Lien, F.S., Keats, A., Hseih, K.J. and D'Amours, R. (2006), Validation of Bayesian inference for emission source distribution using the Joint Urban 2003 and European Tracer Experimnets, Fourth International Symposium on Computational Wind Engineering (CWE2006), Yokohama, Japan, 4 pp.