*Res. Lett. Inf. Math. Sci.*, (2002) **3**, 59–67
Available online at http://www.massey.ac.nz/∼wwiims/research/letters/

# Source Term Estimation of Pollution from an Instantaneous Point Source

P. Kathirgamanathan
*I.F.S., Massey University, Palmerston North, New Zealand*
*P.Kathirgamanathan@massey.ac.nz*
R. McKibbin
*I.I.M.S., Massey University Albany Campus, Auckland, New Zealand*
*R.Mckibbin@massey.ac.nz*
R.I. McLachlan
*I.F.S., Massey University, Palmerston North, New Zealand*
*R.Mclachlan@massey.ac.nz*

**Abstract**
The goal is to develop an inverse model capable of simultaneously estimating the parameters appearing in an air pollution model for an instantaneous point source, by using measured gas concentration data. The approach taken was to develop the inverse model as a non-linear least squares estimation problem in which the source term is estimated using measurements of pollution concentration on the ground. The statistical basis of the least squares inverse model allows quantification of the uncertainty of the parameter estimates, which in turn allows estimation of the uncertainty of the simulation model predictions.

## 1 Introduction

Decision-making about off-site emergency actions in case of an instantaneous gas release incident needs real-time forecasting of the concentration of gas in the atmosphere. The accuracy associated with forecasting of the concentration of gas in the atmosphere is highly dependent on source term parameters such as the location, timing and total amount of release. Inaccuracy in the model source term can lead to differences between estimated and actual concentration.

The process of deducing the source term from observations of airborne concentration reduces to estimating parameters in an air pollution model. Several papers [Edwards, 1993; Kibler, 1977; Mulholland, 1995; Sohier, 1997] have been published in the area. They use different models, but all depend on an intelligent first guess of the parameters and concentration measurements at many locations.

We report on a methodology for identifing the source term based on a non-linear least squares regression and linear regression coupled with the solution of an advection-diffusion equation for an instantaneous point source.This method only depends on the initial guess of the release time and the approximate value of this time can be easily calculated. Furthermore, we find that reliably estimating the parameters requires concentration measurements at a minimum of three downstream locations.

## 2 An Advection-Diffusion Equation

A Cartesian co-ordinate system $(X, Y, Z)$ is used with the $X$-axis orientated in the direction of the mean wind, the $Y$-axis in the horizontal cross-wind direction, and the $Z$-axis in the upwards vertical direction. Instantaneous gas release with a total mass release $Q$ is assumed to occur at time $t = 0$ at a point $(0, 0, H)$ which is at a height $H$ above the ground. The gas particles are subsequently blown by a wind with mean velocity $\mathbf{u} = (U, 0, 0)$. The gas molecules move with the wind in the $X$ direction at the same time as being dispersed by turbulence in the atmosphere. For a cloud of gas particles, the mass concentration $C(X, Y, Z, t)$ in time and space is governed by the equation of mass conservation:

$$\frac{\partial C}{\partial t} = -\nabla \cdot \mathbf{q} \tag{1}$$

where the pollutant mass flux per unit area $\mathbf{q}$ is given by:

$$\mathbf{q} = C\mathbf{u} - \mathbf{K} \otimes \nabla C \tag{2}$$

where $C\mathbf{u}$ is the mean mass advection by the wind and $\mathbf{K}$ is a dispersion tensor, which is assumed to be of the form:

$$\mathbf{K} = \begin{pmatrix} K_x & 0 & 0 \\ 0 & K_y & 0 \\ 0 & 0 & K_z \end{pmatrix}$$

where $K_x$, $K_y$, $K_z$ are eddy diffusivities in the $X$, $Y$ and $Z$ directions respectively. Substitution into Equation (2) gives an expression for the mass flux vector:

$$\mathbf{q} = \left( CU - K_x \frac{\partial C}{\partial X}, -K_y \frac{\partial C}{\partial Y}, -K_z \frac{\partial C}{\partial Z} \right) \tag{3}$$

Substitution of the expression for $\mathbf{q}$ into Equation (1) gives:

$$\begin{aligned} \frac{\partial C}{\partial t} + U \frac{\partial C}{\partial X} &= \frac{\partial}{\partial X}\left( K_x \frac{\partial C}{\partial X} \right) + \frac{\partial}{\partial Y}\left( K_y \frac{\partial C}{\partial Y} \right) \\ &+ \frac{\partial}{\partial Z}\left( K_z \frac{\partial C}{\partial Z} \right) \end{aligned} \tag{4}$$

where $C$ is concentration of the contaminant. Equation (4) is to be solved subject to initial and boundary conditions. The initial conditions are represented by:

$$C(X, Y, Z, 0) = Q\delta(X)\delta(Y)\delta(Z - H), \tag{5}$$

where $\delta$ is the Dirac delta function, which has the following properties:

$$\delta(X) = 0 \text{ for } X \neq 0 \text{ and } \int_{-\infty}^{\infty} \delta(X) dX = 1.$$

The pollutant concentration approaches zero far from the source in the lateral direction and high above the ground and there is zero vertical flux through the ground surface. The boundary conditions are of the form:

$$\begin{aligned} & C \to 0 \text{ as } X, Y \to \pm\infty, Z \to \infty \\ & \tfrac{\partial C}{\partial Z}(X, Y, 0, t) = 0 \end{aligned} \tag{6}$$

## 3 Solution of An Advection-Diffusion Equation

Theoretical models are available to determine the wind velocity $U$, and the eddy diffusivities $K_x$, $K_y$ and $K_z$ as functions of the vertical distance $Z$ [Huang, 1979]. However, the resulting functions are such that they make the analytical solution of Equation (4) under appropriate boundary conditions extremely difficult. To simplify the model here, it is therefore assumed that $\mathbf{u}$, $K_x$, $K_y$ and $K_z$ are constants. Equation (4) then becomes:

$$\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial X} = K_x \frac{\partial^2 C}{\partial X^2} + K_y \frac{\partial^2 C}{\partial Y^2} + K_z \frac{\partial^2 C}{\partial Z^2} \tag{7}$$

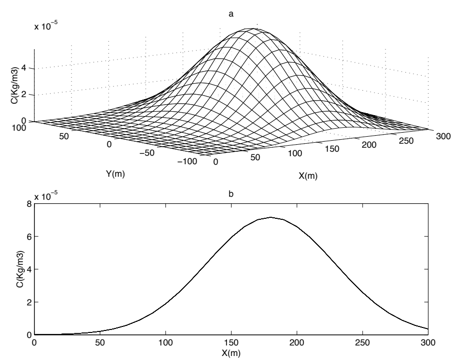


Figure 1: (a) Concentration distribution on the ground (b) Concentration distribution on the ground directly downwind of the release (on $Y = 0$)

which is to be solved subject to the initial and boundary conditions (5) and (6). The solution of (7) can be derived using Laplace and Fourier transforms and is:

$$C = \frac{Q}{8\pi^{\frac{3}{2}} \left(K_x K_y K_z\right)^{\frac{1}{2}} t^{\frac{3}{2}}} e^{-\frac{(X-Ut)^2}{4K_x t} - \frac{Y^2}{4K_y t}} \times \left( e^{-\frac{(Z-H)^2}{4K_z t}} + e^{-\frac{(Z+H)^2}{4K_z t}} \right). \tag{8}$$

Equation (8) is similar to the Gaussian model for an instantaneous point source [Seinfeld and Pandis, 1997]. Further, if we define $\sigma_x^2$=$2K_x t$, $\sigma_y^2$=$2K_y t$ and $\sigma_z^2$=$2K_z t$, the two models are identical. Here the $\sigma$'s are the standard deviations of the concentration distribution in the $X$, $Y$ and $Z$ directions. Therefore, there is a relation between the standard deviation of spread that arises in the Gaussian distribution and the eddy diffusivities in the advection-diffusion equation. The ground distribution of the concentration predicted using Equation (8) for the data values $Q$=1000 kg, $K_x = K_y = 12$ $m^2 s^{-1}$, $K_z = 0.2113$ $m^2 s^{-1}$, $t = 100$ $s$ are shown in Figure 1. Figure 1(a) shows the concentration distribution in the $X - Y$ plane on the ground $Z = 0$, while Figure 1(b) shows the concentration distribution directly downwind (on $Y = 0$), 100 seconds after the release.

## 4 Inverse Modelling

Inverse modelling is the extraction of model parameter information from data. It is a discipline that provides tools for the efficient use of data in the estimation of constants appearing in the mathematical models. In this inverse modelling problem, the structure of the equation is known; measurement of the outputs, time ($t$) and concentration ($C$), are available. Some of the parameters are unknown.

The aim of this section is to obtain the best or optimal estimate of the parameters (e.g, mass release $Q$, lateral eddy diffusivity $K_x$, source height $H$, distance of the source from measuring point $X$, $Y$ and time of the pollutant release relative to the measurement time) appearing in Equation (8) from measurements made at some position(s). The value of $K_z$ can be found by using the theoretical model $K_z = aZ^n$ [Yeh, 1975], and calculating a value at some reference height ($a$ and $n$ are constants depend on atmospheric conditions.

Taking natural logarithms of both sides of the Equation (8) when $Z = 0$ (i.e. for concentration on the ground) gives:

$$
\begin{aligned}
f\ =\ & \ln\left(\frac{Q}{4\pi^{\frac{3}{2}}\left(K_x^2K_z\right)^{\frac{1}{2}}}\right)-\frac{3}{2}\ln\left(T+t_0\right)\\
& -\left(\frac{X^2}{4K_x}+\frac{Y^2}{4K_x}+\frac{H^2}{4K_z}\right)\frac{1}{T+t_0}\\
& -\frac{U^2\left(T+t_0\right)}{4K_x}+\frac{2UX}{4K_x}
\end{aligned} \tag{9}
$$

where $f=\ln(C)$, $T+t_0=t$, $t_0$ is the (unknown) time after pollutant release when the measurement clock was started and $T$ is the time (known) on that clock. It has also been assumed that lateral eddy diffusion in the $X$ and $Y$ directions are equal, $K_y=K_x$. In simple terms, Equation (9) can be written as $f(T;\mathbf{b})$ where $T$ is the independent variable and $\mathbf{b}=[Q,H,X,Y,K_x,t_0]$ is a parameter vector.

## 4.1 Sensitivity Coefficients and Linear Dependence

Sensitivity coefficients are very important because they indicate the magnitude of change of the response $f$ due to perturbations in the values of the parameters. They also provide information about which parameters can or cannot be estimated simultaneously. They are defined by the first derivatives of $f$ with respect to each parameter. The parameters can be simultaneously estimated without ambiguity if the sensitivity coefficients over the range of observations are not linearly dependent. Linear dependence occurs when the relation:

$$
\begin{aligned}
0\ =\ & \alpha_1\frac{\partial f_i}{\partial Q}+\alpha_2\frac{\partial f_i}{\partial H}+\alpha_3\frac{\partial f_i}{\partial Y}+\alpha_4\frac{\partial f_i}{\partial X}\\
& +\alpha_5\frac{\partial f_i}{\partial K_x}+\alpha_6\frac{\partial f_i}{\partial t_0}
\end{aligned}
$$

for each of the observations $f_i$ with not all $\alpha_j$ equal to zero [Beck, 1977]. If we set $\alpha_5=\alpha_6=0$ and $\alpha_1$, $\alpha_2$ , $\alpha_3$, $\alpha_4$ are certain non-zero constants, it can be showned that

$$
\alpha_1\frac{\partial f_i}{\partial Q}+\alpha_2\frac{\partial f_i}{\partial H}+\alpha_3\frac{\partial f_i}{\partial Y}+\alpha_4\frac{\partial f_i}{\partial X}=0
$$

This shows that the parameters in Equation (9) cannot be estimated simultaneously, i.e. parameters cannot be estimated simultaneously if the data is collected at one location. Therefore measurements at more than one location are needed to estimate the parameters. Experimental results in Section 5.2 show that measurement locations cannot lie on a straight line on the ground to get good parameter estimates. Therefore concentration measurements taken from three different locations on the ground were considered to estimate the parameters in the air pollution model given by Equation (8).

Now consider an experiment in which data are generated at three different locations on the ground $P_1=(X_0,Y_0,0)$, $P_2=(X_0+x_1,Y_0+y_1,0)$ and $P_3=(X_0+x_2,Y_0+y_2,0)$. Therefore Equation (9) will become:

$$
\begin{aligned}
f\ =\ & \ln\left(\frac{2Q}{8\pi^{\frac{3}{2}}\left(K_x^2K_z\right)^{\frac{1}{2}}}\right)-\frac{3}{2}\ln\left(T+t_0\right)\\
& -\left(\frac{\left(X_0+x\right)^2}{4K_x}+\frac{\left(Y_0+y\right)^2}{4K_x}+\frac{H^2}{4K_z}\right)\frac{1}{T+t_0}\\
& -\frac{U^2\left(T+t_0\right)}{4K_x}+\frac{2U\left(X_0+x\right)}{4K_x}
\end{aligned}
$$

where $(x, y) = (0, 0)$, $(x_1, y_1)$ or $(x_2, y_2)$. This may be rearranged in the form:

$$\begin{aligned} f \; &= \beta_0 + \beta_1 \frac{x}{T + t_0} + \beta_2 \frac{y}{T + t_0} + \beta_3 \frac{1}{T + t_0} \\ &+ \beta_4 \left( -\frac{\left(x^2 + y^2\right)}{4\left(T + t_0\right)} + \frac{Ux}{2} - \frac{U^2\left(T + t_0\right)}{4} \right) \\ &- \frac{3}{2} \ln\left(T + t_0\right) \end{aligned} \tag{10}$$

where:

$$\begin{aligned} \beta_0 &= \ln\left( \frac{2Q}{8\pi^{\frac{3}{2}}\left(K_x^2 K_z\right)^{\frac{1}{2}}} \right) + \frac{2UX_0}{4K_x}, \\ \beta_1 &= -\frac{X_0}{2K_x}, \quad \beta_2 = -\frac{Y_0}{2K_x} \\ \beta_3 &= \left( \frac{X_0^2 + Y_0^2}{4K_x} + \frac{H^2}{4K_z} \right), \quad \beta_4 = \frac{1}{K_x}. \end{aligned}$$

## 4.2 Computation of Parameters

The output of Equation (10) is a logarithm of pollution concentration as a function of time, space and a set of unknown parameters. On the other hand, pollution concentration measurements are available. The method, then, is to find estimates of the unknown parameters that best fit the measured data. If $f$ is the log of measured concentration and $\hat{f}$ is the log of modelled concentration, the error in the fit of the measurement and the model, $\delta$, is:

$$\delta = \sqrt{\sum_{i=1}^{3n} \left( f_i - \hat{f}_i\left(\mathbf{b}\right) \right)^2}$$

where $3n$ is the number of measurements and $\mathbf{b} = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, t_0]$.

For the best match $\mathbf{b}$ must be varied to minimise $\delta$. This result can be achieved using the Gauss-Newton method. Essentially, the procedure is iterative and requires good starting value estimates for all the parameters. If the starting values are not reasonably good, the iteration may not converge or may converge to a local minimum.

An alternative approach to this problem of parameter estimation is now considered. This is to transform both the data and the function so that there is a multiple linear relationship between the transformed data and transformed unknown coefficients within the minimisation iteration loop. This procedure requires a good starting value of $t_0$ only. This can be calculated using the method outlined later in this section. If the data values are transformed by letting:

$$W = f + \frac{3}{2} \ln\left(T + t_0\right), W_1 = \frac{x}{T + t_0}, W_2 = \frac{y}{T + t_0},$$

$$W_3 = \frac{1}{T + t_0}, \text{ and } W_4 = -\frac{x^2 + y^2}{4\left(T + t_0\right)}$$

then the Equation 10 becomes:

$$W = \beta_0 + \beta_1 W_1 + \beta_2 W_2 + \beta_3 W_3 + \beta_4 W_4 \tag{11}$$

The step then is to form estimates of $\beta$'s using multiple linear regression that best fit the measured values $W_i$. If $\hat{W}_i$ are the modelled values, the error $\delta$ in the fit of measurements and the model is:

$$\delta = \sqrt{\sum_{i=1}^{3n}(W_i - \hat{W}_i)^2} \tag{12}$$

For the best match $t_0$ must be varied in the region $[T_0 - \epsilon, T_0 + \epsilon]$ to minimise $\delta$. Here $T_0$ is the approximation of $t_0$ and $\epsilon$ is an error. This minimisation result can be achieved using $fmin$ in MATLAB. For the new $t_0$ value $\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$ and $\beta_4$ can be calculated from Equation (11). Then, by substituting these values into Equation (10), all the required parameters $H, Q, X_0, Y_0$ and $K_x$ can be calculated.

## 4.3 Calculations of Initial Guess $t_0$

Concentration distributions at the points $P_1$ and $P_2$ can be written as:

$$C_{P_1} = \frac{2Q}{8\,(\pi t)^{\frac{3}{2}}\,(K_x^2 K_z)^{\frac{1}{2}}} e^{-\frac{(X_a - Ut)^2}{4K_x t} - \frac{Y_a^2}{4K_y t} - \frac{H^2}{4K_z t}} \tag{13}$$

$$C_{P_2} = \frac{2Q}{8\,(\pi t)^{\frac{3}{2}}\,(K_x^2 K_z)^{\frac{1}{2}}} e^{-\frac{(X_b - Ut)^2}{4K_x t} - \frac{Y_b^2}{4K_y t} - \frac{H^2}{4K_z t}} \tag{14}$$

where $t = t_0 + T$, $X_a{=}X_0$, $X_b{=}X_0{+}x_1$, $Y_a{=}Y_0$, and $Y_b{=}Y_0{+}y_1$. Dividing Equation (13) by Equation (14) and then differentiating w.r.t. $T$ followed by taking natural logarithms of both sides gives:

$$\ln F + T\frac{F'}{F} = -t_0\frac{F'}{F} - \frac{2x_1}{4K_x}$$

where $F = \frac{C_{P_1}}{C_{P_2}}$ and $F' = \frac{dF}{dT}$. The graph of $\ln F + T\ \frac{F'}{F}$ plotted against $\frac{F'}{F}$ is a straight line, with a gradient of $m = -t_0$ and intercept $= -\ \frac{2x_1}{4K_x}$ (when $T = 0$). Note: The logarithms of concentration distributions at $P_1$ and $P_2$ have to be smoothed using polynomial fits for noisy data before applying the method.

# 5 Modelling Application

## 5.1 Source Term Estimation

To illustrate this inverse modelling application, consider an input of environmental data generated from an instantaneous point source of strength 1000 $kg$ located at (0, 0, 20 $m$) in the Cartesian co-ordinate system. Figure 2 shows the concentration signal against time $T$ at the points $P_1$, $P_2$ and $P_3$, where $P_1 = (5000, 100, 0)$, $P_2 = (5480, 230, 0)$ and $P_3 = (5130, 580, 0)$, i.e. $P_1$, $P_2$ and $P_3$ are on the vertices of a equilateral triangle of side 500 $m$. For illustrative purposes $K_z$ and $U$ are taken as 0.211 and 1.80 respectively. The results of the source term estimation for the pollution concentration in Figure 2(a) are tabulated in Table 1. Then random relative noise of 1%, 2%, 3%, 4% and 5% were added to the simulated signal and the calculation of error in the source term was repeated for one hundred times for each case. Average error values of parameters are tabulated in rows 1 to 5 of Table 2.

## 5.2 Selection of Measurement Locations

The results of numerical experiments show that the accuracy of the parameter estimates depends on the location of the pollution measurement points. To analyse the effect three cases were considered.

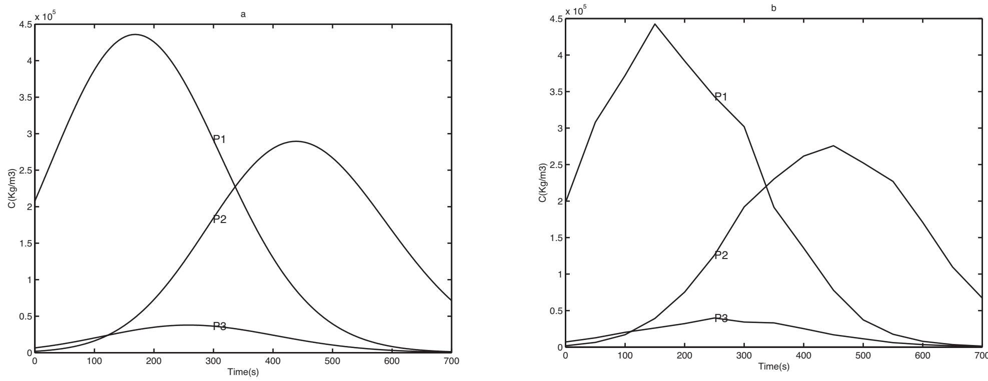


Figure 2: (a) Concentration signal at the points $P_1$, $P_2$ and $P_3$ on the ground with no noise, (b) Concentration signal at the points $P_1$, $P_2$ and $P_3$ on the ground with noise of 5%.

**Table 1.** Source term estimates

| $t_0$ | $K_x$ | $X_0$ | $Y_0$ | $Q$ | $H$ |
|---|---|---|---|---|---|
| 0.72 | 12 | 5000 | 100 | 1000 | 20 |

**Table 2.** Percentage errors in calculated parameters of the source term, for various relative noise levels.

| Noise | $t_0$ | $K_x$ | $X_0$ | $Y_0$ | $Q$ | $H$ |
|---|---|---|---|---|---|---|
| 1 % | 0.8 | 0.7 | 0.8 | 0.4 | 2.7 | 8.5 |
| 2 % | 1.6 | 1.4 | 1.5 | 0.7 | 5.8 | 18.6 |
| 3 % | 2.5 | 2.2 | 2.4 | 1.1 | 8.0 | 25.2 |
| 4 % | 3.1 | 2.7 | 2.9 | 1.6 | 11.0 | 33.0 |
| 5 % | 4.3 | 3.7 | 4.0 | 1.9 | 12.3 | 33.8 |

(i) All three stations $P_1$, $P_2$ and $P_3$ lie on a straight line. (In Figure 4, $P_3$ is on the line $P_1P_2$.)

(ii) Stations $P_1$, $P_2$ and $P_3$ are on the vertices of a perpendicular isoscles triangle. (In Figure 4, $\alpha = 90°$ and $L_1 = L_2$.)

(iii) $P_1$, $P_2$, and $P_3$ are on the vertices of an equilateral triangle. (In figure 4, $\alpha = 60°$ and $L_1 = L_2$.)

In the first case whatever the values of $\theta$, $L_1, L_2$ and $L_3$, the calculated parameter values were wrong even in the case of perfect simulated data. The error in parameter estimates $Q$ and $X_0$ of the other two cases are plotted against the distance between the points in Figures 3(a) and 3(b) respectively. In each case, parameter values were calculated when the angle ($\theta$) between $P_1P_2$ and $X$-axis is equal to 0°,15° and 30°. The above experiments demonstrate how the distance and angle between the measurement locations affects the accuracy of the source term estimation results.

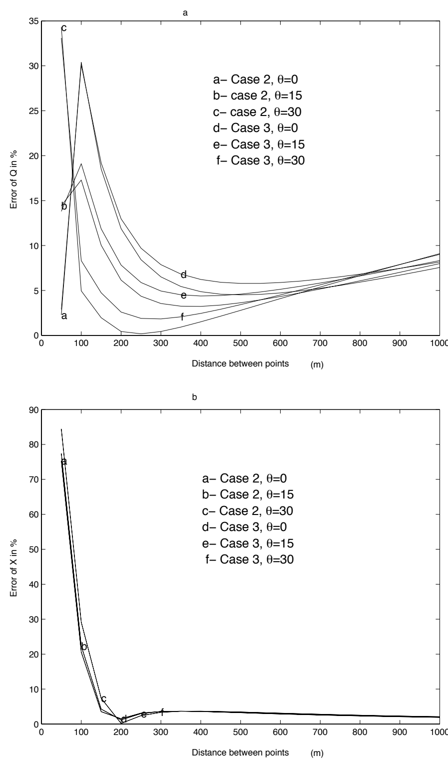


Figure 3: (a) Error in $Q$ Vs distance between points, (b) Error in $X_0$ Vs distance between points.

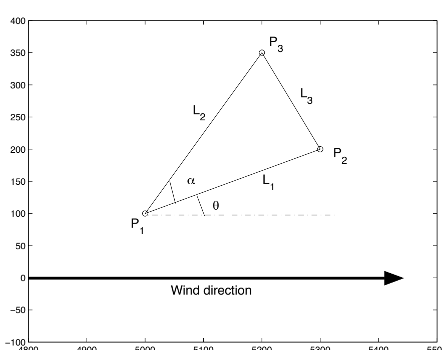


Figure 4: Locations of the points $P_1$,$P_2$ and $P_3$ on the ground.

## 6 Summary and Discussion

The goal of the work presented here was to develop an inverse model capable of simultaneously estimating the parameters appearing in the air pollution model for an instantaneous point source. The approach taken was to develop the inverse model as a non-linear least squares estimation problem in which the source term was estimated using pollution concentration measurements on the ground. The statistical basis of the least square inverse model allows for quantifying the uncertainty of the parameter estimates, which in turn allows for quantifying the uncertainty of the simulation model predictions.

First in the process, it has been demonstrated that data from at least three spatial locations are needed to reliably estimate the parameters in the model. Secondly, we formulated the inverse model as a least squares minimization problem, and then we tested the methodology using artificial data generated from the forward problem.

The accuracy of the calculated parameter values varies with the distance between the measurement locations. Therefore the optimal design of the locations for pollution measurement on the ground is important. This is one possibility for improvement of the model.

This paper is a report of an initial study using both linear and nonlinear least squares estimation techniques for calculating source term parameters from an inverse model. The next phase of this study is to find estimates of source terms of pollution from steady and non-steady point sources of unknown time duration.

## 7 References

Beck, V.J., and Arnold, J.K, Parameter Estimation in Engineering and Science. John Wiley and Sons, New York, 1977.

Edwards, L.L., Freis, R.P., Peters L.G., and Gudiksen P.H., The use of non-linear regression analysis for integrating pollutant concentration measurements with atmospheric dispersion modelling for source term estimation, *Nuclear Technology*, 101, 168–181, 1993.

Huang, C.H., Theory of dispersion in turbulent shear flow, *Atmospheric Environment*, 13, 453–463, 1979.

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

Mulholland, M., and Seinfeld, J.H., Inverse air pollution modelling of urban scale carbon monoxide emissions, *Atmospheric Environment*, 29(4), 497–516, 1995.

Seinfeld, J.H., and Pandis, S.P., *Atmospheric chemistry and physics: from air pollution to climate change.* John Wiley and Sons, New York, 1998.

Sohier, A., Rojas-Palma, C., and Liu, X., Towards a monitoring framework for the source term estimation during the early phase of an accidental release at nuclear power plant, *Radiation protection dosimetry*, 73(1–4), 231–234, 1997.

Yeh, G., and Hung, C., Three-dimensional Air Pollutant modelling in the Lower Atmosphere, *Boundary-Layer Meteorology,* 9, 381-390, 1975.