# Source Parameter Estimation of Atmospheric Pollution from Accidental Gas Releases

**P. Kathirgamanathan**
Dept of Mathematics and Statistics, University of Jaffna, Jaffna, Sri Lanka
pkathirgamanathan@yahoo.co.uk

**Abstract:** This paper presents the development of an inverse model that may be used to estimate the source term parameters for a polluting gas released into the atmosphere from a point above the ground. The model uses measured pollution concentrations at observation sites on the ground as well as meteorological data such as wind speed and cloud cover. The inverse model is formulated as a least- squares minimisation problem coupled with the solution of an advection-dispersion equation. The minimisation problem where the pollutants are released instantaneously is well-posed and the source term is calculated with reasonable accuracy. However, the problem with a non-steady extended release source is ill-posed; consequently, its solution is extremely sensitive to errors in the measurement data. Tikhonov's regularisation, which stabilises the solution process, is used to overcome the ill-posedness of this problem and the regularisation parameter is estimated using the properties of the non-linear L-curve, and Wahhba's leaving-out-one lemma. Finally, the accuracy of the model is examined by imposing normally-distributed relative noise into concentration data generated by the forward model.

***Keywords:*** *Non-linear ill posed problem; Inverse air pollution model; Parameter estimation*

## 1. INTRODUCTION

The analysis process for accidental gas releases are categorised into 4 cases as follows (Kathirgamanathan *et al.*, 2001, 2003a, 2003b): (1) instantaneous release from a known location, (2) instantaneous release from an unknown location, (3) extended release over a period of time from a known location, (4) extended release over a period of time from an unknown location. In this paper we propose a methodology for Case 4 where pollution originates from a point source with an extended release over a period of time from an unknown location. That is, we consider the problem where the transport properties of the medium are assumed to be known but the location and release history of the pollution are unknown.

The methodology for estimating the location and release rates of pollution sources is based on the solution of an advection-dispersion equation and a least squares technique. The least squares technique optimises the agreement between measured and model-predicted concentration by varying the model input parameters within reasonable ranges of uncertainties. In the solution process, the unknown release rate function is discretised into many components. The relationship between the concentration of pollution, $C$, and the discretised components is linear and that between $C$ and the location parameters is non-linear. Therefore the problem involves estimating both non-linear and linear parameters. It has been shown already (Kathirgamanathan *et al.*, 2003) that estimating linear parameters (release rates) for given non-linear parameters (location) is a linear ill-posed problem. Therefore estimating release rates from a source of unknown location is a non-linear ill-posed problem. This problem is further complicated by inexact information. In reality, the data contains measurement errors, so the true solution will not fit the data. Also the meteorological parameters are not known exactly. In this model, these values are assumed to be uniform and constant; however in reality, they do vary in space and time and are difficult to model accurately.

We shall use Tikhonov's regularisation to overcome the ill-posedness of the problem. In this method, the sum of two components is minimised: a norm of data misfit and a norm of linear model parameters. Balancing the two components is controlled by a regularisation parameter. We solve the problem by constructing an iterative procedure for the nonlinear parameters, where at each iteration a linear problem is solved. We have slightly modified the L-curve criterion developed by Hansen (1997) for linear ill-posed inverse problems to our nonlinear problem. We use this criterion as well as Wahhba's (2000) 'leaving-out-one' lemma to estimate the optimal value of the regularisation parameter for this problem.

## 2. THE INVERSE MODEL

A Cartesian co-ordinate system $(X, Y, Z)$ is used with the $X$-axis oriented in the direction of the mean wind, the $Y$-axis in the horizontal cross-wind direction, and the $Z$-axis oriented in the vertical direction. In the estimation of the source term parameter problem, the location and release rate of the pollutant at source are not available, but the concentration of pollutant distribution at some down-stream locations such as $P = (X_0, Y_0, 0)$, $Q = (X_0 + x_1, Y_0 + y_1, 0)$, and $R = (X_0 + x_2, Y_0 + y_2, 0)$ are available, where $X_0$, $Y_0$ are unknown and $x_1$, $x_2$, $y_1$ and $y_2$ are known. Our goal here is to estimate the release rate $q(t)$ of the pollution and its location such as $X_0$, $Y_0$ and $H$. Here, $H$ is a height of the source of a pollution from the ground. We use the concentration measurements at the down-stream location $P$, $Q$, $R$ along with the equation

$$C = \int_0^t \frac{q(\tau)}{8\pi^{\frac{3}{2}}\left(K_x K_y K_z\right)^{\frac{1}{2}}(t-\tau)^{\frac{3}{2}}} \times e^{\left[-\frac{(X-U(t-\tau))^2}{4K_x(t-\tau)} - \frac{Y^2}{4K_y(t-\tau)}\right]} \times \left( e^{\left[-\frac{(Z-H)^2}{4K_z(t-\tau)}\right]} + e^{\left[-\frac{(Z+H)^2}{4K_z(t-\tau)}\right]} \right) d\tau \quad (1)$$

to estimate the source release rate and its location.

### 2.1. The least-squares formulation

It is assumed that $n + 1$ concentration values $C_i = C(X_0, Y_0, 0, t_i)$ are measured at

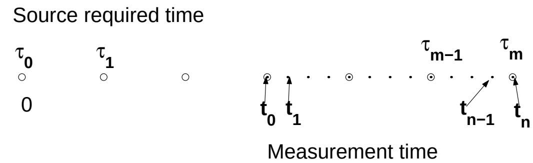


**Figure 1.** Measurement time and source required time.

the point $(X_0, Y_0, 0)$ at equal time intervals between $t_0 = \tau_0$ and $t_n = \tau_m$. The simplest way to proceed is to solve (1) on a mesh with uniform spacing. We suppose that we wish to determine the source release at times $\tau_0 = 0$, $\ldots$, $\tau_m = t_n$, where $m < n$ (see Figure ), since the number of parameters to be estimated should be no greater than the number of data points. Discretising (1) by the trapezoidal rule gives the system of equations

$$\mathbf{c} = \mathrm{A}(\mathbf{p})\mathbf{q} \quad (2)$$

where $\mathbf{c} = [C(0), \ldots, C(t_n)]^T$, $A_{ij} = K(t_i, \tau_j)\beta_{ij}$, $\mathbf{q} = [q(\tau_0), \ldots, q(\tau_m)]^T$ and $\mathbf{p} = [X_0, Y_0, H]^T$, and where $\beta_{ij}$ is a quadrature weight and the kernal

$$K(t, \tau) = \frac{e^{\left[-\frac{(X_0 - U(t-\tau))^2}{4K_X(t-\tau)} - \frac{Y_0^2}{4K_Y(t-\tau)} - \frac{H^2}{4K_Z(t-\tau)}\right]}}{4\pi^{\frac{3}{2}}(K_X K_Y K_Z)^{\frac{1}{2}}(t-\tau)^{\frac{3}{2}}}.$$

Generally, one solves inverse problems by minimising an objective function. Now the problem for estimating the release rate $\mathbf{q}$ and the location $\mathbf{p}$ is

$$\text{minimise} \quad Z(\mathbf{q}, \ \mathbf{p}) = \|A(\mathbf{p})\mathbf{q} - \mathbf{c}\|_2^2, \quad (3)$$

where $A(\mathbf{p})\mathbf{q}$, $\mathbf{c}$ are vectors containing the estimated and measured concentrations respectively, $\mathbf{p}$ is the vector of unknown non-linear parameters identifying the source location, and $\mathbf{q}$ is the vector of unknown linear parameters identifying the source release rates. The estimated concentrations are obtained from the solution of the forward problem using estimates of unknown parameter values.

Since the minimisation problem given in (3) has a combination of linear $\mathbf{q}$ and non-linear parameters $\mathbf{p}$, we separate the solution process into two steps. We find the non-linear parameter $\mathbf{p}$ by constructing an iterative procedure, where at each iteration a linear sub-problem is solved to estimate the linear parameter $\mathbf{q}$ corresponding to that particular value of $\mathbf{p}$. We solved the same problem given in (3) for a known value of $\mathbf{p}$(Kathirgamanathan *et* *al.*, 2003). It was shown that the problem is ill-posed and we therefore used Tikhonov's regularisation to solve the problem. This means that the linear sub-problem inside the nonlinear iteration is an ill-posed problem.

### 2.2. Regularized least squares

Tikhonov's regularization replaces the ill-posed problem with the well-posed problem by imposing a bound on the solution. With Tikhonov's regularisation, we introduce the regularised objective function

$$\begin{aligned} Z(\mathbf{q},\ \mathbf{p}) &= \|A(\mathbf{p})\mathbf{q} - \mathbf{c}\|_2^2 + \lambda^2 \|L\mathbf{q}\|_2^2, \\ &= \phi_d + \lambda^2 \phi_m, \end{aligned} \tag{4}$$

here, $\phi_d = \|A(\mathbf{p})\mathbf{q} - \mathbf{c}\|_2^2$ is the residual norm (or data misfit function), and $\phi_m = \|L\mathbf{q}\|_2^2$ is the solution norm. We will be interested in the function $Z(\mathbf{q}, \mathbf{p})$ and its local and global minima with respect to $(\mathbf{q}, \mathbf{p})$ for different values of the regularisation parameter $\lambda$. Note that the objective function $Z$ is the 2-norm of the following system of equations

$$\begin{bmatrix} A(\mathbf{p}) \\ \lambda L \end{bmatrix} \mathbf{q} = \begin{bmatrix} \mathbf{c} \\ 0 \end{bmatrix}, \tag{5}$$

where $L$ is the regularisation operator and $\lambda$ is the regularisation parameter that controls the relative strength of $L$, i.e. it compromises between the accuracy and the stability of the solution. The most common form of regularisation operator is

$$\|L\mathbf{q}\|^2 \approx \int_0^{t_n} \left( \frac{d^N q}{d\tau^N} \right)^2 d\tau. \tag{6}$$

The most popular choice for obtaining a smooth solution is $N = 2$ (Skaggs, 1994).

## 3. OPTIMAL CHOICE OF $\lambda$

The non-linear problem (4) is different from the linear problem in several ways. First, we cannot use linear algebra alone to determine the minima of $Z$. Second, the non-linear objective $Z$ may have more than one minimum for each value of $\lambda$. In the course of our research we developed a number of different algorithms, each of which is applied to many test cases. The implementation of all the algorithms has a number of features in common. First, we frequently have to determine local minima of $Z(\mathbf{q}, \mathbf{p})$ for a given $\lambda$. In all our algorithms this is done by exploiting the fact that some of the variables, namely $\mathbf{q}$, appear in (4) linearly and hence can be determined using simple linear algebra: for given values of $\mathbf{p}$ and $\lambda$ we define $\mathbf{q}(\mathbf{p}, \lambda)$ as the value which minimizes $Z(\mathbf{q}, \mathbf{p})$. This is computed in directly in *MATLAB* as the least-squares solution of (4). For the computation of the local minimum of $Z(\mathbf{q}, \mathbf{p})$ closest to a given point $(\mathbf{q}, \mathbf{p})$ we then relied exclusively on *MATLAB* 's routine *lsqnonlin*. However, this is speeded up enormously because, after the elimination of $\mathbf{q}$ using linear algebra, only three non-linear variables ($\mathbf{p} = [X_0, Y_0, H]$) remain.

Let us now consider the numerical details of the algorithm to solve (4). The solution is arrived at through four steps. In the first step, we find all or most of the local minimum of (4) for a fixed value of $\lambda$ when it is equal to its lowest value in the sequence. This is done by solving (4) many times, at each time with a different initial value $\mathbf{p} = \mathbf{p}_0$. We choose $\mathbf{p}_0$ randomly using the *MATLAB* function *rand* within a selected interval. There is one important reason for solving (4) for a fixed lowest $\lambda$. Equation (4) contains the error norm ($\phi_d$) and the solution norm ($\phi_m$) where the former is a non-linear part and the latter is quadratic. Therefore, (4) is almost quadratic if $\lambda$ is large, and non-linear if $\lambda$ is small. We also found that the number of local minima of (4) increases with the decreasing values of $\lambda$. Therefore the number of local minima of Equation (4) for the lowest value of $\lambda$ identifies all or most (say $n_l$) of the L-curves for this equation.

In the second step, we take each of the local minima obtained from the first step as the starting value to solve (4) for a sequence value of $\lambda$ from lowest to largest. That means solving (4) $n_l$ (the number of local minima when $\lambda$ is equal to its lowest value) times for a sequence values of $\lambda$. The idea behind this is to compute all or most of the L-curve displaying the error norm (or data misfit) versus the solution norm for a sequence values of $\lambda$.

In the third step, for each $\lambda$, we pick a point that gives the lowest function value to construct the final L-curve. Finally, we smooth the L-curve on a log-log scale by a spline curve similar to the work done by Hansen (1997). The fourth and final step is to pick the optimal point on the curve. The optimal point on each curve is calculated by examining the curvature along the L-curve and using Wahhba's (Farquharson *et al.*, 2000) 'leaving-out-one' lemma to estimate the optimal value of the regularisation parameter for this problem.

## 4. MODELLING APPLICATIONS

In this section, we present numerical calculations to evaluate the accuracy of the methods developed. To do so, we consider an input of concentration data generated from a point source of strength $q(t)$ $kg\ s^{-1}$ located at $(0, 0, H)$ in the Cartesian coordinate system. We simulate the concentration signals at downstream locations $P = (X_0, Y_0, 0)$, $Q = (X_0 + 30, Y_0 + 30, 0)$ and $R = (X_0 + 100, Y_0 + 70, 0)$. We obtain concentration signals by using the forward problem (1) and true parameter values. In order to simulate errors, we corrupt the concentration signals by adding normally distributed random noise. For illustrative purposes, $K_x$, $K_y$, $K_z$ and $U$ are taken as 12, 12, 0.2113 and 1.8, respectively. The purposes of this example is to demonstrate the simultaneous estimation of parameters $X_0$, $Y_0$, $H$, and the source release function $q(t)$

Table 1: 10% noise in the measured signal

| **p** | True value | Estimated value ± Confidence interval |
|---|---|---|
| $X_0$ | 150.00 | $162.00 \pm 9.42$ |
| $Y_0$ | 25.00 | $19.3 \pm 6.83$ |
| $H$ | 12.00 | $11.1 \pm 0.85$ |

In this example we consider a set of data which is corrupted by 10% of different random noise. The results of the source parameter estimation are summarised in Table 1, and Figure . Listed in Table 1 are the true non-linear parameter (location) values along with the reconstructed (or estimated) values and their confidence intervals. The L-curve that has lowest function value is shown in Figure a. Figure b is the curvature of the L-curve as a function of $\lambda$. The peaks in the figure correspond to the corners on the L-curve ($P_1$ & $P_2$ in Figure a). The Figure c illustrates the Wahhba's (Farquharson *et al.*, 2000) 'leaving-out-one' lemma (NGCV) to estimate the optimal point on the L-curve, where the lowest value of the NGCV function is clearly indicated ($P_3$ is a corresponding point in Figure a). Figure d depicts the true error in the solution as a function of $\lambda$. Here, the true error refers to the difference between the reconstructed and simulated (perfect) concentration values. This plot is only possible if we know the true concentration. Figure e depicts the error

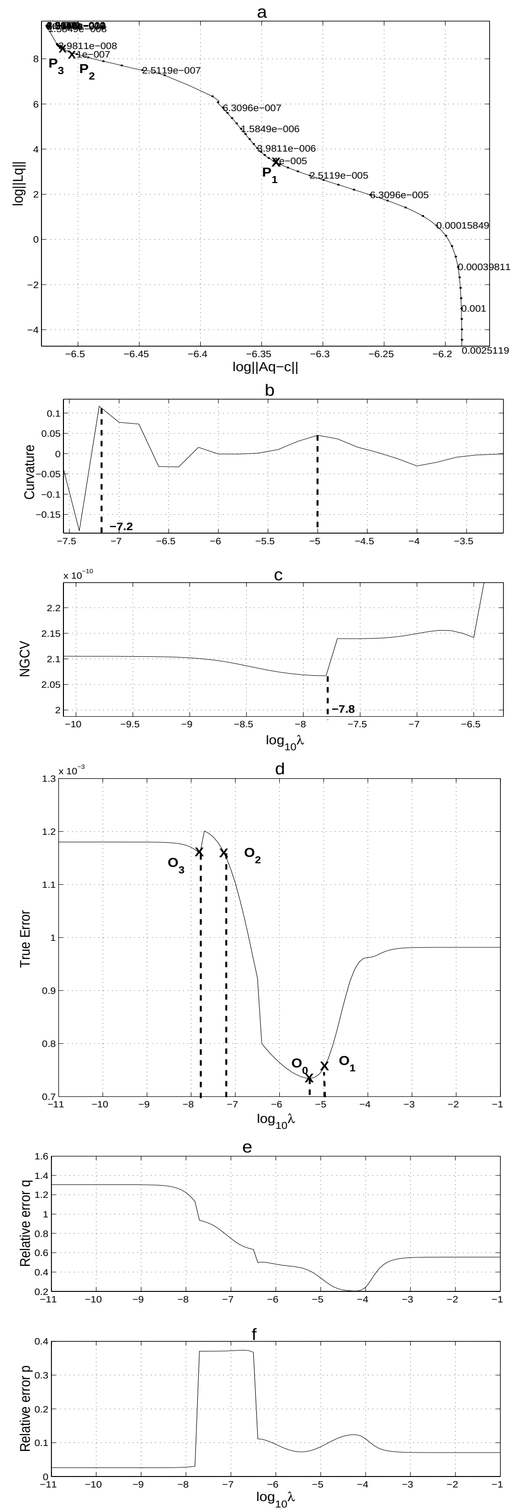


**Figure 2.** (a) non-linear L-curve, (b) curvature *vs* regularisation parameter, (c) NGCV *vs* regularisation parameter, (d) true error *vs* regularisation parameter, (e) relative error in **q** *vs* regularisation parameter, (f) relative error in **p** *vs* regularisation parameter.

in $\mathbf{q}$ (release rates) as a function of $\lambda$. Figure f depicts the error in the reconstructed location $\mathbf{p}$ as a function of $\lambda$.

The results from further numerical simulations and comparisons suggest that the corner which is closer to the origin in the L-curve gives the best approximate solution than other corners in the L-curve. Further we found that the accuracy of the estimation decreases with the following: (i) increasing noise in the data, (ii) decreasing the size of the source function discretisation, (iii) regularisation, and (iv) increasing distances between source and observation sites.

In this paper we only considered data measurements at three locations since it has been already demonstrated (Kathirgamanathan *et al.*, 2001) that data from at least three spatial locations are needed to reliably estimate the parameters in the model. However, the results from the numerical simulations suggest that the error in the reconstructed parameter values decreases only slightly as the number of locations increases.

## 5. SUMMARY AND DISCUSSION

The goal of the work presented here is to develop an inverse model capable of simultaneously estimating the location and release rate of a pollutant gas from a point source. The approach is based on a non-linear least squares estimation using pollutant concentration measurements on the ground. As the problem is ill-posed, we apply Tikhonov's regularisation method to stabilise the solution. The problem is non-linear and therefore we cannot use linear algebra alone to determine the solution. Here we develop an algorithm which we apply to many test cases. In our algorithm we used the fact that some of the parameters are linear and hence can be determined using simple linear algebra. For the computation of non-linear parameters we then relied exclusively on *MATLAB*'s routine *lsqnonlin*. This process is speeded up enormously because, after the elimination of linear parameters, only three non-linear parameters remain.

A example given in the last section describes how the model is able to determine the location and release rate of a pollution source, and how factors such as noise in the data, regularisation and the size of discretisation affect the accuracy of the solution. The results from these examples suggest that the inverse model is capable of estimating the location and the release rate of a pollution source to a reasonable degree of accuracy. Four factors affect the accuracy of the solution are (a) size and randomness of noise in the data, (b) size of discretization of the source function, (c) regularisation, and (d) distance between source and observation sites. From our observations it can be noticed that the major factor induces error in the reconstructed solution is the noise in the data. Therefore we can conclude that the total error in the solution mostly depends on inaccuracies in the data.

## 6. REFERENCES


Farquharson, C. G., and D. W. Oldenburg, Automatic estimation of the trade-off parameter in nonlinear inverse problems using the GCV and L-curve criteria, *SEG 70th Annual Meeting*, Calgary, Alberta, 6 - 11 August 2000.

Hansen, C. P., *Rank-deficient and discrete ill-posed problems.* SIAM, Philadelphia, PA, 1997.

Hansen, C. P., *Regularisation tools: A MATLAB package for analysis and solution of discrete ill-posed problems* Danish Computing Center for Research and Education, 1993.

Kathirgamanathan, P., R. McKibbin, and R. I. McLachlan, Source term estimation of pollution from instantaneous point source, *MODSIM*, 6(2), 1013–1018, 2001.

Kathirgamanathan, P., R. McKibbin, and R. I. McLachlan, Inverse Modelling for Identifying the Origin and Release rate of Atmospheric Pollution- An Optimisation Approach, *MODSIM*, 7(1), 61–66, 2003.

Kathirgamanathan, P., R. McKibbin and R.I. McLachlan, Source release rate estimation of atmospheric pollution from non-steady point source at known location. *Environmental Modelling and Assessment*, 9(1), 33–42, 2004.

Kibler, J.F., and J. T. Suttle, Air pollution model parameter estimation using simulated LIDAR data, *AIAA Journal*, 15(10), 1381–1384, 1977.

Skaggs, T. H., and J. K. Kabala, Recovering the release history of a ground water contaminant using a non-linear least squares method, *Water Resources Research*, 30(1), 71–79, 1994.