Source Parameter Estimation Of Atmospheric pollution from accidental releases of gas

A THESIS PRESENTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

MATHEMATICS

AT MASSEY UNIVERSITY, PALMERSTON NORTH

NEW ZEALAND.

Padmanathan Kathirgamanathan

August 2003

Abstract

This thesis 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-diffusion equation. The least-squares technique allows quantification of the uncertainty of the calculated estimates, which in turn allows estimation of the uncertainty of the simulation model predictions.

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. The optimal value of the regularisation parameter in the problem is estimated using both the linear and non-linear L-curve criterion, and a generalised cross-validation approach. The accuracy of the model is examined by using simulated concentration data (generated by the forward model) to which normally-distributed relative noise has been added, as well as some real experimental data.

Acknowledgements

First of all, I would like to express my sincere gratitude to my chief supervisor, Professor Robert McKibbin, for giving me an interesting real world problem to work on. Robert has been an exceptional advisor, both academically and personally. His support, encouragement, valuable suggestions discussions and comments through out my PhD are magnificent. I was very fortunate to have an opportunity to work with him.

I am equally thankful to my second supervisor Professor Robert McLachlan for his continuous support and the regular weekly meeting. Without Robert's ability, enthusiasm, and moral support the majority of the contents in this thesis could not have completed. His careful reading of this thesis and his very valuable comments made tremendous improvement to its content.

I would also like to thank my third supervisor Associate Professor Igor Boglaev for providing me valuable comments on various parts of my thesis. My thanks also extended to Dr Aroon Parshotam for his comments on several chapters. I am also grateful to my fellow PhD students for their help in particular Jonathan Marshal and Patrick Rynhart.

Thank you to Massey University for the Doctoral Scholarship and to the Institute of Fundamental Sciences for the Graduate Assistant and Teaching Assistant works. I also extend my thanks to the New Zealand Mathematical Society and the Institute of Fundamental Sciences for giving me financial assistance to attend international conferences.

Finally, I would like to express my sincere gratitude to my wife and children for their love and understanding, and my parents for their love and prayers. None of this would have been possible without their support.

Contents

List of Symbols

q	release rate of pollutant gas $[kg\ s^{-1}]$
q_0	release amount of pollutant gas $[kg]$
C	concentration of pollutant gas $[kg\ m^{-3}]$
t	measurement time with respect to release started time $[s]$
T	measurement clock time $[s]$
t_0	measurement started time with respect to release start time $[s]$
S	location of the source
H	height of the source from the ground level $[m]$
X_0	distance in the X-direction between the source and the receptor $[m]$
Y_0	distance in the Y-direction between the source and the receptor $[m]$
λ	regularization parameter
L	regularization operator
n	number of measurements
N	order of regularization
σ	singular values
Σ	singular matrix
W	left singular vector
V	right singular vector
$\mathbf{q}$	mass flux per unit area
$\mathbf{K}$	dispersion tensor
K_x	dispersion coefficient in the X- direction $[m^2\ s^{-1}]$
K_y	dispersion coefficient in the Y- direction $[m^2\ s^{-1}]$
K_z	dispersion coefficient in the Z- direction $[m^2\ s^{-1}]$
U	wind speed in the X-direction $[m\ s^{-1}]$
σ_x	the standard deviation of the concentration in the X-direction
σ_y	the standard deviation of the concentration in the Y-direction
σ_z	the standard deviation of the concentration in the Z-direction
GCV	generalised cross validation
SVD	singular value decomposition
a, b, c, d, e, f, p, r	atmospheric stability parameters
$\mathbf{c}$	measured concentration vector
$\hat{\mathbf{c}}$	estimated concentration vector

Chapter 1

Introduction

1.1 An Overview of the Source Parameter Estimation Problem

The accidental release of gases from industrial plants is not unusual; such classes may spring from relief valves, emergency venting, explosions, sudden failure of storage tanks, reactors, and so on. Post-accident management plans are necessary for public protection from potential accidents resulting from the leakage of dangerous gases. In the event of accidental gas releases, the determination of the release rate and its location are important, since forecasting of the subsequent concentration of gas in the atmosphere is totally dependent on them. Therefore, the estimation of the source's release rate and its location is very useful for post-accident management staff to prioritise off-site evacuation actions.

Atmospheric dispersion modelling may be used in a post-accident management plan to assess the situation. Atmospheric dispersion models describe the transport and dispersion of pollutant gases in the atmosphere. A dispersion model, which is capable of describing the behaviour of air pollutants in the atmosphere, requires the following input data: (1) meteorological data such as wind speed, direction and atmospheric stability, and (2) the source emission rate and its origin.

In reality, meteorological data can be measured using available measuring instruments, but the source release rate and its origin are often unknown. Methods to determine the release rate and the location of the source of the pollutant gas are therefore an important part of modelling atmospheric dispersion. In many cases, the only available information concerning the accidental release of a pollutant gas is the time history of the pollutant gas concentration at several downstream locations. These concentration data and meteorological data can be used along with dispersion models to reconstruct the release rate and location of the pollutant gas.

The estimation of the source release rate is an ill-posed problem. Even if the model parameters are known exactly and the measured data are error-free, the solution to the release rate estimation is not unique. In ill-posed problems, small changes in the data can cause large changes in the solution; i.e. ill-posed problems are extremely sensitive to the measurement error in the data. In reality, data contain measurement errors, which can lead to instability of the solution. In order to stabilise the solution, regularisation methods must be used. Several regularisation methods are available in the literature and may be used to solve such problems.

The estimation of the source release rate is further complicated by inexact information. Parameter values such as dispersion coefficients and wind speed are not known exactly and are very difficult to model accurately. In the dispersion models used in this research, dispersion coefficients and wind speed are assumed to be uniform in space and time.

1.2 The Problem Statement

The source release rate estimation of atmospheric pollutants from a point source using the input of environmental data such as concentration of pollutant gas is considered. The problem involves an unknown release rate of a point source of gas in a three-dimensional medium. We categorise the process under investigation for the source release of pollutant gas as follows:

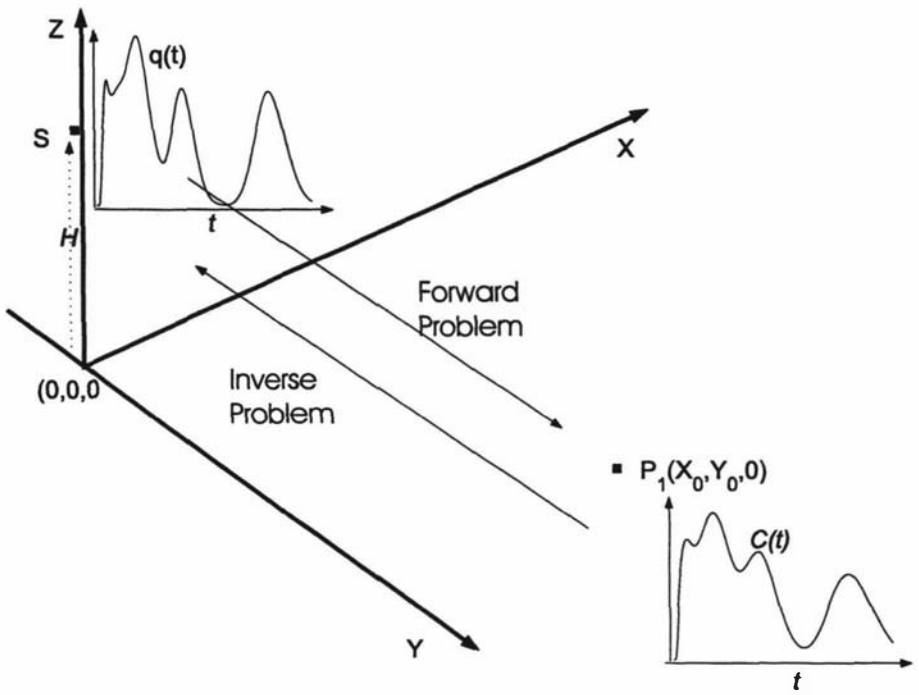

Figure 1.1: Demonstration of source term estimation of pollution problem

1. instantaneous release from a known location;

2. instantaneous release from an unknown location;

3. extended release (non-steady) over a period of time from a known location;

4. extended release (non-steady) over a period of time from an unknown location.

For each of the above categories, the source release rate, which is known, is input into a forward model based on the solution of an advection-dispersion equation to obtain the concentration distribution of the pollutant gas at a later time. These data are sampled at some downstream locations, and used in the inverse model to reconstruct a release rate of pollutant gas.

Figure 1.1 illustrates the set-up of the problem. 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. The release of gas with mass release rate $q(t)$ is assumed to start 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 a horizontal direction at the same time as being dispersed by turbulence in the atmosphere. For a cloud of gas particles, the mass concentration at the point (X, Y, Z) at time t is $C(X, Y, Z, t)$. In this inverse problem we have a concentration profile $C(t)$ at some point $P_1(X_0, Y_0, 0)$ and wish to find the release profile $q(t)$ at the point $S(0, 0, H)$.

Categories 1 and 2 are for instantaneous releases, and therefore release rate function $q(t) = q_0\delta(t)$. The first category is very simple since we know the values of X_0, Y_0 and H, and only a single parameter q_0 is to be estimated. In Category 2, we do not know the location of the source and the parameters to be estimated in this model are X_0, Y_0, H and q_0. In category 3, the location of the source is known but the source release function $q(t)$ is not known, so the function $q(t)$ is to be estimated. In the last category, the location of the source and release rate are not known, therefore the parameters X_0, Y_0, H and the function $q(t)$ are to be estimated.

1.3 The Scope of the Source Term Estimation

This research has relevance to the following:

(i) post-accident management plan;

(ii) leak detection in pipelines and other installations that are used to store gases;

(iii) finding the location of gas leaks detected by monitoring stations.

1.4 The Overview of the Thesis

The aim of this thesis is to present an inverse model capable of estimating the release rate of pollutant gas in the atmosphere using measured pollutant gas concentrations on the ground. The presentation of the thesis is organised into 8 chapters including the present.

Chapter 2 contains the main literature review. It contains background material for the inverse problem and for atmospheric dispersion modelling in general. An overview of the inverse problem, numerical techniques, methods of regularisation and the problem areas in the least-squares estimation are described in the first four sections. Common forms of regularisation parameter selection methods are also explained. The fifth section considers the theory of atmospheric dispersion modelling and the forward problems of pollutant dispersion in the atmosphere. In the last section we briefly describe some of the related problems in the area.

In Chapter 3 we introduce the problem in one dimension. It contains some simple methodologies to estimate the parameters appearing in the model for the cases of instantaneous and extended steady releases of pollutant from a point source. In addition it explains the methodology to estimate the derivative of noisy data. The aim of this chapter is to understand the inverse problem and estimate the difficulties associated with the measurement errors in the data in general.

The main theme of the thesis starts in Chapter 4. The first two categories of the problem, such as instantaneous releases from known and unknown locations, are considered here. This chapter provides a detailed description of an inverse model capable of simultaneously estimating parameters such as location and released amount for an instantaneous point source. In addition, the chapter demonstrates why data from at least three spatial locations are needed to estimate reliably the location of the pollution. Finally, some numerical examples are provided to test the methodology.

Chapter 5 is devoted to the release-rate estimation inverse problem of pollutant gas from a known location. The chapter starts with the description of the forward problem and the data requirements for the inverse problem. Then the methodology of the inverse problem for estimating the release rate of atmospheric pollutant by using measured pollutant concentration data at a single location is presented. The chapter also provides some analysis of the inverse problem using singular value decomposition and condition number with the aid of numerical examples. In the next section, numerical regularisation of linear least squares for the release rate estimation problem using Tikhonov's method is presented. Finally, some numerical examples are given with and without measurement errors to compare the accuracy of the reconstructed solution against different regularisation parameter selection methods, orders of regularisation, and complete and incomplete data sets.

Chapter 6 presents an inverse modelling procedure to estimate the location and release rate of atmospheric pollutant. The chapter starts with the forward problem and the data requirements for the inverse problem. Next, the methodology for the problem is presented, which includes non-linear least-squares formulation, numerical regularisation using Tikhonov's method, and methods for selecting the regularisation parameter. The algorithm for estimating the location and release rate of the pollutant is then given. Next, methodology is provided to estimate the confidence interval for a non-linear ill-posed inverse problem. Finally, some numerical examples are provided to test the methodology using simulated data.

In Chapter 7, some real experimental data are analysed for Categories 3 and 4. This chapter is divided into seven sections. The first two sections describe respectively the geographical setting of the place where the experiment was carried out and the details about the tracer-gas release experiment. Calculations of transport parameters for the inverse model are given in Section three. Section four describes the difficulties associated with inexact input parameters and their effects on the inverse model. Release-rate estimation of source release from a known and unknown

location are given in Sections five and six, respectively, and a summary then follows.

Chapter 8 presents the summary, conclusion and possible extension of this work.

This thesis consists of the papers [23], [24], [25] and [26].

Chapter 2

Inverse Problems and Dispersion Modelling

The mathematical literature relevant to this thesis can be divided into two categories:

(i) on inverse problems;

(ii) on atmospheric dispersion modelling.

2.1 Inverse Problems

2.1.1 Introduction

In a wide variety of science and engineering applications, the quantity of interest cannot be measured directly but inferences have to be made about the quantity on the basis of indirect observations. In such cases, the problem of estimating the unknown parameters by means of the mathematical model between the observations and unknown values is called an inverse problem. The inverse problem has two variants. The most common is that the system and output is known, and we wish to estimate the input. The case where the system is unknown is a model

identification problem. It is also called an inverse problem. This can be explained using Figure 2.1 and Equation (2.1).

Figure 2.1: The structure of forward and inverse Problem

$$K(\mathbf{q}) = \mathbf{y}. \tag{2.1}$$

The forward problem is where the system K and the input $\mathbf{q}$ are known. It is left to predict $\mathbf{y}$. In most cases the solution is unique, and small errors in $\mathbf{q}$ cannot induce large errors in $\mathbf{y}$, which means the problem is well-posed. The objective of the inverse problem is to determine an estimate of $\mathbf{q}$ from observations of $\mathbf{y}$. In reality, the observations are corrupted by measurement noise. To make progress it is necessary to model the measurement errors. The most popular assumptions are that either the absolute or relative errors are independent identically distributed random variables. In the case of additive noise in the data (2.1) may be expressed as:

$$K(\mathbf{q}) + \mathbf{e_1} = \mathbf{y}, \tag{2.2}$$

and in the case of multiplicative noise in the data (2.1) may be expressed as:

$$K(\mathbf{q})(1 + \mathbf{e_2}) = \mathbf{y}, \tag{2.3}$$

where $\mathbf{e_1}$ is a $n \times 1$ vector of normally distributed random error, $\mathbf{e_2}$ is a $n \times 1$ vector of log normally distributed random error and $\mathbf{y}$ is an $n \times 1$ vector of observations. The task is then to estimate parameters $\mathbf{q}$ based on the measurements $\mathbf{y}$.

A variety of techniques have been proposed in the literature ([3], [12], [7], [10], [52], [51], [50], [62]) for solving similar kind of problems. In general these techniques

can be classified in to two broad classes of approaches such as probabilistic and deterministic [1]. In a deterministic approach the parameters $\mathbf{q}$ are treated as non-random(or deterministic) variables, whereas in the probabilistic approach the parameters are treated as random variables. Both approaches are explained in the following two sections even though our approach in this thesis is based on the deterministic approach (least-squares). In section 2.3, the Bayesian method for inverse problems is described briefly and then some comparison is given of the Bayesian method solution with the least-squares method in conjunction with Tikhonov's regularisation solution.

2.2 Deterministic Approach

The inverse estimation procedure in this thesis is based on linear and non-linear least-squares methods. The distinction between linear least-squares and non-linear least-squares is that the linear least-squares method possesses an analytical solution, whereas non-linear least-squares can generally only be solved by numerical methods.

2.2.1 Least-square estimation

Linear least squares

When the system K in (2.1) is linear, (2.2) and (2.3) reduces to

$$\mathbf{y} = K\mathbf{q} + \mathbf{e}, \tag{2.4}$$

where $\mathbf{e} = \mathbf{e_1}$ in case of (2.2) or $\mathbf{e} = K\mathbf{q}\mathbf{e_2}$ in case of (2.3) and K is an $n \times m$ matrix, where n is the number of data points and m is the size of a vector $\mathbf{q}$. The least-squares estimate $\hat{\mathbf{q}}$ of $\mathbf{q}$ (in the case of constant error variance) is the vector that satisfies

$$\nabla_{\hat{\mathbf{q}}} \left[[\mathbf{y} - K\hat{\mathbf{q}}]^T [\mathbf{y} - K\hat{\mathbf{q}}] \right] = \mathbf{0},$$

where ∇ is a gradient. Differentiation produces the expression

$$-2K^T\mathbf{y} + 2(K^TK)\hat{\mathbf{q}} = \mathbf{0}.$$

As a result,

$$\hat{\mathbf{q}} = \left(K^TK\right)^{-1} K^T\mathbf{y}. \tag{2.5}$$

Non-linear least squares

Consider a general non-linear model

$$\mathbf{y} = K(X, \mathbf{p}) + \mathbf{e}, \tag{2.6}$$

where K is function of $\mathbf{p}$ and X. Here $\mathbf{p}$ is $m \times 1$ vector of parameters to be estimated and X is the $n \times m$ design matrix of the model.

The non-linear problem is different from the linear problem because the parameters cannot be estimated by using only linear algebra. All methods for non-linear problems are iterative and require starting values for all the parameters to be estimated. Within the *MATLAB* environment we can use *fminsearch*, *lsqnonlin* subroutines to perform nonlinear least-squares iterations with problem (2.6). *fminsearch* uses the Nelder-Mead simplex search algorithm to estimate the parameter values whereas there is a choice in *lsqnonlin* of using the Trust Region algorithm, or the Levenberg-Marquardt algorithm, or the Gauss-Newton algorithm. In this thesis, we mostly use *lsqnonlin* with the Levenberg-Marquardt algorithm, because it is more efficient than *fminsearch*.

2.2.2 The concept of the ill-posed problem

For a problem to be well posed, it has to satisfy the following requirements [42]: (i) the solution of the problem exists; (ii) the problem has a unique solution; (iii) the solution depends continuously on the initial or boundary data. A problem is well conditioned if points (i) and (ii) hold, as well as a modification of point (iii), a small deviation in the problem parameters result in a small deviations in the solution. The violation of (i) or (ii) or (iii) implies that the problem is ill-posed (or ill-conditioned). We can illustrate the main difficulties associated with ill-posed problems by means of small examples. These are as follows:

Example 1. From linear algebra

Consider the following least square problem [17]

$$\min_{\mathbf{q}} \|A\mathbf{q} - \mathbf{b}\|_2,$$

where $\|.\|_2$ is the Euclidean 2-norm,

$$\mathbf{A} = \begin{pmatrix} 0.16 & 0.10 \\ 0.17 & 0.11 \\ 2.02 & 1.29 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 0.26 \\ 0.28 \\ 3.31 \end{pmatrix},$$

there exists an exact solution $\mathbf{q} = (1,1)^T$. Now, consider the effect of a small perturbation in $\mathbf{b}$ where

$$\min_{\mathbf{q}} \|A\mathbf{q_1} - \mathbf{b_1}\|_2,$$

with

$$\mathbf{b_1} = \begin{pmatrix} 0.16 & 0.10 \\ 0.17 & 0.11 \\ 2.02 & 1.29 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.01 \\ -0.03 \\ 0.02 \end{pmatrix} = \begin{pmatrix} 0.27 \\ 0.25 \\ 3.33 \end{pmatrix}.$$

From the ordinary least square solution $\mathbf{q_1}$ of a system $A\mathbf{q_1} = \mathbf{b}_1$ by means of a QR factorisation of A we get

$$\mathbf{q_1} = \begin{pmatrix} 7.01 \\ -8.40 \end{pmatrix},$$

where Q is an orthogonal and R is an upper triangular matrix. This shows that a relatively small change in a system parameter $\mathbf{b}$ induces a large change in the solution $\mathbf{q}_1$.

Example 2. Condition number

Consider the least squares problem

$$\min_{\mathbf{q}} \|A\mathbf{q} - \mathbf{b}\|_2.$$

Let $A \in \mathbf{R}^{n\times m}$. The condition number of A is defined as [56]

$$cond(A) = \|A\| \frac{\|\mathbf{q}\|}{\|A\mathbf{q}\|},$$

where $\|.\| = \|.\|_2$ and $\|A\| = \max_{\|\mathbf{q}\|\neq 0} \frac{\|A\mathbf{q}\|}{\|\mathbf{q}\|}$. Suppose that A happens to be square and non-singular. Then

$$\frac{\|\mathbf{q}\|}{\|A\mathbf{q}\|} \leq \|A^{-1}\|.$$

Therefore

$$cond(A) \leq \|A\|\|A^{-1}\|,$$

or

$$cond(A) = \alpha\|A\|\|A^{-1}\|,$$

with

$$\alpha = \frac{\frac{\|\mathbf{q}\|}{\|A\mathbf{q}\|}}{\|A^{-1}\|}.$$

For certain choices of $\mathbf{q}$, we have $\alpha = 1$, and consequently

$$cond(A) = \|A\|\|A^{-1}\|.$$

Now suppose the right hand side vector $\mathbf{b}$ is corrupted by noise $\triangle\mathbf{b}$. Then the solution $\mathbf{q}$ is also changed by an amount $\triangle\mathbf{q}$:

$$A(\mathbf{q} + \triangle\mathbf{q}) = \mathbf{b} + \triangle\mathbf{b}, \quad A(\triangle\mathbf{q}) = \triangle\mathbf{b},$$

hence,

$$\triangle\mathbf{q} = A^{-1}\triangle\mathbf{b}.$$

Taking norms, we write

$$\|\triangle\mathbf{q}\| \leq \|A^{-1}\|\|\triangle\mathbf{b}\|.$$

From $\triangle\mathbf{b} = A\triangle\mathbf{q}$, we have $\triangle\mathbf{b} \leq A\triangle\mathbf{q}$. Therefore,

$$\frac{\|\triangle\mathbf{b}\|}{\|A\|} \leq \|\triangle\mathbf{q}\| \leq \|A^{-1}\|\|\triangle\mathbf{b}\|. \tag{2.7}$$

Applying the same reasoning to $A\mathbf{q} = \mathbf{b}$ and $\mathbf{q} = A^{-1}\mathbf{b}$, we get

$$\frac{\|\mathbf{b}\|}{\|A\|} \leq \|\mathbf{q}\| \leq \|A^{-1}\|\|\mathbf{b}\|. \tag{2.8}$$

Taking (2.7) and (2.8) together, we get

$$\frac{1}{cond(A)}\frac{\|\triangle\mathbf{b}\|}{\|\mathbf{b}\|} \leq \frac{\|\triangle\mathbf{x}\|}{\|\mathbf{q}\|} \leq cond(A)\frac{\|\triangle\mathbf{b}\|}{\|\mathbf{b}\|}.$$

It can be seen from the above that the condition number of the matrix is proportional to the error in the solution. A small condition number induces a small upper bound on the relative error, and large condition number induces a large upper bound on the relative error.

Example 3. Fredholm integral equation

Consider the following integral equation [17]

$$\int_a^b K(s,t)f(t)dt = g(s), \quad c \leq s \leq d,$$

where the right-hand side g and the kernel K are given, and f is the unknown solution to be determined. If the solution f is perturbed by

$$\triangle f(t) = \epsilon sin(2\pi pt), \quad p = 1, 2, \ldots, \quad \epsilon = const$$

then the corresponding perturbation of the right-hand side g is given by

$$\triangle g(s) = \epsilon \int_a^b K(s,t)sin(2\pi pt)dt, \quad p = 1, 2, \ldots$$

By the Riemann-Lebesgue lemma, we can write

$$\triangle g \to 0, \quad p \to \infty.$$

Therefore, the ratio $\frac{\|\triangle f\|}{\|\triangle g\|}$ becomes randomly large by choosing the larger value for p. This example shows that small alterations in g can cause larger deviations in the inverse solution of f.

2.2.3 Analysis of ill-conditioned least squares

When least squares problems are encountered, the usual first step is to analyse the condition matrix A (for the linear least squares) or the Jacobian matrix W (for the nonlinear least squares) using the singular value decomposition (SVD).

If A denotes an $n \times m$ matrix and its elements are real, then SVD of A is defined as [16]

$$A = W\Sigma V^T = \sum_{i=1}^{m} w_i \sigma_i v_i^T$$

where

$$W = (w_1, \ldots, w_m) \in \mathbb{R}^{n \times m}, \quad V = (v_1, \ldots, v_m) \in \mathbb{R}^{m \times m}$$

are matrices with

$$W^T W = I_m, \quad VV^T = V^T V = I_m,$$

$$\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_m),$$

where the σ's are the singular values of A, $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_m \geq 0$. The columns of W are the left singular vectors and the columns of V are the right singular vectors of A. The generalised singular value decomposition (GSVD) of the matrix pair (A, L) is a generalisation of the SVD of A so that the generalised singular values of (A, L) are the square roots of the generalised eigenvalues of the matrix pair $(A^T A, L^T L)$ [17].

Let $A \in \mathbb{R}^{n \times m}$ and $L \in \mathbb{R}^{p \times m}$ with $n \geq m \geq p$. The GSVD of A and L is in the form

$$A = W \begin{pmatrix} \Sigma & 0 \\ 0 & I_{m-p} \end{pmatrix} X^{-1}, \quad L = V\,(M, 0)\, X^{-1},$$

where

$$W \in \mathbb{R}^{n \times m} \text{ and } V \in \mathbf{R}^{p \times p},$$

$$X \in \mathbb{R}^{m \times m} \text{ is nonsingular,}$$

$$\Sigma \text{ and } M \in \mathbb{R}^{p \times p},$$

with

$$\Sigma = diag(\sigma_1, \ldots, \sigma_p), \quad M = diag(\mu_1, \ldots, \mu_p),$$

$$0 \leq \sigma_1 \leq \ldots \leq \sigma_p \leq 1, \quad 1 \geq \mu_1 \geq, \ldots \geq \mu_p \geq 0,$$

and normalised so that

$$\sigma_i^2 + \mu_i^2 = 1, \quad i = 1, \ldots, p.$$

The generalised singular values γ_i of (A, L) are defined as the ratios

$$\gamma_i = \frac{\sigma_i}{\mu_i}, \quad i = 1, \ldots, p.$$

According to [16], two classes of ill-conditioned problems based on the properties of their coefficient matrices can be distinguished:

(i) **rank-deficient problems**: The properties of their coefficient matrices are

(a) have a small cluster of small singular values,

(b) have a well-determined gap between large and small singular values, implying that one or more rows and columns of the coefficient matrix are a nearly linear combination of some or all of the remaining rows and columns,

(c) usually imply that there exists a reformulation that will eliminate the ill conditions.

(ii) **discrete ill-posed problems**: The properties of their coefficient matrices

(a) have a larger cluster of small singular values,

(b) have singular values σ_i that all decay gradually to zero with no clear gap in the singular spectrum,

(c) have singular vectors $\mathbf{W_i}$ and $\mathbf{V_i}$ of the coefficient matrix that tend to have more sign changes in their elements as the index i increases,

(d) usually have no reformulation that can change the ill conditioning.

2.2.4 Numerical regularisation methods

Numerical regularisation techniques seek to approximate the exact solution of an ill-condition problem by the solution of a related well-conditioned problem that includes information about the desired solution to the original problem.

Rank-deficient problems

Let

$$A \in \mathbb{R}^{n \times m}$$

where

$$rank(A) = k \quad \text{if} \quad \sigma_k \neq 0 \quad \text{and}$$

$$\sigma_{k+1} = \sigma_{k+2} \ldots \sigma_m = 0.$$

The regularisation method for rank-deficient problems has two steps:

(i) replace A by a matrix of rank k

$$A_k = \sum_{i=1}^{k} u_i \sigma_i v_i^T \tag{2.9}$$

so that when A is replaced by A_k, we obtain the new least squares problem

$$\min \|A_k \mathbf{q} - \mathbf{b}\|_2^2$$

(ii) compute the minimum-norm least squares solution to

$$\min \|A_k \mathbf{q} - \mathbf{b}\|_2 \quad \text{subject to} \quad \min \|\mathbf{q}\|_2, \tag{2.10}$$

and the solution is given by

$$\mathbf{q}_k = \sum_{i=1}^{k} \frac{\mathbf{u}_i^T}{\sigma_i} \mathbf{v}_i.$$

This solution is referred to as the truncated SVD (TSVD) solution. However, it is common in regularisation problems to use a more general side constraint $\|L\mathbf{q}\|_2$, where L is the derivative operator defined in the next subsection. In this case, the solution is obtained by computing the least square solution to

$$\min \|A_k\mathbf{q} - \mathbf{b}\|_2 \quad \text{subject to} \quad \min \|L\mathbf{q}\|_2.$$

The solution $\mathbf{q}_{Lk}$ is given by [17]

$$\mathbf{q}_{Lk} = \mathbf{q}_k - \mathbf{V}_k^o \left(L\mathbf{V}_k^o\right)^{-1} L\mathbf{q}_k, \text{ where } \mathbf{V}_k^o = (v_{k+1}, \ldots, v_m).$$

This is known as the truncated GSVD solution.

The Tikhonov's regularisation (which will be described in the next section) can also be applied successfully to rank-deficient problems despite the fact that this method does not seem to involve the rank of the matrix.

Discrete ill-posed problems

(i) **Tikhonov's Regularisation**

With the Tikhonov's regularisation [55], the problem of finding a solution to the least square problem of $\min \|A\mathbf{q} - \mathbf{b}\|_2^2$ is transformed into the following minimisation problem:

$$\begin{aligned} \min \quad \Phi(\mathbf{q}) &= \|A\mathbf{q} - \mathbf{b}\|^2 + \lambda^2 \|L\mathbf{q}\|^2 \\ &= \Phi_d + \lambda^2 \Phi_m \end{aligned} \tag{2.11}$$

The function $\Phi(\mathbf{q})$ is the global objective function, Φ_d is the data misfit function and Φ_m is the model objective function. The parameter λ is a regularisation parameter that determines how well the solution fits the data. As λ becomes large, the solution fits the data less well and as λ becomes very

small the solution starts to fit the noise. The regularisation parameter λ has to be adjusted in such way that the solution fits the data in some optimal way. This will be discussed in the next section.

The solution to the minimisation problem (2.11) can be estimated by differentiating the minimisation function with respect to each component q_i of $\mathbf{q}$ and setting each partial derivative to zero. This gives:

$$\left(A^T A + \lambda L^T L\right) \mathbf{q} = A^T \mathbf{b}. \tag{2.12}$$

An alternative formulation to (2.12) is

$$\begin{bmatrix} A \\ \lambda L \end{bmatrix} \mathbf{q} = \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix}. \tag{2.13}$$

Here L is the regularisation operator and is defined by [52]

$$\|L\mathbf{q}\|^2 = \int_{t_1}^{t_n} \left(\frac{d^N \mathbf{q}}{d\tau^N}\right)^2 d\tau. \tag{2.14}$$

Useful choices of the regularisation operator L are:

(a) when $N = 0$ (zero order regularisation), the norm of the model is minimised by

$$L_0 = I_m,$$

where I_m is the $m \times m$ identity matrix and m is the number of columns in A.

(b) when $N = 1$ (first-order regularisation), then the first derivative of the

model is minimised by

$$L_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & \dots \\ 0 & -1 & 1 & 0 & 0 & \dots \\ 0 & 0 & -1 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \in \mathbb{R}^{m-1 \times m}$$

(c) when $N = 2$ (second-order regularisation), then the second derivative of the model is minimised (or the smoothness of the model is maximised) by

$$L_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \dots \\ 0 & 1 & -2 & 1 & 0 & \dots \\ 0 & 0 & 1 & -2 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \in \mathbb{R}^{m-2 \times m}$$

The Tikhonov's regularisation method can be applied successfully to discrete ill-posed problems. It can also be applied to rank-deficient problems, despite the fact that this method does not seem to involve the numerical rank of the matrix.

(ii) **The least squares method with a quadratic constraint**

This regularisation method is almost equivalent to Tikhonov's method, and solves the following least squares problem with a quadratic inequality constraint [16]

$$\min \|A\mathbf{q} - \mathbf{b}\| \text{ subject to } \|L\mathbf{q}\| \leq \alpha,$$

or

$$\min \|L\mathbf{q}\| \text{ subject to } \|A\mathbf{q} - \mathbf{b}\| \leq \delta,$$

where α and δ are non-zero parameters that are playing the role of the regularisation parameter. This method requires prior knowledge of the noise level.

(iii) **The conjugate gradient method**

Another approach to solving the least squares problem is to apply the conjugate gradient method to the normal equation:

$$A^T A\mathbf{q} = A^T\mathbf{b}.$$

Here the coefficient matrix $A^T A$ is symmetric and positive semi definite. The conjugate gradient algorithm consists of a loop of five steps, as below [17]:

For $k = 1, 2, \ldots$

$$\begin{aligned}
\alpha_k &= \frac{\|A^T\mathbf{r}^{k-1}\|^2}{\|A\mathbf{d}^{k-1}\|^2}, \\
\mathbf{q}^k &= \mathbf{q}^{k-1} + \alpha_k\mathbf{d}^{k-1}, \\
\mathbf{r}^k &= \mathbf{r}^{k-1} - \alpha_k A\mathbf{d}^{k-1}, \\
\beta_k &= \frac{\|A^T\mathbf{r}^k\|^2}{\|A^T\mathbf{r}^{k-1}\|^2}, \\
\mathbf{d}^k &= A^T\mathbf{r}^k + \beta_k\mathbf{d}^{k-1},
\end{aligned}$$

where $\mathbf{r}^k$ is the residual vector $\mathbf{r}^k = \mathbf{b} - A\mathbf{q}^0$. The algorithm initialised with the starting vector $\mathbf{q}^0$, $\mathbf{r}^0 = \mathbf{b} - A\mathbf{q}^0$, and $\mathbf{d}^0 = A^T\mathbf{r}^0$. The number of iterations k plays the role of the regularisation parameter. It is necessary to estimate the regularisation parameter k very accurately.

2.2.5 Methods for computing the regularisation parameter

A good regularisation parameter should yield a fair balance between the perturbation error and the regularisation error in the regularised solution. The selection of the optimal regularisation parameter is based on minimising the total error. This situation can be described by the Figure 2.2. The regularisation error approaches zero as $\lambda \to 0$, while the error due to noise in the data increases as $\lambda \to 0$. Two major classes can be characterised for computing the regularisation parameter as

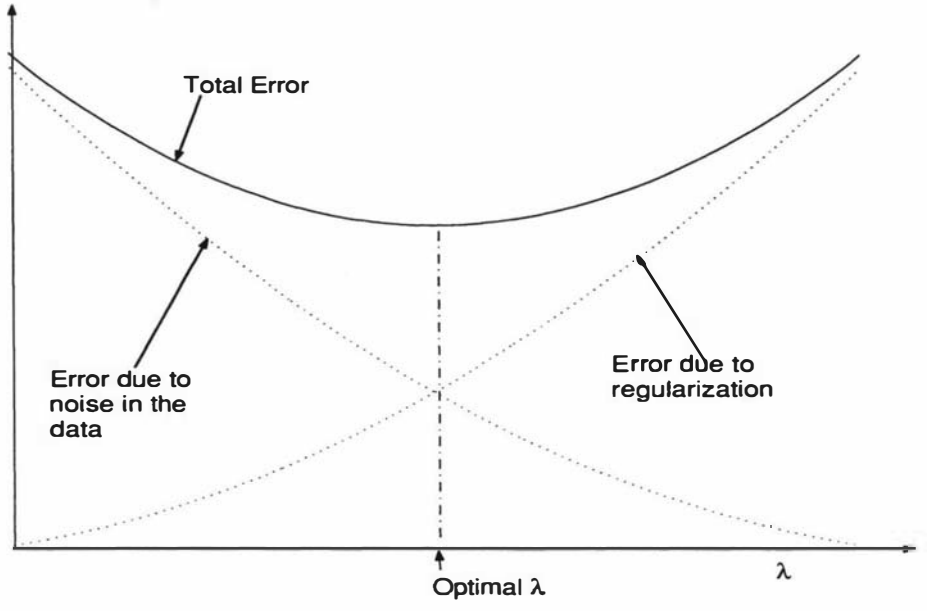

Figure 2.2: Total error in the regularised least square problems

follows:

(i) methods based on knowledge or a good guess of the noise level;

(ii) methods that do not depend on an *a priori* knowledge of the noise level.

A well-known method which relies on prior information on the noise level is the Discrepancy Principle. On the other hand, methods such as Generalised Cross Validation (GCV), L-curve and Quasi-optimality criterion are the most popular methods which do not require any information about the noise level.

The Discrepancy principle

The only method from class (i) is the Discrepancy principle [35]. The idea is simply to select the regularisation parameter λ so that the residual norm is equal to an *a priori* bound δ_- of the norm of the error ε in the right hand side, i.e.

$$\|A\mathbf{q}_\lambda - \mathbf{b}\| = \delta_- \text{ where } \|\varepsilon\| \leq \delta_- \tag{2.15}$$

The solution is obtained by solving following minimisation problem

$$\min \|L\mathbf{q}\| = \delta_{-} \tag{2.16}$$

subject to

$$\|A\mathbf{q} - \mathbf{b}\| \leq \delta_{-}.$$

Generalised cross validation (GCV)

Regularisation parameter estimation using generalised cross validation was first introduced by Wahaba [57]. This method does not require any prior information on the noise variance. The GCV method is based on the idea of removing one of the data points (b_i) from a given array of data ($\mathbf{b}$) and then solving the inverse problem using all the remaining data. The resulting inverse solution $\mathbf{q}$ will predict the right-hand vector $\mathbf{b}$, which can be used to estimate the missing element b_i.

Let the error in the misfit is

$$\|b_i - \hat{b}_i\|,$$

where $\hat{b}_i$ is the estimated value. For a fixed value of λ, repeat this process for n times (number of data points) and sum the error misfits. We then have

$$\sum_{i=1}^{n} \left(b_i - \hat{b}_i\right)^2.$$

If we repeat the whole process for a sequence of λ values, we then have the function

$$V(\lambda) = \sum_{i=1}^{n} \left(b_i - \hat{b}_{\lambda i}\right)^2.$$

The best value of λ can then be defined as the one that gives the smallest sum of the misfits. For the linear inverse problem, this corresponds to minimising the

generalised cross validation (GCV) function [57]. Here the GCV is defined by:

$$GCV(\lambda) = \frac{\|A\mathbf{q}_{reg} - \mathbf{b}\|_2^2}{(trace(I - AA^I))^2} \tag{2.17}$$

where A^I is a matrix which produces the regularised solution $\mathbf{q}_{reg}$, i.e. $\mathbf{q}_{reg} = A^I\mathbf{b}$.

The GCV method has proven its effectiveness in numerous applications [32]. However, we indicate two difficulties associated with this method:

(i) the minimum of the GCV is often very flat and, therefore, difficult to locate numerically,

(ii) the method may fail to compute the optimal value of λ when errors are highly correlated [17].

L-curve

Another popular tool for choosing the regularisation parameter that does not require any side information, is the L-curve method, introduced by Hansen [16]. The L-curve is simply a logarithmic plot of residual norm $\|A\mathbf{q}_\lambda - \mathbf{b}\|^2$ versus the solution norm $\|L\mathbf{q}_\lambda\|^2$ for a set of admissible regularisation parameter. In this way, the L-curve displays the compromise between the minimisation of these two terms. The 'corner' of the L-curve is defined as the point on the L-curve

$$(\zeta(\lambda), \mu(\lambda)) = (\log \|A\mathbf{q}_{reg} - b\|, \log \|L\mathbf{q}_{reg}\|) \tag{2.18}$$

with the maximum curvature. Here the curvature k is given by

$$k(\lambda) = \frac{\zeta'\mu'' - \zeta''\mu'}{((\zeta')^2 + (\mu')^2)^{\frac{3}{2}}}, \tag{2.19}$$

where differentiation is with respect to λ. The part of the L-curve to the left of the corner contains a region where the regularisation parameter is becoming smaller and the error between the original and reconstructed signal is dominated by the

perturbation errors. The solution norm is very sensitive to small changes in the regularisation parameter indicating a noisy solution. The right-hand side of the L-corner is a region where the regularisation parameter is gradually increasing and the residual norm is most sensitive to the changes in the regularisation parameter. In the right-hand side of the corner, the reconstructed solution is over smoothed.

Every method has advantages and disadvantages. The advantages of the L-curve are:

(i) robustness, and

(ii) the ability to treat perturbation defined by correlated noise.

The disadvantages are:

(i) concerned with the reconstruction of very smooth solution. The reconstructed solution $\mathbf{q}$ for which the corresponding SVD coefficients $\mathbf{u}_i^T\mathbf{q}$ decay fast to zero is dominated by the first few SVD components. For such solutions, [15] showed that the L-curve would fail and the smoother the solution the worse the estimation of λ computed by the L-curve,

(ii) related to its asymptotic behaviour as the problem size n increases. Vogel [59] showed that the regularisation parameter λ computed by the L-curve may not behave consistently with the optimal parameter λ_{opt}, as n increases.

The Quasi-optimality criterion

Another method which do not require any information of the noise level is the Quasi optimality criterion. This is defined for a continuous regularisation parameter λ by minimising the function [16]

$$Q(\lambda) = \lambda \|\nabla_\lambda \mathbf{q}\|_2. \tag{2.20}$$

2.3 Probabilistic Approach

The main difference between the deterministic and probabilistic approaches is that in the deterministic approach the aim is to find a reliable solution that fits the observed data, whereas in the probabilistic approach the parameters to be estimated are considered as random variables. The most popular probabilistic approach is the Bayesian method. In recent years the Bayesian method has become widespread in inverse problems ([10], [37], [53], [62], [64]). It provides a means to include a *priori* information on the unknown parameters in the procedure. The procedure is based on a formulation of the problem in terms of probability distributions. Here these probability distributions are often assumed to be Gaussian and a *posterior* estimate is derived.

2.3.1 Inverse problem using the Bayesian approach

We consider a data set of input vectors $\mathbf{x} = (x_i)_{i=1}^n$ along with corresponding output vector $\mathbf{y} = (y_i)_{i=1}^n$. We generally assume that $\mathbf{y}$ are corrupted with some noise and therefore we write (2.2), (2.3) as

$$\mathbf{y} = K(\mathbf{x}; \mathbf{q}) + \mathbf{e} \tag{2.21}$$

where $\mathbf{e}$ is the noise vector and $\mathbf{q}$ is a vector of adjustable parameters. The measurements $\mathbf{y}$ and the unknown parameters $\mathbf{q}$ are treated as random variables and we start with some prior understanding or expectation [44]. Imperfect prior understanding is quantified as a probability density function over the model space and then updated using new information. Measurements are quantified as a probability density function since the data usually have some measurement error. The key idea is to know how the measurement's probability density function maps into the model space [46], i.e. the conditional probability density $P(\mathbf{q}|\mathbf{y})$ of the unknown parameters have to be constructed for a given data $\mathbf{y}$. This is achieved using the

prior information and Bayes's theorem [46].

Bayes's theorem

Bayes's Theorem is a simple mathematical formula used for calculating conditional probabilities. It is given by [30], [44]

$$P(\mathbf{q}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{q})P(\mathbf{q})}{P(\mathbf{y})} \propto P(\mathbf{y}|\mathbf{q})P(\mathbf{q})$$

where $P(\mathbf{q})$ is called the *prior density* which describes our knowledge and assumptions about $\mathbf{q}$ prior to any measurements, $P(\mathbf{y})$ is the marginal density of $\mathbf{y}$, and $P(\mathbf{y}|\mathbf{q})$ is the marginal density of the measurement $\mathbf{y}$ given the parameters $\mathbf{q}$, called the *likelihood density*. The likelihood density can also be viewed as a model of how to simulate the measurements when $\mathbf{q}$ is known. $P(\mathbf{q}|\mathbf{y})$ can be regarded as a statistical solution to the inverse problem. In practice, the *posterior density* $P(\mathbf{q}|\mathbf{y})$ is used for the computation of different point estimates.

2.3.2 Estimation procedure

In the probabilistic framework the key idea is to construct the conditional probability density of the unknown parameters for given data. The main steps in the process of constructing the conditional probabilities are given here:

(i) If we assume that the noise values e_i are normally distributed with zero mean and variance σ^2, then

$$\begin{aligned} P(e_i|\sigma^2) &= N(e_i|0,\sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{e_i^2}{2\sigma^2}\right). \end{aligned} \tag{2.22}$$

(ii) From (2.21) and (2.22), it follows that the conditional distribution of the target variable, given the input variable and the parameters, is a Gaussian

distribution and therefore

$$P(y_i|\mathbf{q},\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(K(x_i;\mathbf{q}) - y_i)^2}{2\sigma^2}\right). \tag{2.23}$$

(iii) Since the data points are independent, the joint probability of the whole data set, given $\mathbf{q}$ and σ^2, is given by the product over all data points

$$\begin{aligned} L(\mathbf{q}) &= P(\mathbf{y}|\mathbf{q},\sigma^2) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum_{i=1}^{n}(K(x_i;\mathbf{q}) - y_i)^2}{2\sigma^2}\right). \end{aligned} \tag{2.24}$$

and is called the likelihood function of $\mathbf{q}$.

(iv) One popular technique for estimating $\mathbf{q}$ is called *maximum likelihood* and involves setting $\mathbf{q}$ to the value which maximises the function $L(\mathbf{q})$. We denote the value $\mathbf{q}$ which maximises (2.24) by $\mathbf{q}_{ML}$. Similarly we maximise (2.24) w.r.t σ^2, which gives the result

$$\frac{1}{\sigma^2_{ML}} = \frac{1}{n}\sum_{i=1}^{n}(K(x_i;\mathbf{q}_{ML}) - y_i)^2 \tag{2.25}$$

This provides us with an estimate of the noise level associated with the data under the assumed model.

2.3.3 Ill-posed problems

In the Bayesian approach we characterise the uncertainty in $\mathbf{q}$ through a probability distribution $P(\mathbf{q})$. Observation of data points modify this distribution by virtue of Bayes's theorem, with the effect of the data being mediated through the likelihood function. The following steps describe the process for constructing the solution.

(i) We define a prior distribution $P(\mathbf{q})$ which expresses our uncertainty in $\mathbf{q}$

taking account of all information apart from the data itself,

$$P(\mathbf{q}|\lambda) \propto \exp\left(-\lambda\Omega(\mathbf{q})\right) \tag{2.26}$$

where λ can be regarded as a hyper-parameter (parameter of the prior distributions).

(ii) We choose a Gaussian distribution for $P(\mathbf{q}|\lambda)$ of the form

$$P(\mathbf{q}|\lambda) = \left(\frac{\lambda}{2\pi}\right)^{M/2} \exp\left(-\frac{\lambda}{2}\|\mathbf{q}\|_2^2\right). \tag{2.27}$$

where M is the length of the vector $\mathbf{q}$.

(iii) We can now use Bayes's theorem to express the posterior distribution for $\mathbf{q}$ as the product of prior distribution and likelihood function

$$P(\mathbf{q}|\lambda, \sigma^2) \propto P(\mathbf{q}|\lambda)L(\mathbf{q}), \tag{2.28}$$

where $L(\mathbf{q}) = P(\mathbf{y}|\mathbf{q}, \sigma^2)$.

(iv) In the Bayesian method we make predictions of the mean value of $\mathbf{q}$ by integrating the posterior distribution with respect to $\mathbf{q}$. This mean value is given by [44]

$$E[\mathbf{q}] = \int \mathbf{q}P(\mathbf{q}|\lambda, \sigma^2)d\mathbf{q}. \tag{2.29}$$

(v) If we use the posterior distribution to find a point estimate for $\mathbf{q}$, then we choose to do this by finding the value of $\mathbf{q}$ which maximises the posterior distribution [44]. From (2.26–2.28), we see that maximising the log of posterior distribution is equivalent to minimising

$$\frac{1}{2\sigma^2}\|K(\mathbf{x};\mathbf{q}) - \mathbf{y}\|_2^2 + \frac{\lambda}{2}\|\mathbf{q}\|_2^2, \tag{2.30}$$

with respect to $\mathbf{q}$. This is the same as the regularised error function of the deterministic model. Thus there is very close similarity between the Bayesian approach and the conventional regularisation approach. The key difference between the two models is that in the Bayesian approach we make predictions by integrating over the distribution of the model parameters (see (2.29)), rather than by using a specific estimated value of $\mathbf{q}$.

2.3.4 Minimum Relative Entropy method

In the Minimum Relative Entropy(MRE) method the final posterior density function is chosen such that relative entropy between *prior* and *posterior* density function is minimised subject to the data constraint that the *posterior* mean solution matches the measured data [62]. According to Kullback [28], if a *prior* distribution is available, the *posterior* distribution that avoids bias is the pdf that minimises the relative entropy. The relative entropy between *prior* and *posterior* density function is defined as

$$H(p_1, p_2) = \int p_1(\mathbf{q}) \ln \left[\frac{p_1(\mathbf{q})}{p_2(\mathbf{q})}\right] d\mathbf{q} \tag{2.31}$$

where $\mathbf{q}$ is a model vector, p_1 is a *posterior* estimate and p_2 is a *prior* estimate.

2.3.5 Comparison of deterministic and probabilistic approach

In inverse problems, both Tikhonov's regularisation and Bayesian methods have received much interest in the recent literature. Skaggs & Kabala [52] addressed the release rate estimation inverse problem using Tikhonov's regularisation and Woodbury & Ulrych [62] addressed the same problem using minimum relative entropy and Bayesian methods. Later in 1998, both Kabala & Skaggs [21] and Woodbury & Ulrych [63] presented several arguments against each others methods. Then in 2000, Neupauer, Borchers, Wilson [40] found that many of these arguments were

not properly addressed using same test cases and provided some unbiased comparison. They estimated the 1-D ground-water release rate from a known location using both the Tikhonov's regularisation and the minimum relative entropy with the Bayesian method. They compared the methods based on the four categories of a proper comparison of the methods, the appropriate error model, uniqueness of the solutions, and the confidence interval on the solution. They found that the results obtained using both methods were reasonable when either multiplicative or additive noise is used and therefore the solution of each method does not depend on the type of error in the measurement data.

Both methods require some prejudiced inputs. For Tikhonov's method we have to select the order of regularisation and the method for calculating the regularisation parameter. Neupauer *et al.* [40] calculated the regularisation parameter using two different methods and found that both methods produced a similar result. For the order of regularisation, first and second order produced nearly the same results but zero order regularisation produced unsatisfactory results. The prejudged inputs for the Bayesian method include upper and lower bounds and prior expected values of the source concentration. The Bayesian solution is not sensitive to the prior expected value but is sensitive to the upper and lower bounds. They judged that the Tikhonov's method is less biased than Bayesian method.

Confidence intervals can be estimated for both Tikhonov's and Bayesian methods. As mentioned by Woodbury and Ulrych [63], Neupauer *et al.* [40] also found that the confidence intervals for the least squares solution with Tikhonov's regularisation are affected by the choice of the regularisation parameter and therefore the resulting confidence interval estimation is biased. In the Bayesian method the confidence intervals are based on the posterior probability function which is depend the measurement data and the dimension of the solution vector. Neupauer *et al.* [40] found that the confidence interval estimation using Tikhonov's method was less accurate than the Bayesian method in general.

Finally Neupauer *et al.* [40] compared the relative effectiveness of Tikhonov and Bayesian solution at reproducing different types of input source history functions. For a peaked function and a delta function the Bayesian method was more robust than Tikhonov's method when the data were either error free or noisy with known noise level. But for an unknown magnitude of noisy data Tikhonov's method performed better than the Bayesian method. On the other hand the Bayesian method had the advantage of identifying the region where the data gave no information about the source history function. But in this case Bayesian method will default to the prior expected value with a wide range of confidence interval. Several other articles [10], [37] also have been published to compare the solutions obtained using Bayesian and Tikhonov's regularisation. In general the results using both approaches are fairly similar and the extent to which the solutions obtained using these two methods differ depends on the magnitude of uncertainties in the data and the influence of the priori distributions and the regularisation operator in the calculation. Therefore the selection of the method is entirely up to the nature and primary aim of the problem to be solved. While both methods have their adherents, it is true that the development of both approaches are continuing vigorously today.

2.4 Problem Areas in Least Squares

2.4.1 Common problems

Three major difficulties in least square problems relate to failure of the basic assumptions of normality, common variance, and independence of errors [45]. In some situations it is not correct to assume normality of the error vector. The normal distribution of error vector is not required for the evaluation of parameters. Normality is only needed to construct the confidence intervals. The confidence interval creation is based on the t-distribution, which necessitates the underlying random variables to be normally distributed [45]. Confidence interval estimates can be

affected by non-normality, particularly when the underlying distribution is highly skewed or has fixed boundaries.

The assumption of constant variance is a key part in ordinary least squares. The minimum variance property of least squares estimation is based on this assumption. Equal weighting, as in ordinary least squares, does not give the minimum variance estimation if the variances are not equal. Therefore, the straight impact of mixed variances in ordinary least squares is a bias in the estimated values [45].

Correlations among the error vector may arise from many sources. It is common for the data collected in a time succession to have correlated errors. The error associated with an observation at one point in time will tend to be correlated with the errors of the immediately prior observations. The impact of correlated errors on ordinary least squares is loss in precision in the estimation and an invalid confidence interval estimation.

2.4.2 Error model

Two popular error models in the inverse problem literature [52], [62] are additive(Equation (2.2)) and multiplicative(Equation (2.3)) error models. Additive errors are normally distributed, uncorrelated, with zero mean and constant variance. Multiplicative error is log normally distributed and does not have constant variance across the full range of predictor.

Skaggs and Kabala [52] used the multiplicative error model to test their 1-d release rate estimation problem in ground-water modelling. The method developed by Skaggs and Kabala [52] is based on linear least squares with the help of Tikhonov's regularisation. Woodbury and Ulrych [62] raised several concerns about the approach. Then these concerns were further discussed by Kabala and Skaggs [21]. In the discussion they mentioned that (a) the multiplicative model is closer to the reality than additive model, (b) artificial data sets created with additive errors are simply unrealistic because the measurements near the edges of the sampling grid

do not tend toward zero and are frequently negative, (c) their model does not make any assumptions about structure of measurement error and the multiplicative form of error is consistent with the method, (d) additive form of error tends to create unrealistic data sets that are inconsistent with physically based transport models and with the non-negativity constraints with their methodology, (e) it is unlikely that true plume measurement errors follow a simple form, and consequently any technique that relies heavily on the assumption of a simple structure is lesser value. In the same journal Woodbury and Ulrych [63] replied to the comments of Kabala and Skaggs [21].

Following these two heated arguments Neupauer *et al.* [40] addressed the issue of the comparison of the error models to reconstruct the solution using same norm of multiplicative and additive error vector. During the comparison process they found that the inverse methods produce reasonable results and the results do not depend on the type of error model used. In most of the situations they could not justify that either error model is more appropriate than the other.

2.4.3 Confidence interval estimation

Let us assume the errors in (2.6) are random and normally distributed with zero mean and constant variance σ^2. To estimate the parameter $\mathbf{p}$ we find

$$\min_{\mathbf{p}} \sum_{i=1}^{n} \left(y_i - K(x_i; \mathbf{p})\right)^2 . \tag{2.32}$$

We can estimate the reliability by using confidence intervals. The $(1-\alpha)100\%$ confidence interval for $\mathbf{p}$ is given by [45], [22]

$$\mathbf{p} \pm t_{n-m,(1-\alpha)/2} \sqrt{E_{ii}}, \tag{2.33}$$

where

$$E = \sigma^2 \left(J(\hat{\mathbf{p}})^T J(\hat{\mathbf{p}})\right)^{-1},$$

J is the Jacobian w.r.t $\mathbf{p}$ of $y_i - K(x_i;\mathbf{p})$ and $t_{n-m,(1-\alpha)/2}$ is obtained from a t distribution table.

If the errors do not have the same variance, then instead of solving the problem (2.32) we should solve

$$\min_{\mathbf{p}} \sum_{i=1}^{n} \left(\frac{y_i - K(x_i;\mathbf{p})}{\sigma_i} \right)^2, \tag{2.34}$$

and therefore

$$E = \left(J_v(\hat{\mathbf{p}})^T J_v(\hat{\mathbf{p}}) \right)^{-1},$$

where J_v is the Jacobian w.r.t $\mathbf{p}$ of $(y_i - K(x_i;\mathbf{p}))/\sigma_i$. If the σ_i's are unknown then this is impossible and in practice (2.32–2.33) is used instead. However, the resulting estimate for $\mathbf{p}$ and the confidence interval will be biased to the extent that the σ's are not constant. In practice we estimate the value of the σ as

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - K(x_i;\mathbf{p}) \right)^2}{n-m}}. \tag{2.35}$$

In the process of checking the reliability of the confidence interval, we considered couple of small examples. In each example the data are corrupted using equal error norm of additive and multiplicative errors. We constructed data $y_i = K(x_i;\mathbf{p}) + e_i$ and using the above mentioned equations to calculate the parameter values and confidence interval. The parameter values, confidence intervals are calculated 10,000 times each time with different error vector and assuming the error variance are not known. For the 95% confidence region, approximately 95% of the times true parameter values are with in a confidence interval region for the data created using additive noise. But for the data created using multiplicative noise, only about 80-90% of the times true value are inside the confidence interval region. Therefore the confidence interval estimation for the multiplicative estimation is too small. This can be seen clearly from Table 2.1. Table 2.1 summarise the comparison of mean parameter estimates and its confidence interval(CI) for equal norm of multiplicative and additive errors.

Considered Examples	True value a	Multiplicative error mean(a)	Multiplicative error mean(CI)	Additive error mean(a)	Additive error mean(CI)
$y = ax + b$	10.00	10.0003	0.1132	9.9997	0.2266
$y = a\exp(x) + b$	10.00	10.0056	0.0847	10.0005	0.1056

Table 2.1: Comparison of estimation: multiplicative and additive error

2.5 Atmospheric Dispersion Modelling

2.5.1 Introduction

Atmospheric dispersion modelling is used to predict the local impacts of individual sources of pollution, e.g., an incinerator stack, or road traffic, accidental gas explosion, etc. It is one of the tool for establishing emission controls legislation; evaluating proposed emission controls techniques and strategies; selecting locations of future sources of pollutants, in order to minimise their environmental impacts; planning the control of air pollution episodes; and assessing responsibility for existing air pollution levels [48]. Dispersion modelling was initially developed in the 1950s for predicting surrounding pollutant concentrations from chimneys. Since then, the dispersion modelling of pollutants have been studied experimentally and theoretically by several researchers [5], [43], [48], [58]. These studies were based on gradient transfer theory, statistical theory and random walk theory. However, gradient transfer theory and statistical theory (Gaussian model) models were popular. These models vary in mathematical complexity and in their treatment of meteorological parameters.

Atmospheric dispersion models describe the transport and dispersion of pollutant gases. Pollutants are carried in wind by turbulent eddies. Small particles are also subject to random molecular motions but these motions are small for turbulent eddies and may be neglected in estimating particle dispersal [5]. Therefore any model that describes the dispersion of pollutants, requires appropriate information on wind and turbulence.

Common assumptions of these models, which include meteorological conditions, are unchanged over the time of the transport of pollutants and no chemical transformations. Their use at regional scales is limited because their underlying assumptions are not valid at those scales.

The input data for the dispersion models are source, meteorological and site data. The source may be classified into point, line, area and volume sources. Stack emissions and explosions are generally considered to be point source emissions; emissions from roadways with heavy movement of motor vehicles are classified as line sources; the area sources include a cluster of minor emissions that are difficult to estimate if considered individually; hazardous sites with considerable depth sources are called volume source. The most important data source is information on the amount of contaminant released from its location. The amount released may be measured as mass per unit of time (in the case of continuous release) and its total quantity (in the case of instantaneous release).

Meteorological data play a very important role in the dispersion of pollution, and typical meteorological data collected for dispersion calculations are wind velocity, percentage cloud cover, net radiation, temperature and stability.

2.5.2 Transportation of pollutant in the atmosphere

Transport of contaminants in the atmosphere is mainly controlled by advection, dispersion and transformation. Advection describes the movement of contaminant along with the bulk movement of air in the atmosphere. Dispersion describes the spreading of the contaminant as it moves through the atmosphere. Transformation is the process that converts the contaminant into another form of particles through chemical reactions. Most of the atmospheric dispersion model assumptions include no chemical transformation at all. Therefore, only the processes of advection and dispersion are considered here.

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. A source with release rate $q(t)$ kg/s is assumed to start at time $t = 0$ at a point $(0, 0, H)$ that is at a height H above the ground. The released 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:

$$\frac{\partial C}{\partial t} = -\nabla \cdot \mathbf{q} + S, \tag{2.36}$$

where $S(X, Y, Z, t)$ is the source release rate in $(kg/s)/(m^3)$ and the pollutant mass flux per unit area $\mathbf{q}$ is given by

$$\mathbf{q} = C\mathbf{u} - K\nabla C. \tag{2.37}$$

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

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

where K_x, K_y, K_z are eddy diffusivities in the X, Y and Z directions, respectively. Substitution into (2.37) gives an expression for the mass flux vector

$$\mathbf{q} = \left(CU - K_x \frac{\partial C}{\partial X}, -K_y \frac{\partial C}{\partial Y}, -K_z \frac{\partial C}{\partial Z} \right). \tag{2.38}$$

Substitution of the expression for **q** into (2.36) gives:

$$\frac{\partial C}{\partial t} + U\frac{\partial C}{\partial X} = \frac{\partial}{\partial X}\left(K_x\frac{\partial C}{\partial X}\right) + \frac{\partial}{\partial Y}\left(K_y\frac{\partial C}{\partial Y}\right) + \frac{\partial}{\partial Z}\left(K_z\frac{\partial C}{\partial Z}\right) + q(t)\delta(X)\delta(Y)\delta(Z-H). \tag{2.39}$$

The pollutant concentration approaches zero far from the source in the lateral directions and high above the ground, and there is zero vertical flux through the ground surface. The boundary and initial conditions may therefore be written in the form

$$C(X,Y,Z,0^-) = 0, \tag{2.40}$$

$$C(\pm\infty,Y,Z,t) = 0, \tag{2.41}$$

$$C(X,\pm\infty,Z,t) = 0, \tag{2.42}$$

$$C(X,Y,+\infty,t) = 0, \tag{2.43}$$

$$\frac{\partial C(X,Y,0,t)}{\partial Z} = 0, \tag{2.44}$$

where the domain of the problem is $-\infty < X < \infty$, $-\infty < Y < \infty$ and $0 < Z < \infty$.

2.5.3 Forward problem

If the contaminant source location (X_0, Y_0, H) and release rate $q(t)$ are known, the set of Equations (2.39 - 2.44) can be solved using analytical or numerical methods. Here X_0 and Y_0 are the distances in X and Y direction of the source from the receptor (pollutant measuring point). Finding $C(X,Y,Z,t)$ is the forward problem.

Instantaneous source release

Now consider the case of instantaneous release, e.g., from an explosion. The governing equation for contaminant transport in the atmosphere for an instantaneous source release in a three-dimensional domain is (2.39). The source term is repre-

sented by

$$q(t) = q_0\delta(t),$$

and the boundary, initial conditions are represented by (2.40 – 2.44). 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 [19]. However, the resulting functions are such that they make the analytical solution of (2.39) unobtainable. To simplify the model here, it is therefore assumed that $\mathbf{u}$, K_x, K_y and K_z are constants. Equation (2.39) then becomes:

$$\frac{\partial C}{\partial t} + U\frac{\partial C}{\partial X} = K_x\frac{\partial^2 C}{\partial X^2} + K_y\frac{\partial^2 C}{\partial Y^2} + K_z\frac{\partial^2 C}{\partial Z^2} + q_0\delta(t)\delta(X)\delta(Y)\delta(Z-H), \quad (2.45)$$

which is to be solved subject to the initial and boundary conditions. The solution of (2.45) can be derived using Laplace and Fourier transforms, and is:

$$C = \frac{q_0}{8\pi^{\frac{3}{2}}(K_xK_yK_z)^{\frac{1}{2}}t^{\frac{3}{2}}}\exp\left[-\frac{(X-Ut)^2}{4K_xt} - \frac{Y^2}{4K_yt}\right] \times\left(\exp\left[-\frac{(Z-H)^2}{4K_zt}\right] + \exp\left[-\frac{(Z+H)^2}{4K_zt}\right]\right). \quad (2.46)$$

Gaussian Model

The Gaussian model is the basic method to calculate air pollution concentrations from point or line source. The use of the model began to be popular when Pasquil [43] published his dispersion rates for plumes over open level terrain. The Gaussian model has achieved popularity because it is easy to use and most measured data fit the model reasonably well [61], [49]. It assumes that in the area of the emission and the wind speed remains the same with or without pollutant. This type of model applies to gases with the same density as air, or to very dilute gases irrespective of their density. The equations governing the dispersion phenomenon may be resolved by making the following assumptions [58]:

(i) during the transport of pollutants from source to receptor, the mass that is emitted from the source is assumed to remain in the atmosphere, i.e. none of the material is removed through chemical reaction,

(ii) the meteorological conditions are assumed to remain unchanged with time, at least over the period of transport from source to receptor,

(iii) the time-averaged concentration profiles at any distance in the cross-wind direction, horizontal are well represented by this distribution, and concentration profiles in the vertical direction are also well represented by a Gaussian, or normal, distribution,

(iv) molecular diffusion, which is $10^3 - 10^4$ times less than the turbulent diffusion can be ignored,

(v) the wind field is uniform as a function of space and time,

(vi) distances generally exceeding 100 meters from the source,

(vii) turbulence is homogeneous and isotropic.

The Gaussian model equation for the forward problem described by (2.46) can be written as following [41]:

$$C = \frac{q_0}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\sigma_z} \exp\left[-\frac{(X-Ut)^2}{2\sigma_x^2} - \frac{Y^2}{2\sigma_y^2}\right] \times \left(\exp\left[-\frac{(Z-H)^2}{2\sigma_z^2}\right] + \exp\left[-\frac{(Z+H)^2}{2\sigma_z^2}\right]\right). \tag{2.47}$$

where σ_i's are the standard deviation of the concentration distributions. Equations (2.46) and (2.47) are similar to each other. 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 model equations are identical. Therefore, there is a relation between the standard deviation of spread that arises in the Gaussian distribution and the eddy diffusivities in the advection dispersion equation.

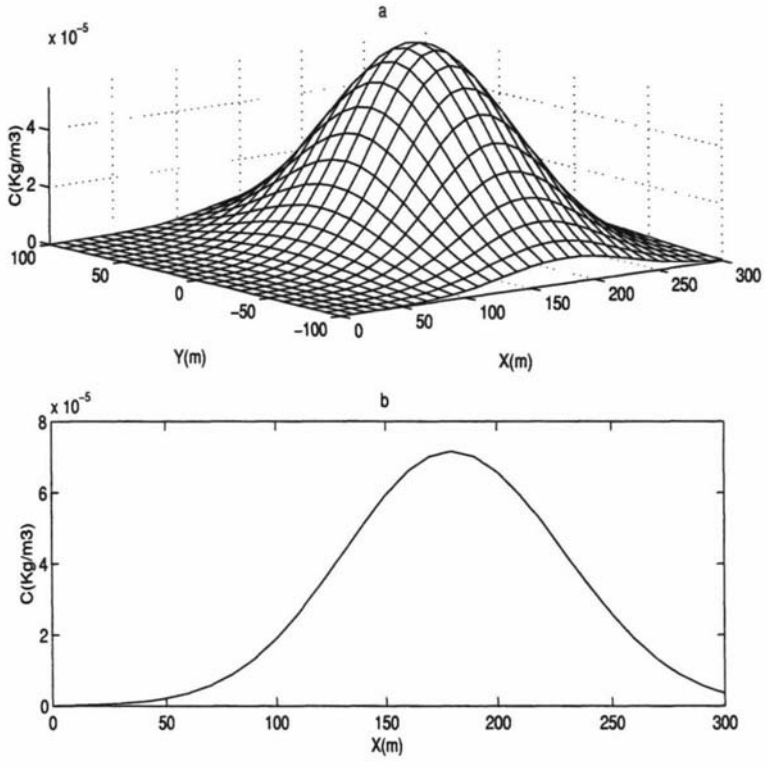

Figure 2.3: (a) Concentration distribution on the ground; (b) Concentration distribution on the ground directly downwind of the release (on $Y = 0$)

The ground distribution of the concentration predicted using Equation (2.46) for the data values $q_0 = 1000\ kg$, $K_x = K_y = 12\ m^2 s^{-1}$, $K_z = 0.2113\ m^2 s^{-1}$, $t = 100\ s$ are shown in the Figure (2.3). Figure (2.3a) shows the concentration distribution in the $X - Y$ plane, while Figure (2.3b) shows the concentration distribution directly downwind of the release (on $Y = 0$).

Continuous source release

A release rate $q(t)$, can be considered as a result of an infinite number of overlapping instantaneous releases [41]:

$$q(t)dt \tag{2.48}$$

where t is the release time for any elementary puffs. Therefore

$$dC = \frac{q(t)dt}{8\pi^{\frac{3}{2}}(K_xK_yK_z)^{\frac{1}{2}}t^{\frac{3}{2}}}\exp\left[-\frac{(X-Ut)^2}{4K_xt}-\frac{Y^2}{4K_yt}\right] \times\left(\exp\left[-\frac{(Z-H)^2}{4K_zt}\right]+\exp\left[-\frac{(Z+H)^2}{4K_zt}\right]\right). \tag{2.49}$$

Since the total concentration, C is given by $\int dC$, the concentration at the point (X_0, Y_0, Z_0) is

$$C = \int_0^t \frac{q(\hat{t})}{8\pi^{\frac{3}{2}}(K_xK_yK_z)^{\frac{1}{2}}(t-\hat{t})^{\frac{3}{2}}}\exp\left[-\frac{\left(X_0-U(t-\hat{t})\right)^2}{4K_x(t-\hat{t})}-\frac{Y_0^2}{4K_y(t-\hat{t})}\right] \times\left(\exp\left[-\frac{(Z_0-h)^2}{4K_z(t-\hat{t})}\right]+\exp\left[-\frac{(Z_0+h)^2}{4K_z(t-\hat{t})}\right]\right)d\hat{t}. \tag{2.50}$$

As with the Gaussian model, the concentration at the point (X_0, Y_0, Z_0) is defined by

$$C = \int_0^t \frac{q(\hat{t})}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\sigma_z}\exp\left[-\frac{\left(X_0-U(t-\hat{t})\right)^2}{2\sigma_x^2}-\frac{Y_0^2}{2\sigma_y^2}\right] \left(\exp\left[-\frac{(Z_0-h)^2}{2\sigma_z^2}\right]+\exp\left[-\frac{(Z_0+h)^2}{2\sigma_z^2}\right]\right)d\hat{t}. \tag{2.51}$$

2.5.4 Dispersion coefficients and standard deviation of distributions

Theoretical models are available to estimate the dispersion coefficients and wind speed as a function of vertical height. But these models make impossible to solve the advection dispersion equation analytically. To reduce the difficulties, Brustarert & Yeh [4] developed new models for estimating dispersion coefficients and wind speed using power law approximation. They showed that the power law approximations for dispersion coefficients and wind speed are good ones in the lower atmosphere by matching their model with the available theoretical model. Using this approach,

Huang [19] derived the following formula for dispersion coefficients and wind speed

$$U(Z) = aZ^p, \quad K_z = bZ^r, \quad K_y = \frac{1}{2}U\frac{d\sigma_y^2}{dX}, \tag{2.52}$$

where a, b, p and r are constants depending on the atmospheric stability and σ_y^2 is the mean square particle displacement along Y-axis. In practical applications, it is usually assumed that $K_x = K_y$ [48]. Knowing the stability condition of the atmosphere, one can determine the parameters a, b, p and r from the tables given by Brutsaert and Yeh [4].

The standard deviation of concentration distribution for the Gaussian model at a certain downwind distance X from the source can be calculated as follows

$$\sigma_x = \sigma_y = cX^d, \quad \sigma_z = eX^f. \tag{2.53}$$

Knowing the stability condition of the atmosphere, one can determine the parameters c, d, e and f from the tables given by Turner [58].

2.5.5 Stability classes

Atmospheric stability is important for the dispersion of gas as it decides how fast the dispersion in the air occurs. Dispersion is caused by a vertical temperature gradient in the atmosphere causing a buoyancy force on the dispersed gas. This atmospheric stability depends upon wind and sunlight. Strong sunlight heats the surface of the earth. The hot air rises, allowing the pollutant gas to mix up and disperse. Strong winds tend to dilute the pollutant gas and blow them out.

To solve an advection-dispersion equation, the values of wind speed and dispersion coefficients have to be known. But these values vary in space and have to be modelled using (2.52–2.53). The values of constants a, b, c, d, e, f, p and r in these equations are different for various stability classes, which depend on the meteorological conditions. Therefore it is very important to know the correct stability class

of the modelling environment when we solve the advection-dispersion equation.

Atmospheric stability has been divided into six classes, arbitrarily labelled A through F, with A being the most unstable. Turner [58] has suggested a way to estimate the stability classes of the atmosphere based on the angle of the sun, the extent cloud cover, and the surface wind speed.

2.6 Related Problems

2.6.1 Atmospheric modelling

The process of deducing the source term parameters from observations of airborne concentration reduces to estimating parameters in an air pollution model. In 1977, Kibler and Suttles [27] attempted to approximate the parameters in the air pollution model by means of simulated LIDAR (laser radar) data. An air pollution transport model developed by Smith is used as representative of the pollutant transport process. During the estimation process LIDAR sensor data are compared with the dispersion model out put using least-squares estimator to obtain a set of model parameters that closely fit the simulated data.

In 1991, Edwards *et al.* [6] developed a methodology for estimating source release parameters using non-linear regression analysis. This method is coupled with the existing three-dimensional MEDIC/MATHEW/ADPIC dispersion models. The regression technique optimises the agreement between the measured and model predicted concentration by varying the model parameters. Several other articles [36], [54] also have been published for similar problems in the same area. In these approaches air pollution transport models for the steady-state point source are used as representative of the pollution transport process, and none of these are based on the advection-dispersion model.

A variety of numerical and analytical techniques have been proposed to solve similar problems in the area of ground-water modelling. Because of the physical

and mathematical similarities between the mass transfer in water and air, inverse techniques used in ground-water modelling are equally relevant to the problem of air pollution modelling.

2.6.2 Ground-water modelling

One of the popular methods used to identify the source location in ground-water modelling is to run forward simulations and check the solutions with the measured data. This method follows an optimisation technique to obtain the best-fitted solution. One of the first group's researchers to handle the pollution source identification problem using the optimisation technique was Govelick *et al.* [11]. They modelled the ground-water pollution source identification as forward simulations in conjunction with an optimisation model based on linear programming and multiple regressions. The method used is restricted by the case where data are available in the form of breakthrough curves and it is also assumed that the physical parameters are known exactly. The breakthrough curves are defined to be the relationship between the contaminant concentration vs time.

Wanger [60] developed a methodology for simulation model parameter estimation and source characterisation. This inverse model combines ground-water flow and contaminant transport simulation with nonlinear maximum likelihood estimation to determine optimal estimates of unknown model parameters based on measurements of contaminant concentration.

Bagtzoglou [2] attempted to the solve advection-dispersion equation backwards in time without using optimisation techniques. In this work, the reversed time transport equation is modelled using random walk particle methods, for which the advective part of the transport model is reversed while the dispersion part is left unchanged. A probabilistic framework to identify the sources in heterogeneous media is presented in [2].

In 1994, Skaggs and Kabala [52] attempted to reconstruct the history of the

ground-water contaminant plume from measurements of its current spatial distribution in a one-dimensional medium. In this study, Tikhonov's regularisation is used in numerical experiments to recover the release history of a plume that has originated from known single site. In 1995 [51], Skaggs and Kabala used Quasi-reversibility methods for the same problem and found this approach was less accurate than the Tikhonov's method.

In 1996, Woodbury and Ulrych [62] applied minimum relative entropy (MRE) methods to Skaggs and Kabala's problem. They found that for the noise-free data, MRE was able to reconstruct the plume evolution history well, but for data with noise, the MRE solution is less accurate than the solution based on Tikhonov's method. Then in 2000, Neupauer, Borchers and Wilson [40] evaluated the relative effectiveness of Tikhonov's regularisation and minimum relative entropy to reconstruct two source history functions: a three-peaked Gaussian and a step function in a one-dimensional medium. They addressed the issues of the reproducibility of the solution and the appropriateness of the models for simulating random measurement error. They found that Tikhonov's regularisation and minimum relative entropy are effective in reconstructing the smooth source function. The second-order regularisation in Tikhonov's method attempts to fit a smooth function and therefore is less accurate than minimum relative entropy in reproducing the step function. They also found that if the noise level is known exactly, both methods perform equally well; however, if the noise level is underestimated, the minimum relative entropy method does not reproduced the solution well. Therefore the reconstruction using Tikhonov's regularisation is more accurate when the data contain a measurement error of unknown magnitude.

Mahar and Datta [33], [34] developed a methodology based on nonlinear optimisation to identify the unknown ground-water contaminant sources. They used measured values of pollutant concentrations at selected locations as the input data for the model. This method directly incorporates the governing equations for the

physical process in the optimisation model.

Skaggs and Kaballa [50] used nonlinear least squares methods (NLS) to recover the release history without regularisation. In their study, a flow system is assumed to be one dimension, with the plume originating from a known single site. They found that the solution was very sensitive to noise and to the extent to which the plume is dissipated. They also found that recovering the contaminant release history using Tikhonov's regularisation method is more effective than the NLS method but, computationally, the NLS method is much cheaper than Tikhonov's method.

Chapter 3

Pollutant Modelling in a One-dimensional Medium

3.1 Introduction

This chapter introduces the problem of finding the source term parameters of a pollutant in a 1-d medium (e.g., pollutant modelling in a river). The problem involves a point source of pollutant at an unknown location in a river. The spatial distribution of the pollutant concentration are sampled at stations along the river. The sampled data are used to reconstruct the source term parameters. We consider two types of sources

(i) a pollutant from an instantaneous point source;

(ii) a pollutant from an extended (continuous) point source of constant release rate.

The main objective of this chapter is to understand and identify the difficulties associated with solving the inverse problem.

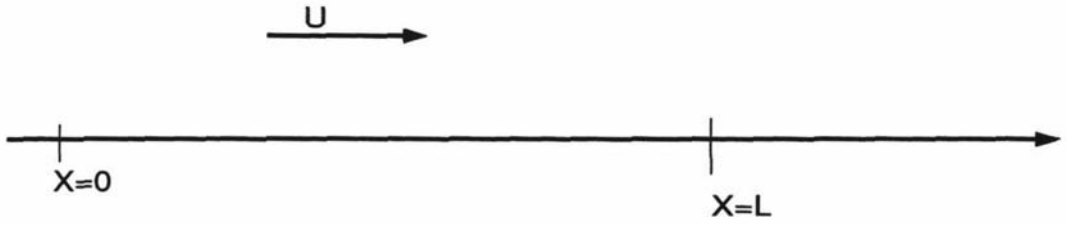

Figure 3.1: Forward model

3.2 An Instantaneous Point Source

3.2.1 The forward problem

It has been assumed that a pollutant of amount q_0 kg is introduced at $X = 0$ when $t = 0$. At $X = L$ the pollution concentration C is measured. The mathematical formulation of the problem can be written down as

$$\frac{\partial C}{\partial t} + U\frac{\partial C}{\partial X} - K_x\frac{\partial^2 C}{\partial X^2} = q_0\delta(X)\delta(t); \quad C(X,0) = 0,\ X \neq 0 \tag{3.1}$$

where U, K_x, q_0, and $C = C(X,t)$ are the velocity of flow, dispersion coefficient, amount of pollutant release and concentration, respectively, and δ 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 solution of (3.1) can be derived using Fourier and Laplace transforms, and is:

$$C = \frac{q_0}{\sqrt{4\pi K_x t}} \exp\left[-\frac{(X - Ut)^2}{4K_x t}\right]. \tag{3.2}$$

The forward problem is concerned with determination of the concentration in a medium, when the boundary and initial conditions, dispersion coefficients and the physical properties of the medium are known. The distribution of the concentration predicted using (3.2) when $t = 50,\ 500,\ 1500$ s, for the data values $q_0 = 1000$ kg,

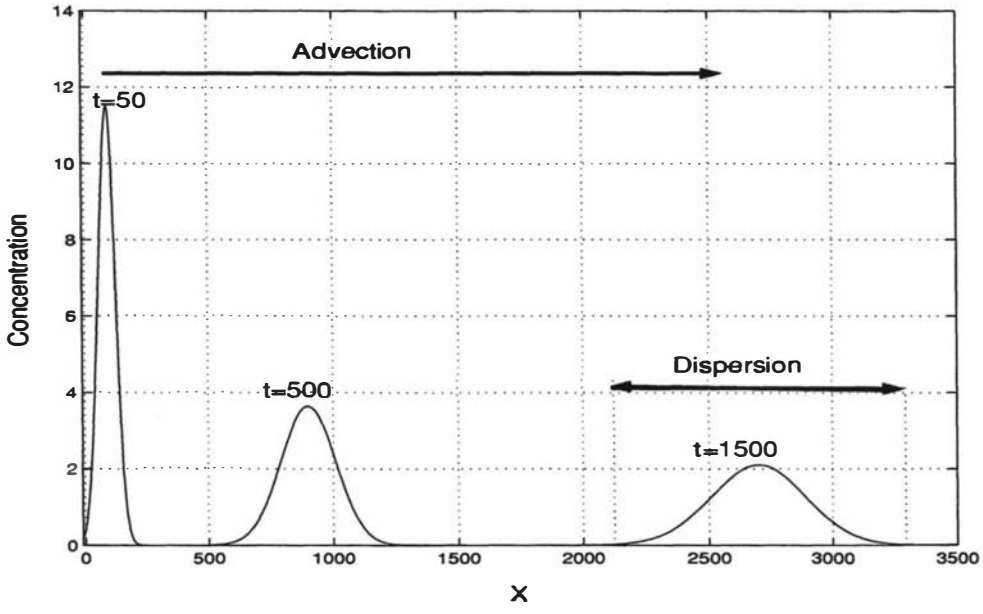

Figure 3.2: Advection and dispersion in 1-d medium

$K_x = 12\ m^2 s^{-1}$, $U = 1.8\ ms^{-1}$ is shown in Figure 3.2. This figure clearly shows the advection and dispersion of pollutant in 1-d medium.

3.2.2 The inverse problem

An inverse problem is concerned with the estimation of at least one of the quantities (boundary conditions, initial condition, dispersion coefficients, and the physical properties of the medium) by using the concentration measurements taken within that medium. Here we consider, the amount of pollutant release q_0 at $X = 0$, time of the release, dispersion coefficient K_x, and the distance $L = X_0$ from the source are regarded as unknowns. The additional information obtained from concentration measurements taken at a location $X = L$ at times t_i, $i = 0, 1, \ldots, n$ is then used for the estimation of unknowns. If t_0 is the unknown time after the pollutant is released when the concentration measurement at $X = X_0$ is started and T is the measurement clock time, then (3.2) gives:

$$C = \frac{q_0}{\sqrt{4\pi K_x (T + t_0)}} \exp\left[-\frac{(X_0 - U(T + t_0))^2}{4K_x(T + t_0)}\right]. \tag{3.3}$$

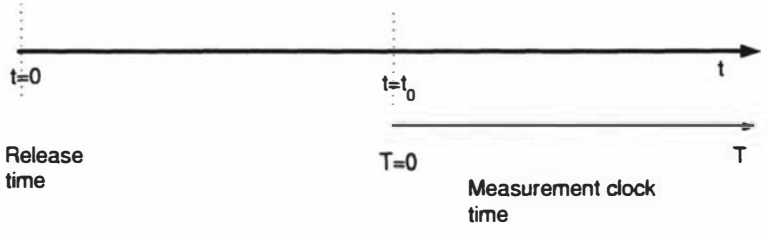

Figure 3.3: Release, Measurement clock time

In this section we explore the use of different techniques (approach) to find the unknown parameters in (3.3).

Approach 1

Taking natural logarithms on both sides of the Equation (3.2) and differentiation with respect to t gives

$$\frac{U^2}{K_x} = \frac{X_0^2}{K_x t^2} - \frac{2}{t} - 4F, \tag{3.4}$$

where $F = \frac{\dot{C}}{C}$. First and second order differentiation of Equation (3.4) with respect to t gives

$$\frac{2X_0^2}{K_x} = 2t - 4t^3\dot{F}, \tag{3.5}$$

and

$$2 - 4t^3\ddot{F} - 12t^2\dot{F} = 0 \tag{3.6}$$

respectively. If Δt is the time between two consecutive concentration measurements and concentrations at times $t-3\Delta t$, $t-2\Delta t$, $t-\Delta t$, t, $t+\Delta t$, $t+2\Delta t$ and $t+3\Delta t$ are available, then Equation (3.6) can be solved for t. $\dot{F}$ and $\ddot{F}$ are calculated using finite difference methods. Here we used the central difference scheme of the finite difference method. Then, by substituting the value of t in Equations (3.4) and (3.5), X_0 and K_x can be estimated if U is known.

Approach 2

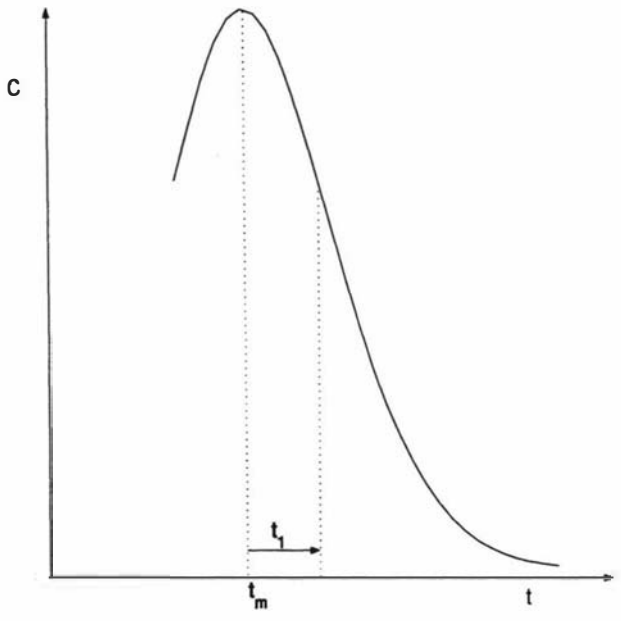

Figure 3.4: Approach 2, Measurement time

Assume that $t = t_m$ (see Figure 3.4) when $\dot{C} = 0$. Then (3.4) gives

$$\frac{U^2}{K_x} = \frac{X_0^2}{K_x t_m^2} - \frac{2}{t_m}. \tag{3.7}$$

Let $t = t_m + t_1$, where t_1 is the measurement clock time relative to the time t_m which corresponds to the peak of the concentration signal (see Figure 3.4). Then (3.4) gives

$$\frac{U^2}{K_x} = \frac{X_0^2}{K_x(t_m + t_1)^2} - \frac{2}{t_m + t_1} - 4F. \tag{3.8}$$

The time t_m can be estimated from (3.7), (3.8), and is

$$t_m = -t_1 + \frac{-t_1 U^2 + \sqrt{t_1 U^4 + 8K_x F(K_x t_1 + t_1^2 U^2)}}{4K_x F}, \tag{3.9}$$

provided the values U, K_x are known.

Approaches 1 and 2 involve at one stage differentiation of the measured data. The measured data always have errors, which lead to instability of the solution.

Therefore the Approaches 1 and 2 are only valid for data without any measurement errors unless special regularisation algorithm is used. The following approaches (Approach 3 and 4) are valid for the perfect data as well as for the noisy data. In both approaches, the parameter estimation problem is formulated as least squares minimisation model. The minimisation model optimises the agreement between the measured and model predicted concentrations by varying the model parameters within reasonable ranges of uncertainties.

Approach 3

It is assumed that n concentration values $C(X_0, T_i) = c_i$ $(i = 1, 2, \ldots, n)$ are available at the point $X = X_0$ at equal time intervals with some measurement uncertainties. The method is to find unknown parameters that best fit the data. Data fitting depends on varying a set of model parameters (such as X_0, t_0, K_x and q_0) to minimise the objective function of data misfit. Let $\mathbf{r} = [t_0, X_0, K_x, q_0]$, then $\hat{c}_i(\mathbf{r})$ be the model-computed concentrations in the same units as the measurements. We then define the data misfit objective function as

$$\|\mathbf{c} - \hat{\mathbf{c}}(\mathbf{r})\|_2^2, \tag{3.10}$$

where

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{c}}(\mathbf{r}) = \begin{pmatrix} \hat{c}_1(\mathbf{r}) \\ \hat{c}_2(\mathbf{r}) \\ \vdots \\ \hat{c}_n(\mathbf{r}) \end{pmatrix}.$$

According to the method of least squares, the optimum values of the model parameters $\mathbf{r} = [t_0, X_0, K_x, q_0]$ are obtained by minimising the objective function with respect to each of the parameters simultaneously. In this problem we use *MATLAB*'s optimisation routine *lsqnonlin* to calculate the optimal values of $\mathbf{r}$. One of the difficulties with this problem, however, is that there may be more than one

local minimum within a reasonable range of values of the parameters.

Approach 4

Taking natural logarithms on both sides of (3.3) gives

$$\ln(C) = \ln\left(\frac{q_0}{\sqrt{4\pi K_x}}\right) + \frac{UX_0}{2K_x} - \frac{1}{2}\ln(T+t_0) - \frac{X_0^2}{4K_x}\frac{1}{T+t_0} - \frac{U^2}{4K_x}(T+t_0), \quad (3.11)$$

which may be rearranged in the form

$$\ln(C) + \frac{1}{2}\ln(T+t_0) = \ln\left(\frac{q_0}{\sqrt{4\pi K_x}}\right) + \frac{UX_0}{2K_x} - \frac{X_0^2}{4K_x}\frac{1}{T+t_0} - \frac{U^2}{4K_x}(T+t_0). \quad (3.12)$$

Let $\beta_0 = \ln\left(\frac{q_0}{\sqrt{4\pi K_x}}\right) + \frac{UX_0}{2K_x}$, $\beta_1 = -\frac{X_0^2}{4K_x}$ and $\beta_2 = -\frac{U^2}{4K_x}$. Then (3.12) becomes

$$\ln(C) + \frac{1}{2}\ln(T+t_0) = \beta_0 + \beta_1\frac{1}{T+t_0} + \beta_2(T+t_0). \quad (3.13)$$

Let

$$\mathbf{f_1}(t_0) = \begin{pmatrix} \ln(c_1) + \frac{1}{2}\ln(T_1+t_0) \\ \ln(c_2) + \frac{1}{2}\ln(T_2+t_0) \\ \vdots \\ \ln(c_n) + \frac{1}{2}\ln(T_n+t_0) \end{pmatrix}, \qquad \mathbf{f_2}(\mathbf{b}, t_0) = \begin{pmatrix} \beta_0 + \beta_1\frac{1}{T_1+t_0} + \beta_2(T_1+t_0) \\ \beta_0 + \beta_1\frac{1}{T_2+t_0} + \beta_2(T_2+t_0) \\ \vdots \\ \beta_0 + \beta_1\frac{1}{T_n+t_0} + \beta_2(T_n+t_0) \end{pmatrix},$$

where $\mathbf{b} = [\beta_0, \beta_1, \beta_2]^T$ is vector of unknown linear parameters. Therefore, the objective function that provides minimum variance estimates is the ordinary least square function

$$\text{minimize} \quad \|\mathbf{f_1}(t_0) - \mathbf{f_2}(t_0, \mathbf{b})\|_2^2. \quad (3.14)$$

Since the minimisation function given in (3.14) has a combination of linear $\mathbf{b}$ and nonlinear t_0 parameters, we separate the solving process into two steps. We find the nonlinear parameter t_0 by constructing an iterative procedure where at each

iteration a multiple linear regression subproblem is solved to estimate the linear parameter **b** corresponds to that particular value of t_0. We use *MATLAB*'s optimisation routine *lsqnonlin* to determine t_0.

Table 3.1: Comparison of Estimates- Noise 0%

True values $t_0 = 150.0$, $X_0 = 500.0$, $q_0 = 1000.0$, $K_x = 12.0$

Parameters	Initial Values	Estimated Values	
		Approach-3	Approach-4
t_0	Any values	150.0	150.0
X_0	"	500.0	500.0
q_0	"	1000.0	1000.0
K_x	"	12.0	12.0

In Approach 3 the relationship between data and the unknown coefficients of the function is non-linear. Therefore we use an iterative procedure to estimate the unknown coefficients. It may converge at the local minimum or global minimum. Firstly, we found all or most of the local minima by solving the minimisation problem (3.10) many times as possible, at each time with a different initial value. Parameter values at all function minima are tabulated in Tables 3.1, 3.2 and 3.3 for the data corrupted by the noise 0%, 10% and 25% respectively. It can be seen that when the data is perfect or contain smaller amount of noise the minimisation function considered has only one minimum. When the data is corrupted by larger amount of noise the considered minimisation function has more than one minimum.

In Approach 4, we transformed both data and the function so that the function has linear relationship between $\mathbf{f}_1$ and unknown coefficients for a given value of t_0. Again here the procedure is iterative but only depends on the initial value of t_0. Similar to the previous Approach, firstly we found all the local minima of (3.14). The parameter values at each function minimum are tabulated in the Tables 3.1, 3.2 and 3.3 for the data corrupted by the noise 0%, 10% and 25% respectively. It can be seen clearly that the minimisation function (3.14) has only one minimum

Table 3.2: Comparison of Estimates- Noise 10%

True values $t_0 = 150.0$, $X_0 = 500.0$, $q_0 = 1000.0$, $K_x = 12.0$

Trials	Parameters	Initial Values	Estimated Values	
			Approach-3	Approach-4
1	t_0	366.0	162.3	149.4
	X_0	1417.0	521.6	498.5
	q_0	3919.0	1016.8	1000.7
	K_x	20.0	11.2	12.0
2	t_0	20.5	162.3	149.4
	X_0	1328.0	521.6	498.5
	q_0	3620.0	1016.8	1000.7
	K_x	6.0	11.2	12.0
3	t_0	123.0	162.5	149.4
	X_0	113.0	521.9	498.5
	q_0	887.0	1016.7	1000.7
	K_x	9.0	11.2	12.0

Table 3.3: Comparison of Estimates- Noise 25%

True values $t_0 = 150.0$, $X_0 = 500.0$, $q_0 = 1000.0$, $K_x = 12.0$

Trials	Parameters	Initial Values	Estimated Values	
			Approach-3	Approach-4
1	t_0	1005.0	185.6	148.7
	X_0	1138.0	562.8	496.9
	q_0	2305.0	1042.2	996.3
	K_x	9.0	9.9	12.1
2	t_0	122.8	4.5	148.7
	X_0	887.0	564.0	496.9
	q_0	1832.0	999.6	996.3
	K_x	6.1	18.9	12.1
3	t_0	19.9	185.6	148.7
	X_0	1192.0	562.8	496.9
	q_0	4081.0	1042.2	996.3
	K_x	19.1	9.9	12.1

for any data. Further the estimated solution using Approach 4 is much closer to the true value than the solution using Approach 3. Therefore Approach 4 might be better choice than Approach 3.

Approximate value of K_x

By taking logarithms on both sides of (3.2), we get

$$\ln(C) = \ln\left(\frac{q_0}{\sqrt{4\pi K_x}}\right) + \frac{UX_0}{2K_x} - \frac{1}{2}\ln(t) - \frac{X_0^2}{4K_x}\frac{1}{t} - \frac{U^2}{4K_x}t. \tag{3.15}$$

The graph of $\ln C$ as a function of t (Figure 3.5b) is approximately linear function

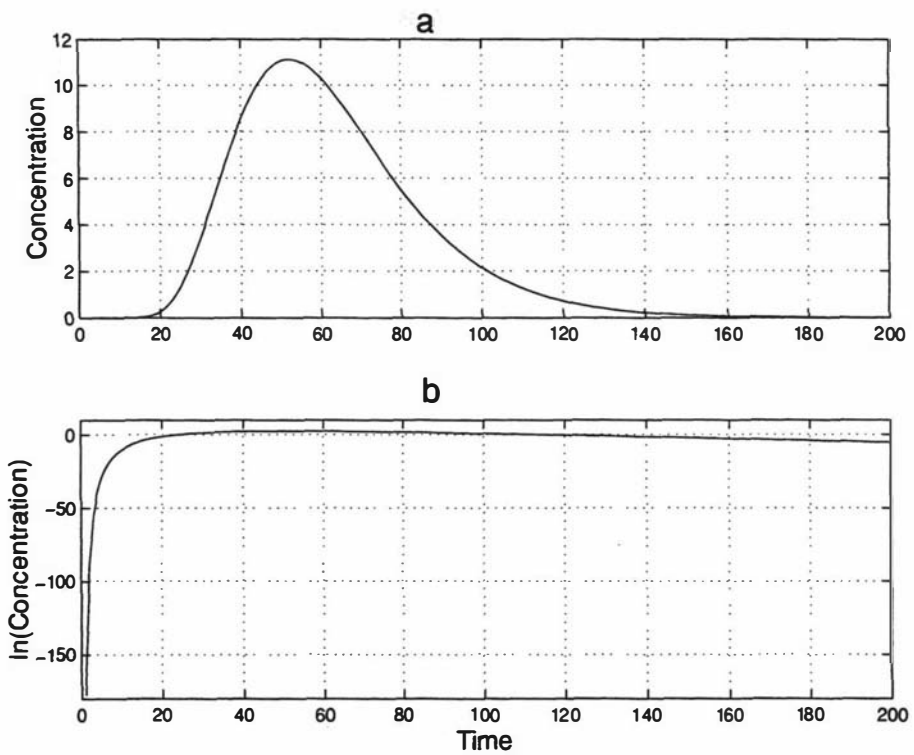

Figure 3.5: Approximation of K_x: (a) time *vs* concentration, (b) time *vs* ln(concentration)

for large values of t. Therefore the gradient of the linear part is approximately equal to

$$m = -\frac{U^2}{4K_x}. \tag{3.16}$$

The dispersion coefficient K_x can be approximated by Equation (3.16), provided U is known.

3.3 Extended Point Source Release

3.3.1 The forward problem

For this case, it has been assumed that a pollutant source of amount $q_0\ kgs^{-1}$ starts at $t = 0$ and stops at $t = t_f$ at a point $X = 0$. At $X = X_0$, as shown in Figure 3.1, the pollutant concentration C is measured. A mathematical model of this problem can be formulated as

$$\frac{\partial C}{\partial t} + U\frac{\partial C}{\partial X} - K_x\frac{\partial^2 C}{\partial X^2} = q_0\delta(X)\left[H(t) - H(t - t_f)\right], \quad C(X, 0) = 0, \tag{3.17}$$

where δ is the Dirac delta function and H is the Heaviside function, which have the following properties:

$$H(t) = 1 \quad \text{for} \quad t \geq 0 \quad \text{and} \quad H(t) = 0 \quad t \leq 0. \tag{3.18}$$

The solution of (3.17) at $X = X_0$ under the boundary and initial conditions is given by

$$C(t) = \int_0^t \frac{q_0}{\sqrt{4\pi K_x(t - \hat{t})}} \exp\left[-\frac{(X_0 - U(t - \hat{t}))^2}{4K_x(t - \hat{t})}\right] d\hat{t}, \quad \text{if } t \leq t_f, \tag{3.19}$$

$$C(t) = \int_0^{t_f} \frac{q_0}{\sqrt{4\pi K_x(t - \hat{t})}} \exp\left[-\frac{(X_0 - U(t - \hat{t}))^2}{4K_x(t - \hat{t})}\right] d\hat{t}, \quad \text{if } t > t_f. \tag{3.20}$$

3.3.2 The inverse problem

In this section, we formulate an inverse problem for the case when $t \leq t_f$. Let $\theta = t - \hat{t}$, then Equation (3.19) becomes

$$\begin{aligned} C(t) &= -\int_t^0 \frac{q_0}{\sqrt{4\pi K_x \theta}} \exp\left[-\frac{(X_0 - U\theta)^2}{4K_x\theta}\right] d\theta \\ &= \int_0^t \frac{q_0}{\sqrt{4\pi K_x \theta}} \exp\left[-\frac{(X_0 - U\theta)^2}{4K_x\theta}\right] d\theta \end{aligned} \tag{3.21}$$

Differentiating both sides with respect to t gives

$$\frac{dC}{dt} = \frac{q_0}{\sqrt{4\pi K_x t}} \exp\left[-\frac{(X_0 - Ut)^2}{4K_x t}\right]. \tag{3.22}$$

If t_0 is the unknown time after pollutant release when the concentration measurement at $X = X_0$ is started, and T is the measurement clock time, then the Equation (3.22) becomes

$$\frac{dC}{dt} = \frac{q_0}{\sqrt{4\pi K_x (T + t_0)}} \exp\left[-\frac{(X_0 - U(T + t_0))^2}{4K_x(T + t_0)}\right] \tag{3.23}$$

This equation is similar to the Equation (3.3) and therefore the unknown parameters can be estimated using the methods described in the section 3.2. But the processing of (3.23) involves the differentiation of measured data. Therefore this approach is only valid if the measured data is free from measurement error unless we use any regularisation methods. In the next section, we describe why the ordinary numerical methods cannot be used to construct the derivative of the noisy data, and the treatment using Tikhonov's regularisation method.

3.4 Differentiation of Measured Data

Many inverse problems involve the numerical differentiation of measured data. The mathematical operations of differentiation and integration are inverses of each

other. Therefore the differentiation of data might be considered as an inverse problem. If the derivative of the data is calculated using ordinary numerical methods, only an alternating-sign, tooth function of high amplitude can be obtained since the experimental data always contain errors, i.e. errors in the data lead to instability of the ordinary numerical methods of calculating the derivative. Special regularisation algorithms, which stabilise the solution process, must be used to construct the derivative of experimental data.

3.4.1 Difficulties with the numerical differentiation

If integration is an inverse operation with respect to differentiation, then the integral equation for the derivative can be expressed as

$$\int_{t_0}^{t} f'(x)dx = f(t) - f(t_0). \tag{3.24}$$

Let f be a differentiable function and f_δ a corrupted version of f with

$$\|f - f_\delta\| \leq \delta.$$

If we use the central difference approximation with step size h, then the Taylor's series expansion gives

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + O(h^2).$$

For the corrupted data

$$\frac{f_\delta(x+h) - f_\delta(x-h)}{2h} \approx \frac{f(x+h) - f(x-h)}{2h} + \frac{\delta}{h},$$

and therefore the error in the approximation E is defined as

$$E = O(h^2) + \frac{\delta}{h}.$$

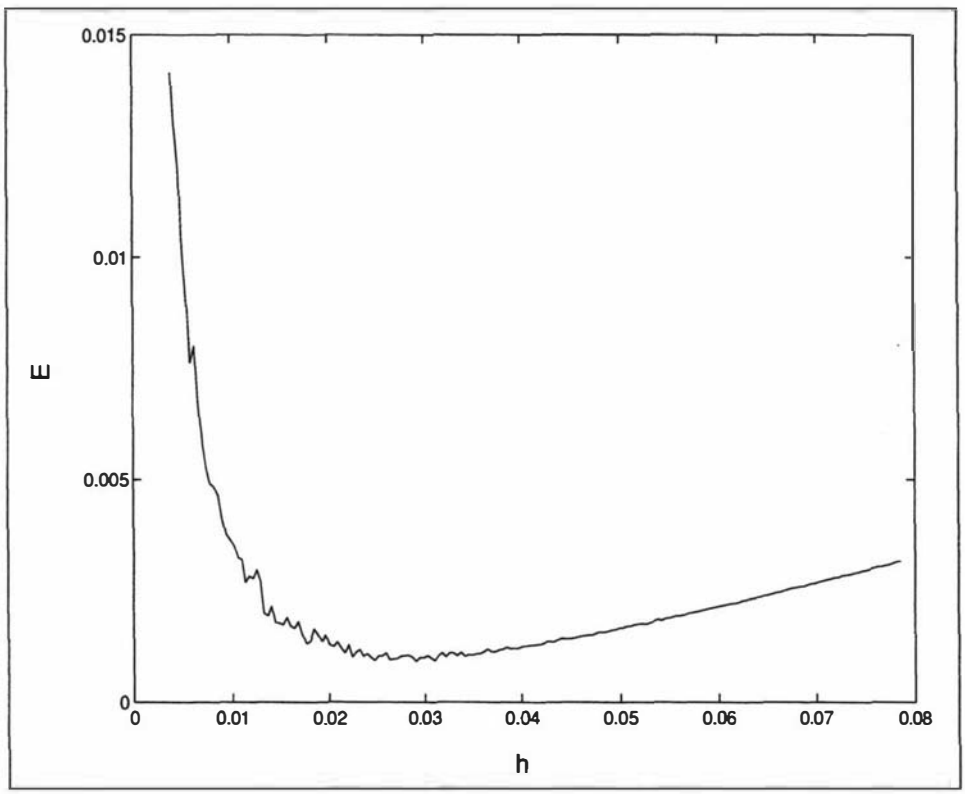

Figure 3.6: Total error when the relative noise in the data is 0.001%

Consider the function

$$f = \sin(x) \text{ for } x \in \left[0, \frac{\pi}{2}\right].$$

For a fixed error level $\delta = 0.001\%$, the graph of $E = \|f' - \frac{f_\delta(x+h) - f_\delta(x-h)}{2h}\|$ against h looks as it is in Figure 3.6. If h becomes too small, the total error E increases due to the error term δ/h. If h is too large, then the approximation error becomes too large. There is an optimal value of $h = h_0$ for the minimum error. But this is difficult to calculate since it depends on unavailable information on the exact experimental data.

3.4.2 Derivative construction with Tikhonov's regularisation

If $t = t_0 + ih$, $i = 0, 1, \ldots, n$ and h is the time difference between two data measurements, then using discretisation and the composite trapezium rule, (3.24) can be written as the following system of linear equations

$$\begin{pmatrix} 0 & 0 & 0 & 0 \ldots & 0 \\ 1/2 & 1/2 & 0 & 0 \ldots & 0 \\ 1/2 & 1 & 1/2 & 0 \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \; \vdots & \vdots \\ 1/2 & 1 & 1 & 1 \ldots & 1/2 \end{pmatrix} \begin{pmatrix} y(t_0) \\ y(t_1) \\ y(t_2) \\ \vdots \\ y(t_n) \end{pmatrix} = \begin{pmatrix} 0 \\ f(t_1) - f(t_0) \\ f(t_2) - f(t_0) \\ \vdots \\ f(t_n) - f(t_0) \end{pmatrix} \tag{3.25}$$

$$K\mathbf{y} = \mathbf{f},$$

where $\mathbf{y}$'s are the derivative of the right hand side. Now the minimisation problem for estimating $\mathbf{y}$ is formulated as

$$\min_{\mathbf{y}} \|K\mathbf{y} - \mathbf{f}\|_2^2.$$

In practice, the right-hand side is never known exactly but only up to an error $\delta > 0$. Furthermore, the coefficient matrix K in Equation (3.25) is ill-conditioned and therefore small variations of the right-hand side cause large variations in the solution. Special regularisation methods that stabilise the solution should be used to solve the problem.

Using Tikhonov's regularisation method, the new minimisation problem for estimating $\mathbf{y}$ is

$$\min_{\mathbf{y}} \left[\|K\mathbf{y} - \mathbf{f}\|_2^2 + \lambda^2 \|L\mathbf{y}\|_2^2 \right], \tag{3.26}$$

where λ is a regularisation parameter and L is a regularisation operator (see Equation (2.14)).

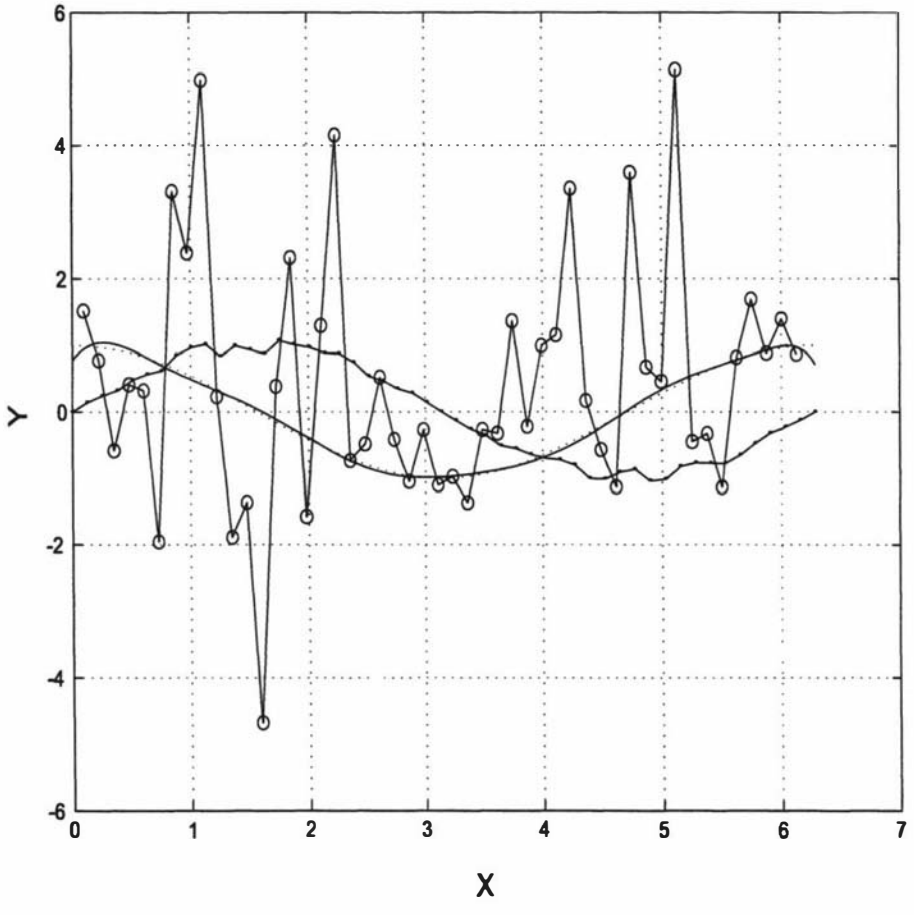

Figure 3.7: Derivative construction of noisy data

3.4.3 Numerical experiments

As an example, consider data generated from a 'noisy' *sine* function on the interval $X \in [0, 2\pi]$. In this example, $Y = \sin(X)$ is corrupted by adding normally distributed random noise of 15%. Figure 3.7 shows the derivative construction of noisy data. On the figure, simulated noisy data are shown by a solid line with dots, derivatives calculated by means of Tikhonov's regularisation are shown by solid line, derivatives constructed by means of ordinary numerical method (central difference) are shown by a solid line with empty markers, and *cosine* function is shown by dotted lines. It can be seen that the approach based on Tikhonov's regularisation reconstructs the derivatives well. The main input parameter of the Tikhonov's method is the order (N) of regularisation. The accuracy of the solu-

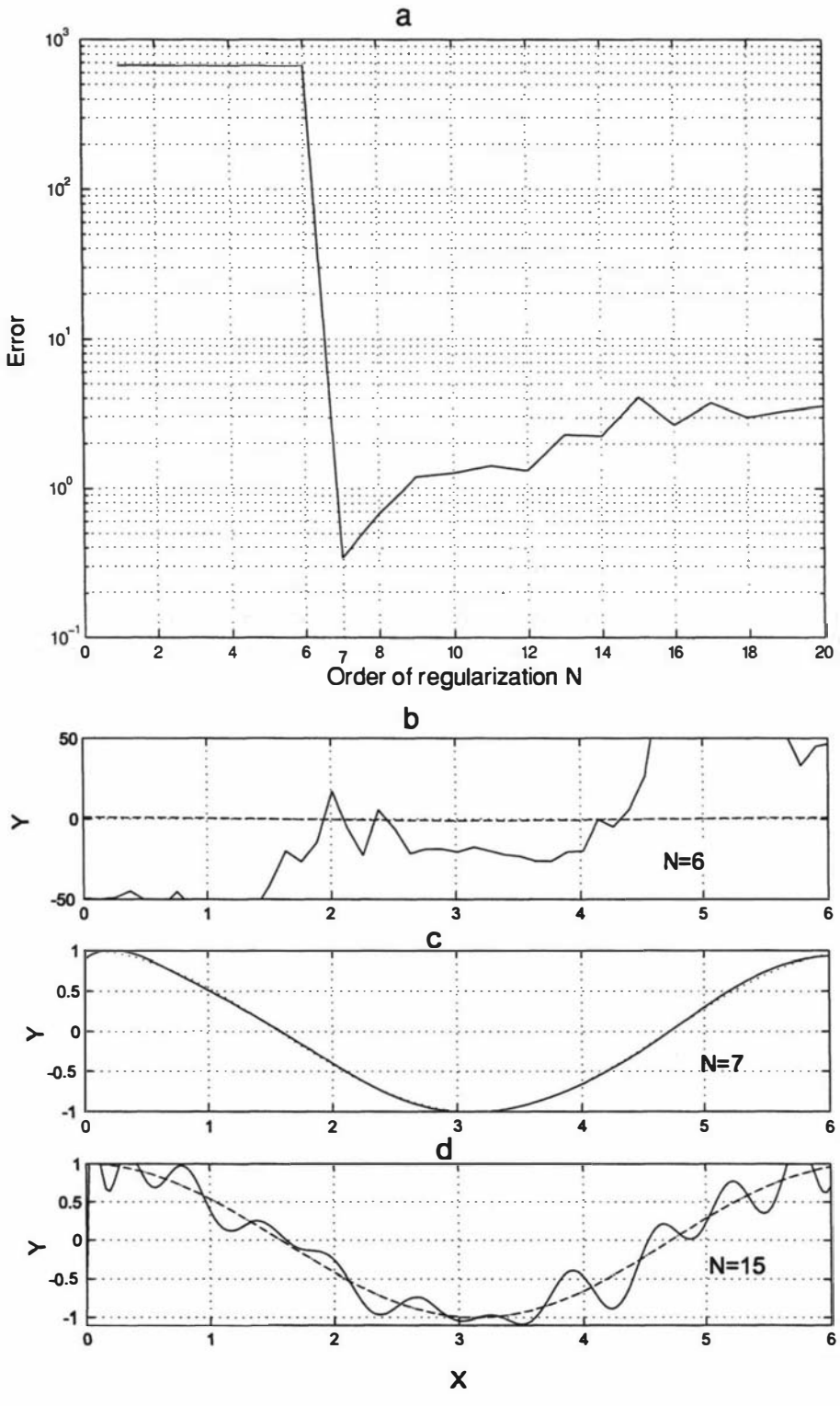

Figure 3.8: Comparison of derivative construction for different values of N

tion can be adjusted by choosing different values for N. Here N is the order of regularisation. A comparison of the derivatives, obtained by means of Tikhonov's regularisation for different values of N, are shown in Figure 3.8. Figure 3.8a depicts the true error in the reconstructed solution for different values of N. Figures 3.8b, c and d depict the reconstructed derivatives for the three values of N corresponding to under-smoothing ($N = 6$), approximate smoothing ($N = 7$) and over-smoothing ($N = 15$), respectively.

Choosing the value of N in a real world problem depends on the type of function to be estimated. It is hard to tell in advance about the value of N unless we know any information about the function to be estimated. Generally we use higher order regularisation if the function is smooth and lower order regularisation if the function is non smooth[52] .

Chapter 4

Pollutant Release from an Instantaneous Point Source

4.1 Introduction

The aim of this chapter is to develop an inverse model capable of estimating the released amount of atmospheric pollutant from an instantaneous point source at a known and unknown location.

In this problem, source information is not available, but measurements of the distribution of pollutant on the ground at some location are available over a time period. For the first case, we assume the source location and transport properties of the medium are known, but the release amount of the pollutant is unknown and the objective is to estimate its release amount. For the second case, only the transport properties of the medium and pollutant distribution over a time period are available, and the release amount and the location of the source are unknown. The objective here is to estimate the release amount and location of the source. The input to this model for the first case requires measured pollutant concentration data at a minimum of one location on the ground, whereas input to the model for the second case requires measured pollution concentration data at a minimum of

three locations on the ground.

The methodology for estimating the source release is based on nonlinear least squares regression, and linear regression coupled with the solution of an advection-diffusion equation for an instantaneous point source. 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.

4.2 The Forward Problem

Assume that a gas release with a total mass q_0 is released instantaneously at time $t = 0$ at a point $(0, 0, H)$ in a Cartesian co-ordinate system. The concentration profile at any point is given by the Equation (2.46) in Chapter 2. We can find the concentration of the pollutant at any point at any time using (2.46). This is the forward problem.

4.3 The Inverse Problem

The inverse problem 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 mathematical models. In this inverse problem, the structure of the equation is known; and measurements 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 optimal estimates of the parameter (q_0) for the Cases 1 and 2. Here Cases 1 and 2 refers to the first two categories of the problem under investigation which were described in Chapter 1.

4.3.1 Case 1- Known location

If t_0 is the time after pollutant release when the measurement clock is started and T is the time on that clock, then the concentration at the point $(X_0, Y_0, 0)$ is

$$C = q_0 f_1(t_0, T), \tag{4.1}$$

where

$$f_1(t_0, T) = \frac{1}{4\pi^{\frac{3}{2}} (K_x K_y K_z)^{\frac{1}{2}} (t_0 + T)^{\frac{3}{2}}} \exp\left[-\frac{(X_0 - U(t_0 + T))^2}{4K_x(t_0 + T)} - \frac{Y^2}{4K_y(t_0 + T)}\right] \times \exp\left[-\frac{H^2}{4K_z(t_0 + T)}\right].$$

For a given value of t_0, there is a linear relationship between C and f_1 and therefore q_0 is equal to the gradient of the straight line graph of f_1 vs C. The methodology for estimating q_0 for the case where t_0 is unknown is based on a nonlinear least-squares technique. In this case we separate the solving process into two steps. We first find t_0 by constructing an iterative procedure where the linear problem is solved to find q_0 at each iteration. The algorithm for estimating q_0 for the unknown t_0 is as follows:

(i) initialisation; choose $t_0 = t_0^0$,

(ii) Solve linear problem to estimate q_0^n,

(iii) compute new t_0^{n+1} from optimisation routine,

(iv) stop if $|t_0^{n+1} - t_0^n| < \delta$, otherwise goto step 1.

For the iteration, MATLAB's optimisation routine such as *lsqnonlin* is used.

4.3.2 Case 2- Unknown location

Taking natural logarithms of both sides of (4.1) when $Z = 0$ (i.e. for concentration on the ground) gives

$$\ln C = \ln\left(\frac{q_0}{4\pi^{\frac{3}{2}}(K_x K_y K_z)^{\frac{1}{2}}}\right) - \frac{3}{2}\ln(T+t_0) - \left(\frac{X^2}{4K_x} + \frac{Y^2}{4K_y} + \frac{H^2}{4K_z}\right)\frac{1}{T+t_0} - \frac{U^2(T+t_0)}{4K_x} + \frac{2UX}{4K_x}. \tag{4.2}$$

and $T + t_0 = t$, t_0 is the (unknown) time after pollutant release when the measurement clock is started and T is the time (known) on that clock. Taking logarithms of dimensional quantities is not exactly right. Therefore we divided (4.1) on both sides by a constant to non-dimensionalise the equation before taking logarithms. In simple terms, the right-hand side of (4.2) can be expressed as $f(T, \mathbf{b})$, where T is the independent variable and $\mathbf{b} = [q_0, H, X, Y, t_0]$ is a parameter vector.

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 for each parameter. The sensitivity coefficients for the model (4.2) are given respectively as

$$\frac{\partial f}{\partial q_0} = \frac{1}{q_0}, \quad \frac{\partial f}{\partial H} = \frac{-H}{2K_z(T+t_0)},$$

$$\frac{\partial f}{\partial Y} = \frac{-Y}{K_y(T+t_0)}, \quad \frac{\partial f}{\partial X} = \frac{-X}{K_x(T+t_0)} + \frac{U}{2K_x},$$

$$\frac{\partial f}{\partial t_0} = \frac{-3}{2(T+t_0)} + \left(\frac{X^2}{4K_x} + \frac{Y^2}{4K_y} + \frac{H^2}{4K_z}\right)\frac{1}{(T+t_0)^2} - \frac{U^2}{4K_x}.$$

The parameters can be estimated simultaneously without ambiguity if the sensitivity coefficients over the range of observations are not linearly dependent. Linear

dependence occurs when the relation

$$\alpha_1 \frac{\partial f}{\partial q_0} + \alpha_2 \frac{\partial f}{\partial H} + \alpha_3 \frac{\partial f}{\partial Y} + \alpha_4 \frac{\partial f}{\partial X} + \alpha_5 \frac{\partial f}{\partial t_0} = 0,$$

for each of the observations with not all α_j equal to zero [3]. If we set $\alpha_5 = 0$ and α_1, α_2 , α_3, α_4 are certain non-zero constants, it can be shown that

$$\alpha_1 \frac{\partial f}{\partial q_0} + \alpha_2 \frac{\partial f}{\partial H} + \alpha_3 \frac{\partial f}{\partial Y} + \alpha_4 \frac{\partial f}{\partial X} = 0.$$

This shows that the parameters in (4.2) cannot be estimated simultaneously, i.e. parameters cannot be estimated simultaneously if the data are collected at one location. Therefore measurements at more than one location are needed to estimate the parameters.

Now consider an experiment in which data are generated at more than one locations on the ground such as $P_1 = (X_0, Y_0, 0)$, $P_2 = (X_0 + x_1, Y_0 + y_1, 0)$, $P_3 = (X_0 + x_2, Y_0 + y_2, 0)$, ... Therefore (4.2) will become:

$$\begin{aligned} f = \ln & \left(\frac{q_0}{4\pi^{\frac{3}{2}} (K_x K_y K_z)^{\frac{1}{2}}} \right) - \frac{3}{2} \ln (T + t_0) \\ & - \left(\frac{(X_0 + x)^2}{4K_x} + \frac{(Y_0 + y)^2}{4K_y} + \frac{H^2}{4K_z} \right) \frac{1}{T + t_0} \\ & - \frac{U^2 (T + t_0)}{4K_x} + \frac{2U (X_0 + x)}{4K_x}, \end{aligned} \tag{4.3}$$

where $(x, y) = (0, 0)$, (x_1, y_1), (x_2, y_2), ... The equation (4.3) may be rearranged in the form:

$$f = \beta_0 + \beta_1 \frac{x}{T+t_0} + \beta_2 \frac{y}{T+t_0} + \beta_3 \frac{1}{T+t_0} + \frac{1}{K_x}\left(-\frac{(x^2+y^2)}{4(T+t_0)} + \frac{Ux}{2} - \frac{U^2(T+t_0)}{4}\right) - \frac{3}{2}\ln(T+t_0), \qquad (4.4)$$

where

$$\beta_0 = \ln\left(\frac{q_0}{4\pi^{\frac{3}{2}}(K_xK_yK_z)^{\frac{1}{2}}}\right) + \frac{2UX_0}{4K_x}, \quad \beta_1 = -\frac{X_0}{2K_x}, \quad \beta_2 = -\frac{Y_0}{2K_y}$$

$$\beta_3 = \left(\frac{X_0^2+Y_0^2}{4K_x} + \frac{H^2}{4K_z}\right).$$

Now, the sensitivity coefficients for the model (4.4) are given respectively as

$$\frac{\partial f}{\partial \beta_0} = 1, \quad \frac{\partial f}{\partial \beta_1} = \frac{x}{(T+t_0)},$$

$$\frac{\partial f}{\partial \beta_2} = \frac{y}{(T+t_0)}, \quad \frac{\partial f}{\partial \beta_3} = \frac{1}{(T+t_0)},$$

$$\frac{\partial f}{\partial t_0} = \frac{-\beta_1 x}{(T+t_0)^2} - \frac{\beta_2 y}{(T+t_0)^2} - \frac{-\beta_3}{(T+t_0)^2} + \frac{1}{K_x}\left(\frac{x^2+y^2}{4(T+t_0)^2} - \frac{U^2}{4}\right) - \frac{3}{2(T+t_0)}.$$

Linear dependence occurs when the relation

$$\alpha_1 \frac{\partial f}{\partial \beta_0} + \alpha_2 \frac{\partial f}{\partial \beta_1} + \alpha_3 \frac{\partial f}{\partial \beta_2} + \alpha_4 \frac{\partial f}{\partial \beta_3} + \alpha_5 \frac{\partial f}{\partial t_0} = 0$$

is satisfied, where not all of the α_i are zero. If we assume all the measurement locations are on a straight line, then it can be shown that the above linear dependence equation is true for some non-zero α_i. This result shows that the linear dependence occurs if we have only two points, and therefore we have to consider at least three location which are not co-linear.

Computation of parameters

The output of (4.4) 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 those estimates of the unknown parameters that best fit the measured data. If f is the log of measured concentration and $\hat{f}$ is the logarithm of modelled concentration, the error in the fit of the measurement and the model δ is

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

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

For the best match, $\mathbf{b}$ must be varied to minimise δ. This result can be achieved using the *MATLAB*'s optimisation routine *lsqnonlin*. Essentially, the procedure is iterative and requires good starting estimates for all 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:

$$W = f + \frac{3}{2} \ln \left(T + t_0 \right) - \frac{1}{K_x} \left(- \frac{\left(x^2 + y^2 \right)}{4 \left(T + t_0 \right)} + \frac{Ux}{2} - \frac{U^2 \left(T + t_0 \right)}{4} \right),$$

$$w_1 = \frac{x}{T + t_0}, \; w_2 = \frac{y}{T + t_0}, \; w_3 = \frac{1}{T + t_0}.$$

then (4.4) becomes

$$W = \beta_0 + \beta_1 w_1 + \beta_2 w_2 + \beta_3 w_3. \tag{4.5}$$

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

$$\delta = \sqrt{\sum_{i=1}^{3n} (W_i - \hat{W}_i)^2}. \tag{4.6}$$

For the best match, t_0 must be varied in the region $[T_0 - \epsilon, T_0 + \epsilon]$ to minimise δ. Here T_0 is the approximation of t_0 and ϵ is an error. The short algorithm for the estimating process is as follows:

(i) initialisation; choose $t_0 = t_0^0$,

(ii) estimate β's,

(iii) compute new t_0^{n+1} from optimisation routine,

(iv) stop if $|t_0^{n+1} - t_0^n| < \delta$, otherwise go to step 1.

For the iteration, a *MATLAB*'s optimisation routine such as *fminsearch* or *lsqnonlin* is used. For the new t_0, values β_0, β_1, β_2 and β_3 can be calculated from (4.5). Then, by substituting these values into (4.4), all the required parameters H, q_0, X_0 and Y_0 can be calculated.

Calculations of initial guess t_0

From (2.46), concentration distributions at the points P_1 and P_2 can be written as

$$C_1 = \frac{q_0}{4\,(\pi t)^{\frac{3}{2}}\,(K_x K_y K_z)^{\frac{1}{2}}} \exp\left[-\frac{(X_a - Ut)^2}{4K_x t} - \frac{Y_a^2}{4K_y t} - \frac{H^2}{4K_z t}\right], \tag{4.7}$$

$$C_2 = \frac{q_0}{4\,(\pi t)^{\frac{3}{2}}\,(K_x K_y K_z)^{\frac{1}{2}}} \exp\left[-\frac{(X_b - Ut)^2}{4K_x t} - \frac{Y_b^2}{4K_y t} - \frac{H^2}{4K_z t}\right], \tag{4.8}$$

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 (4.7) by (4.8) and then differentiating with respect to 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 $F = \frac{C_1}{C_2}$ and $F' = \frac{dF}{dT}$. The graph of $\left(\ln F + T\frac{F'}{F}\right)$ plotted against $\frac{F'}{F}$ is a straight line with the gradient $m = -t_0$ and the intercept at $-\frac{2x_1}{4K_x}$ (when $T = 0$). Note that 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. Some applications of this approach will be considered in section 4.4.

4.4 Modelling Applications

4.4.1 Source term estimation

In this section we evaluate the accuracy of the developed method to reconstruct the source term for the following cases:

(i) complete sampling (i.e. where the sampling would capture both the front and end edges of the plume. In this case $t_0 = 0$) of plume with no measurement error,

(ii) complete sampling with measurement error,

(iii) incomplete sampling of the plume with measurement error.

In all three cases we used a source strength q_0 kg located at $(0, 0, H)$ in the Cartesian co-ordinate system, and the sampling of concentration measurements is taken at $P_1(X_0, Y_0, 0)$, $P_2(X_0 + 480, Y_0 + 130, 0)$ and $P_3(X_0 + 130, Y_0 + 480, 0)$.

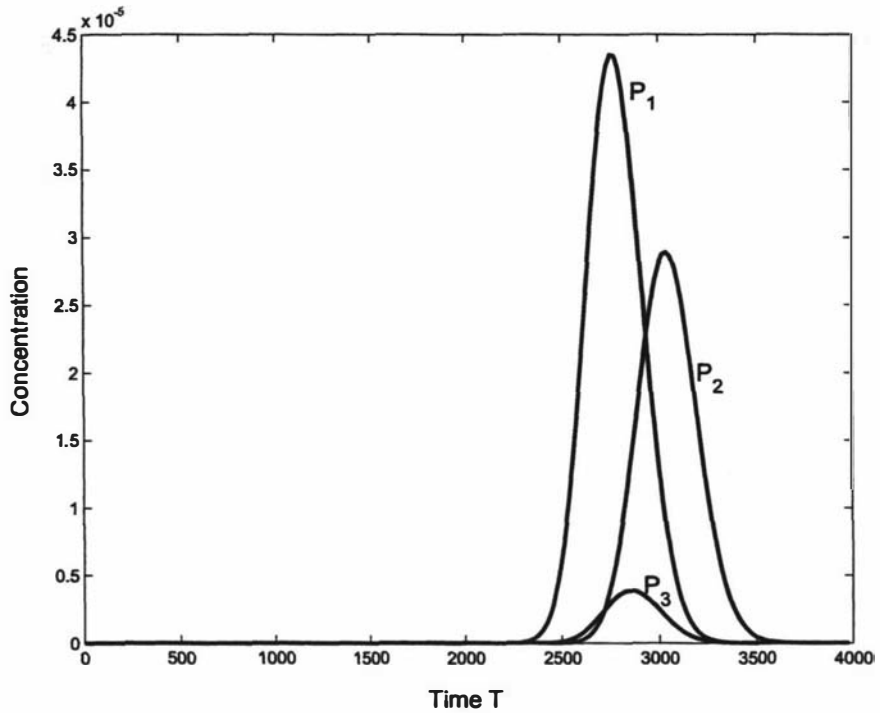

Figure 4.1: Complete sampling at P_1, P_2 and P_3 without measurement error

Case (i)

For the first case, complete samplings of the plume at the points P_1, P_2 and P_3 are shown in Figure 4.1. Transport parameters are $U = 1.8\ ms^{-1}$, $K_x = K_y = 12\ m^2s^{-1}$ and $K_Z = 0.2113\ m^2s^{-1}$. The results of reconstructed source term parameters are given in Table 4.1.

Table 4.1: Source term estimates-case (i)

X_0	Y_0	H	q_0
5000	100	20	1000

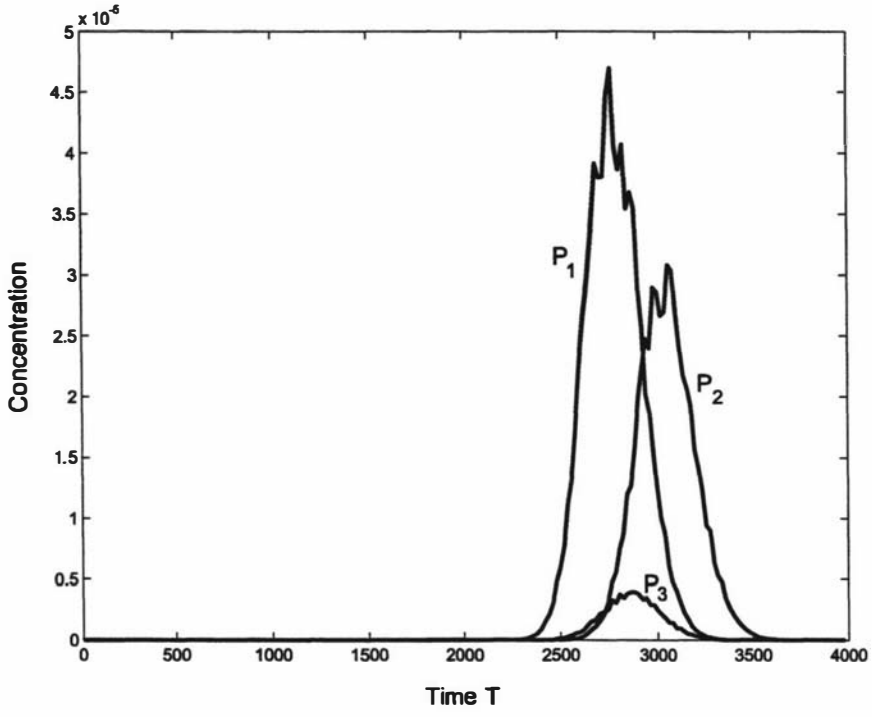

Figure 4.2: Complete sampling at P_1, P_2 and P_3 with measurement error

Table 4.2: Source term estimates-case (ii)

	X_0		Y_0		H		q_0	
True values	5000		100		20		1000	
Noise	$\bar{X}_0$	$\frac{\sigma(X_0)}{\sqrt{N}}$	$\bar{Y}_0$	$\frac{\sigma(Y_0)}{\sqrt{N}}$	$\bar{H}$	$\frac{\sigma(H)}{\sqrt{N}}$	$\bar{q}_0$	$\frac{\sigma(q_0)}{\sqrt{N}}$
2 %	5000	0.1	100	0.1	19.6	0.3	999	1
4 %	5000	0.1	100	0.1	18.7	0.7	998	11
6 %	5000	0.2	99	0.2	19.7	0.8	1020	18
8 %	5000	0.3	99	0.3	21.4	0.8	1020	19
10 %	5000	0.3	101	0.4	22.9	1.0	1040	27

Case (ii)

Sampled experimental data always contain measurement errors, and inverse models must be able to handle noisy data. We used the following error model:

$$data = (1 + \epsilon R) data_{exact},$$

where *data* is the measured concentration at any location, $data_{exact}$ is the true concentration at the same location, ϵ is the noise level and $R = 2(rand(n, 1) - 0.5)$ is the random deviate. Here *rand* is a *MATLAB* function that generates uniformly distributed random numbers in the interval (0.0, 1.0).

To evaluate the accuracy of the inverse model developed here for reconstructing the source term with noisy data, we created sets of data measurements with noise. These data at the points P_1, P_2 and P_3 are shown in Figure 4.2 and contain measurement errors with $\epsilon = 10\%$ relative noise. We then estimate the source term parameter values. This is done one hundred times, each time with different error vector for the noise levels 2%, 4%, 6%, 8% and 10%. The listed values in the Table 4.2 are the mean and standard error of the parameter estimates for the noise levels 2%, 4%, 6%, 8% and 10% along with the true values. It can be seen from the table that the error in the mean value of H, q_0 and its standard deviations are relatively higher than the other parameter values. In general, the standard deviation of the parameter estimates increases with the increasing noise level in the data and the accuracy of the estimated values are in accord with the error in the data, and the parameter estimates can be seen to be unbiased.

Case (iii)

In practical situations it is often not possible to sample the entire plume: for example, if we detect a pollutant, and then start to collect samples from the time of first detection onwards. In this situation, we might miss some measurements and do not know the time we start measurements (t_0) relative to the time of the start of release of the pollutants. To analyse the data in this situation, we create some sets of data measurements with noise. These data at the points P_1, P_2 and P_3 are shown in Figure 4.3. Data in this Figure contain a measurement error with $\epsilon = 10\%$ relative error.

Similar to the previous case we calculate the source term estimates one hundred

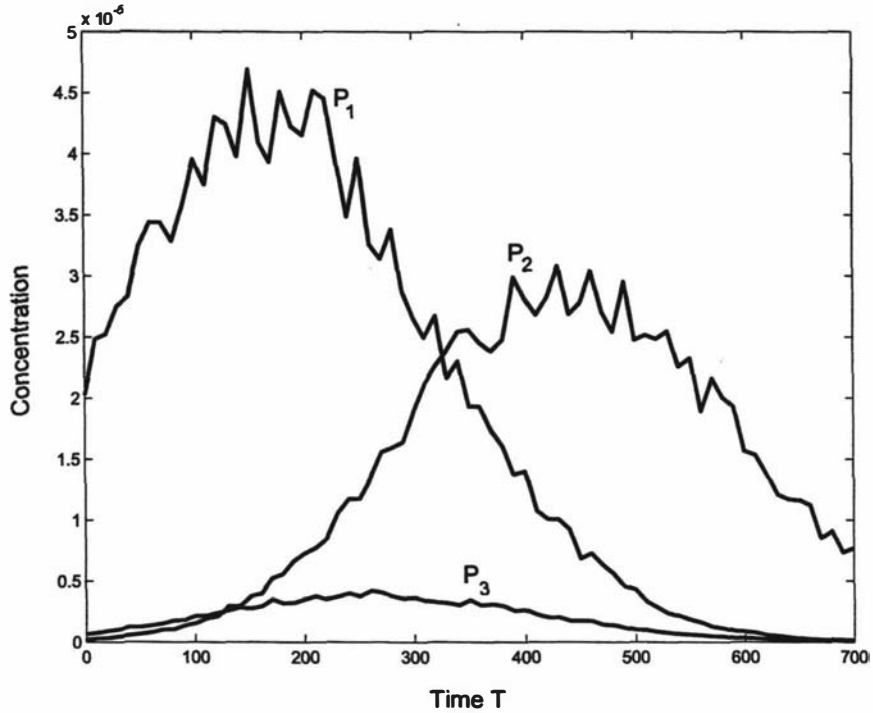

Figure 4.3: Incomplete sampling at P_1,P_2 and P_3 with measurement error

Table 4.3: Source term estimates-case (iii)

	t_0		X_0		Y_0		H		q_0	
True values	2600		5000		100		20		1000	
Noise	$\bar{t}_0$	$\frac{\sigma(t_0)}{\sqrt{N}}$	$\bar{X}_0$	$\frac{\sigma(X_0)}{\sqrt{N}}$	$\bar{Y}_0$	$\frac{\sigma(Y_0)}{\sqrt{N}}$	$\bar{H}$	$\frac{\sigma(H)}{\sqrt{N}}$	$\bar{q}_0$	$\frac{\sigma(q_0)}{\sqrt{N}}$
2 %	2556	4	4940	7	99	0.1	19.3	0.3	992	5
4 %	2520	7	4840	14	99	0.2	18.7	0.6	984	9
6 %	2376	11	4590	18	98	0.2	18.6	0.7	959	14
8 %	2268	11	4350	21	98	0.3	17.2	0.7	919	14
10 %	2088	14	4100	26	98	0.4	19.9	0.8	942	22

Table 4.4: Correlation coefficients

	t_0	X_0	Y_0	H	q_0
t_0	1				
X_0	0.9962	1			
Y_0	0.8408	0.8597	1		
H	-0.0943	-0.0092	0.2358	1	
q_0	0.8164	0.8660	0.8760	0.4913	1

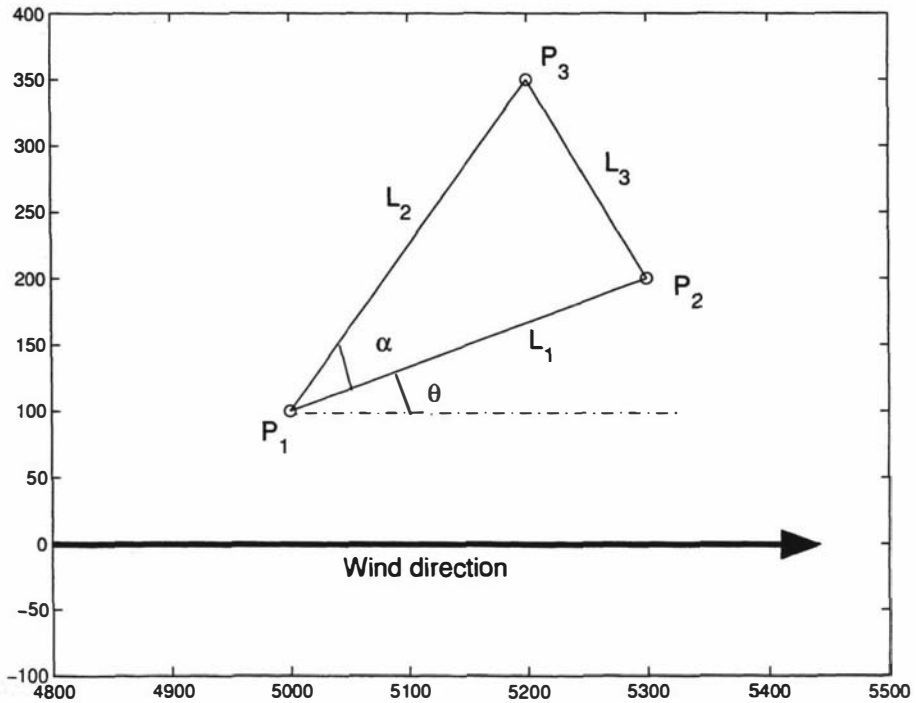

Figure 4.4: Locations of the points P_1,P_2 and P_3 on the ground

times, each time with different error vector for different noise levels. Listed in the Table 4.3 are the mean and standard error of the parameters t_0, X_0, Y_0, H and q_0 along with their true values. The values in the table shows that the estimated values are little biased and the error in the parameter value t_0 affects the other parameter values a lot, in particular X_0. We have to calculate the correlation coefficients between the parameters to compare how the error in one parameter value affects other parameter values. Table 4.4 shows the correlation coefficients between the parameters t_0, X_0, Y_0, H and q_0 for the estimated values in the Table 4.3. In cases where two variables have a correlation coefficient of 0.8 or greater (in absolute terms), may signal a high multi co-linearity problem [22]. In Table 4.4 correlation coefficients between t_0 and other parameters except H are greater than 0.8. These values clearly shows that there is a multi co-linearity in the problem. Nevertheless, the introduction of the non-linear parameter t_0 clearly makes it much harder to estimate the parameters.

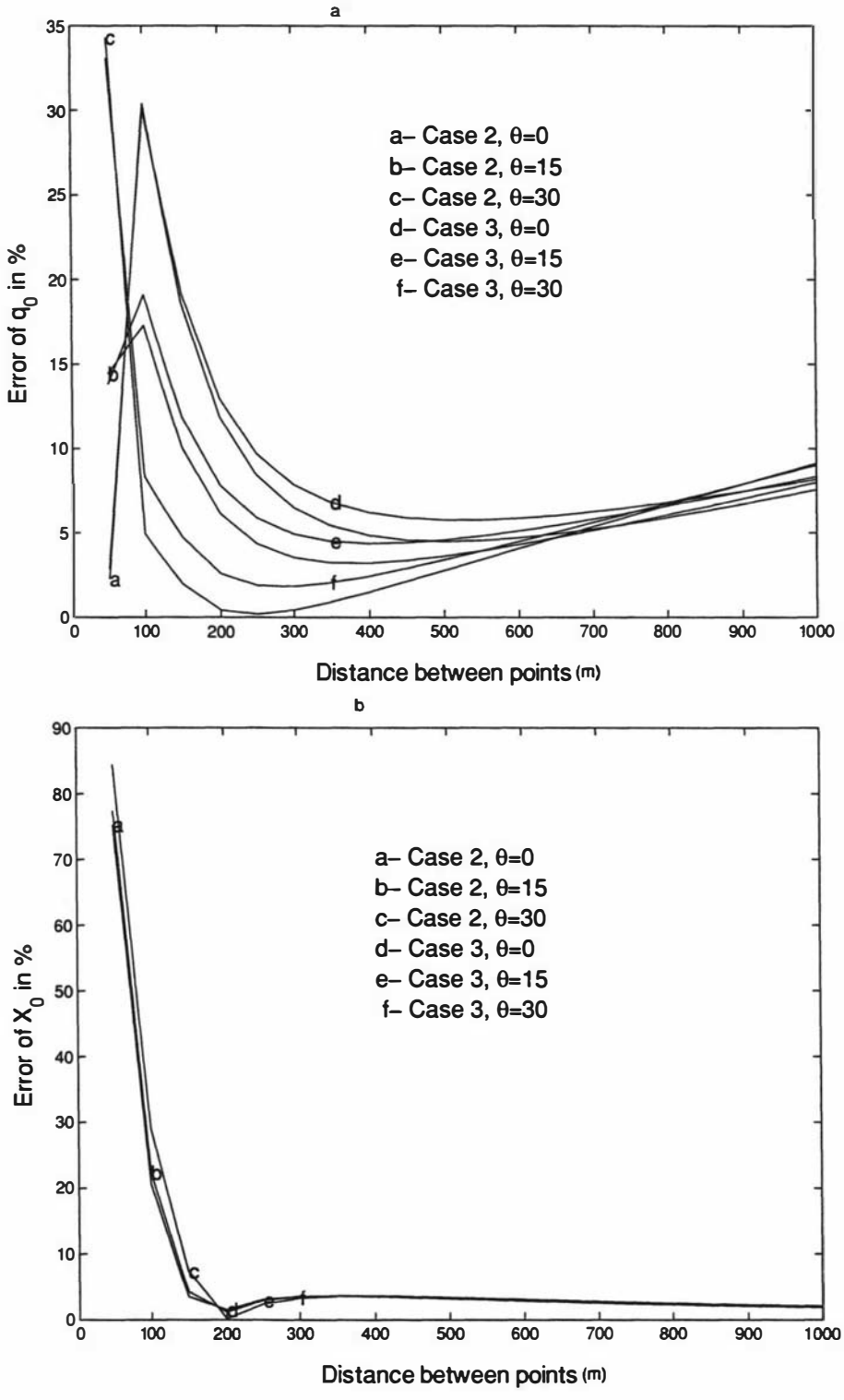

Figure 4.5: (a) Error in q_0 vs distance between points, (b) Error in X_0 vs distance between points

4.4.2 Measurement locations

The results of numerical experiments show that the accuracy of the parameter estimates depends on the locations of pollutant measurement points. We consider the three cases:

(i) all three locations P_1, P_2 and P_3 lie on a straight line (i.e. $\alpha = 0$). In Figure 4.4, P_3 is on the line P_1P_2.

(ii) stations P_1, P_2 and P_3 are on the vertices of a right angle isosceles triangle. In Figure 4.4, $\alpha = 90°$ and $L_1 = L_2$.)

(iii) P_1, P_2, and P_3 are the vertices of an equilateral triangle. In Figure 4.4, $\alpha = 60°$ and $L_1 = L_2$.

In the first case, whatever the values of θ, L_1, L_2 and L_3, the calculated parameter values are very far from the expected values, even in the case of perfectly simulated data. The errors in parameter estimates for q_0 and X_0 in the other two cases are plotted against the distance between the points in Figure 4.5 (a) and (b), respectively. In each case, parameter values are calculated when the angle (θ) 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 affect the accuracy of the estimation of the source term parameters.

It is clear that the quality of the source parameter estimation of pollutant is related to the location of the measurement stations. Therefore it is important that we carefully design measurement networks in order to maximise the accuracy of the model prediction and minimise the costs associated with data collection. In a real situation we should avoid the following in order to maximise accuracy of the estimates (a) all three measurement points are on a straight line, (b) any two-measurement points are parallel or perpendicular to the wind direction, (c) shorter distances between measurement locations. Transport in pollutants in the atmosphere is mainly controlled by advection and dispersion. The dispersion has

the effects in X, Y and Z direction and therefore we can maximise the available information about the dispersion of the pollutant by choosing any two measurement points which are not parallel or perpendicular to the wind direction.

4.5 Summary and Discussion

The goal of the work presented here is to develop an inverse model capable of simultaneously estimating the parameters appearing in an air pollution model with an instantaneous point source. The approach taken is to develop the inverse model as a nonlinear least squares estimation problem in which the source term parameters are estimated using pollutant concentration measurements on the ground. The least squares inverse model allows us to quantify the uncertainty in the parameter estimates, which in turn, allows us to quantify the uncertainty in the simulation model predictions.

First, we demonstrated that data from at least three spatial locations are necessary to estimate reliably the parameters in the model. Second, we formulated the inverse model as a least-squares minimisation problem, and tested the methodology using artificially generated data from the forward problem. The method is able to reconstruct the source term parameters with reasonable accuracy for the complete and incomplete noisy data, but the accuracy is high when the data is complete.

The accuracy of the estimated parameter values varies with the distance between the measurement locations. Therefore the optimal design of the locations for pollutant measurements on the ground is important. This is one possibility for improvements to the model.

The next chapter will be devoted to finding estimates of the release rates of source of non-steady continuous point source from a known location.

Chapter 5

Release Rate Estimation of Pollution from a Non-steady Point Source at a Known Location

5.1 Introduction

The aim of this Chapter is to develop an inverse model that is capable of estimating the release rate of an atmospheric pollutant gas from a time-dependent point source at a known location. In this problem, source information is not available, but measurements of the pollutant distribution are available over a known period of time. We assume that the source location and transport properties of the medium are known, but the release rate of the pollutant gas is unknown. Therefore, the goal is to construct the release rate function $q(t)$. The input to this model requires measured pollutant gas concentration data at a minimum of one location on the ground, and meteorological conditions such as wind speed and percentage cloud cover. To estimate the unknown release rate function $q(t)$, we use (2.50) along with the concentration measurements $C(X_0, Y_0, 0, t_i)$, where t_i, $i = 1, 2 \ldots n$, is the i-th sampling time, $(X_0, Y_0, 0)$ is the sampling location on the ground and n is the

number of sampled data.

As for many other inverse problems, the reconstruction of the source release rate from measurement data is a linear ill-posed inverse problem [52]. It can be either a rank-deficient or a discrete ill-posed problem. Several regularisation methods exist to solve such problems. In this chapter, Tikhonov's regularisation is mainly used to obtain a solution since it can be applied to both rank-deficient and discrete ill-posed problems. We use L-curve and *GCV* criteria to estimate the optimal regularisation parameter.

5.2 The Forward Problem

Assume that a gas release $q(t)$ kg/s is started at time $t = 0$ at a point $(0, 0, H)$ in a Cartesian co-ordinate system. Here the forward problem is concerned with solving (2.45) subject to the boundary and initial conditions (2.39–2.44) and its solution is given by the Equation (2.50)

5.3 The Inverse Model

In this inverse modelling problem, the structure of the equation (2.45) is known and the measurements of the inputs, time (t), concentration (C), and the source location, are available. The function $q(t)$ is unknown. The aim of this section is to develop the model to estimate the function $q(t)$.

The concentration of pollutant at a point $(X_0, Y_0, 0)$ can be expressed as

$$C(X_0, Y_0, 0, t) = \int_0^t K(t, \tau) q(\tau) d\tau, \tag{5.1}$$

where the kernel $K(t, \tau)$ is given by

$$K(t, \tau) = \frac{\exp\left[-\frac{(X_0 - Ut)^2}{4K_x(t-\tau)} - \frac{Y_0^2}{4K_y(t-\tau)} - \frac{H^2}{4K_z(t-\tau)}\right]}{4\pi^{\frac{3}{2}}(K_x K_y K_z)^{\frac{1}{2}}(t-\tau)^{\frac{3}{2}}} \quad \text{for } \tau < t.$$

5.3.1 Linear least-squares formulation

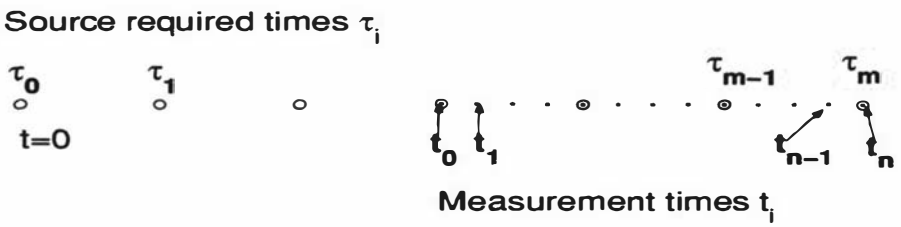

Figure 5.1: Source and data points

It is assumed that $n+1$ concentration values $C_i = C(X_0, Y_0, 0, t_i)$ are measured at the point $(X_0, Y_0, 0)$ at equal time intervals between t_0 and t_n relative to the release started time $t = 0$, as shown in Figure 5.1. The simplest way to proceed is to solve the Equation (5.1) on a mesh with uniform spacing. We subdivide the time interval $[0, t_n]$ into m such that $m \leq n$. Then by using the trapezoidal rule for the $n+1$ concentration values, (5.1) can be expressed as the following system of linear equations

$$\mathbf{c} = A\mathbf{q}, \tag{5.2}$$

where

$$A = \begin{pmatrix} \frac{K_{11}}{2} & \dots & \frac{K_{1l}}{2} & 0 & \dots & 0 \\ \frac{K_{21}}{2} & \dots & K_{2l} & \frac{K_{2(l+1)}}{2} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{K_{n1}}{2} & \dots & K_{nl} & \dots & \dots & 0 \\ \frac{K_{(n+1)1}}{2} & \dots & K_{(n+1)l} & \dots & \dots & \frac{K_{(n+1)(m+1)}}{2} \end{pmatrix} \Delta\tau,$$

$$\mathbf{c} = \begin{pmatrix} C_{t_0} \\ C_{t_1} \\ \vdots \\ C_{t_n} \end{pmatrix} \quad \text{and} \quad \mathbf{q} = \begin{pmatrix} q_{\tau_0} \\ q_{\tau_1} \\ \vdots \\ q_{\tau_m} \end{pmatrix}.$$

Here we used the Trapezoidal rule because it weights all the required source times equally with uniform time intervals and it is simpler to implement. It is also possible

to use other integration rules. Now the minimisation problem for estimating the release rate $\mathbf{q}$ is

$$\text{minimise} \quad Z(\mathbf{q}) = \|A\mathbf{q} - \mathbf{c}\|_2^2. \tag{5.3}$$

When we come across least squares problems a usual recommendation is not to believe the computed solution since we do not know the properties of the coefficient matrix A, which might or might not be ill conditioned. The first step towards the solution process should be to include an analysis of the coefficient matrix A. In the following two subsections we shall study the coefficient matrix A using singular value decomposition and the condition number.

5.3.2 Singular value decomposition analysis

An important tool for the analysis of the least square problem is the singular value decomposition (SVD) of the coefficient matrix. As discussed in Chapter 2, the SVD of A can be expressed as

$$A = W\Sigma V^T,$$

where the columns of W are the left singular vectors; the columns of V are the right singular vectors of A; and Σ is the matrix of singular values. We will now provide some numerical examples in order to illustrate the properties of the coefficient matrix A. These examples are created using the forward problem (2.50) for the values $n = 600$, $X_0 = 800, 1800, 2800, 3800$, $Y_0 = 100$, $H = 20$, $t_0 = 550$, $K_x = K_y = 12$ and $K_z = 0.2$. Source function discretisation size=100 and time between two data points=20 are used.

The singular value spectrum of these examples are shown in Figures 5.2. We can see that when the distance between the source and receptor is small the singular values decay to zero with a clear gap in the spectrum and smaller cluster of small singular values. Whereas when the distance is large the singular values gradually decay to zero with no clear gap in the spectrum and larger cluster of small singular

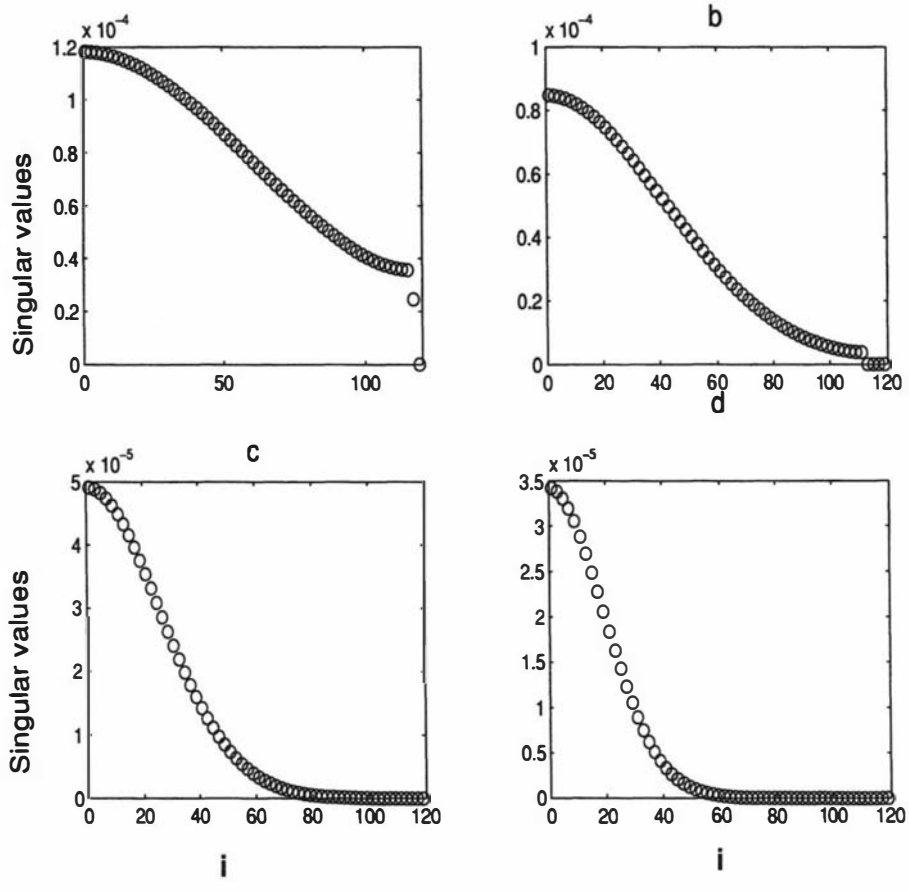

Figure 5.2: Singular values spectrum(a) $X_0 = 800$, (b) $X_0 = 1800$, (c)$X_0 = 2800$, (d) $X_0 = 3800$

values. Based on the definitions given in Chapter 2, the problem we are facing is either a discrete ill-posed problem or a rank-deficient problem.

The least squares solution of (5.3) using the SVD is given by

$$\mathbf{q} = \sum_{i=1}^{m} \frac{\mathbf{w}_i^T \mathbf{c}}{\sigma_i} \mathbf{v}_i, \tag{5.4}$$

where $\mathbf{w}_i$ and $\mathbf{v}_i$ are the column vectors of W and V respectively, and σ_i is the i-th singular value. This solution can be taken as products $\mathbf{w}_i^T \mathbf{c}/\sigma_i$ of each of the elements of vector $\mathbf{v}_i$. We will now use one of the considered examples to examine $\mathbf{w}_i^T \mathbf{c}/\sigma_i$ for perfect and noisy data. The results are shown in Figure 5.3. For perfect

data, the factors $\mathbf{w}_i^T\mathbf{c}$ decay as did the singular values (Figure 5.3 a). When dealing with noisy data, the factors $\mathbf{w}_i^T\mathbf{c}$ do not decay to zero; in fact they stabilise around the noise level (Figure 5.3 c). Figure 5.3 d shows $\mathbf{w}_i^T\mathbf{c}/\sigma_i$ 'blow up' as the singular values decay. This shows that high frequency components dominate in the solution and therefore the solution becomes meaningless.

5.3.3 The condition number of A

The degree of ill conditioning of the coefficient matrix A is proportional to the condition number that determines the sensitivity of the solution to small perturbations in the data. The coefficient matrix A in our problem depends mostly on the values of X, Y and the size of the source discretisation function. With the aid of specific numerical examples, we show how the condition number of the matrix A varies with the distance between source, observation sites and the discretisation size of the source function.

Figure 5.4 shows that the condition number of the coefficient matrix increases sharply with increasing distance between source, observation sites and increasing ratio between data points, source points. In the figure *ra* refers to the ratio between the data and source points.

5.4 Regularisation of the Least Square Problem

The coefficient matrix of problem (5.3) always has a very large condition number and therefore it is ill posed. This problem is either a rank-deficient or a discrete ill-posed problem and depends on the distance between the source, observation sites and the source discretisation size. In order to solve an ill-posed problem, well posedness must be restored by restricting the class of admissible solutions. This can be achieved using regularisation methods. Several regularisation methods have been discussed in Chapter 2. Tikhonov's regularisation is used in this research

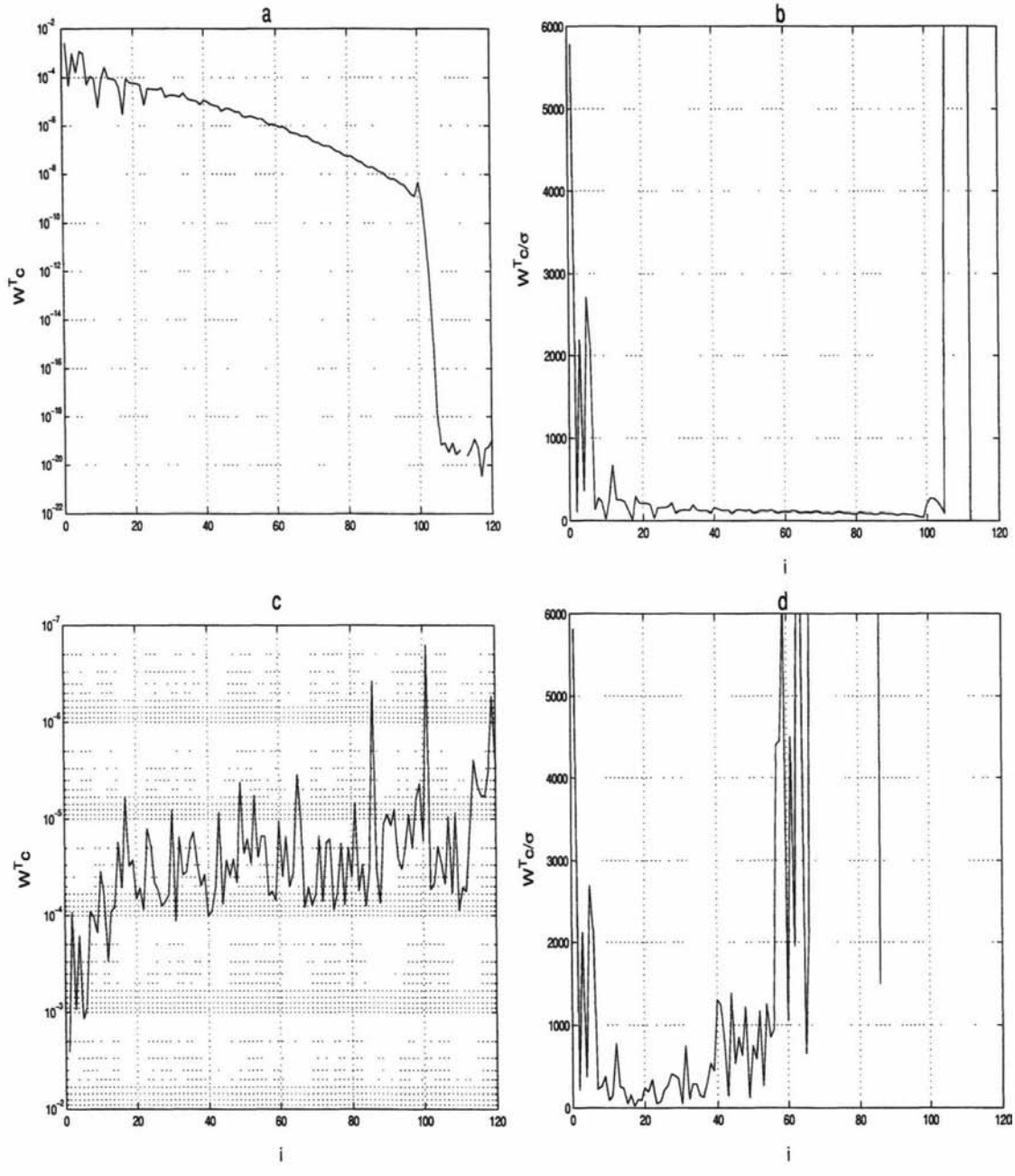

Figure 5.3: Examining the factors in the SVD-solution for data with and without noise (a)$\mathbf{w_i^T c}$ for perfect data, (b) $\frac{\mathbf{w_i^T c}}{\sigma_i}$ for perfect data, (c)$\mathbf{w_i^T c}$ for noisy data, (d) $\frac{\mathbf{w_i^T c}}{\sigma_i}$ for noisy data

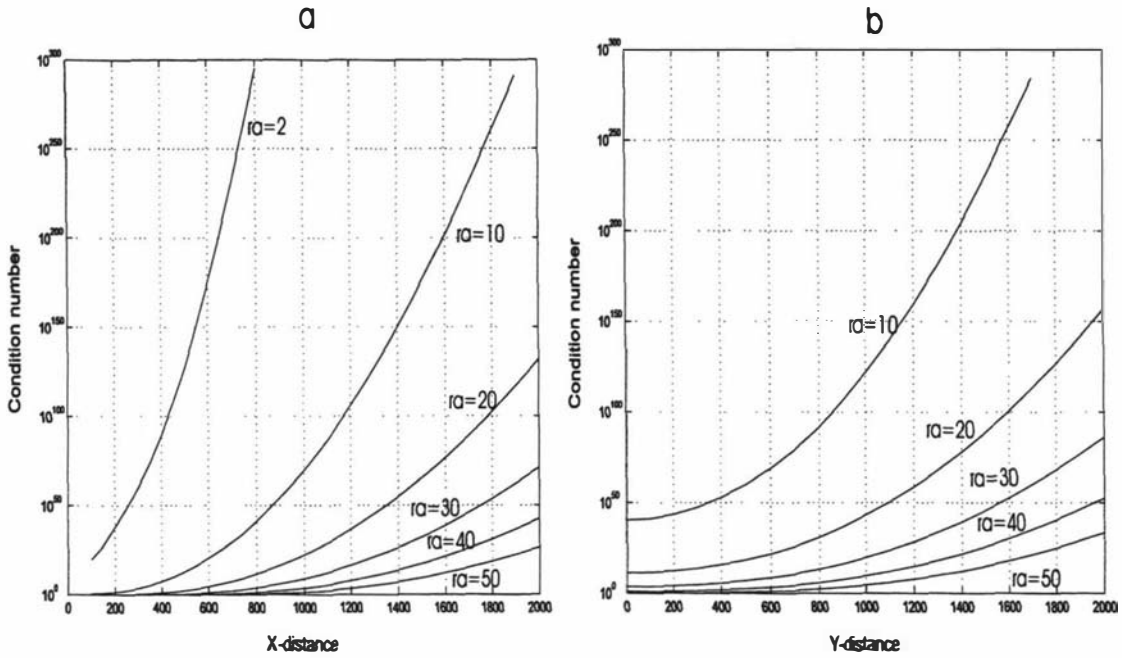

Figure 5.4: (a) X- distance *vs* condition number, (b) Y- distance *vs* condition number

since it can be applied to both rank deficient and discrete ill-posed problems. With Tikhonov's regularisation, we introduce the regularised objective function

$$\begin{aligned} Z(\mathbf{q}) &= \|A\mathbf{q} - \mathbf{c}\|_2^2 + \lambda^2 \|L\mathbf{q}\|_2^2, \\ &= \phi_d + \lambda^2 \phi_m, \end{aligned} \tag{5.5}$$

where ϕ_d is the residual norm (or data misfit function), and ϕ_m is the solution norm. We will be interested in the minimisation of the function $Z(\mathbf{q})$ for different values of λ. Note that the objective function Z is the 2-norm of the following system of equations

$$\begin{bmatrix} A \\ \lambda L \end{bmatrix} \mathbf{q} = \begin{bmatrix} \mathbf{c} \\ 0 \end{bmatrix},$$

where L is the regularisation operator and λ is the regularisation parameter.

5.5 Numerical Evaluations

5.5.1 Evaluation of parameter selection methods

The minimisation problem (5.5) depends on the optimal value of the regularisation parameter, λ. A number of techniques have been discussed in Chapter 2 for estimating optimal λ when the observations are contaminated with measurement error.

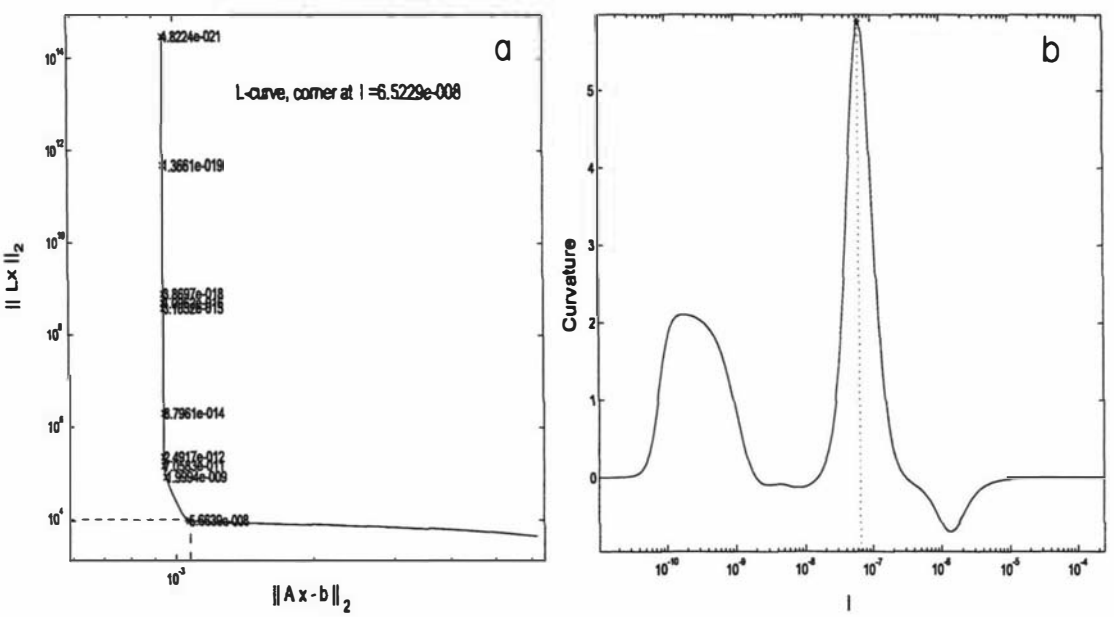

Figure 5.5: (a) L-curve: Estimation of optimal value of regularisation parameter λ, (b) Curvature estimation of L-curve

Two widely used methods for estimating regularisation parameters are the L-curve [16] and *GCV* [57] criteria. The L-curve method is assumed to have an L shape in a logarithmic axis scale, and an optimum value can be found to be the point where this curve has maximum curvature. The *GCV* method is based on a graph of a cross-validation function of λ, and on choosing the minimum GCV value. In general, each method will find different optimal values of λ. First, we illustrate with the aid of numerical examples how these two methods estimate the optimal value of the regularisation parameter, λ. We then compare the accuracy

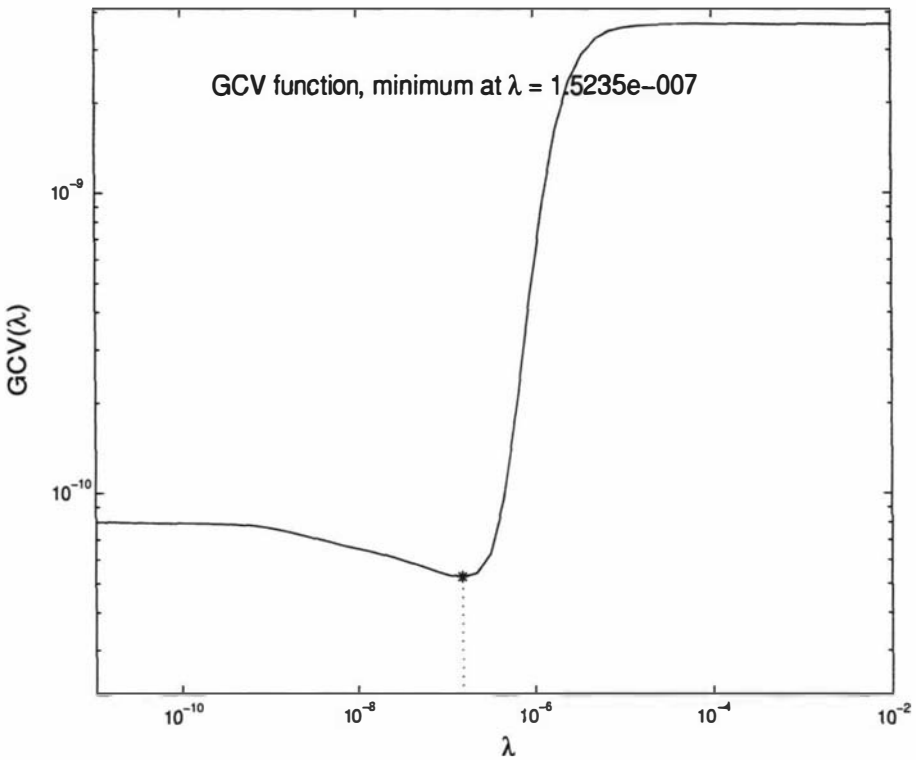

Figure 5.6: GCV: Estimation of optimal value of regularisation parameter (λ)

of the estimates obtained using GCV and L-curve methods.

According to [7], the method for optimally choosing the regularisation parameter is based on minimising the total error as shown in Figure 2.2. But the total error in that figure are not computable unless we know the exact solution. Since we use the simulated data and therefore it is possible to compute the figure and hence the parameter estimation methods can be comparable. Accuracy of the parameter choice methods such as discrepancy principle (DP), minimum bound (MB) and GCV have been compared in the literature by Lukas [20]. In this paper, the author compared the methods with respect to the function

$$E_1(\lambda) = \|\mathbf{q}_\lambda - \hat{\mathbf{q}}\|$$

and

$$E_2(\lambda) = \|A\mathbf{q}_\lambda - \hat{\mathbf{c}}\|,$$

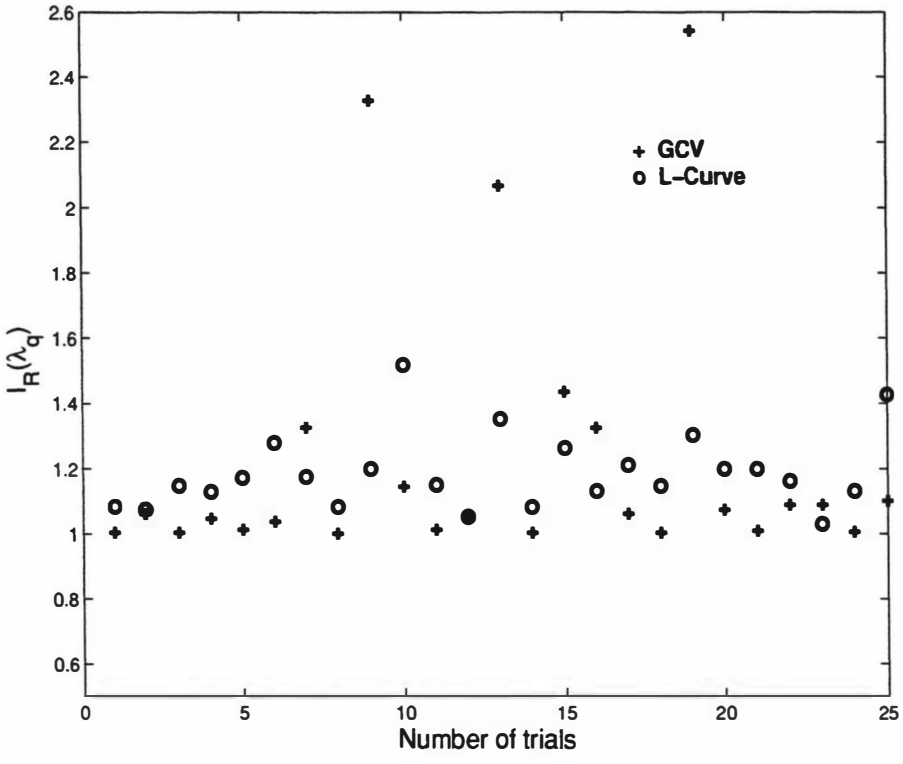

Figure 5.7: Comparison of inefficiencies for L-curve, GCV)

where $\hat{\mathbf{q}}$ and $\hat{\mathbf{c}}$ are the true release rate and the true concentration respectively. The author defined the inefficiency of the parameter value λ_q with respect to $E_1(\lambda)$ and $E_2(\lambda)$ to be

$$I_W(\lambda_q) = \frac{E_1(\lambda_q)}{\min_\lambda E_1(\lambda)},$$

$$I_R(\lambda_q) = \frac{E_2(\lambda_q)}{\min_\lambda E_2(\lambda)},$$

and stated that good parameter choice methods should consistently give a parameter estimate with inefficiency I_W or I_R close to 1. Based on this approach, we compare the methods of GCV and L-curve in this section.

Figures 5.5a and 5.5b illustrate the L-curve method to estimate the regularisation parameter. Here we used Example 1 of section 5.3.2 with a noisy *sine* function. Figure 5.5a is the L-curve, where the corner is clearly indicated, and Figure 5.5b is the curvature of the L-curve, as a function of λ. The largest peak in Figure 5.5b

Example	Source function	Noise in the data
1	Smooth function	0 %
2	Smooth function	20 %
3	Square function	0 %
4	Square function	20 %
5	Sharp function	0 %
6	Sharp function	20 %

Table 5.1: Six different examples

corresponds to the corner on the L-curve. Figure 5.6 illustrates the GCV method to estimate the regularisation parameter for the same problem, where the lowest value of the GCV function is clearly indicated. Since we know the true solution, it is possible to compare the accuracy of the values obtained using the L-curve and GCV. Figure 5.7 illustrates the comparison. We calculated λ using L-curve and GCV methods and for each estimate, inefficiencies $I_R(\lambda)$ are calculated. This is done 25 times, each time with different error vector, and plotted these values in Figure 5.7. It can be seen that the large proportion 21/25 of inefficiencies is between 1 and 1.4 for both methods. According to [20], this is an excellent result and both methods are good to estimate the regularisation parameter for the problem considered in this thesis with unknown noise component in the data. If we compare the inefficiencies of L-curve and GCV, the larger proportion 18/25 of inefficiencies of GCV is less than L-curve. Therefore we can conclude that, GCV gives a better approximation than L-curve in most of the times.

To analyse the sensitivity of the reconstructed solution with the parameter selection method, we consider six examples to compare the reconstructed solutions. These examples are given in Table 5.1. In all these cases, two hundred samplings are used at $X_0 = 3800$, $Y_0 = 100$ between the time interval $t = 1600, 7600$. Transport parameters K_x, K_y, K_z and U are taken as 12, 12, 0.2 and 1.8 respectively. The considered three input source functions are the

(i) smooth function

$$q(t) = \begin{cases} 1000 \sin\left(\frac{\pi(t-1000)}{4000}\right), & 1000 \leq t \leq 5000, \\ 0 & \text{otherwise} \end{cases}$$

(ii) square function

$$q(t) = \begin{cases} 1000, & 1000 \leq t \leq 4500, \\ 0 & \text{otherwise} \end{cases}$$

(iii) sharp function

$$q(t) = 1000 \exp\left[-\frac{(t-200)^2}{5t}\right] + 1000 \exp\left[-\frac{(t-600)^2}{10t}\right] + 1000 \exp\left[-\frac{(t-1000)^2}{20t}\right] + 1000 \exp\left[-\frac{(t-2400)^2}{100t}\right], \quad t \geq 0$$

The source function is discretised into 100 uniformly spaced interval between $t = 0$ and $t = 7600$. The concentration signal at $P_1 = (3800, 100, 0)$ is simulated using the forward problem (5.1) with actual parameter values. In order to simulate errors, data are corrupted by adding normally distributed random noise of 20%. The perfect and corrupted sampled data for the smooth, square and peak function sources are shown in Figure 5.8. The percentage of error in the reconstructed source, along with λ values, are tabulated in Tables 5.2, 5.3 for case (i), Tables 5.4 5.5 for case (ii), and Tables 5.6, 5.7 for cases (iii). The estimates of the release rates are shown in Figures 5.9, 5.10 for case (i), in Figures 5.11, 5.12 for case (ii), and in Figures 5.13 and 5.14 for case (iii).

5.5.2 The Evaluation of the regularisation order

The regularised solution depends mostly on the value of the regularisation parameter and the order of regularisation. In this section we analyse the effects of regularisation order on the reconstructed solution. We use the same test cases discussed in the previous section to analyse the order of regularisation, and its results

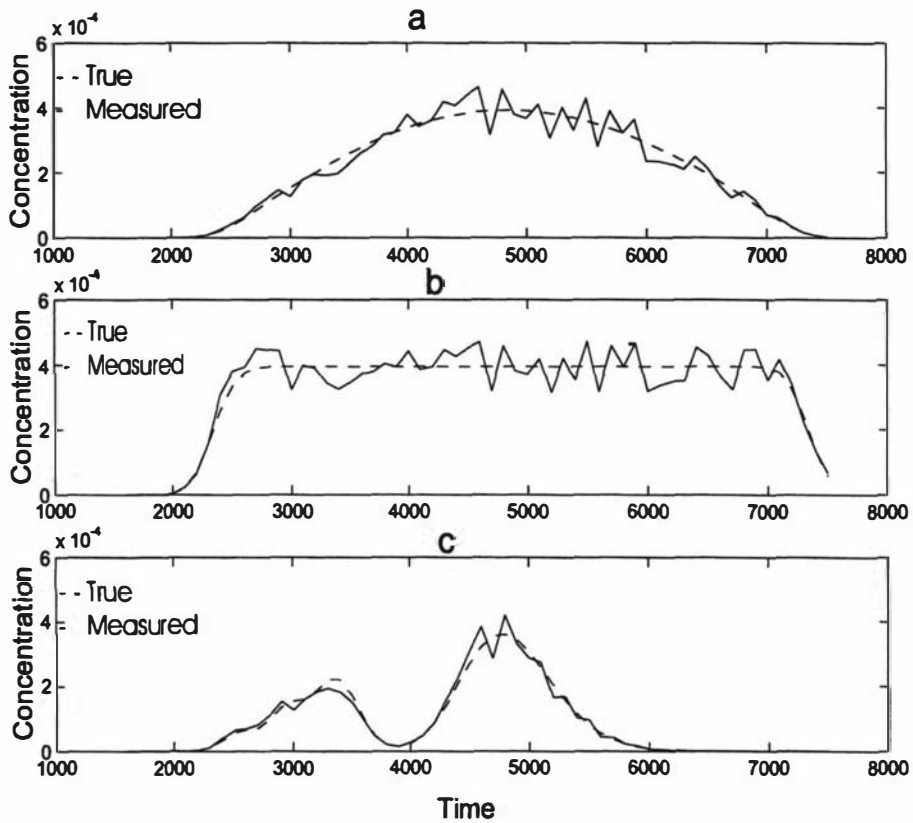

Figure 5.8: Measured concentration: (a) concentration at (3,800, 100, 0) from smooth input source, (b) concentration at (3,800, 100, 0) from square input source, (c) concentration at (3,800, 100, 0) from peak function input source

Order (N)	Regularisation parameter λ		Relative error in %	
	L-curve	GCV	L-curve	GCV
0	7.9×10^{-7}	1.0×10^{-14}	45	0.46
1	7.9×10^{-7}	3.9×10^{-12}	0.7	0.12
2	7.9×10^{-7}	1.7×10^{-14}	0.5	0.31
3	7.9×10^{-7}	1.5×10^{-14}	0.4	0.35
4	7.9×10^{-7}	1.5×10^{-14}	0.4	0.11
5	7.9×10^{-7}	5.3×10^{-14}	0.4	0.12

Table 5.2: Smooth source- perfect data: relative error in the solution

are also given in Tables 5.2–5.7 and Figures 5.9–5.14. These results show that in general the error in the reconstructed solution decreases with the increasing order of regularisation for L-curve method. From further investigation, we found that the error in the reconstructed solution decreases up to the order 9 and then increases

Order (N)	Regularisation parameter λ		Relative error in %	
	L-curve	GCV	L-curve	GCV
0	7.9×10^{-7}	6.9×10^{-8}	44.0	6.6
1	7.9×10^{-7}	8.0×10^{-7}	3.0	3.0
2	7.9×10^{-7}	2.0×10^{-6}	4.2	2.9
3	7.9×10^{-7}	4.6×10^{-6}	4.8	2.9
4	7.9×10^{-7}	1.9×10^{-5}	5.3	2.9
5	7.9×10^{-7}	2.0×10^{-5}	5.6	3.2

Table 5.3: Smooth source- measurement error 20 %: relative error in the solution

Order (N)	Regularisation parameter λ		Relative error in %	
	L-curve	GCV	L-curve	GCV
0	7.9×10^{-7}	1.0×10^{-14}	46	0.79
1	7.9×10^{-7}	1.2×10^{-12}	10	7.5
2	7.9×10^{-7}	2.5×10^{-14}	8.6	3.9
3	7.9×10^{-7}	9.5×10^{-15}	8.3	4.3
4	7.9×10^{-7}	3.0×10^{-14}	8.2	4.5
5	7.9×10^{-7}	8.8×10^{-14}	8.1	4.7

Table 5.4: Square source- perfect data: relative error in the solution

Order (N)	Regularisation parameter λ		Relative error in %	
	L-curve	GCV	L-curve	GCV
0	7.9×10^{-7}	6.9×10^{-8}	45	11
1	7.9×10^{-7}	3.9×10^{-7}	11.0	10.4
2	7.9×10^{-7}	7.0×10^{-7}	10.3	10.4
3	7.9×10^{-7}	1.1×10^{-6}	10.4	10.4
4	7.9×10^{-7}	1.9×10^{-6}	10.6	10.5
5	7.9×10^{-7}	3.5×10^{-6}	10.8	10.5

Table 5.5: Square source- measurement error 20 %: relative error in the solution

with increasing order. It can be noted from these Tables and Figures that the error in the reconstructed solution is very large when the order of regularisation is zero for the L-curve method. But for the GCV-based solution we could not identify any correspondence between the order of regularisation and the error in the solution. In general, the error is minimum when the order of regularisation is two and the GCV-based method performs little better in most of the situations than L-curve method.

Order (N)	Regularisation parameter λ		Relative error in %	
	L-curve	GCV	L-curve	GCV
0	7.9×10^{-7}	2.9×10^{-15}	57.5	2.9
1	7.9×10^{-7}	1.0×10^{-15}	36.0	3.3
2	7.9×10^{-7}	2.5×10^{-15}	34.6	8.5
3	7.9×10^{-7}	9.5×10^{-15}	29.2	10.5
4	7.9×10^{-7}	3.0×10^{-14}	24.6	11.4
5	7.9×10^{-7}	8.8×10^{-14}	24.6	11.6

Table 5.6: Non-smooth source- perfect data: relative error in the solution

Order (N)	Regularisation parameter λ		Relative error in %	
	L-curve	GCV	L-curve	GCV
0	7.9×10^{-7}	4.9×10^{-8}	57.0	24.3
1	7.9×10^{-7}	9.8×10^{-8}	36.1	26.4
2	7.9×10^{-7}	1.2×10^{-7}	34.9	26.3
3	7.9×10^{-7}	1.9×10^{-7}	33.7	27.8
4	7.9×10^{-7}	3.3×10^{-7}	32.1	29.8
5	7.9×10^{-7}	6.2×10^{-7}	32.8	32.4

Table 5.7: Non-smooth source- measurement error 20 %: relative error in the solution

5.5.3 Effects due to the inaccuracy of the parameters

We have only considered a problem in which the transport properties of the medium and source location are known. Sometimes in an actual situation, only the approximate values of the parameters are known. Therefore, it is appropriate to analyse the effects of the inaccuracy of the transport parameters and location. We considered six different test cases for X_0, Y_0, H, U, K_x and K_z with incorrect values. In each case, one of these values is changed by $\pm 20\%$, while all other parameters are unchanged. Perfect and complete concentration measurements are used for this purpose.

Figure 5.15 shows the reconstructed release values along with the true values. Figure 5.15a is the simulated release rate when X is overestimated or underestimated by 20%. When X is underestimated, the reconstructed release rate is shifted to a later time and overestimated. This is because the concentration signal has as-

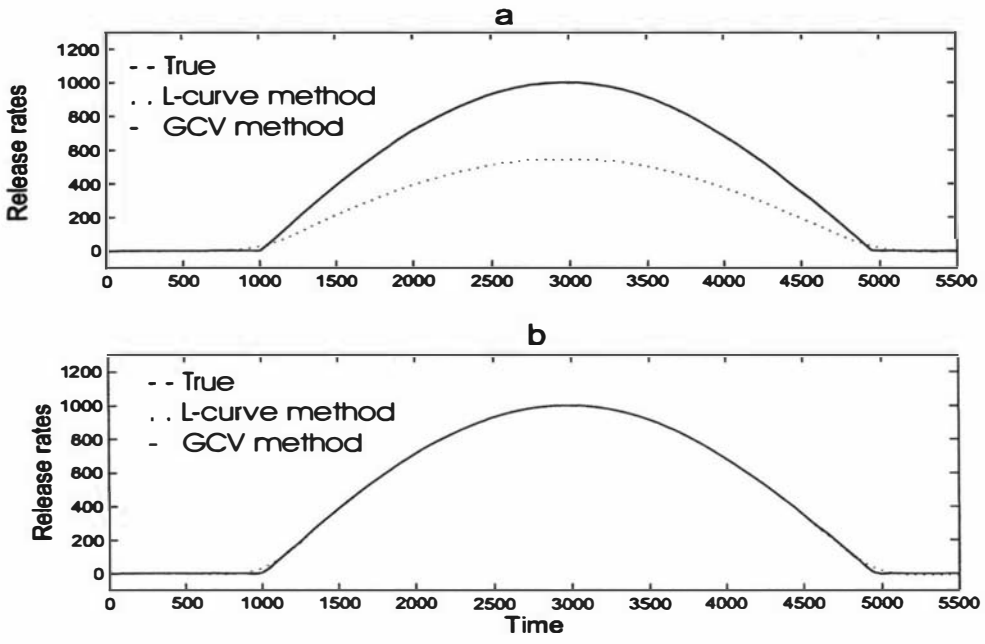

Figure 5.9: Smooth source- Release rate estimation using perfect data: (a) Regularised solution, order $N = 0$, (b) Regularised solution, order $N = 2$

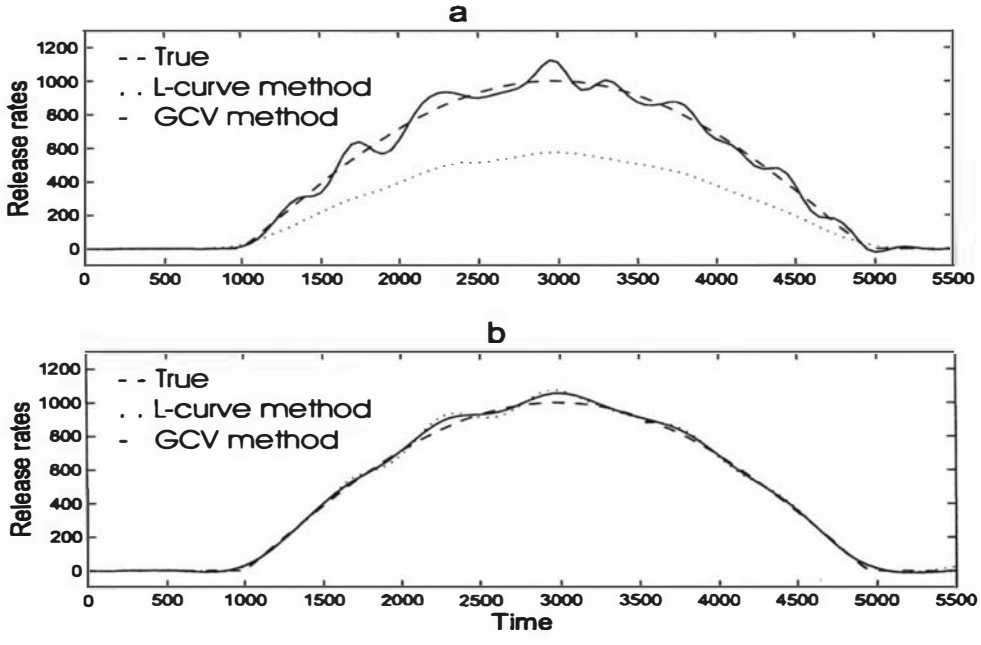

Figure 5.10: Smooth source- release rate estimation using data which are corrupted by 20 % of measurement noise: (a) regularised solution, order $N = 0$, (b) regularised solution, order $N = 2$

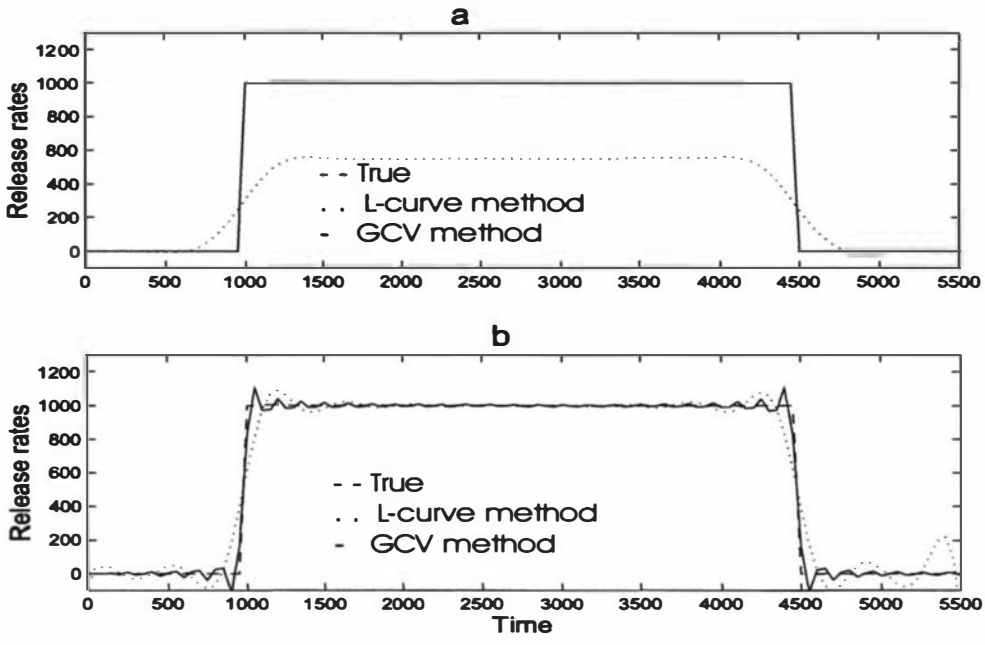

Figure 5.11: Square source- Release rate estimation using perfect data: (a) Regularised solution, order $N = 0$, (b) Regularised solution, order $N = 2$

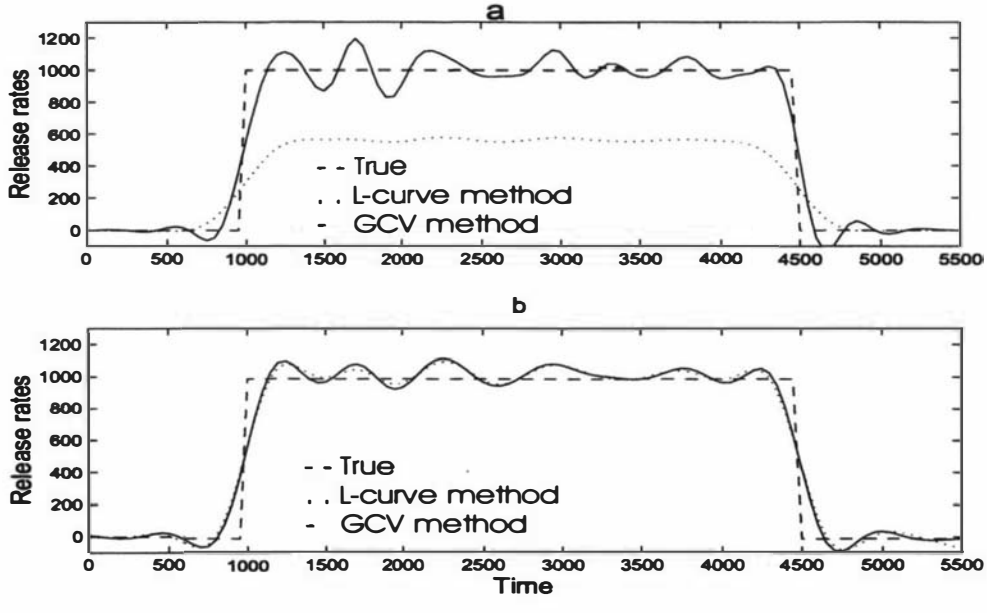

Figure 5.12: Square source- release rate estimation using data which are corrupted by 20 % of measurement noise:(a) regularised solution, order $N = 0$, (b) regularised solution, order $N = 2$

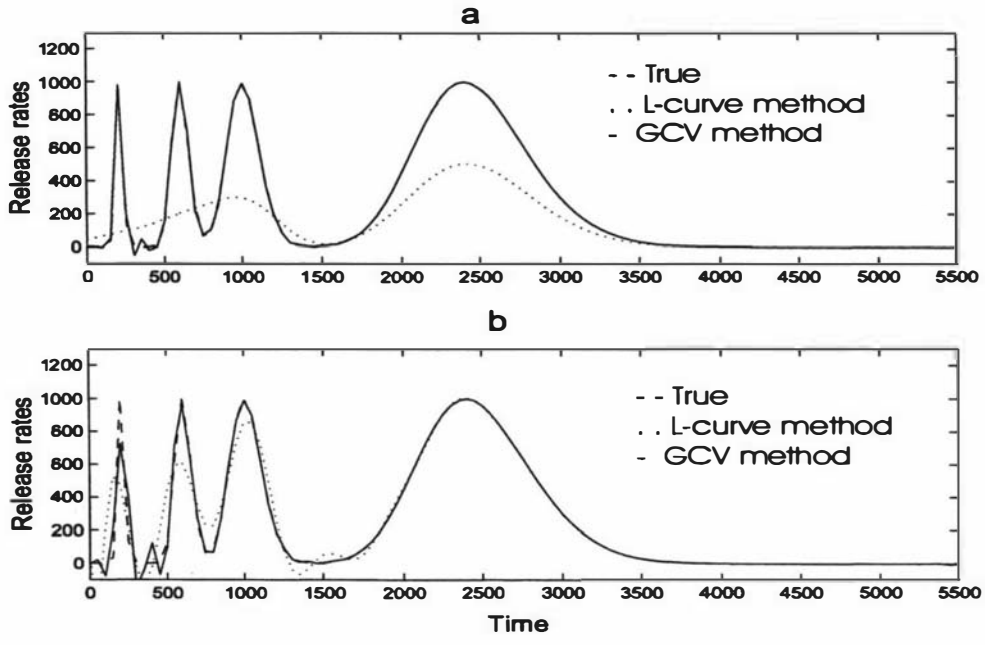

Figure 5.13: Non-smooth source- release rate estimation using perfect data: (a) regularised solution, order $N = 0$, (b) regularised solution, order $N = 2$

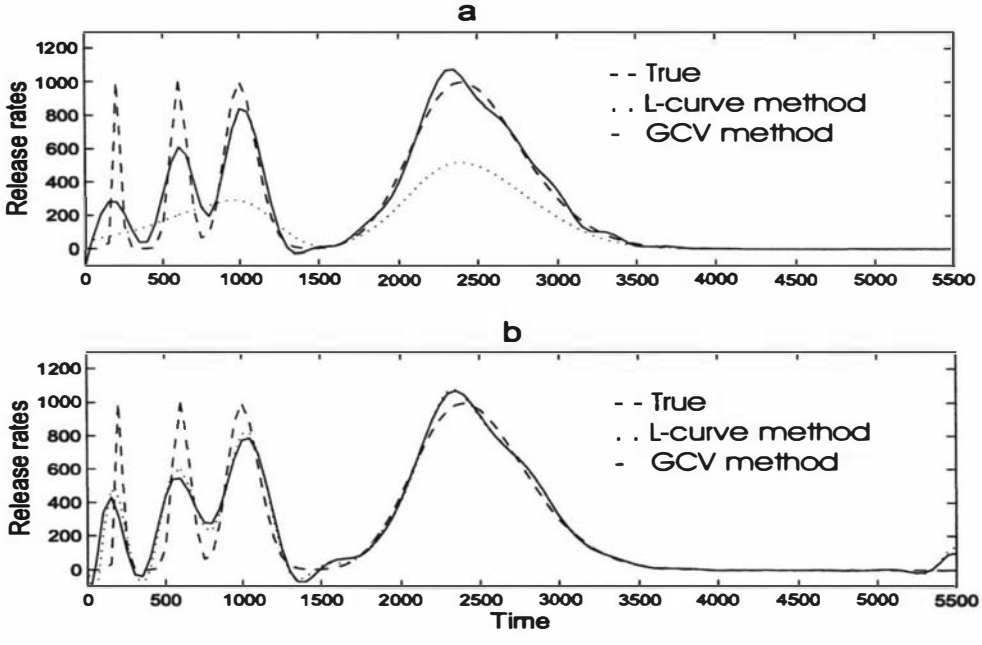

Figure 5.14: Non-smooth source- release rate estimation using data which are corrupted by 20 % of measurement noise: (a) regularised solution, order $N = 0$, (b) regularised solution, order $N = 2$

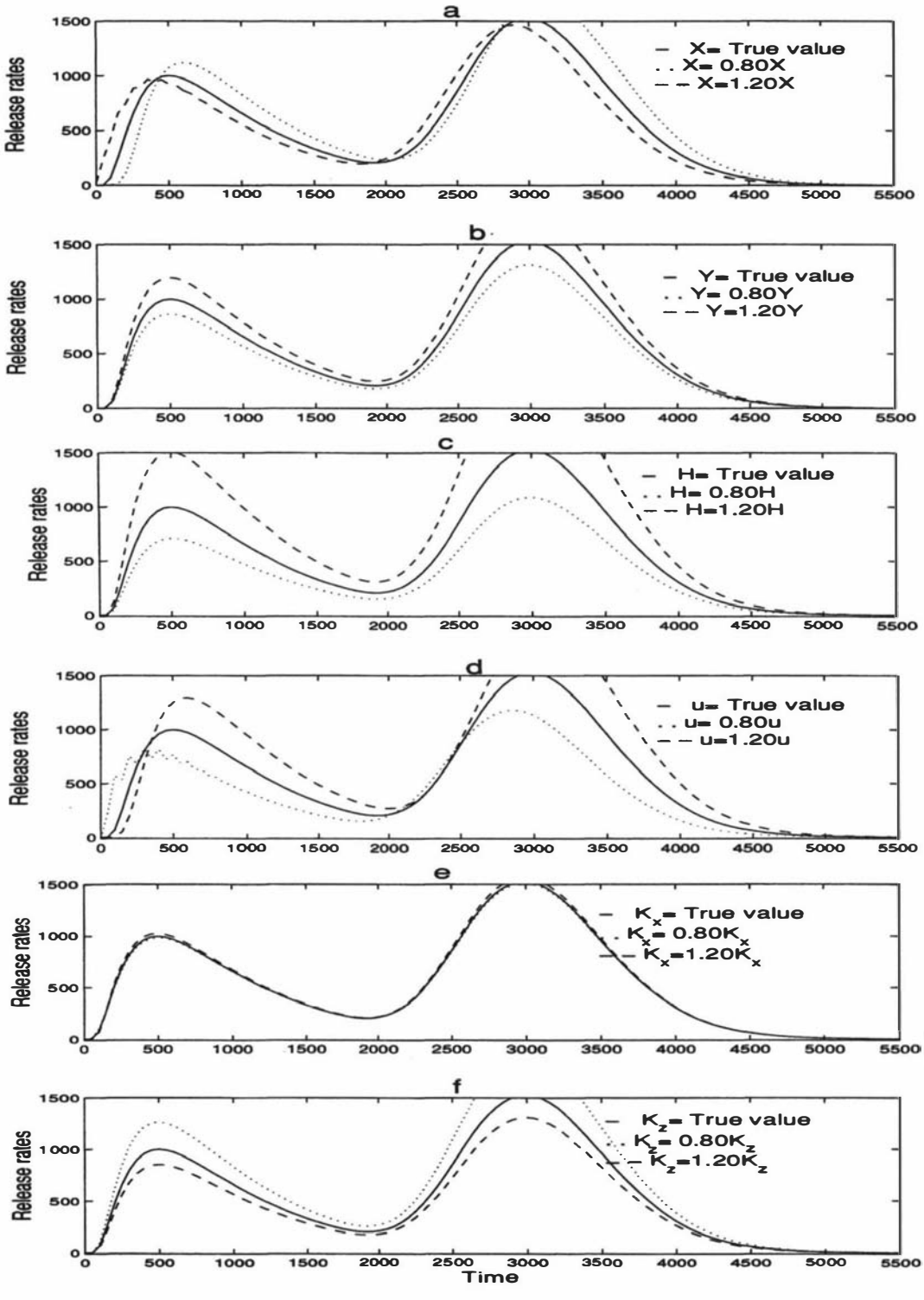

Figure 5.15: Effects due to individual factors: (a) error in X, (b) error in Y, (c) error in H, (d) error in U, (e) error in K_x, (f) error in K_z

sumed that the travel time is lower than it actually is. When X is overestimated, the reconstructed release rate is shifted to an earlier time and underestimated. This is because the concentration signal is assumed, the travel time is longer than it actually is.

Figures 5.15b and c respectively, are the simulated release rates when Y and H are inaccurate by $\pm 20\%$. When Y and H are underestimated, the reconstructed release rate is more dispersed, and when Y and H are overestimated the reconstructed release rate is less dispersed. This is because, in the Y and Z directions there is no advection by the wind and only dispersion is taking place.

Figure 5.15d is the result of the simulation where U is either underestimated or overestimated by 20%. When U is overestimated, the release rate is shifted to a later time. This is because the signal is assumed to travel faster than it actually does. The reconstructed release rate is shifted to an earlier time when U is underestimated. This is because the signal is assumed to travel slower than it actually does.

Figures 5.15e and 5.15f illustrate the simulated release rates when K_x and K_z are inaccurate by $\pm 20\%$. When K_x and K_z are underestimated, the reconstructed release rate is less dispersed; when K_x and K_z are overestimated, the reconstructed release rate is more dispersed. But it can be noted from Figure 5.15f that the effect due to the inaccuracy in K_z is much higher than the error due to K_x.

5.5.4 Evaluating the validity of the model

In this section, numerical calculations are presented to evaluate the accuracy of the developed model. The following five different situations are considered: (i) complete sampling (i.e. the sampling would capture both the trailing and leading edges of the signal) of signal with no measurement error (Figure 5.16a); (ii) complete sampling with measurement error (Figure 5.17a); (iii) incomplete sampling with measurement error (Figure 5.18a); (iv) complete sampling with measurement error for a source

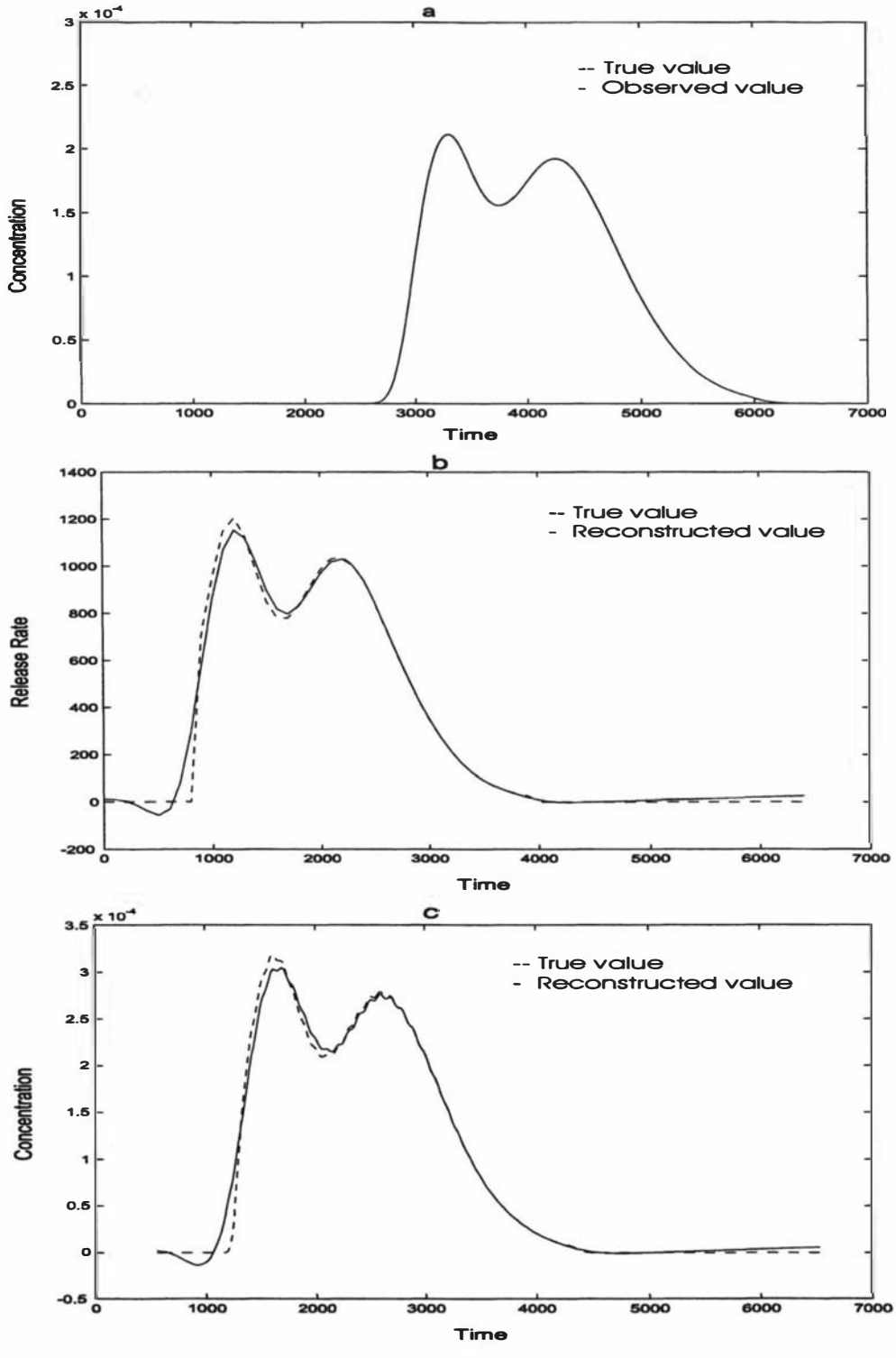

Figure 5.16: Recovering pollution history with perfect complete data: (a) measured pollution concentration at X=3800 m, Y=100 m, (b) recovered source release rate and (c) reconstructed pollution concentration history at X=800 m, Y=100 m

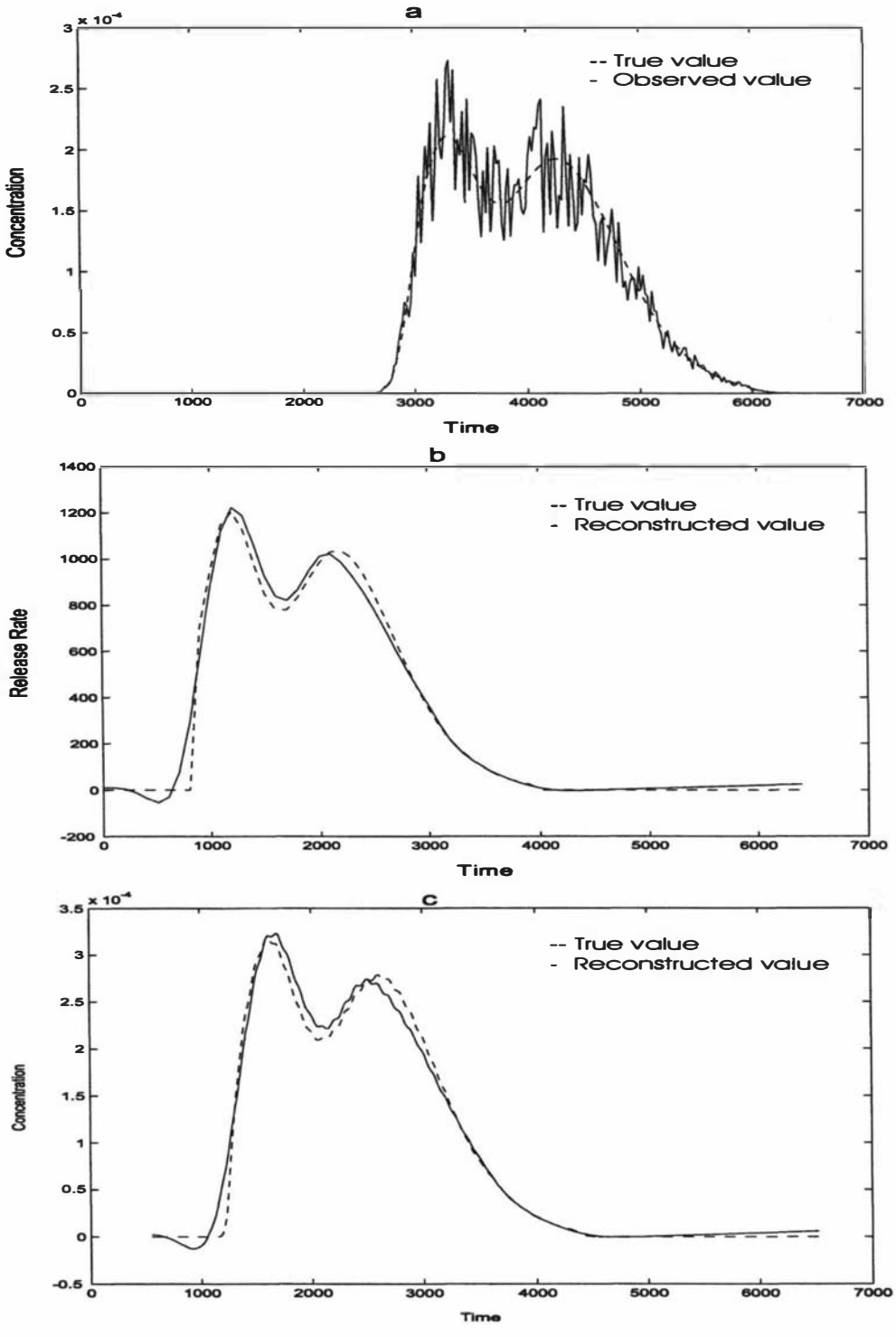

Figure 5.17: Recovering pollution history with complete noisy data of 30 % relative noise: (a) measured pollution concentration at X=3800 m, Y=100 m, (b) recovered source release rate and (c) reconstructed pollution concentration history at X=800 m, Y=100 m

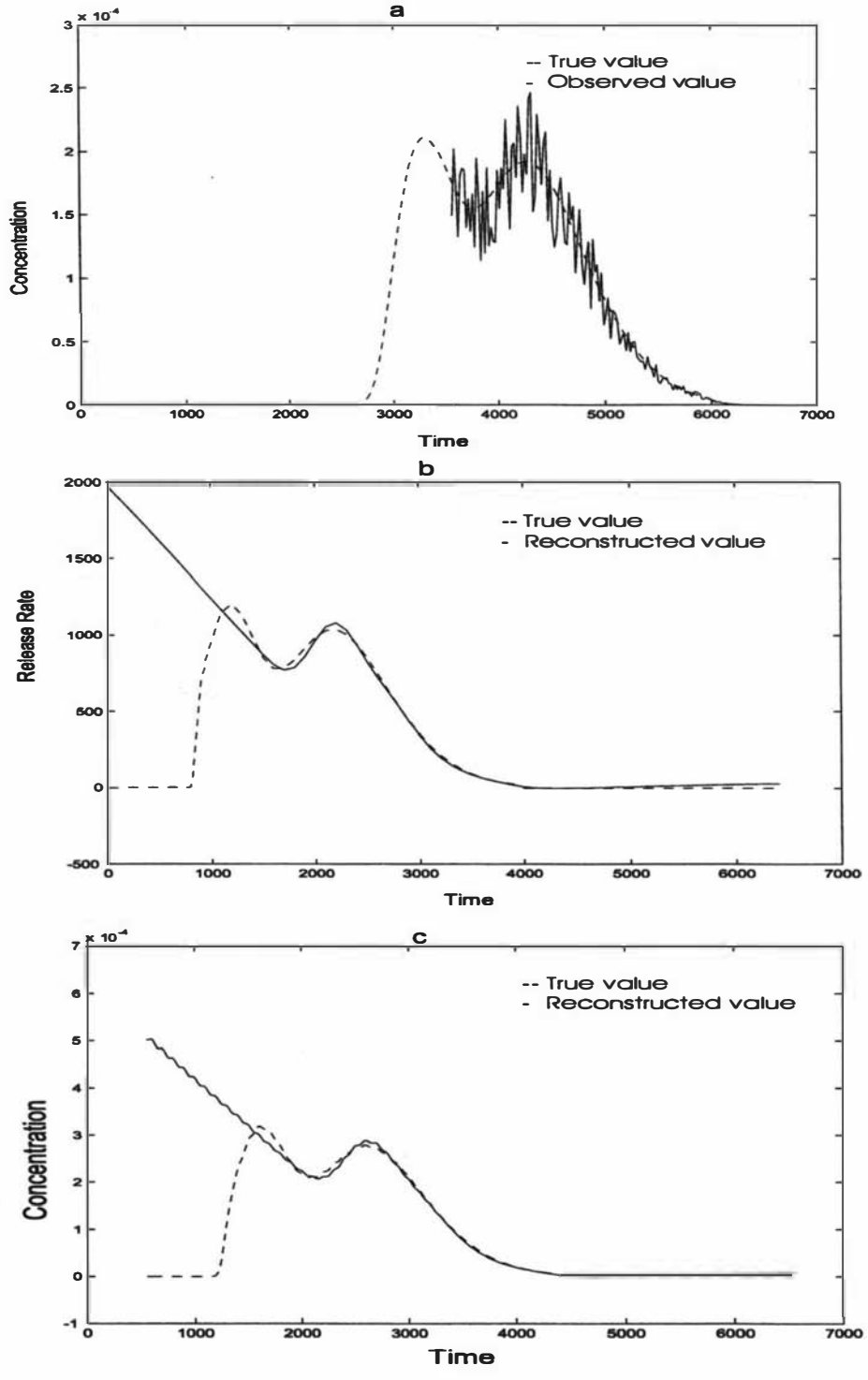

Figure 5.18: Recovering pollution history with incomplete (at the beginning) noisy data of 30 % relative noise: (a) measured pollution concentration at X=3800 m, Y=100 m, (b) recovered source release rate and (c) reconstructed pollution concentration history at X=800 m, Y=100 m

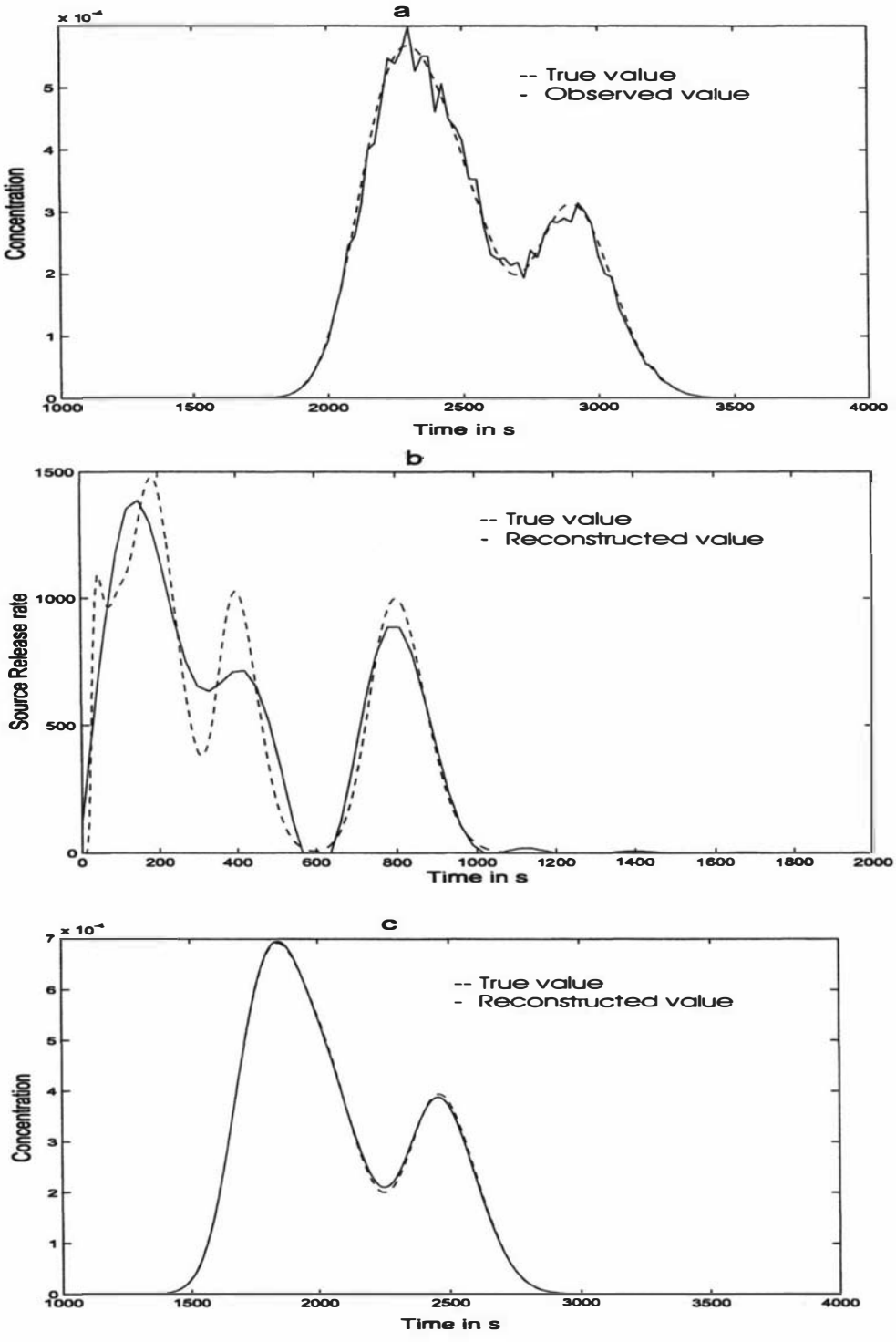

Figure 5.19: Recovering pollution history with complete noisy data of 15 % relative noise: (a) measured pollution concentration at X=3800 m, Y=100 m; (b) recovered source release rate and (c) reconstructed pollution concentration history at X=3800 m, Y=100 m

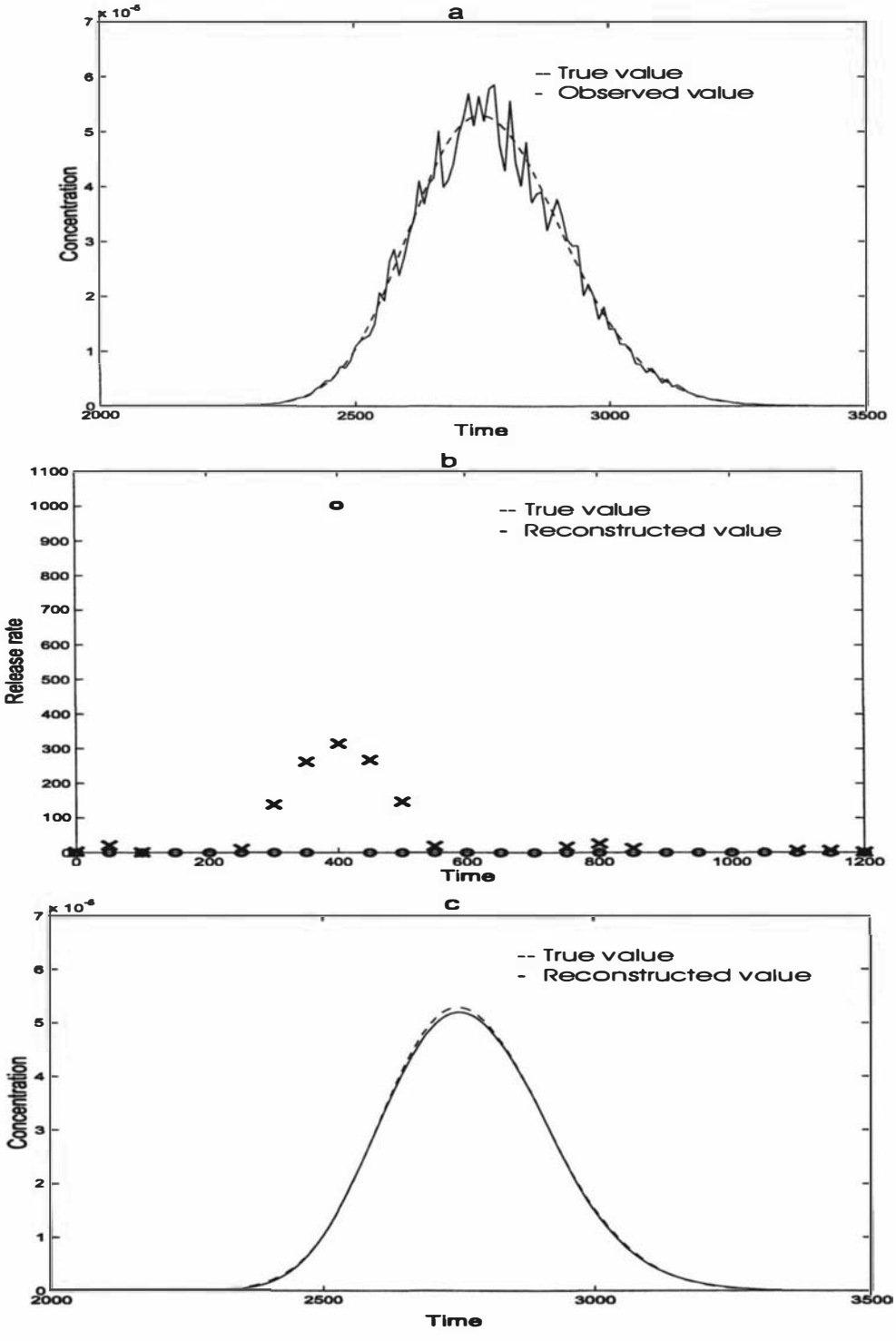

Figure 5.20: Recovering pollution (δ function release) history with complete noisy data of 15 % relative noise: (a) measured pollution concentration at X=3800 m, Y=100 m, (b) recovered source release rate and (c) reconstructed pollution concentration history at X=3800 m, Y=100 m

function with sharp peaks (Figure 5.19a); (v) complete sampling with measurement error for a δ source function (Figure 5.20a).

For Tikhonov's regularised solution, the second-order regularisation method is used and the parameter is selected using the GCV method. The release rate estimation results are shown in Figures 5.16b–5.19b, and the concentration history at the measurement location is shown in Figures 5.16c–5.19c, for each situation.

For situations (i) and (ii), the true solution and the reconstructed solution are very close. The timing and magnitudes of the peaks are well reproduced. For situation (iii), the results are well produced only just before the first peak. The reconstructed solution at early times indicates the data are unable to provide correct information about the source at earlier times. This is because the plume is more dispersed, and information about the plume released at the earlier time is lost. For situation (iv), the magnitude of the peaks of release rate are not well produced but the concentration at $(3800, 100, 0)$ are well matched with the true values. For situation (v), the timing of the δ function is well produced but the magnitude is not. Concentration history at a point (3800, 100, 0) is well matched with the true value, even though the reconstructed release rates vales are not well matched with the true values.

Noise in	Order of regularisation		
the data	0	1	2
0 %	0.465	0.111	0.090
5 %	0.467	0.112	0.092
10 %	0.469	0.115	0.094
20 %	0.473	0.120	0.100
30 %	0.477	0.125	0.107
50 %	0.485	0.138	0.123

Table 5.8: Relative residual norm in %

The relative residual norms of the solutions for situation (ii) with different noise levels in the data and the order of the regularisation parameter are given in Table 5.8. This shows the residual is minimum when $N = 2$ and the residual increases

with increasing noise in the data. However, compared with the increase of noise in the data, the increase of the residual is negligible.

5.6 Summary and Discussion

The goal of the work presented in this chapter is to develop an inverse model capable of estimating the release rate and release history of air pollution from a non-steady point source. The approach taken is to develop an inverse-model as a linear least-square problem in which the release rate and history are estimated by using the pollutant concentration measurement from one location on the ground.

First, we demonstrate that the problem of finding the release rate is a rank-deficient or a discrete ill-posed inverse problem that depends on the properties of the coefficient matrix. Second, the well-posed inverse model for estimating the release rate is formulated using Tikhonov's regularisation method. The developed model is tested using simulated data from the forward problem.

The method described in this chapter is able to recover the release history and release rate accurately for complete sampling data, but is less accurate if the sampling set is incomplete (see Figure 5.18). Furthermore, the accuracy of the solution does not depend much on the amount of noise in the measured data. The method described in this chapter is an effective approach to recover the source history of an atmospheric pollutant from complete noisy concentration data.

In real life situations, the method developed in this chapter is useful to assess the environmental consequences of accidental releases of pollutant gas. By estimating the release start-time, release stop-time, release pattern and the total amount released into the atmosphere, we may be able to:

(i) predict the current and future concentration of pollutants patterns in other areas to take necessary actions,

(ii) estimate the amount of gas left in the storage tank and the cost resulting

from an accident if the pollutant release is from a storage tank,

(iii) identify appropriate locations to build monitoring stations to observe the factories which handle toxic gases. For example the death and injury toll in the Bhopal tragedy could have been minimised if there had been such monitoring stations. In a situation like the Bhopal accident, an observer in the monitoring stations can model the concentration of gas distribution in the surrounding atmosphere using the measured data and then pass that information to the post-accident management staff to undertake a necessary action plan.

The next chapter describes the release rate estimation of an atmospheric pollutant originating from a non-steady point source of an unknown location.

Chapter 6

Release Rate Estimation of Pollution from a Non-steady Point Source at an Unknown Location

6.1 Introduction

This chapter presents an inverse modelling procedure to estimate the location and release rate of an atmospheric pollutant. The input to this model requires measured pollutant concentration at a minimum of three observation sites on the ground together with meteorological conditions such as wind speed and percentage cloud cover. The methodology for estimating the location and release rates of pollution sources uses 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 pollutant C and the discretised components is linear, and the relationship between C and the location parameters is nonlinear. Therefore the

problem involves estimating both nonlinear and linear parameters. In Chapter 5 it was shown that estimating linear parameters (release rates) for given nonlinear parameters (location) is a linear ill-posed problem. Therefore estimating release rates from a source with an unknown location is a nonlinear ill-posed problem. This problem is further complicated by inexact information. In reality, the data contain measurement errors, so the true solution will not fit the data. The meteorological parameters are also not known exactly. In this model, these values are assumed to be constant; however, in reality, they vary in space and time.

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 [16] for linear ill-posed inverse problems to our nonlinear problem. We use this criterion as well as Wahhba's [8] 'leaving-out-one' lemma to estimate the optimal value of the regularisation parameter for this problem.

6.2 The Inverse Model

Here the inverse problem is concerned with the estimation of release rate and its location using the measured concentration data at some down stream locations such as $P = (X_0, Y_0, 0)$, $Q = (X_0 + x_1, Y_0 + y_1, 0)$, and $R = (X_0 + x_2, Y_0 + y_2, 0)$ as shown in Figure 6.1, 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 (2.50) to estimate the source release rate and its location.

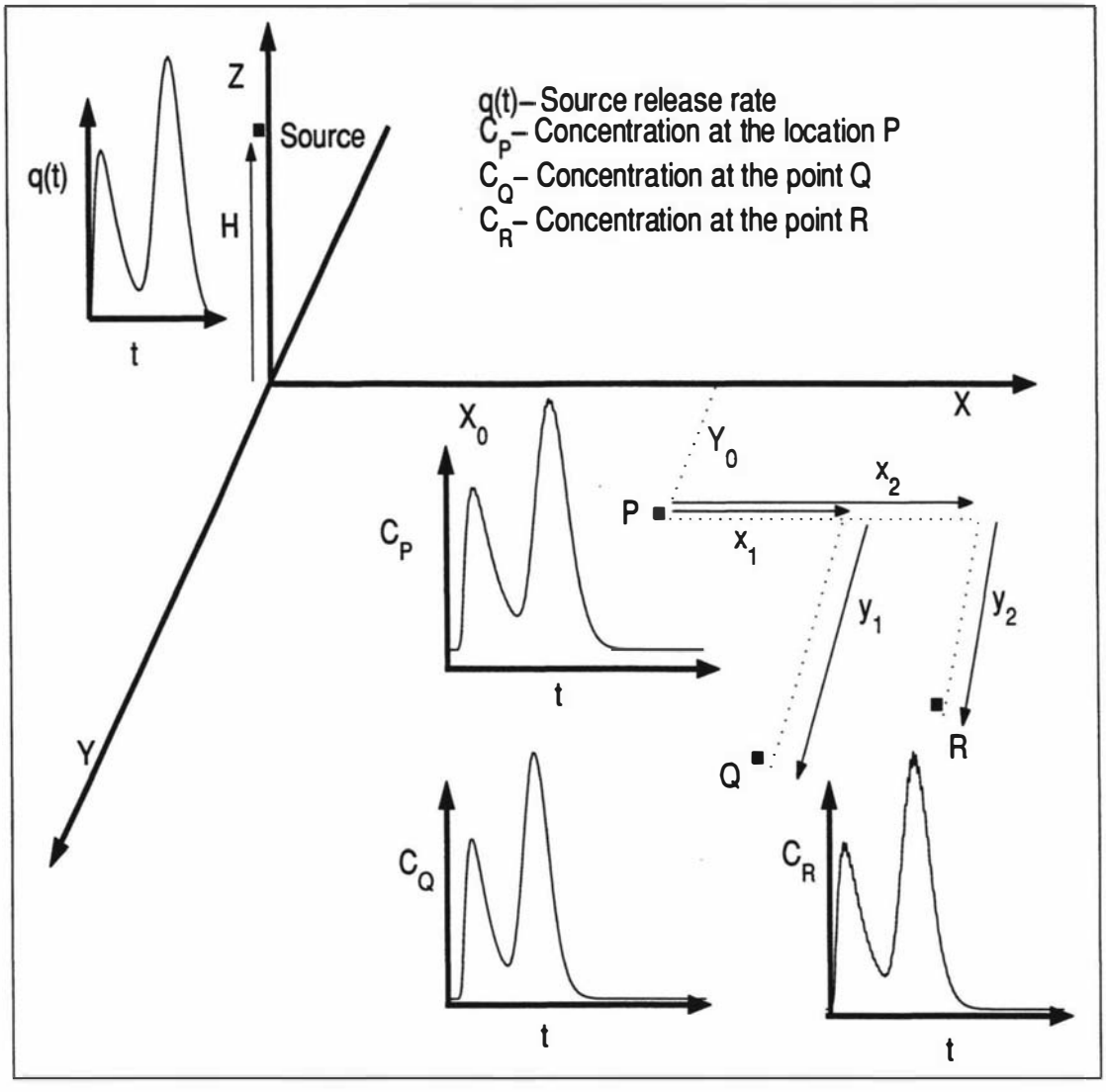

Figure 6.1: Illustration of the inverse problem

This is an inverse problem. Source release rate estimation is an ill-posed problem. Even if the measured data are error free, the solution to the release rate estimation problem is not stable. In order to stabilise the solution, regularisation methods must be used. Measured concentration data usually contain measurement errors. As a result, $q(t)$, X_0, Y_0 and H recovered by this inverse analysis are not exact and are only estimates within the measurement and regularisation errors.

The concentration of pollutant at a point $(X_0, Y_0, 0)$ can be expressed as

$$C(X_0, Y_0, 0, t) = \int_0^t K(t, \tau) q(\tau) d\tau, \tag{6.1}$$

where the kernel $K(t,\tau)$ is

$$K(t,\tau) = \frac{\exp\left[-\frac{(X_0-U(t-\tau))^2}{4K_x(t-\tau)} - \frac{Y_0^2}{4K_y(t-\tau)} - \frac{H^2}{4K_z(t-\tau)}\right]}{4\pi^{\frac{3}{2}}(K_xK_yK_z)^{\frac{1}{2}}(t-\tau)^{\frac{3}{2}}}. \tag{6.2}$$

6.2.1 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 the point $(X_0, Y_0, 0)$ at equal time intervals between $t_0 = \tau_0$ and $t_n = \tau_m$ (see Figure 5.1). The simplest way to proceed is to solve (6.1) on a mesh with uniform spacing. We suppose that we wish to determine the source release at times $\tau_0 = 0$, ..., $\tau_m = t_n$, where $m < n$ since the number of parameters to be estimated should be no greater than the number of data points. Discretising (2.50) by the trapezoidal rule gives the system of equations

$$\mathbf{c} = \mathrm{A}(\mathbf{p})\mathbf{q} \tag{6.3}$$

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 β_{ij} is a quadrature weight. Generally, minimising an objective function solves inverse problems. Now the problem for estimating the release rate $\mathbf{q}$ and the location $\mathbf{p}$ is

$$\min_{\mathbf{p},\mathbf{q}} \left[Z(\mathbf{q},\ \mathbf{p}) = \|A(\mathbf{p})\mathbf{q} - \mathbf{c}\|_2^2\right], \tag{6.4}$$

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 (6.4) 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 $\mathbf{p}$. In Chapter 5 we solved the same problem given in (6.4) for a known value of $\mathbf{p}$. 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.

6.2.2 Regularised least squares

Tikhonov's regularisation 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

$$\begin{aligned} Z(\mathbf{q},\ \mathbf{p}) &= \|A(\mathbf{p})\mathbf{q} - \mathbf{c}\|_2^2 + \lambda^2\|L\mathbf{q}\|_2^2, \\ &= \phi_d + \lambda^2\phi_m, \end{aligned} \tag{6.5}$$

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 λ. Note that the objective function Z is the 2-norm of the following system of equations

$$\begin{bmatrix} A(\mathbf{p}) \\ \lambda L \end{bmatrix} \mathbf{q} = \begin{bmatrix} \mathbf{c} \\ 0 \end{bmatrix}, \tag{6.6}$$

where L is the regularisation operator and λ 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

$$\|L\mathbf{q}\|^2 \approx \int_0^{t_n} \left(\frac{d^N q}{d\tau^N}\right)^2 d\tau. \tag{6.7}$$

The most popular choice for obtaining a smooth solution is $N = 2$ [52]. We also found in previous chapters that $N = 2$ gives the best combination of good fit to the data and the smooth solution. When $N = 2$, the second derivative of the solution is minimised and therefore L is given by

$$L = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \dots \\ 0 & 1 & -2 & 1 & 0 & \dots \\ 0 & 0 & 1 & -2 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \in \mathbf{R}^{(m-1)\times(m+1)}.$$

The total true error of this method has the form shown in Figure 2.2. The error due to regularisation approaches zero as $\lambda \to 0$, while the error due to noise in the data increases as $\lambda \to 0$. Selection of the optimal regularisation parameter is based on minimising the total error. In a real situation, we do not know the true error and therefore the curves in Figure 2.2 cannot be computed unless we know the exact solution. Therefore we have to find a strategy for the choice of the regularisation parameter. In the following section we explore the use of different techniques to find the regularisation parameter.

6.3 Selection of the regularisation parameter

The non-linear problem (6.5) is different from the linear problem (5.5) in several ways. First, we can't use only linear algebra to determine the minima of Z. Second, the non-linear objective Z may have more than one minimum for each value of λ. In the course of my research I 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 λ. In all my algorithms this is done by exploiting the fact that some of the variables, namely $\mathbf{q}$, appear in (6.5) linearly and hence can be determined using simple linear algebra: for given values of $\mathbf{p}$ and λ we define $\mathbf{q}(\mathbf{p}, \lambda)$ as the value which minimises $Z(\mathbf{q}, \mathbf{p})$. This is computed in directly in *MATLAB* as the least-squares solution of (6.5). For the computation of the local minimum of $Z(\mathbf{q}, \mathbf{p})$ closest to a given point $(\mathbf{q}, \mathbf{p})$ I 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.

6.3.1 Method 1- variable λ

As discussed earlier, for computing the non-linear parameter $\mathbf{p}$, we solve (6.5) by constructing an iterative minimisation procedure, within each iteration of which a linear problem is solved to find $\mathbf{q}$. We use the *MATLAB* optimisation routine *lsqnonlin* for the iteration of the minimisation procedure, and at each iteration we make use of the L-curve criterion from the linear inverse theory to estimate the regularisation parameter. However, by experiment we found that this approach leads to an oscillation in values of λ, as shown in Figure 6.2, and most of the time the approach does not give the desired solution. To overcome this problem we solve the solution process by fixing the value of λ at each iteration.

6.3.2 Method 2- fixed λ

First, we introduce an L-curve intended for the non-linear problem (6.5). We will call this the non-linear L-curve. The idea is to compute the solution path displaying norm of the linear solution against the residual norm for a sequence of λ values. The short algorithm for computing the L-curve goes as follows:

(i) carry out steps (ii) and (iii) for a sequence of λ values, as illustrated in Figure 6.3;

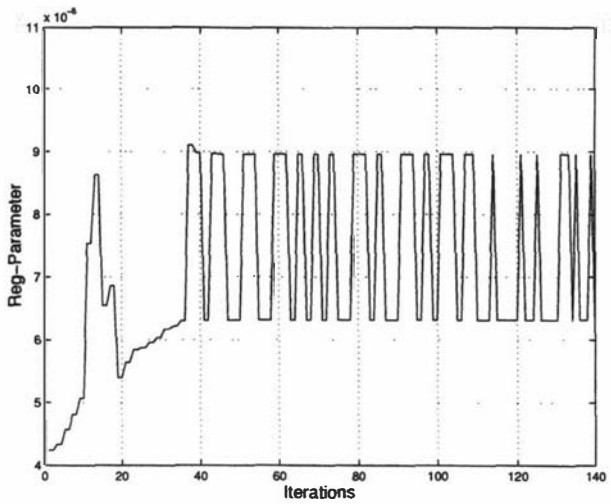

Figure 6.2: Use of linear L-curve for nonlinear problem

(ii) solve (6.5) using *MATLAB*'s optimisation routine *lsqnonlin* for a fixed value of λ;

(iii) take the solution $\mathbf{p}$ from the previous step as an initial value to the next problem with a new λ (see Figure 6.3);

(iv) plot the curve of residual norm $\log \|A\mathbf{q} - \mathbf{c}\|$ *vs* solution norm $\log \|L\mathbf{q}\|$.

This curve also exhibits a corner (see Figure 6.4a). As the regularisation parameter decreases towards the corner, the solution norm changes very little, while the error norm is reduced to a comparatively large extent. A decrease of the regularisation parameter further than the corner results in larger increases in the solution norm, with very little decrease in the error norm.

Method 2a Maximum curvature

Numerically, the point of maximum curvature of the curve locates the corner. Figures 6.4a and b illustrates the procedure. Figure 6.4a is the non-linear L-curve where the corner is clearly indicated and Figure 6.4b is the curvature of the L-curve as a function of λ. The peak in Figure 6.4b corresponds to the corner of the

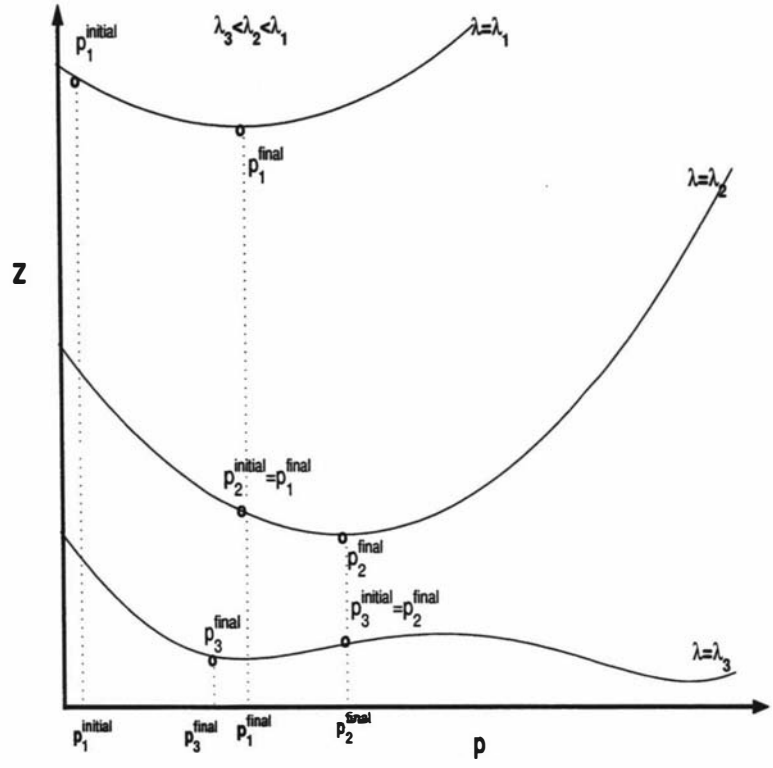

Figure 6.3: Illustration of the minimisation procedure for a sequence of λ values:

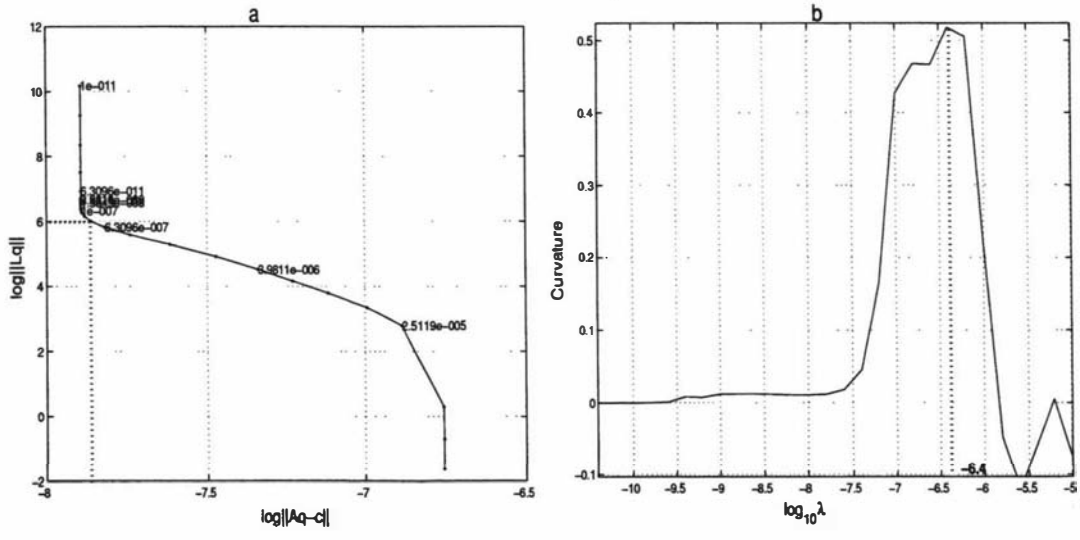

Figure 6.4: This figure demonstrates the use of maximum curvature method to estimate the optimal value of a regularisation parameter: (a) non-linear L-curve, (b) regularisation parameter *vs* curvature

L-curve. The corner is defined as the point on the L-curve:

$$(\varsigma(\lambda), \zeta(\lambda)) = (\log \|A\mathbf{q} - \mathbf{c}\|, \log \|L\mathbf{q}\|) \tag{6.8}$$

with maximum curvature. Here the curvature k is defined as:

$$\kappa(\lambda) = \frac{\varsigma'\zeta'' - \varsigma''\zeta'}{((\varsigma')^2 + (\zeta')^2)^{\frac{3}{2}}} \tag{6.9}$$

where the differentiation is with respect to λ. Unlike the linear problem, the non-linear objective Z may have more than one minimum for each value of λ and this leads to several non-linear L-curves. Therefore, the solution process of (6.5) should include

(i) finding all L-curves,

(ii) finding the best optimal point on the L-curves.

To deal with these problems we have developed an algorithm to determine the best L-curve and its corner.

The Parameter Estimation Algorithm -1

Let us now consider the numerical details of the algorithm to solve (6.5). The solution is arrived at through three steps. In the first step, we find all or most of the local minimum of (6.5) for a fixed value of λ when it is equal to its lowest value in the sequence. This is done by solving (6.5) 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 (6.5) for a fixed lowest λ. Equation (6.5) contains the error norm (ϕ_d) and the solution norm (ϕ_m) where the first is a non-linear part and the later is quadratic. Therefore, (6.5) is almost quadratic if λ is large, and non-linear if λ is small. We also found that the number of local minima of (6.5) increases with the decreasing values of λ.

Therefore the number of local minima of Equation (6.5) for the lowest value of λ 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 (6.5) for a sequence value of λ from lowest to largest. That means solving (6.5) n_l (number of local minima when λ is equal to its lowest value) times for a sequence values of λ. 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 λ. One such example is shown in Figure 6.8a.

The third and final step is to pick the optimal point on each curve and then select the best among those. The optimal point on each curve is calculated by examining the curvature along the L-curve. This can also be calculated using the rotated L-curve or NGCV methods that are described in the next subsections.

Method 2b- Non-linear GCV

Here we adapt Wahhba's [57] 'leaving-out-one' lemma to our non-linear problem. This lemma is based on the idea of removing one of the data points (c_i) from a given array of data $\mathbf{c}$ and then solving (6.5) using all the remaining data. The resulting inverse solutions $\mathbf{p}$ and $\mathbf{q}$ will predict the right hand vector $\mathbf{c}$, which can be used to estimate the omitted element c_i. Let the error in the misfit be equal to $\|c_i - \hat{c}_i\|$, where $\hat{c}_i$ is the estimated value. For a fixed value of λ, we repeat this process n times (number of data points) and sum the error misfits. We then have

$$\sum_{i=1}^{n} (c_i - \hat{c}_i)^2 .$$

If we repeat the whole process for a sequence of λ values, we then have the function

$$V(\lambda) = \sum_{i=1}^{n} (c_i - \hat{c}_{\lambda i})^2 .$$

The best value of λ can then be defined as the one that gives the smallest sum

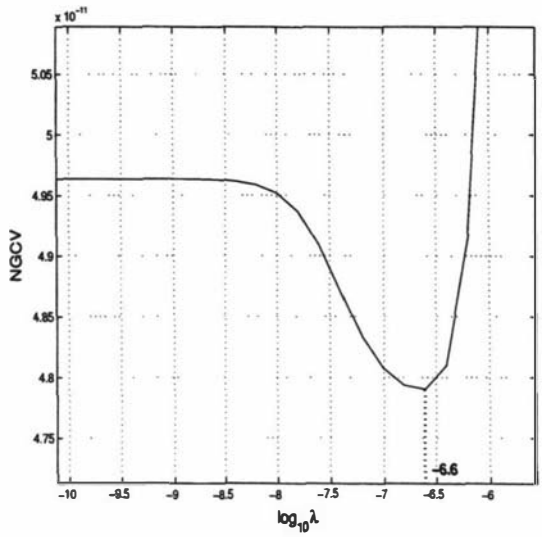

Figure 6.5: parameter selection- NGCV method

of misfits. We will call this method the NGCV. Computationally, NGCV is very expensive. The basic approach behind the NGCV and the GCV is the same. The GCV defined in (2.19) of Chapter 2 is for linear problem where the matrix A is same for all values of λ but in the non-linear problem A is a function of $\mathbf{p}$. Therefore we cannot use (2.19) directly to estimate the optimal value of λ. Figure 6.5 shows the estimation of λ using NGCV method for the same simulated data considered in the last subsection, where the lowest value of NGCV function is clearly indicated. The maximum curvature method and NGCV method use different approaches and therefore, each method will in general find a different optimal value of λ. Computationally, the maximum curvature method is much cheaper than NGCV.

Method 2c Rotated L-curve

The optimal point on the L-curve can also be detected by another simple method [13], [14] which proceeds by defining the rotated L-curve (see Figure 6.6):

$$\psi(\theta) = \tau + \zeta, \quad \theta = \tau - \zeta \tag{6.10}$$

where $\tau = \log\|A(\mathbf{p})\mathbf{q} - \mathbf{q}\|$ is the error norm function and $\zeta = \log\|L\mathbf{q}\|$ is the solution norm function. Then, any three neighbouring points

$$(\theta_j, \psi_j), j = 1, \ldots, n \tag{6.11}$$

will define a parabola. If $\psi_{j-1} > \psi_j$, $\psi_j < \psi_{j+1}$ for any j, then ψ_j will be a candidate for minimum ψ. The value of λ corresponding to ψ_j may be used as an approximation to an optimal value of the regularisation parameter. Figure 6.7a shows the rotated L-curve of Figure 6.4a. It can be seen clearly that none of the points on the curve had formed a parabola. This is because of the dominance of the solution norm over the error norm. This difficulty can be overcome by scaling the axis. But the main problem here is to find the exact value for the scaling factor. Rotated L-curves for different scaling factors are shown in Figure 6.7. It shows that the estimated optimal λ varies slightly with the different scaling factor. It is very hard to calculate the exact scaling factor, only an approximate value of the scaling factor can be obtained from the L-curves. Therefore optimal λ obtained using rotated L-curve method described in this section is only a rough approximation. Alternatively, the corner can be located using the derivative of the L-curve. At the corner, the derivative of the L-curve is approximately equal to -1, i.e. the angle between the tangent at the corner and the + ve axis of the data misfit function τ is 135 °. But in our problem magnitude of the solution norm is much larger than the error norm and therefore at the corner the value of the derivative does not have to be -1 unless we use the exact scaling.

The Table 6.1 shows the comparison of the optimal value of λ obtained for the non-linear L-curve shown in Figure 6.4a using different methods described in this section.

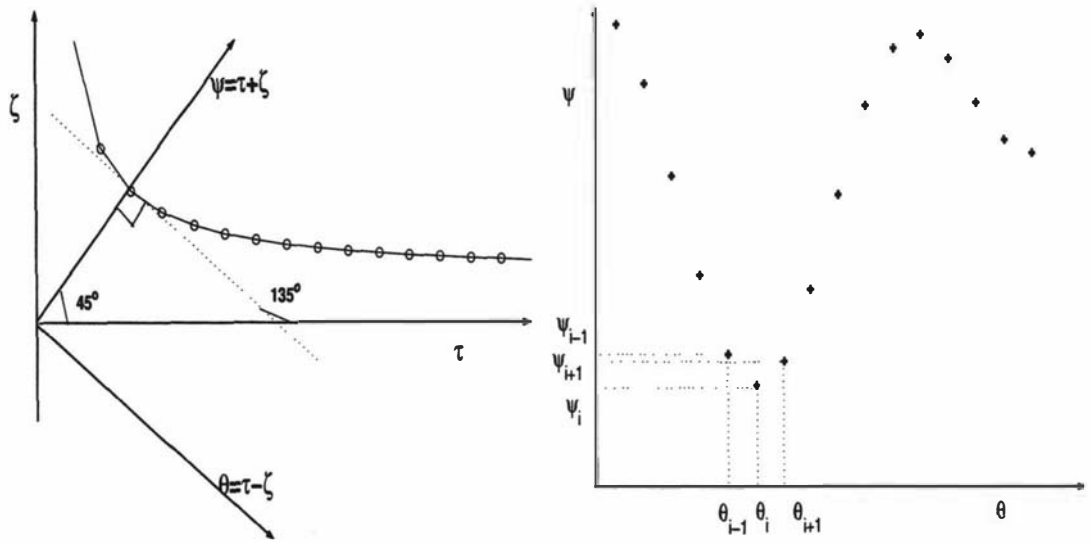

Figure 6.6: Rotation of L-curve

Table 6.1: Comparison of parameter selection methods

Method	λ
Maximum curvature	3.9811×10^{-7}
NGCV	2.5119×10^{-7}
Rotated L-curve	
scale 1:14	1.5849×10^{-7}
scale 1:12	2.5119×10^{-7}
scale 1:8	2.5119×10^{-7}
scale 1:1	none

Difficulties

Now consider the example given in Figure 6.8. All L-curves and their respective corners are clearly shown in Figure 6.8a. Figures 6.8b and c are the log of the norm of the error in **p**, **q** against log of λ respectively. Figure 6.8d is the error norm of the L-curves against log of λ. From Figures 6.8b, c and d it can be seen that the corner (L-curve 3) with the smallest error norm gives the lowest error in the **p** and **q** estimates than other corners. This criterion works well. However, with some extensive simulations, we found that roughly one out of four failures occurs. Figure 6.9 shows the example in which the criterion selects a bad corner, P_1, on the

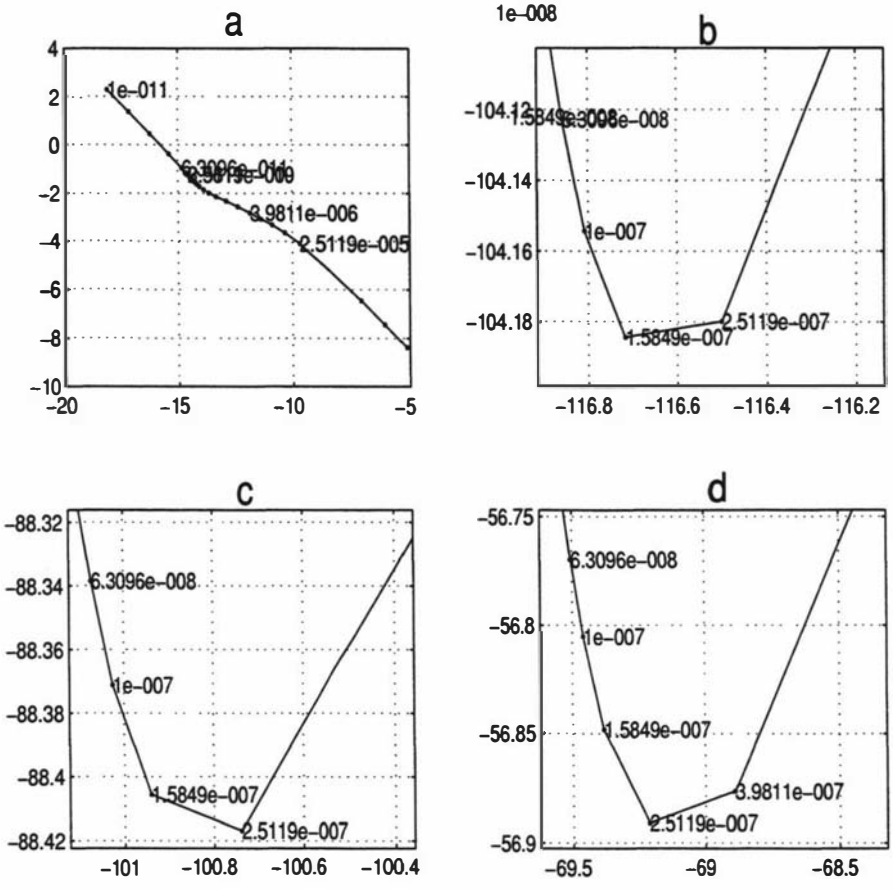

Figure 6.7: Rotated L-curves: (a) scaling factor=1, (b) scaling factor=14, (c) scaling factor=12, (d) scaling factor=8

L-curve-1. Figure 6.9a shows all the L-curves and their corners. P_1 is the corner on the L-curve-1, P_2 and P_3 are the corners on the L-curve-2. Here we used the maximum curvature method to locate the corners. Relative error measures in $\mathbf{p}$ and $\mathbf{q}$ are shown in Figures 6.9b and c respectively in a log scale. These figures clearly shows that the error in $\mathbf{p}$ and $\mathbf{q}$ estimates at the corner P_3 is smaller than the errors at the corners P_1 and P_2. Therefore the corner P_3 should be the better choice than other corners but the criterion we use failed to locate P_3. To overcome these problems we have slightly modified our present algorithm to determine the best L-curve and its corner.

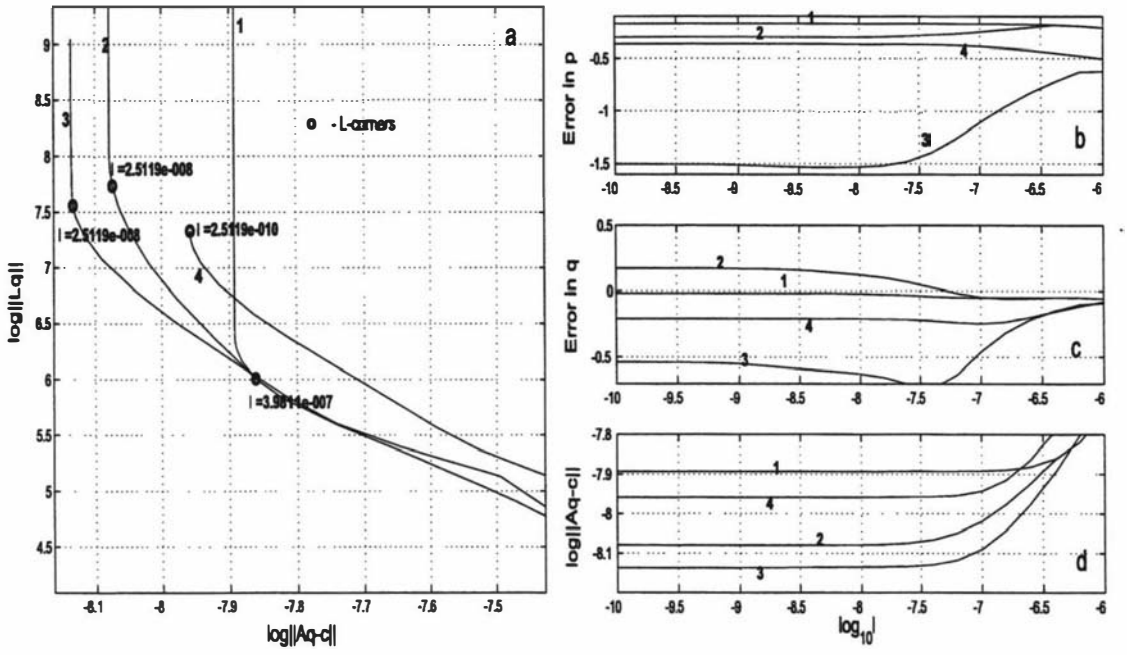

Figure 6.8: Good example: (a) non-linear L-curves, (b) relative error in **p** *vs* regularisation parameter, (c) relative error in **q** *vs* regularisation parameter, (d) data misfit *vs* regularisation parameter

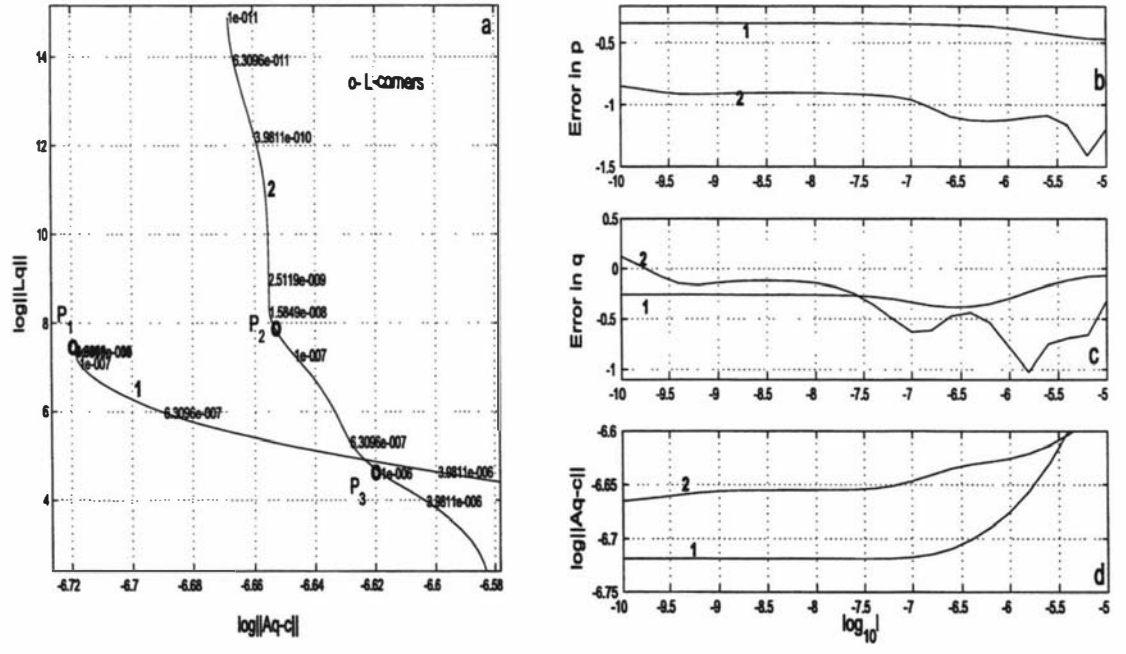

Figure 6.9: Bad example: (a) non-linear L-curves, (b) relative error in **p** *vs* regularisation parameter, (c) relative error in **q** *vs* regularisation parameter, (d) data misfit *vs* regularisation parameter

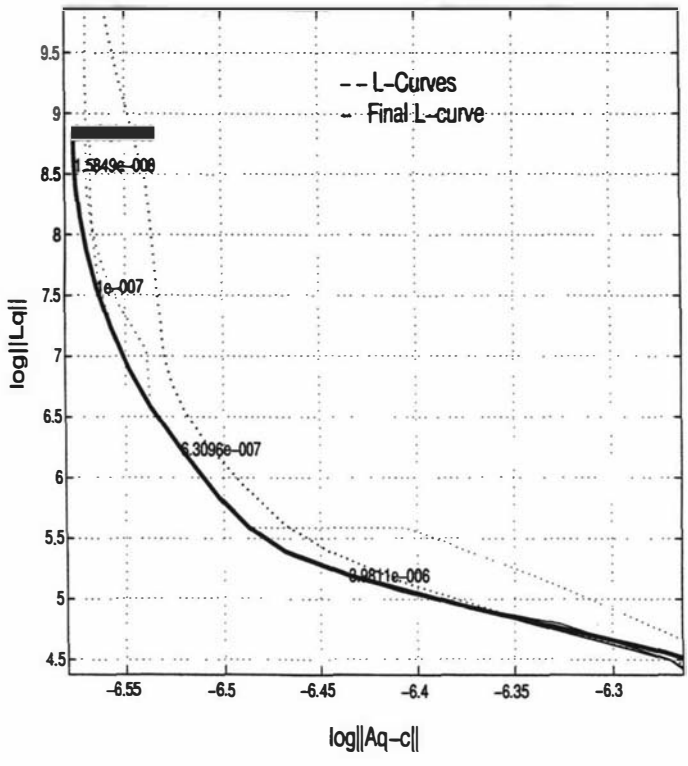

Figure 6.10: L-curves

The Parameter Estimation Algorithm-2

Let us now consider the numerical details of the algorithm to solve (6.5). The solution is arrived at through four steps. The first two steps are similar to the previous algorithm. The idea behind the first two steps 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 λ. One such example is shown in Figure 6.10 by dashed lines. In the third step, for each λ, 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 [17]. It is shown by the solid thick line in Figure 6.10. The fourth and final step is to pick the optimal point on the curve. This can be calculated either by using the rotated L-curve or by examining the curvature along the L-curve or NGCV method.

Finally, we summarize this algorithm as follows:

1. Initialisation

   (i) choose the upper (**UB**) and lower (**LB**) bounds of nonlinear parameter **p**;

   (ii) form an array of λ values from 10^{-12} to 10^{0} of size m.

2. Calculate all or most of the local minima of (6.5) when $\lambda = \lambda(1)$ by solving (6.5) as many times as possible and let the number of trials $= n_l$

   (i) for $i = 1, 2, \ldots, n_l$

   - select the initial value $\mathbf{p} = \mathbf{p_0}$ randomly within $[UB, LB]$;
   - solve the non-linear problem (6.5) for $\lambda = \lambda(1)$;
   - save the solution;

   (ii) let the number of local minima $= LK$;

   (iii) initialise the matrices A_1, A_2 and P of size $LK \times m$ to zero, matrices A_1, A_2 and P store the values of error norm, solution norm and the nonlinear solution **p**, respectively;

   (iv) compute $A_1(1 : LK, 1)$, $A_2(1 : LK, 1)$ and $P(1 : LK, 1)$ from step (i).

3. Compute all the L-curves

   (i) for $i = 1, 2, \ldots, LK$

   - $\mathbf{p_0} = P(i, 1)$;
   - for $k = 2, 3, \ldots, m$;
     . $\lambda = \lambda(k)$;
     . solve the non-linear problem (6.5) using the solution obtained from the previous step as start value, let the solution $= \mathbf{p}$;
     . $P(i, k) = \mathbf{p}$, compute $A_1(i, k)$, and $A_2(i, k)$;
   - plot all the L-curves ($log\|A(\mathbf{p})\mathbf{q} - \mathbf{c}\|$ *vs* $log\|L\mathbf{q}\|$);

4. For each λ, select the point on the L-curve that gives the function minimum, and plot the final L-curve.

5. Select the optimal point on the L-curve

   (i) identify all the possible optimal points by using

   - the maximum curvature (select bigger corners as well as smaller corners) along the L-curve;
   - the minimum of NGCV function;

   (ii) select the point that is closer to the origin.

6.3.3 Method 3 - fixed λ (Cheaper Method)

In this section, we build up an algorithm as a cheaper alternate to the method 2. The inspiration behind this algorithm is based on the work of [8], [9]. In this work, the optimal λ value of the nonlinear minimisation problem

$$\begin{aligned} \min_{\mathbf{x}} \ \phi(\mathbf{x}) &= \|F(\mathbf{x}) - \mathbf{c}\|^2 + \lambda^2 \|L\mathbf{x}\|^2 \\ &= \phi_d(\mathbf{x}) + \lambda^2 \phi_m(\mathbf{x}) \end{aligned} \tag{6.12}$$

is projected using linear inverse theory (L-curve or GCV). In 6.12, ϕ_d is an error norm, ϕ_m is a solution norm and λ is a regularisation parameter that balances the two components. In this problem the relationship between the data and the parameters $\mathbf{x}$ is non-linear. The non-linear relationship is overcome by constructing an iterative procedure in which the non-linear minimisation problem is replaced at each iteration by its linearised approximation. Differentiation of (6.12) with respect to $\mathbf{x}$ and equating the result to zero gives

$$\frac{\partial \phi}{\partial \mathbf{x}} = \lambda^2 L^T L\mathbf{x} + J(\mathbf{x})^T \left(F(\mathbf{x}) - \mathbf{c}\right) = 0, \tag{6.13}$$

where $J(\mathbf{x}) = \frac{\partial F}{\partial \mathbf{x}}$. Then we linearsed (6.13) by

$$\lambda^2 L^T L(\mathbf{x} + \delta\mathbf{x}) + J(\mathbf{x})^T \left(F(\mathbf{x}) + J(\mathbf{x})\delta\mathbf{x} - \mathbf{c}\right) = 0.$$

At the n-th iteration, we solve the linear system of equation for $\delta\mathbf{x}$ is

$$\left(J(\mathbf{x_n})^T J(\mathbf{x_n}) + \lambda^2 L^T L\right) \delta\mathbf{x} = J(\mathbf{x_n})^T \left(\mathbf{c} - F(\mathbf{x_n})\right) - \lambda^2 L^T L \mathbf{x_n}. \tag{6.14}$$

By letting $\mathbf{x_n} = \mathbf{x_{n-1}} + \delta\mathbf{x}$, (6.14) gives

$$\left(J(\mathbf{x_{n-1}})^T J(\mathbf{x_{n-1}}) + \lambda^2 L^T L\right) \mathbf{x_n} = J(\mathbf{x_{n-1}})^T \left(\mathbf{c} - F(\mathbf{x_{n-1}})\right) + J(\mathbf{x_{n-1}})\mathbf{x_{n-1}}.$$

This problem is identical to the linear least squares problem

$$\begin{bmatrix} J(\mathbf{x_{n-1}}) \\ \lambda L \end{bmatrix} \mathbf{x_n} = \begin{bmatrix} c - F(\mathbf{x_{n-1}}) + J(\mathbf{x_{n-1}})\mathbf{x_{n-1}} \\ 0 \end{bmatrix}. \tag{6.15}$$

At the n-th iteration, we solve the linear least squares problem (6.15). In the development of the solution, the value of λ starts from a larger value and then decreases from one iteration to the next slowly in conjunction with the equation

$$\lambda_{n+1} = \max(c\lambda_n, \lambda^*), \tag{6.16}$$

where $0.01 \leq c \leq 0.5$, λ^* is an optimal value of regularisation parameter obtained either using the l inear L-curve or GCV at the n-th iteration for the linear problem (6.15), and λ_n, λ_{n+1} are the values of λ at n-th, $(n+1)$-th iteration, respectively. The value of c is chosen from within a range 0.01 to 0.5 because these values efficiently imposes a steady decrease on λ values and giving consistent algorithm for the non-linear inverse problem. Iteration is carried out until λ, the error norm and the solution norm all have steady state. At steady state, optimal value of regularisation

parameter, λ^*, obtained using L-curve (or GCV) for the linear problem (6.15) is equal to the final λ value. i.e. $\lambda^* = \lambda_{n+1}$. It has been demonstrated [8], [9] [18] using simulated data sets that at the final iterations the obtained value of regularisation parameter was a good estimate of what was expected for a given noise level in the data.

Difficulties

Here, we propose the above approach of estimating optimal λ to our problem (6.5). The idea is to solve (6.5) for a sequences of λ's (each λ acts as the only regularisation parameter for the problem) so that this sequence approaches a steady state as it proceeds. This algorithm is as follows:

1. Choose the starting point $\mathbf{p} = \mathbf{p_0}$, index $I = 1$, and initial value of regularisation parameter $\lambda_I = \lambda_0$, which is very large.

2. Until convergence of λ, $\|A(\mathbf{p})\mathbf{q} - \mathbf{c}\|$, $\|L\mathbf{q}\|$ do

   (i) solve (6.5) for fixed λ_I, and find the solution $\mathbf{p} = \mathbf{p_I}$;

   (ii) obtain the optimal value of $\lambda = \lambda^*$ for the linear problem

   $$\min_{\mathbf{q}} \|A(\mathbf{p_I})\mathbf{q} - \mathbf{c}\| + \lambda^2 \|L\mathbf{q}\|$$

   using the linear L-curve;

   (iii) update the λ value, $\lambda_{I+1} = \max(c\lambda_I, \lambda^*)$;

   (iv) take the solution $\mathbf{p_I}$ as the starting point to the next problem with $\lambda = \lambda_{I+1}$;

   (v) I=I+1.

By numerical experiments we have found that this approach does not lead to our desired solution. In order to better explain this, we provide the non-linear L-curves

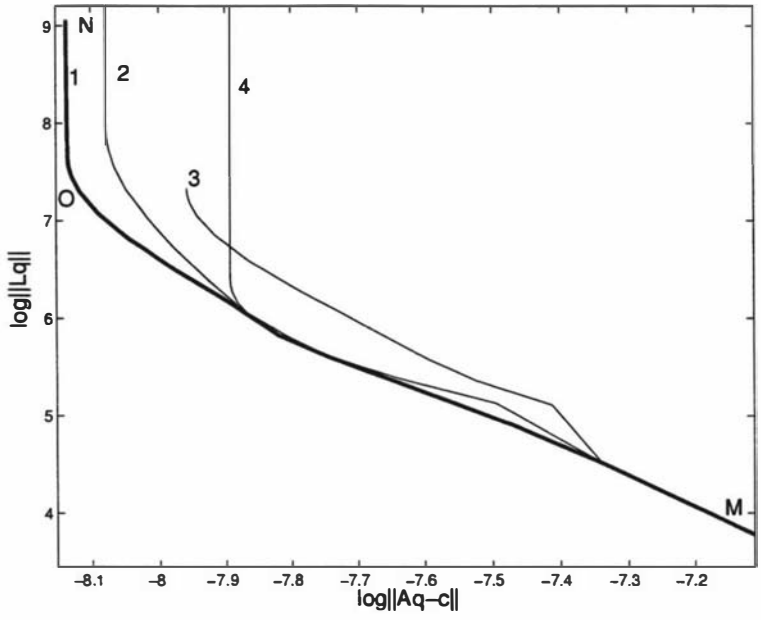

Figure 6.11: Non-linear L-curves

(solution curve or path) of (6.5). When λ is small, the solution of (6.5) may have several local minima and this leads to several non-linear L-curves, as shown in Figure 6.11. Suppose that curve 1 is the path of global minimal solution to (6.5), and curves 2, 3 and 4 are the corresponding local minimal solutions for various λ values. The value of λ at M is large and therefore gives only one minimum for (6.5). If we start from M, there is no guarantee the solution path of (6.5) will be curve 1 in Figure 6.11. It might be curves 2 or 3 or 4 in Figure 6.11. This example clearly demonstrates that the above approach is not very useful for our problem. To overcome the difficulty, we have to modify this approach slightly.

The Modified Method

Here, we propose a new methodology for solving (6.5). The approach is similar to the above algorithm given in section 6.3.3, but here we start with a small regularisation parameter and then slowly increase its value. In the first step, we find all or most of the local minima of (6.5) for the starting value of λ. At this stage we have to be very careful to avoid rank deficiency when we choose a small value of λ.

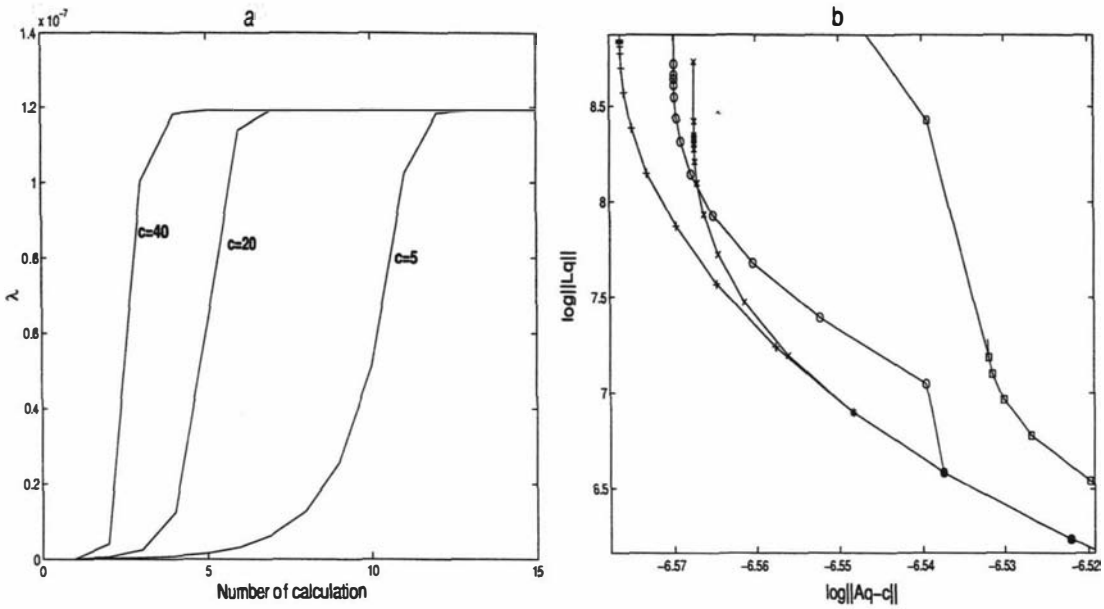

Figure 6.12: (a) Convergence of λ for different values of c, (b) minima solution paths

In the second step, we take each of the local minima obtained from the first step as the starting value to solve (6.5) for various λ from small to large. We increase λ slowly using the formula $\lambda_{I+1} = \min(c\lambda_I, \lambda^*)$, where $c > 1$, λ^* is the optimal value of the regularisation parameter obtained using linear L-curve for the sub-problem

$$\min_{\mathbf{q}} \|A(\mathbf{p_I})\mathbf{q} - \mathbf{c}\| + \lambda^2 \|L\mathbf{q}\|,$$

and $\mathbf{p}_I$ is the global minimal solution for λ_I. We repeat this process until λ, the error norm and the solution norm have achieved steady state (or convergence). When c is small, it takes a longer time to converge on λ, but the gap between two points on the minima solution curves are close enough and it therefore avoids jumping from one curve to other (see Figures 6.12a & b). Therefore use of a small value for c is recommended.

The Parameter Estimation Algorithm - 3

The summary of the algorithm goes as follows

1. Initialisation

   (i) choose upper (UB) and lower (LB) bounds of nonlinear parameter $\mathbf{p}$;

   (ii) choose the small value of $\lambda = \lambda_{small}$.

2. Calculate all or most of the local minima of (6.5) for λ_{small} by solving (6.5) as many times as possible (say n_l),

   (i) for $i = 1, 2, \ldots, n_l$

   - select the initial value $\mathbf{p_0}$ randomly within [UB LB];
   - solve the nonlinear problem (6.5) for a λ_{small};

   (ii)
   - let the number of local minima = LK;
   - store all the local minima in the vector $\mathbf{pp}(1, \ldots, LK)$;
   - select a solution $\mathbf{p} = \mathbf{p}^*$ from $\mathbf{pp}$, which gives the global minimum;
   - using the linear L-curve, find the optimal value of $\lambda = \lambda^*$ for the linear problem

$$\min_{\mathbf{q}} \|A(\mathbf{p}^*)\mathbf{q} - \mathbf{c}\|_2^2 + \lambda^2 \|L\mathbf{q}\|_2^2,$$

   - $\lambda = \min(c\lambda_{small}, \lambda^*)$, where $c > 1$.

3. Compute the path of the global minimum

   (i) while the convergence of λ is achieved

   - for $i = 1, 2, \ldots, LK$
     - $\mathbf{p_0} = \mathbf{pp}(i)$;
     - solve (6.5) using $\mathbf{p_0}$ as a starting point;
     - $\mathbf{pp}(i)$=solution;

- select a solution $\mathbf{p} = \mathbf{p}_I$ from $\mathbf{pp}$ that gives the global minimum;
- using the linear L-curve, find the optimal value of $\lambda = \lambda^*$ for the linear problem

$$\min_{\mathbf{q}} \|A(\mathbf{p}_I)\mathbf{q} - \mathbf{c}\|_2^2 + \lambda^2 \|L\mathbf{q}\|_2^2;$$

- $\lambda_{I+1} = min(c\lambda_I, \lambda^*)$.

4. The convergent value of λ is the optimal value of the regularisation parameter for (6.5), and the value of $\mathbf{p}$ corresponding to that λ is the optimal value of the location.

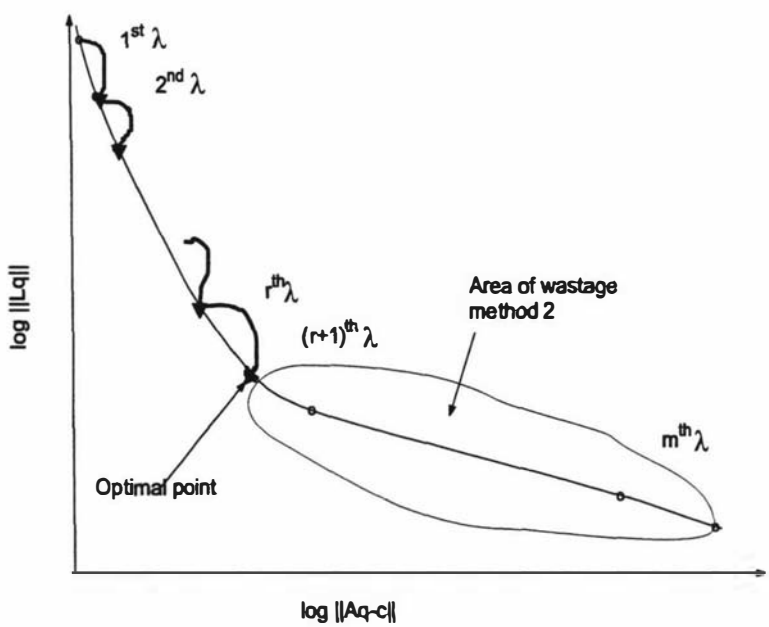

Figure 6.13: Nonlinear solution path

A sample solution path of minimisation problem (6.5) is shown in the Figure 6.13. From this figure it is clear that some of the work done using the method described in section 6.3 is a waste. Therefore the method described in this section is computationally more efficient (cheaper) than the previous method. Both methods are safer since we increase λ slowly, and for each λ we find all local minima and select

the minimal solution that gives the objective function minimum. Therefore, there is no way to trap in a local minimum. These two methods use different approaches to approximate the optimal value, and therefore each method may find different optimal value of λ.

6.4 Confidence Interval Estimation

We have already discussed confidence intervals for least-squares problems in Chapter 2. In this section we develop a formula to calculate the confidence interval for a problem given by (6.4). For any nonlinear model $\mathbf{c} = K(\mathbf{z}) + \mathbf{e}$, the 95 % confidence limit of the parameter estimates $\mathbf{z}$ is given by

$$\mathbf{z} \pm t_{0.025,\alpha} S_{\mathbf{z}} \tag{6.17}$$

and is based on the use of t-distribution [39], where α = degrees of freedom, $S_{\mathbf{z}}$ = square roots of the diagonal elements the variance-covariance matrix $s^2(W^T W)^{-1}$, s^2 is the residual mean square given by

$$s^2 = \sum_{i=1}^{n+1} \frac{[C_i - K(\hat{\mathbf{z}})]^2}{n+1-r},$$

r is the length of the vector $\mathbf{z}$ and W is a matrix of partial derivatives of K.

Now we will use the above properties to develop a method for constructing confidence intervals for the problem given in (6.4). Let

$$\mathbf{z} = [\mathbf{q}, \mathbf{p}]^T, \ \mathbf{p} = [X_0, Y_0, H]^T, \ \mathbf{q} = [q_1, q_2, \ldots, q_m]^T.$$

Therefore the matrix of partial derivatives is $W = [A \ J]$, where J is the $n \times 3$ matrix,

$$J_{i1} = \sum_j \frac{\partial A_{ij}}{\partial X_0} q_j, \ J_{i2} = \sum_j \frac{\partial A_{ij}}{\partial Y_0} q_j, \ J_{i3} = \sum_j \frac{\partial A_{ij}}{\partial H} q_j,$$

and A is given by (6.4). Since the least-squares problem given in (6.4) is ill-posed, we should therefore include regularisation in computing the confidence intervals. If we don't include the regularisation in computing the confidence intervals, then the confidence intervals will be just as large as those that we would get from simple least squares [47], [10]. If we assume

$$K(\mathbf{z}) = A(\mathbf{p})\mathbf{q},$$

then (6.5) can be written as

$$\min_{\mathbf{z}} \left[\|K(\mathbf{z}) - \mathbf{c}\|^2 + \lambda^2 \left\| \begin{bmatrix} L & 0 \end{bmatrix} \begin{bmatrix} \mathbf{q} \\ \mathbf{p} \end{bmatrix} \right\|^2 \right]. \tag{6.18}$$

Therefore the variance-covariance matrix of $\hat{\mathbf{z}}$ with regularisation is

$$var(\hat{\mathbf{z}}) = s^2 \left(w^T w\right)^{-1}, \quad w = \begin{bmatrix} W \\ \lambda \begin{bmatrix} L & 0 \end{bmatrix} \end{bmatrix}.$$

The confidence interval for the problem (6.5) can be estimated by replacing W with w in (6.17).

6.5 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 $R = (X_0 + 100, Y_0 + 70, 0)$ for all examples except Example 5 in this section. In Example 5, distances between source and measurement locations are very large compared with other examples, and we

therefore simulate concentration signals at downstream locations $P = (X_0, Y_0, 0)$, $Q = (X_0 - 300, Y_0 - 30, 0)$ and $R = (X_0 + 200, Y_0 + 70, 0)$. We obtain concentration signals by using the forward problem (6.1) 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.

6.5.1 Method 2

We consider five examples to demonstrate the developed method for finding the release rate $q(t)$ and the location (X_0, Y_0, H) using method 2. The purposes of these examples are:

(i) to demonstrate the simultaneous estimation of parameters X_0, Y_0, H, and the source release function $q(t)$;

(ii) to demonstrate the accuracy of the method to handle different situations such as noise in the data, source type, the discretisation size of the source function, larger distances between source and measurement points, and so on.

Example 1: Measured data corrupted by 10% of two different types of random noise

In this example we consider two sets of data where each set is corrupted by 10% of different random noise. The results of the source-term estimation are summarised in Table 6.2, and Figures 6.14 and 6.15. Listed in Table 6.2 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 Figures 6.14a and 6.15a. Figures 6.14b–6.15b are the curvature of the L-curves as a function of λ. The peaks in the figure correspond to the corners on the respective L-curves (P_1 & P_2 in Figure 6.14a, and P_1 & P_3 in Figure 6.15a). The Figures

6.14c and 6.15c illustrate the NGCV method 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 6.14a, and P_2 is a corresponding point in Figure 6.15a). Figures 6.14d and 6.15d depict the true error in the solution as a function of λ. 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. Figures 6.14e and 6.15e depict the error in $\mathbf{q}$ (release rates) as a function of λ. Figures 6.14f and 6.15f depict the error in the reconstructed location $\mathbf{p}$ as a function of λ.

Table 6.2: Example 1: Different random noise of 10% in the measured signal

		data 1	data 2
$\mathbf{p}$	True value	Estimated value $\pm$	Confidence interval
X_0	150.00	162.00 ± 9.42	153.20 ± 10.16
Y_0	25.00	19.3 ± 6.83	21.00 ± 8.21
H	12.00	11.1 ± 0.85	11.75 ± 0.72

For the first set of data, P_1, P_2 and P_3 are the three possible candidates for the optimal point on the L-curve (Figure 6.14a), and the points O_1, O_2 and O_3 respectively (Figure 6.14d) are their corresponding true errors. O_0 is the point where the true error is minimum. The point O_1 is closer to the minimum O_0 than the other two points, and therefore the regularisation parameter corresponding to a point O_1 is more suitable value for the problem than other values. Similarly for the second set of data, the three candidates are P_1, P_2, P_3 (Figure 6.15a) and the points O_1, O_2 and O_3 respectively (Figure 6.15b) are their corresponding true errors. O_1 is closer to the minimum than other two points, and therefore the regularisation parameter corresponding to point O_1 is more suitable value for the problem than other values.

It can be seen from Figures 6.14d and 6.15d that (i) the true errors are not exactly same and the minimum occurred at two different values of λ, (ii) in both cases the error in the reconstructed solution is dominated by regularisation rather

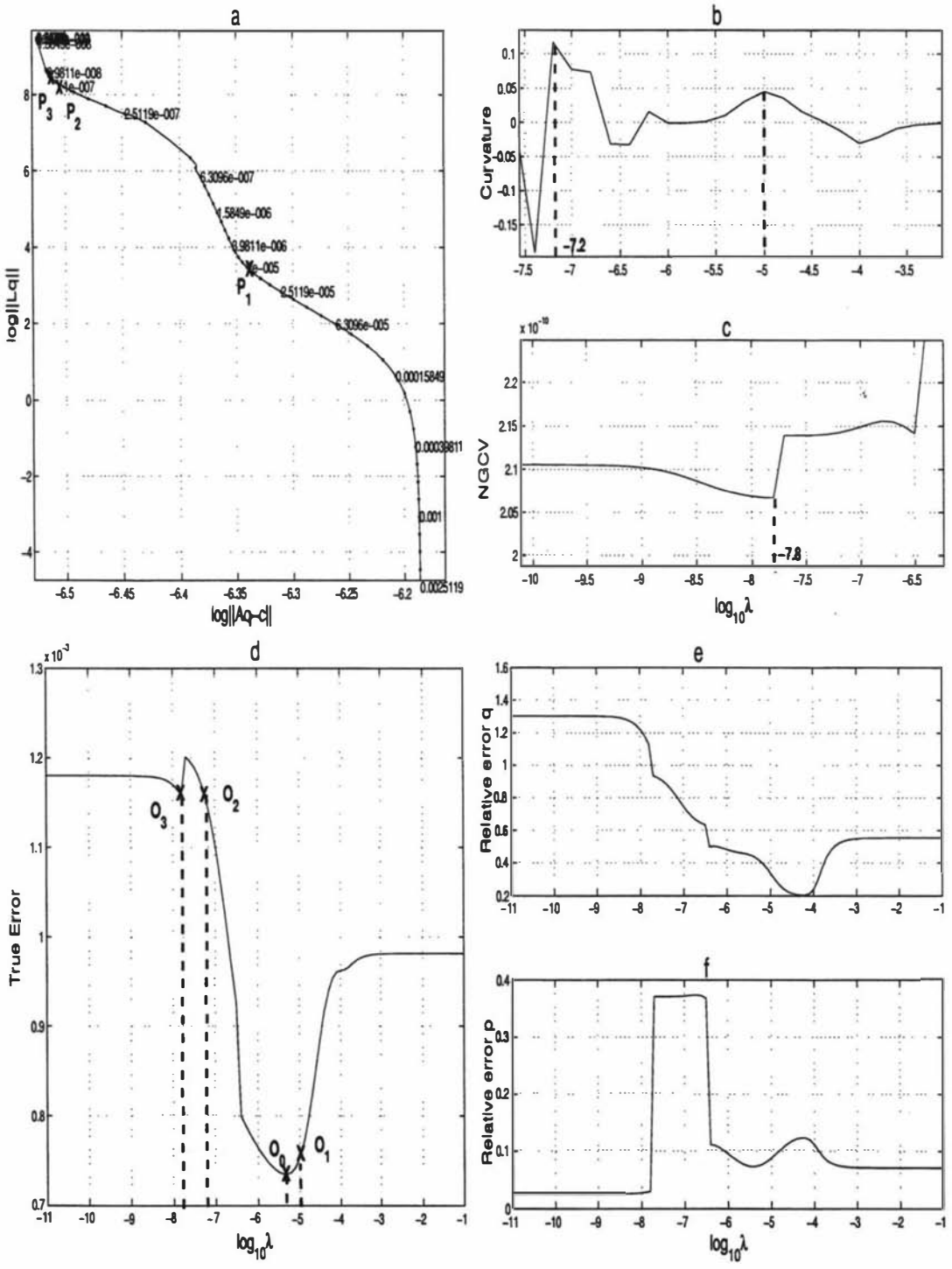

Figure 6.14: Example 1- data 1: (a) non-linear L-curve, (b) curvature *vs* regularisation parameter, (c) NGCV *vs* regularisation parameter, (d) true error *vs* regularisation parameter, (e) relative error in **q** *vs* regularisation parameter, (f) relative error in **p** *vs* regularisation parameter

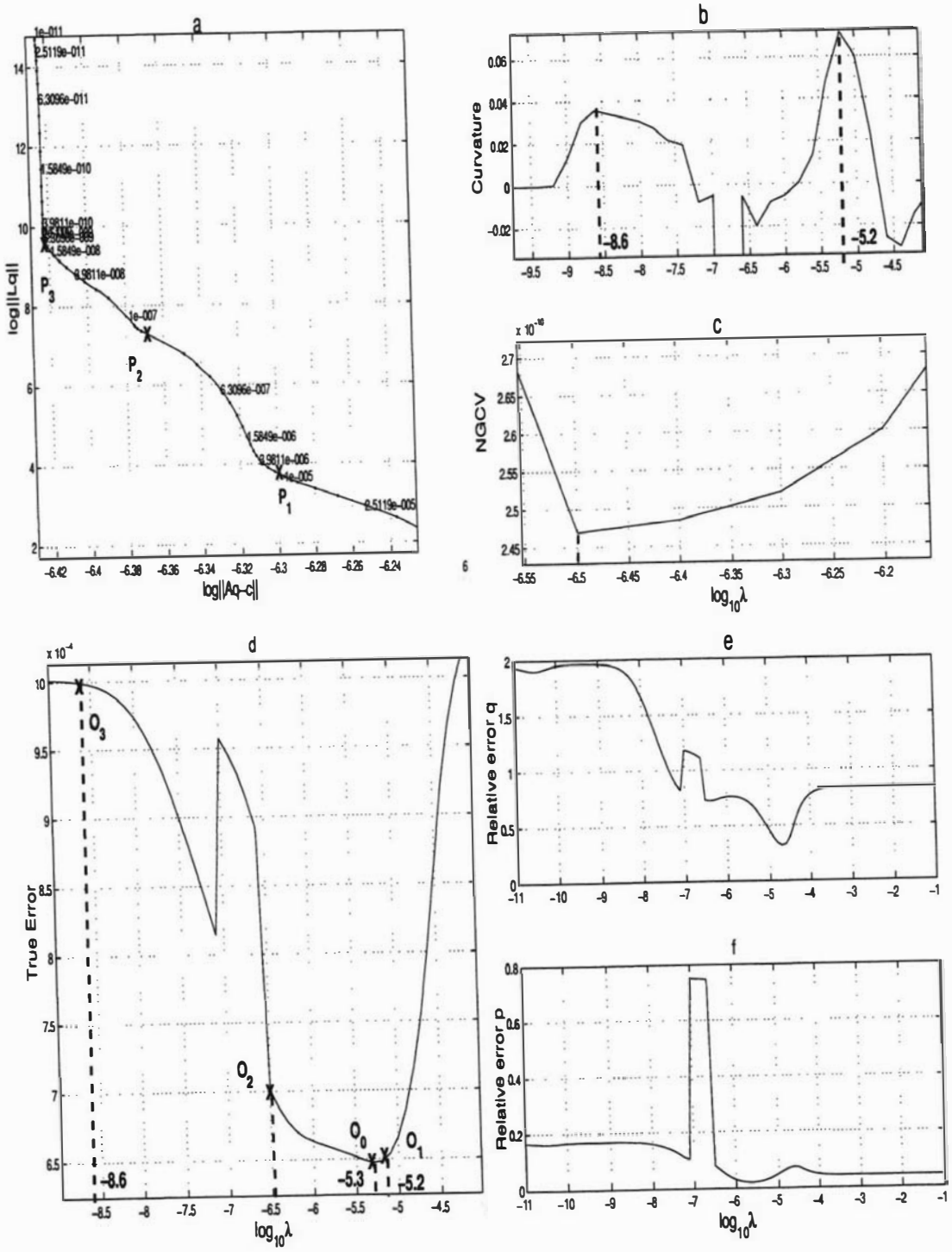

Figure 6.15: Example 1- data 2: (a) non-linear L-curve, (b) curvature *vs* regularisation parameter, (c) NGCV *vs* regularisation parameter, (d) true error *vs* regularisation parameter, (e) relative error in **q** *vs* regularisation parameter, (f) relative error in **p** *vs* regularisation parameter

Table 6.3: Case 2: Different size of same random noise (0%, 1%, 5%, and 10%) in the measured signal

		Noise in the data			
		0%	1%	5%	10%
p	True value	Estimated value ± Confidence interval			
X_0	200.00	202.33 ± 0.37	202.73 ± 2.0	206.5 ± 8.65	183.6 ± 19.20
Y_0	30.00	30.2 ± 0.05	31.2 ± 1.70	34.2 ± 5.49	37.15 ± 9.80
H	10.00	9.98 ± 0.01	10.09 ± 0.09	10.65 ± 0.51	10.95 ± 0.75

than by noise in the data since the estimated optimal points lie on the right of the true minimum, and (iii) the error in the solution for the first data set is larger than for the second data set. Since both data sets are corrupted by the same size of noise, we can therefore conclude that the error in the reconstructed solution depends on the data pattern (randomness).

Example 2: Measured data corrupted by the different size of the same random noise

In this example we consider four sets of data: perfect data, and data corrupted by 1%, 5% and 10% of the same random noise. Twenty data samples are taken between the times t=0.01 and t=300, and the release rate function is discretised into ten equal parts over the same time interval. The results of the source term estimation are summarised in Table 6.3 and Figures 6.16–6.19. Listed in the Table 6.3 are the true non-linear parameter (location) values along with the reconstructed values and their confidence intervals. The L-curve, which contains the function minimum, is shown in Figures 6.16a–6.19a for each data set. Figures 6.16b–6.19b are the curvature of the L-curve as a function of λ. The peaks in each figure correspond to the corner on the corresponding L-curves (P_1 in Figure 6.16a, 6.17a, 6.19a and P_1, P_2 in Figure 6.18a). Figures 6.16c–6.19c illustrate the NGCV method to estimate the optimal point on the L-curve, where the lowest value of the NGCV function is clearly indicated. The point P_2 is their corresponding point on the L-curve in

Figures 6.17a, 6.19a, and P_3 is in Figure 6.18a. In Figure 6.16c, the minimum of the NGCV function is very flat and therefore it is not possible to locate the minimum. Figures 6.16d–6.19d depict the true error in the solution against regularisation parameter. Figures 6.16e–6.19e depict the true linear parameter (release rates), along with its reconstructed value and confidence interval.

For the first set of data, P_1 is the only candidate for the optimal point on the L-curve (Figure 6.16a), and the point O_1 (Figure 6.16b) is its corresponding true error. The error at the point O_1 is not zero, even though the data are exact. This shows the effect of regularisation on the accuracy of the solution. For the second set of data, P_1 and P_2 are the candidates for the optimal point on the L-curve (Figure 6.17a), and the points O_1, O_2 (Figure 6.17b) are their corresponding true errors. Here the point O_1 is our choice, since it is closer to the true minimum (O_0) than O_2. The error at O_1 is larger than the true minimum and is on the right of O_0 (true minimum), i.e. the regularisation error dominates rather than the noise in the total error. For the third set of data, P_1, P_2 and P_3 are the candidates for optimal points on the L-curve (Figure 6.18a), and the points O_1, O_2 and O_3 (Figure 6.18d) are their corresponding true errors. The point O_1 is closer to the true minimum (O_0) than other two points and therefore the regularisation parameter corresponding to the point P_1 is more suitable value for the problem than other values. The error at point P_1 is larger than the true minimum, and is on the left side of O_0 (true minimum). This shows that the total error term is dominated by the noise in the data. Finally, for the fourth set of data, P_1 and P_2 are the candidates for optimal points on the L-curve (Figure 6.19a), and the points O_1 and O_2 (Figure 6.19d) are their corresponding true errors. The point O_1 is our choice, since it is closer to the true minimum than O_2. The error at point O_1 is larger than the true minimum (O_0), and is on the left side of O_0 (true minimum). This shows the total error term is dominated by the noise in the data.

It can be seen from Figures 6.16–6.19 that (i) the true error increases with

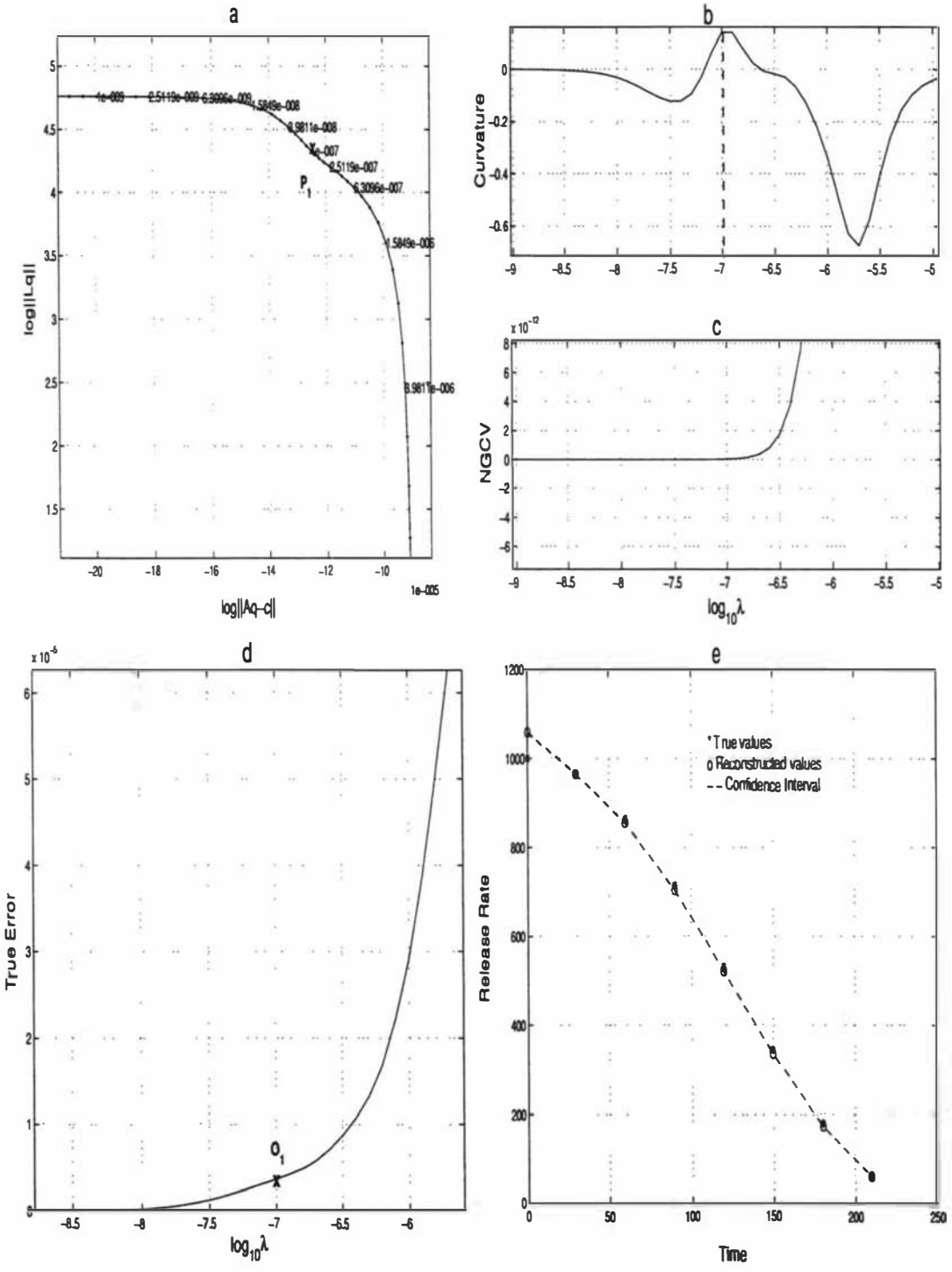

Figure 6.16: Example 2- data 1: (a) non-linear L-curve, (b) curvature *vs* regularisation parameter, (c) NGCV *vs* regularisation parameter, (d) true error *vs* regularisation parameter, (e) release rates *vs* time

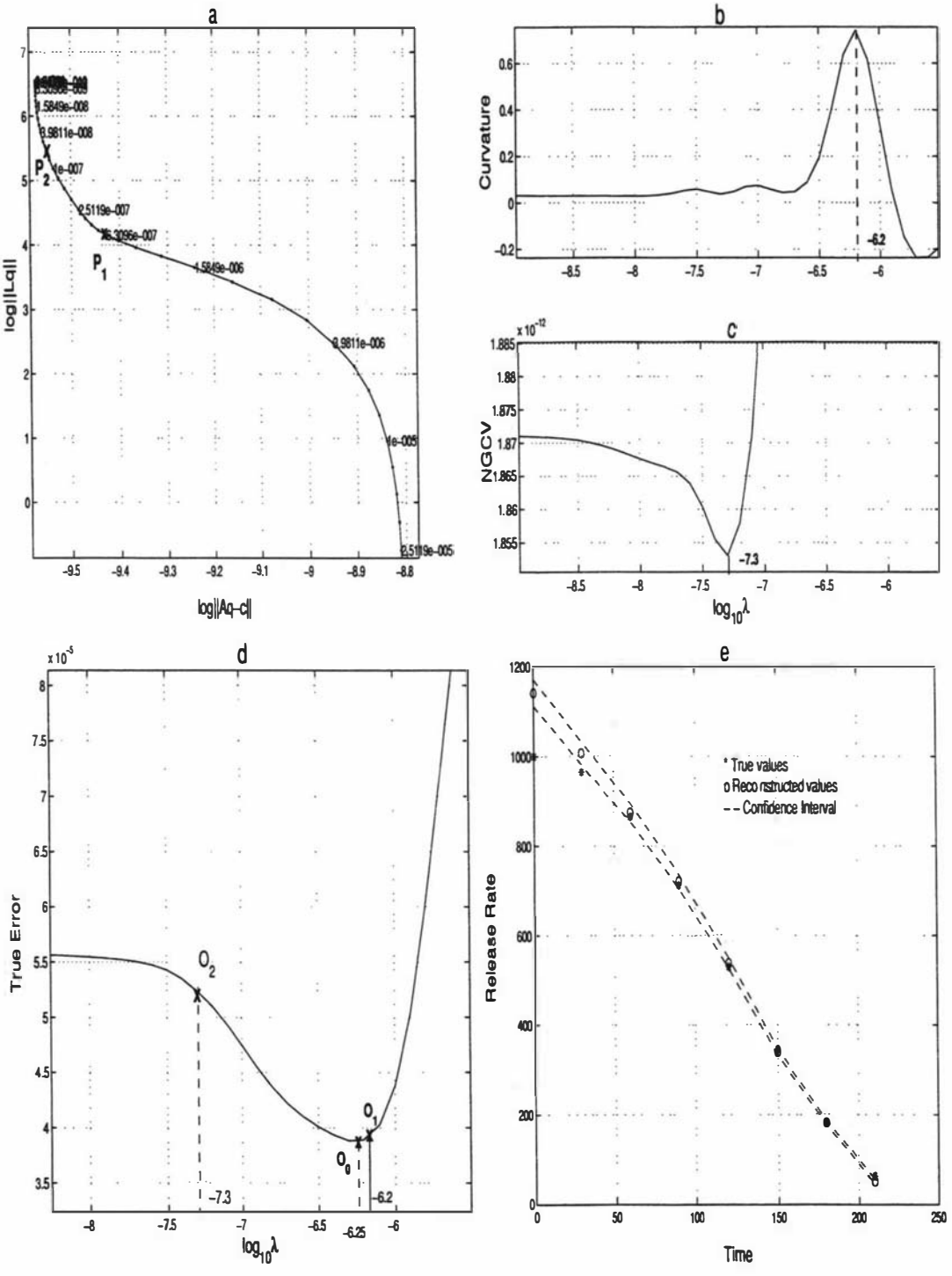

Figure 6.17: Example 2- data 2: (a) non-linear L-curve, (b) curvature *vs* regularisation parameter, (c) NGCV *vs* regularisation parameter, (d) true error *vs* regularisation parameter, (e) release rates *vs* time

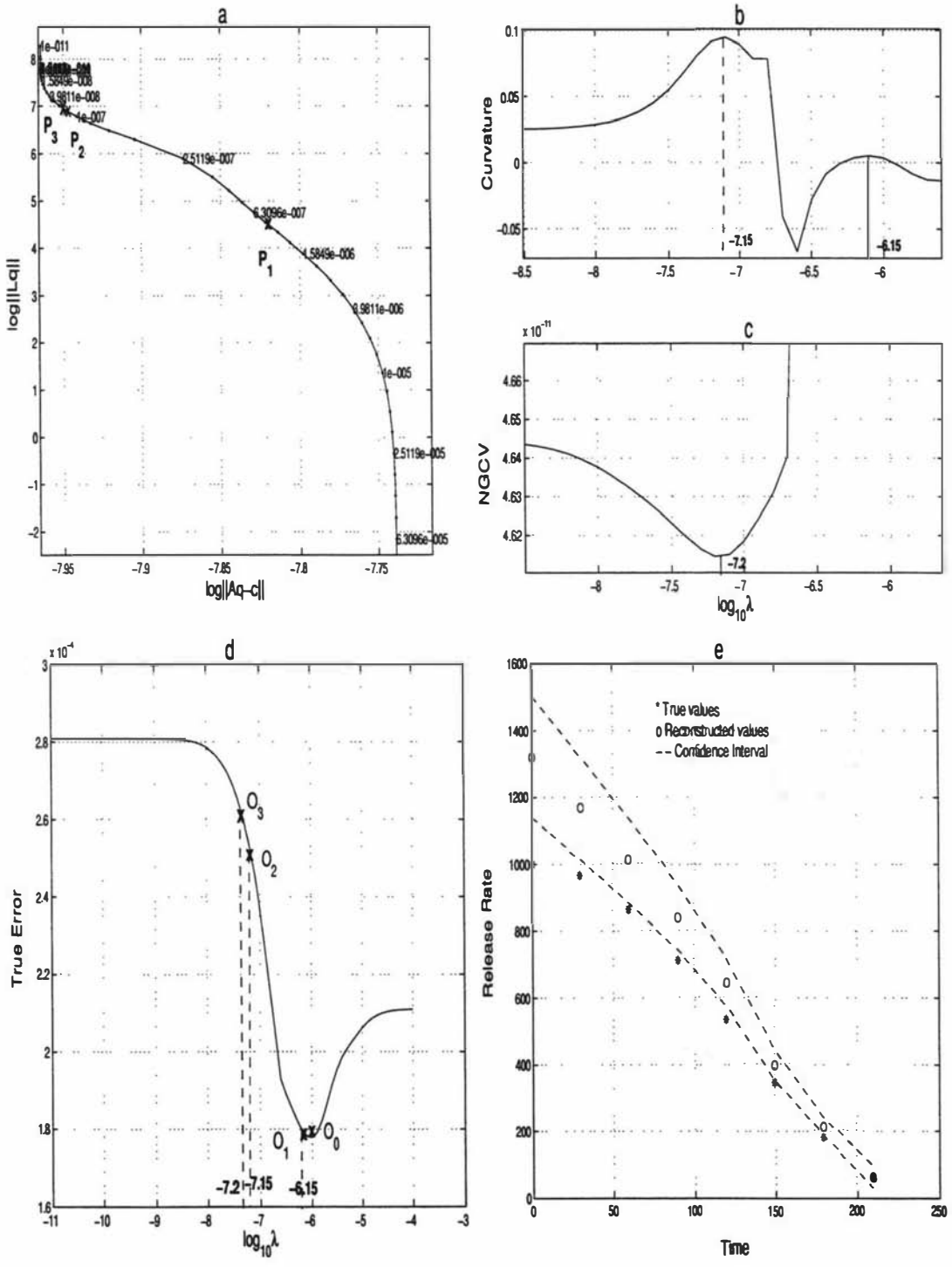

Figure 6.18: Example 2- data 3: (a) non-linear L-curve, (b) curvature *vs* regularisation parameter, (c) NGCV *vs* regularisation parameter, (d) true error *vs* regularisation parameter, (e) release rates *vs* time

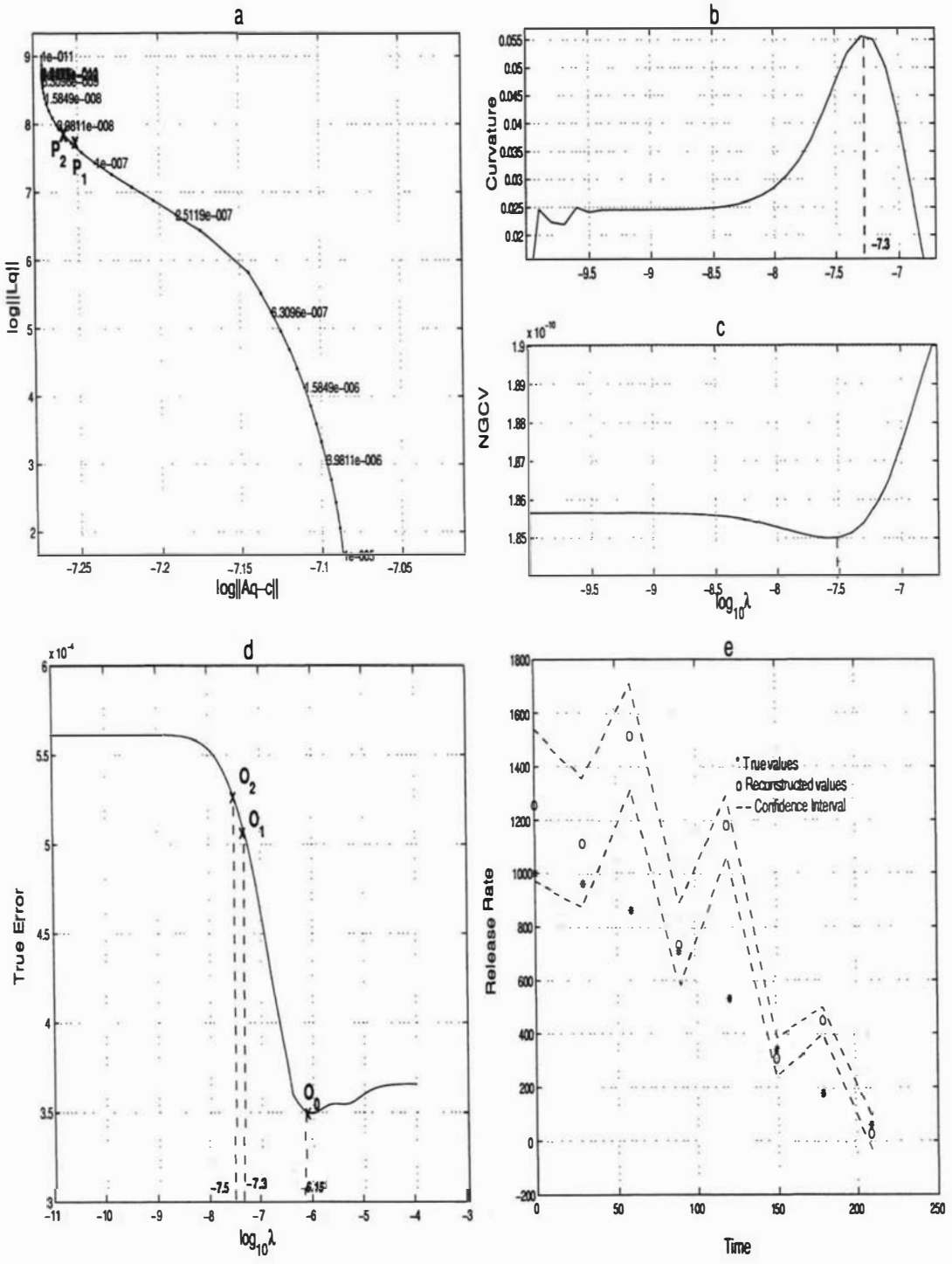

Figure 6.19: Example 2- data 4: (a) non-linear L-curve, (b) curvature *vs* regularisation parameter, (c) NGCV *vs* regularisation parameter, (d) true error *vs* regularisation parameter, (e) release rates *vs* time

Table 6.4: Case 3: Different number of source, data points

		Noise in the data		
		0%	1%	5%
p	True value	Estimated value ± Confidence interval		
X_0	200.00	200.75 ± 0.21	202.29 ± 1.80	200.2 ± 8.0
Y_0	30.00	30.06 ± 0.02	31.08 ± 1.38	35.2 ± 5.89
H	10.00	9.99 ± 0.001	10.10 ± 0.14	10.68 ± 0.51

increasing noise in the data, (ii) the size of the confidence interval increases with increasing noise in the data, (iii) the errors in the first two reconstructed solutions are dominated by the regularisation error, while the last two are dominated by noise in the data.

Example 3: Different numbers of source points m, data points n

In this example we consider the first three of the four data sets from the previous example. Twenty data samples are taken over the time interval $t_0 = 0.01$ to $t_n = 300$, and the release-rate function is discretised into fifteen equal parts over the same time interval. The results of the source-term estimation are summarised in Table 6.4, Figures 6.20–6.22. Listed in the table are the true non-linear parameter (location) values along with the reconstructed values and confidence interval estimates. The L-curves, which have function minimum, are shown in Figures 6.20a–6.22a for each set of data. Figures 6.20b–6.22b depict the error in the linear parameters (release rates) as a function of λ. Figures 6.20c–6.22c depict the error in the reconstructed location $\mathbf{p}$ as a function of λ. Figures 6.20d–6.22d depict the true error in the solution against different regularisation parameters. Figures 6.20e-6.22e depict the true linear parameter (release rates) along with its reconstructed value and confidence interval.

For the first set of data, P_1 is the only corner on the L-curve (Figure 6.20a), and point O_1 (Figure 6.20b) is its corresponding true error. The error at point O_1 is not zero, even though the data are perfect. This shows the effect of regularisation

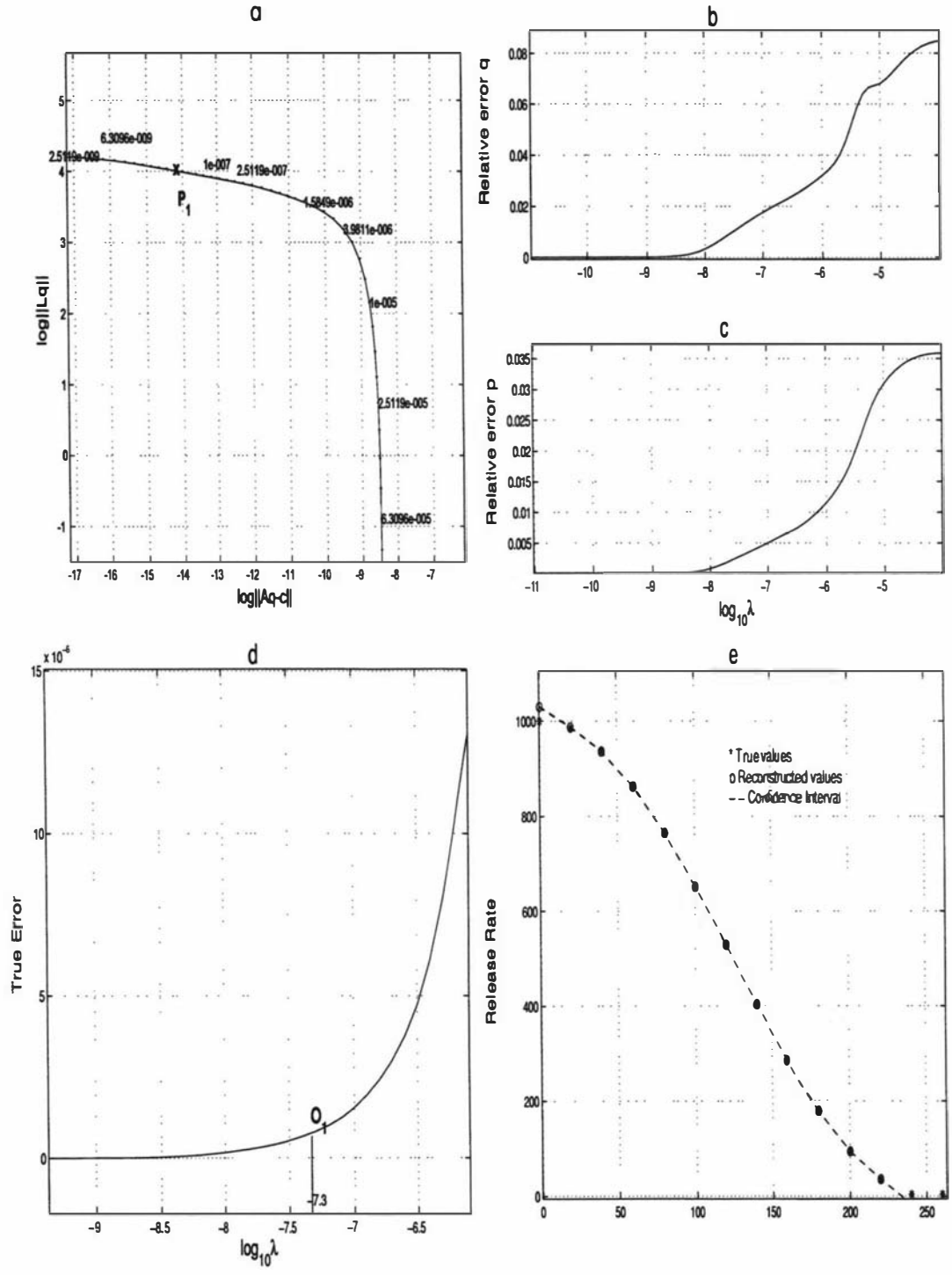

Figure 6.20: Example 3- data 1: (a) non-linear L-curve, (b) relative error in **q** *vs* regularisation parameter, (c) relative error in **p** *vs* regularisation parameter, (d)true error *vs* regularisation parameter, (e) release rates *vs* time

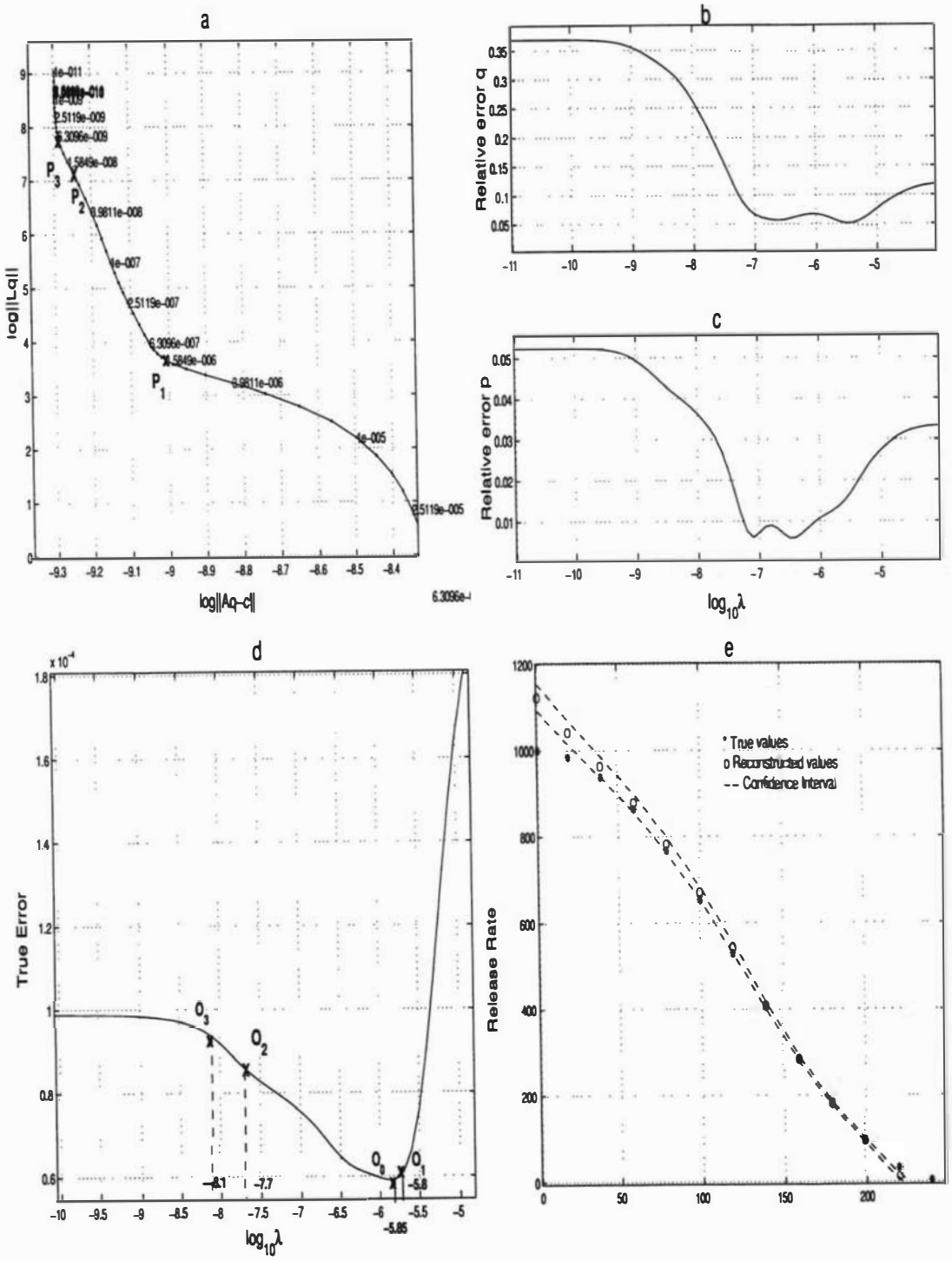

Figure 6.21: Example 3- data 2: (a) non-linear L-curve, (b) relative error in **q** *vs* regularisation parameter, (c) relative error in **p** *vs* regularisation parameter, (d) true error *vs* regularisation parameter, (e) release rates *vs* time

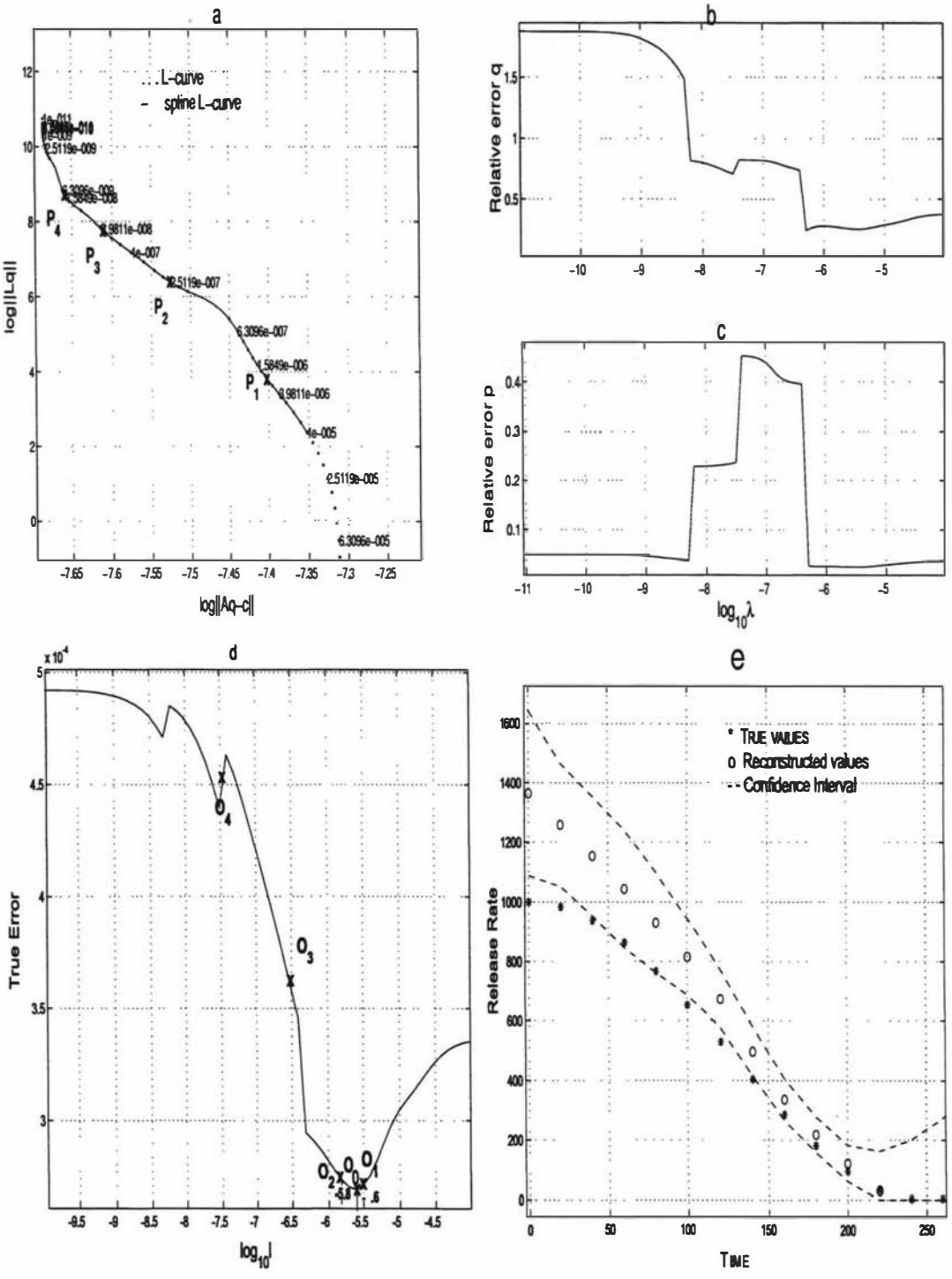

Figure 6.22: Example 3- data 3: (a) non-linear L-curve, (b) relative error in $\mathbf{q}$ *vs* regularisation parameter, (c) relative error in $\mathbf{p}$ *vs* regularisation parameter, (d) true error *vs* regularisation parameter, (e) release rates *vs* time

in the reconstructed solution. For the second set of data, P_1, P_2 and P_3 are the candidates for the optimal point on the L-curve (Figure 6.21a), and points O_1, O_2 and O_3 (Figure 6.21d) are their corresponding true errors. Point O_1 is closer to the true minimum than the other two points and therefore O_1 is our choice for the optimal point. For the third set of data, P_1, P_2, P_3 and P_4 are the four candidates for the optimal point on the L-curve (Figure 6.22a), and points O_1–O_4 (Figure 6.22d) are their corresponding true errors. Point O_1 is closer to the true minimum (O_0) and therefore the regularisation parameter that corresponds to the point P_1 is taken as the optimal value.

Important observations from these solutions are similar to the previous example, but the following differences can be observed between Examples 2 and 3:

(i) when the data are perfect, error in the reconstructed solution decreases with the increasing number of source points (decreasing the discretisation size for the source function), because the total error also contains a numerical error due to the discretisation of the integral equation (6.1). By decreasing the size of the partition of the source function the numerical error due to discretisation can be minimised;

(ii) when the data are corrupted by noise, the error in the reconstructed solution increases with the increasing discretisation parameter, because the condition number of a system matrix increases with the decrease of the partition size of the source function.

Further, Groetsh [12] has shown that increasing the discretisation parameter (i.e. decreasing the size of the partition) creates a large condition number, and the upper bound on the relative solution error increases dramatically. The condition of a system matrix is a measure of ill conditioning, and we can therefore say that the accuracy of the solution depends on the degree of ill conditioning.

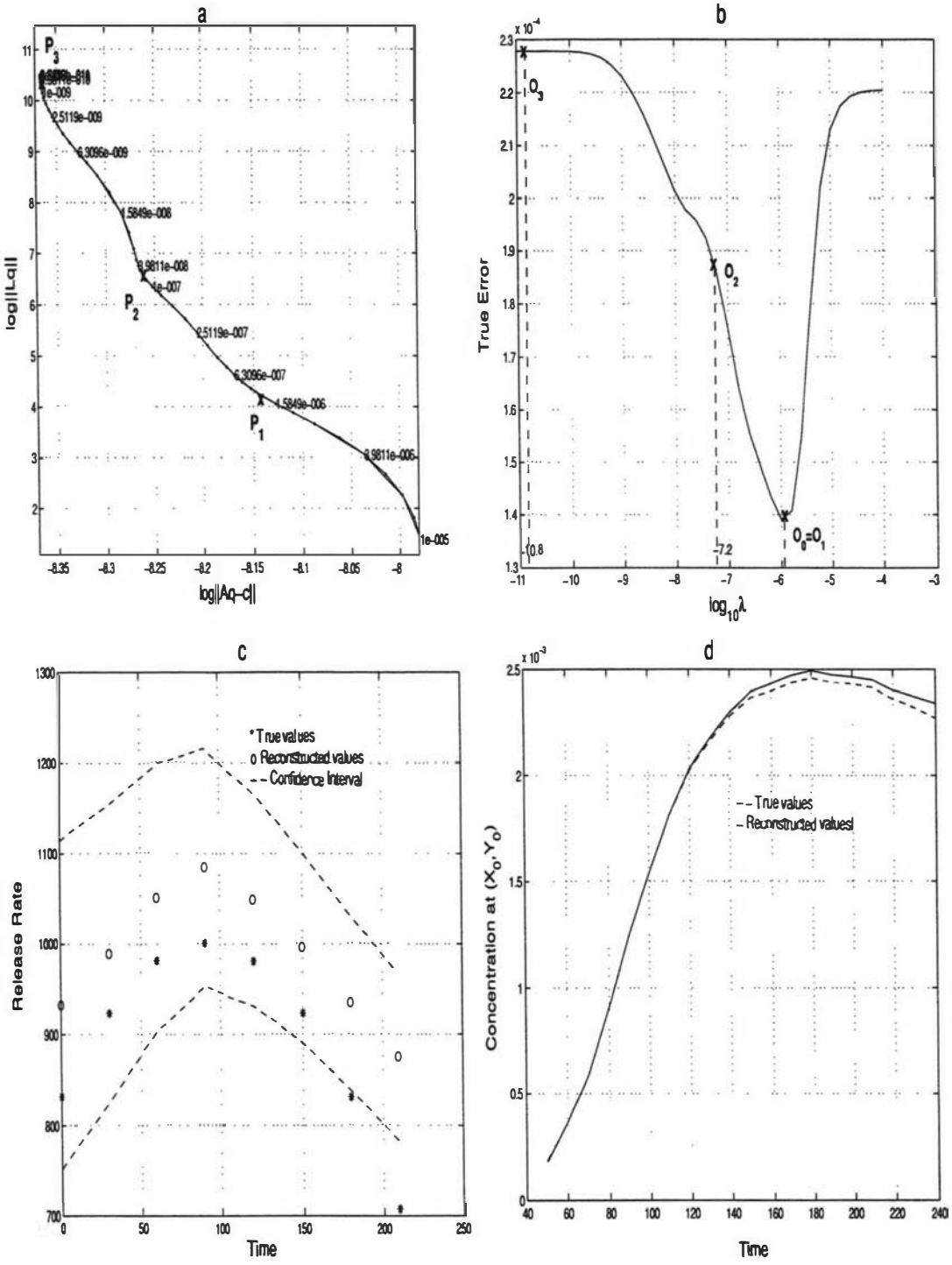

Figure 6.23: Example 4- data 1: (a) Non-linear L-curve, (b) true error *vs* Regularisation parameter, (c) release rates *vs* time, (d) concentration history at the point P=$(X_0, Y_0, 0)$

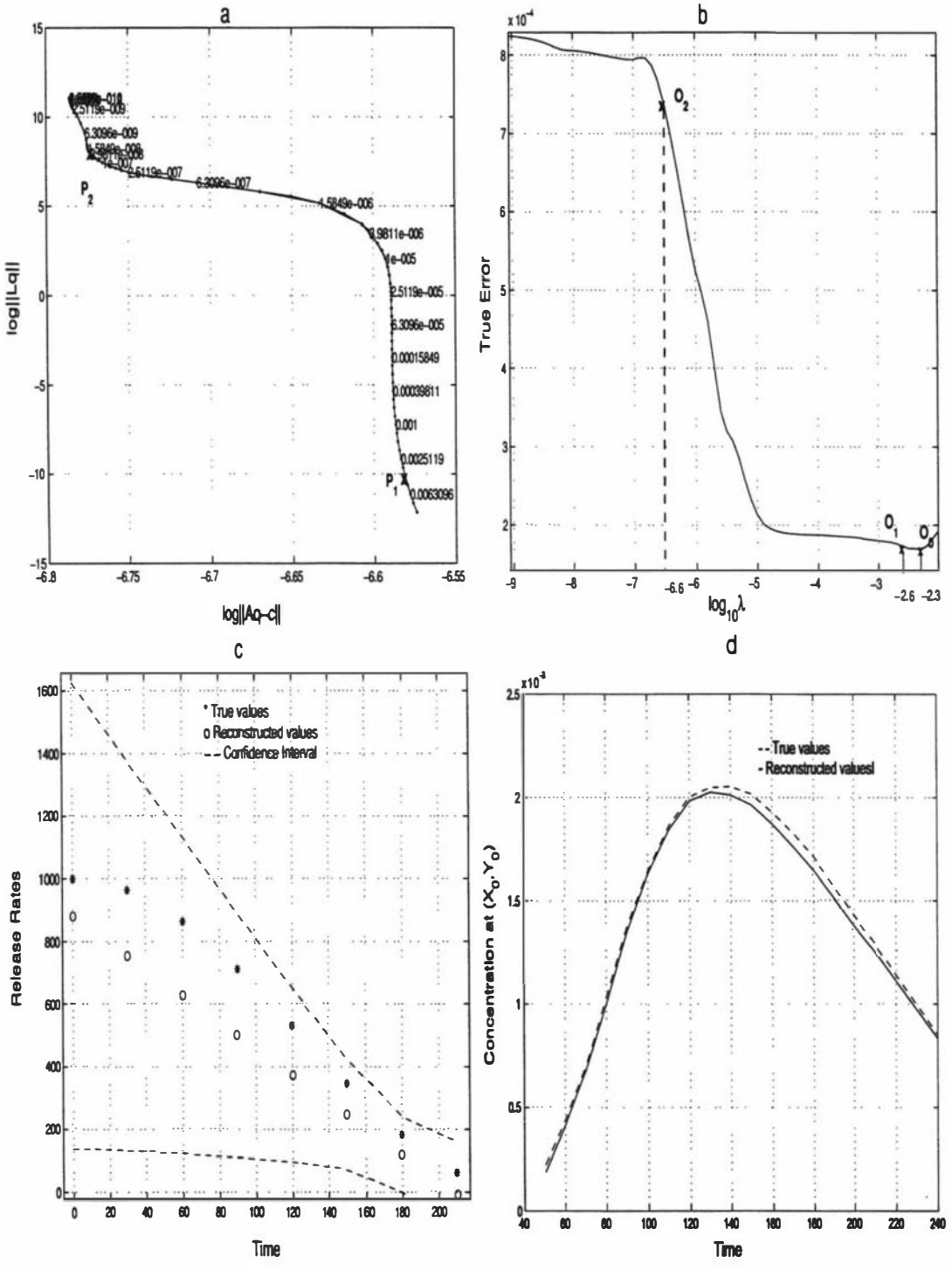

Figure 6.24: Example 4- data 2: (a) Non-linear L-curve, (b) true error *vs* regularisation parameter, (c) release rates *vs* time, (d) concentration history at the point P

Example 4: Two different source functions

In this example we consider two data sets created by using two different source functions shown by stars in the Figures 6.23c and 6.24c respectively. The results of the inverse model estimation are summarised in Table 6.5, Figures 6.23 and 6.24. Listed in the table are the true non-linear parameter (location) values along with the reconstructed values and confidence interval estimates. The L-curve, which has a lowest function value, is shown in Figures 6.23a and 6.24a for each set of data. Figures 6.23b and 6.24b depict the true error in the solution against different regularisation parameters. Figures 6.23c and 6.24c depict the true linear parameter (release rates) along with its reconstructed value and confidence interval. Figures 6.23d and 6.24d depict the true concentration history at the location P along with its reconstructed history.

For the first set of data, P_1, P_2 and P_3 are the three candidates for the optimal point on the L-curve (Figure 6.23a) and points O_1, O_2 and O_3 (Figure 6.23b) are their corresponding true errors. The error at the point P_1 is at the true minimum ($O_0 = O_1$) where both error due to regularisation and error due to noise are equal. Therefore P_1 is more suitable for an optimal point than other points. For the second set of data, P_1, P_2 are the candidates for the optimal point (Figure 6.24a) and points O_1, and O_2 (Figure 6.24b) are their corresponding true errors respectively. Point O_1 is closer to the true minimum and therefore the regularisation parameter that corresponds to point P_1 is more suitable for the problem than other points. The important observation from this example is that the errors in the reconstructed solutions are not equal even though the data are corrupted by the same size of the same random noise.

Example 5: Source at different locations

In this example we first simulate three sets of perfect data using the same source at three different locations. Then we corrupt each set of data by adding 2%, 4%

Table 6.5: Case 4: Comparison of two different source functions when the measured data is corrupted with 25% random noise

	True value	Estimated value ± Confidence interval	
		data 1	data 2
X_0	150.00	151.30 ± 7.74	161.14 ± 14.80
Y_0	25.00	26.8 ± 4.10	19.0 ± 22.6
H	12.00	12.15 ± 0.35	11.50 ± 2.24

Table 6.6: Case 5: Source at different locations

Noise in data		2	4	8
p	True value	Reconstructed vale ± Confidence interval		
X_0	800	799.67 ± 0.95	798.91 ± 1.88	795.25 ± 3.99
Y_0	100	102.86 ± 1.59	104.44 ± 2.97	100.42 ± 6.52
H	12	12.03 ± 0.23	12.17 ± 0.45	13.07 ± 0.98
		Relative error estimates in **q**		
q		0.038	0.065	0.080
p	True value	Reconstructed vale ± Confidence interval		
X_0	2800	2801.6 ± 12.96	2803.20 ± 26.10	2815.40 ± 56.63
Y_0	100	103.81 ± 20.77	106.54 ± 41.5	105.98 ± 77.83
H	12	13.91 ± 1.61	15.03 ± 3.0	14.70 ± 6.63
		Relative error estimates in **q**		
q		0.062	0.096	0.107
p	True value	Reconstructed vale ± Confidence interval		
X_0	3800	3801.23 ± 17.54	3811.20 ± 35.8	3833.31 ± 51.00
Y_0	100	103.28 ± 33.50	106.16 ± 58.30	103.90 ± 105.20
H	12	14.95 ± 3.49	19.16 ± 7.10	20.13 ± 8.30
		Relative error estimates in **q**		
q		0.155	0.216	0.256

and 8% of normally distributed relative noise. The results of the source location estimation and the relative error in the release rate estimation using these corrupted data are summarised in Table 6.6. The important observations from Table 6.6 are

(i) the confidence interval of the location estimates increases with increasing source distance for the same amount of noise,

(ii) the percentage of error in the release rates increases with increasing source distance.

In Chapter 5, we learned that the condition number of the coefficient matrix increases with the increasing distance between source and observation sites. But the condition number is proportional to the ill-conditioning of the coefficient matrix. Therefore the reason for the above two observation is due to ill-conditioning of the problem i.e. ill-conditioning of the problem increases with increasing distance between the source and the observation sites.

6.5.2 Method-3

We consider two examples to demonstrate the developed method for finding the release rate $q(t)$ and the location (X_0, Y_0, H). The purposes of this example are:

(i) to demonstrate the simultaneous estimation of parameters X_0, Y_0, H and the source release function $q(t)$;

(ii) to compare the solutions obtained using methods 2 & 3

Example 6

In this example we consider a set of data that is corrupted by 10% of random noise. The results of the source-term estimation are summarised in Table 6.7 and Figure 6.25. Listed in Table 6.7 are the true non-linear parameter (location) values along with the reconstructed values and their confidence interval estimates. The graph

of the regularisation parameter *vs* the number of calculations is shown in Figure 6.25a. This figure clearly shows that after a few calculations the regularisation parameter converges to 2.34×10^{-6}. Figures 6.25b and 6.25c depict the relative error in $\mathbf{q}$ and $\mathbf{p}$ respectively as a function of λ. Figure 6.25d depicts the true linear parameters (release rates), along with their reconstructed values and confidence interval estimation. Figure 6.25e depicts the true and reconstructed concentration history at the location P.

Source location ($\mathbf{p}$) estimations using method 2 are also listed in Table 6.7. This table also contains the optimal value of λ, the relative error in the source location estimation ($\mathbf{p}$), the relative error in the release rate ($\mathbf{q}$) estimation using each method, and a value of λ where the true error function is minimum. Since we know the true error minimum and it is therefore possible to compare the accuracy of the values obtained using methods 2 and 3. The value the regularisation parameter obtained using method 3 is closer to the value of λ where it has the lowest true error, than the value obtained using method 2. Further, the size of the confidence interval is small for method 3. Therefore, we can say that method 3 performs better than method 2. But if we consider the relative error estimates of $\mathbf{p}$ and $\mathbf{q}$, it is difficult to judge the accuracy of the methods against each other.

Table 6.7: Example 6: Comparison of estimates using Methods 2 and 3

		Method 2	Method 3
$\mathbf{p}$	True value	Estimated values $\pm$	Confidence interval
X_0	300.0	304.0 ± 14.2	305.0 ± 12.6
Y_0	50.0	47.8 ± 10.1	47.9 ± 9.7
H	12.0	11.4 ± 0.7	11.4 ± 0.7
		Regularisation parameter estimates	
λ	1.0×10^{-5}	1.58×10^{-6}	2.34×10^{-6}
		Relative error estimates	
$\mathbf{p}$		0.016	0.018
$\mathbf{q}$		0.122	0.120

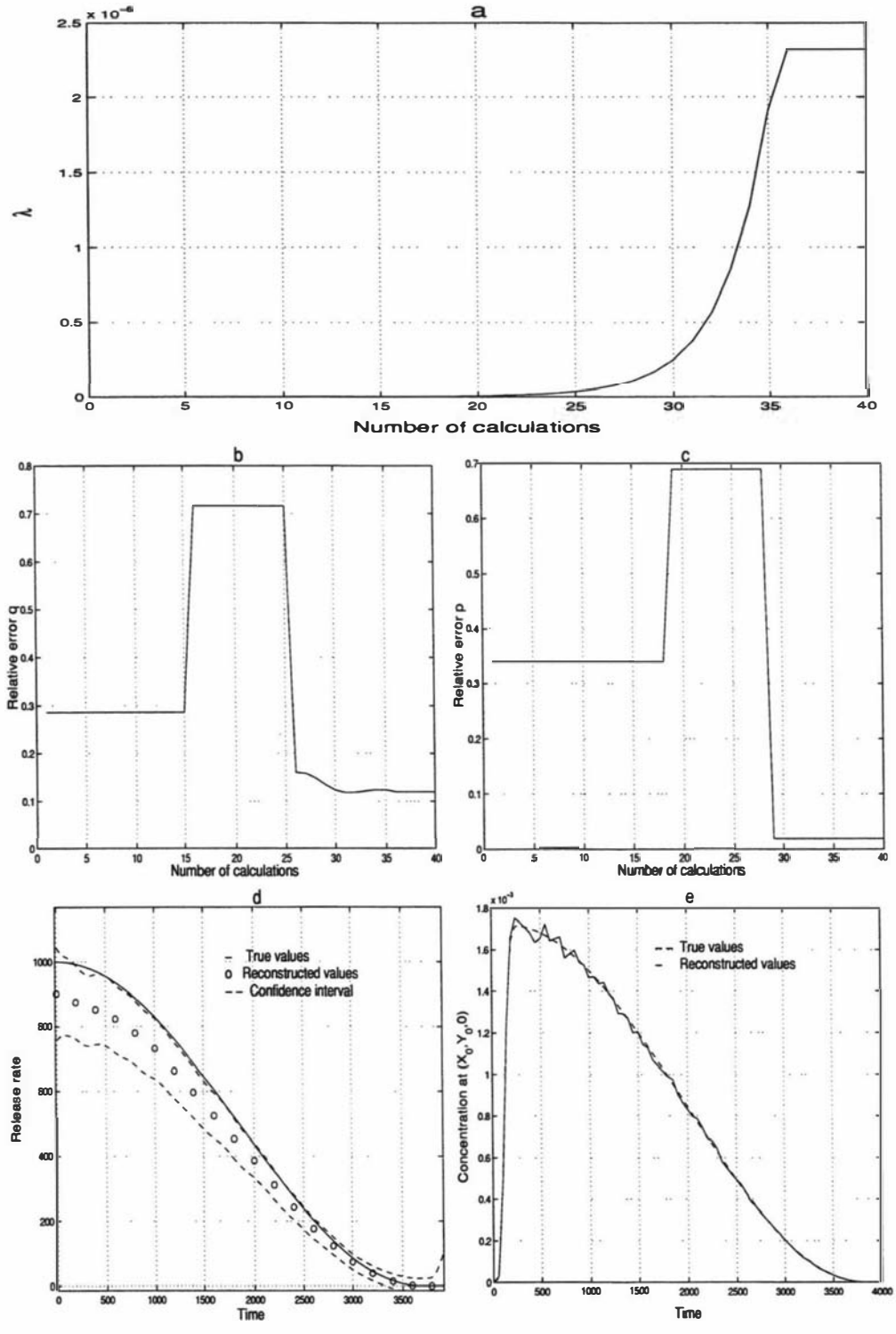

Figure 6.25: Method 3: (a) convergence of λ, (b) relative error in $\mathbf{q}$, (c) relative error in $\mathbf{p}$, (d) release rates *vs* time (e) concentration history at the point P=$(X_0, Y_0, 0)$

Example 7

In this example we consider the first two examples of section 6.5.1 to compare the relative error in the location **p** and the release rates **q** estimates using methods 2 and 3. A comparison of relative errors for these examples using these methods are given in Table 6.8. The relative error in the **p** and **q** estimates using Method 2 and Method 3 are listed in column 4 and column 5 of Table 6.8 respectively. Based on this table it is very difficult to compare the accuracies of Method 2 and 3. We can only say that both methods are in the same level of accuracy.

Table 6.8: Example 7: Comparison of relative error estimates Methods 2 & 3

Example No	Data set no	Parameter	Method 2	Method 3
Example 1	Data 1	**p**	0.088	0.120
		q	0.340	0.288
	Data 2	**p**	0.035	0.07
		q	0.51	0.40
Example 2	Data 1	**p**	0.012	0.000
		q	0.030	0.002
	Data 2	**p**	0.015	0.032
		q	0.075	0.112
	Data 3	**p**	0.038	0.034
		q	0.275	0.335
	Data 4	**p**	0.088	0.085
		q	0.526	0.340

6.6 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 only linear algebra to determine the solution. In the process, we developed three different algorithms, each of which is applied to many test cases. In all of our algorithms 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 made more efficient enormously because, after the elimination of linear parameters, only three non-linear parameters remain.

A series of examples 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:

(i) size and randomness of noise in the data,

(ii) size of discretisation of the source function,

(iii) regularisation, and

(iv) distance between source and observation sites.

From our observations it can be noticed that the major factor that 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.

We employ basically two methods (method 2 and method 3) to find the optimal value of the regularisation parameter. In method 2 we solved the problem for a sequence of λ values and then we employed three different methods described in section 6.3 to picked the optimal point on the L-curve. In method 3 we also solved the problem for a sequence of λ values but here we start from smaller λ and then decrease its value until it reaches a steady value. Both methods perform well, and one particular approach does not always predict a better result than the

other method. In our experience, therefore both methods are equally useful. If we consider only the amount of computation time, then method 3 is a better choice than method 2.

Chapter 7

Evaluation From Field Data

We now move from a simulated data environment to a field data environment. The objective of this chapter is to verify how the developed model works against field data. The data for this study are taken from a tracer gas experiment carried out by the New Zealand Meteorological Service at Upper Hutt valley, New Zealand, in 1979 [65].

7.1 Geographical Setting of Upper Hutt

The Upper Hutt Basin (Figure 7.1) is approximately 60m above sea level and is oriented West-Southwest to East-Northeast. It is approximately 12km in length and 2.5km wide. In the main part of the basin the floor slopes about 25m in 8km. To the North-West, North-East and South-West, hills rise steeply to 350-400m, the Rimutaka Range to the South-East rises eventually to over 800m, and North of Te Marua and Kaitoke the Tararua Range rises to over 1300m. The locations described are shown in Figure 7.2.

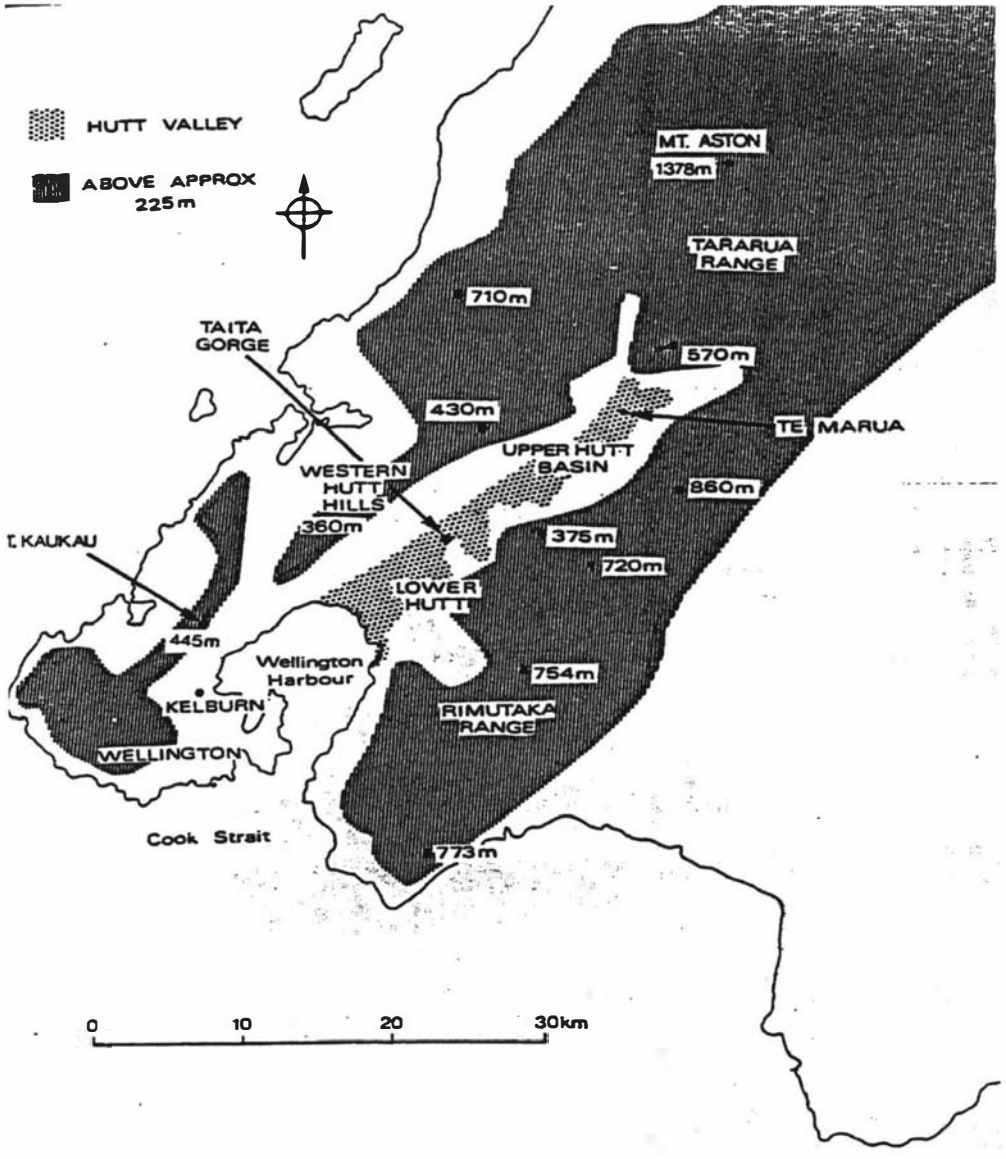

Figure 7.1: Location map of the Wellington region showing the Upper Hutt Basin and surrounding topography schematically

7.2 Tracer Gas Experiment

In the experiment sulphur hexafluoride (SF_6), used as an inert-gas tracer, was released at a steady rate of 3.7 gs^{-1} through a flow meter and pipe at about two metres above the ground. Concentration samples were collected at sites down valley from the release point, as nearly as possible on circular arcs of radius 1, 2, 3, 4,

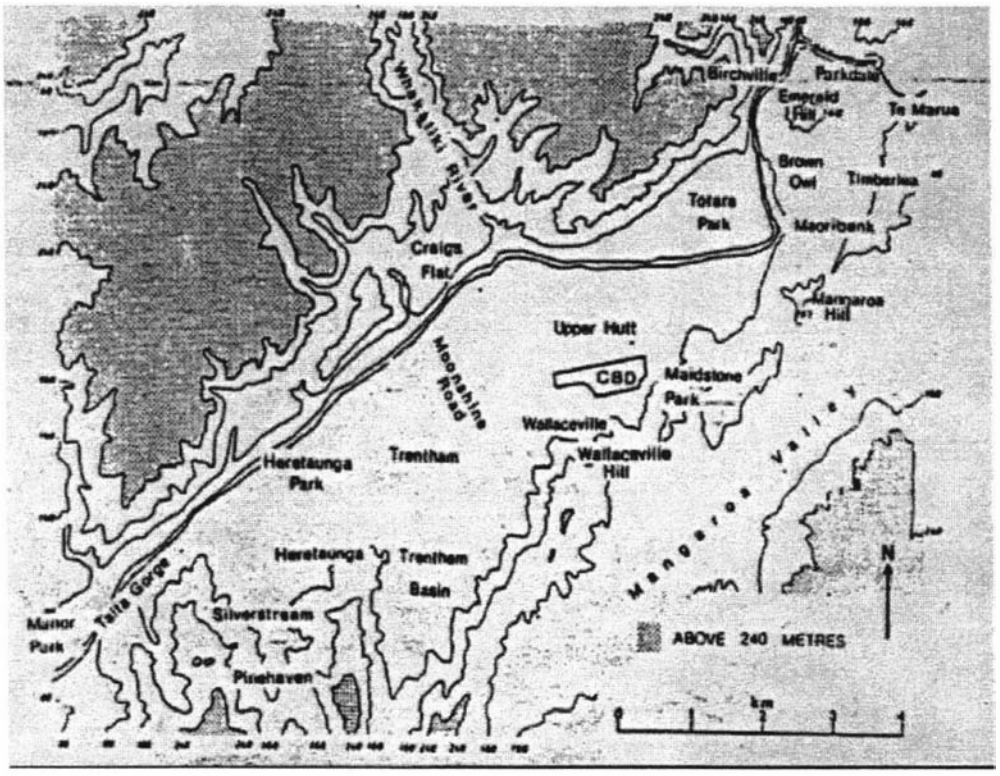

Figure 7.2: Location map of places in the Upper Hutt Basin

5, 6, 7, 8, 9 and 10 km from the release point, as shown in Figure 7.3. Sampled concentrations are given in Table 7.1. These samples were collected over a ten-minutes period, starting at the indicated times. Listed concentrations in the table are 10-minutes averages of measurements. This data set is still probably the best set of data available in New Zealand for dispersion from a ground level point source.

7.3 Estimation Of Atmospheric Parameters

For the inverse model described in this thesis it is assumed that the atmospheric parameters such as dispersion coefficients, mean wind speed and its direction are constant and known. In the case where these values are not known they can be approximated using the available observed concentration values[38]. First, in this process mean wind direction is approximated using the contour distribution

	TIME															
	5:00	5:15	5:30	5.45	6:00	6:15	6:30	6:45	7:00	7:15	7:30	7:45	8:00	8:15	8:30	8:45
1a		4		0		0		0		0						
1b		7		27				5550		4000						
1c	4	17	6620	12170	13540	10700		6930		1370		282				
1d	4	0	866	1360	507	4		4		5		66				
1e	3	4	3	4	3	4		0		0		0				
2a		6		400		3450		4380		355						
2b		8		201		842		411		1240						
2c		17		5		59		16		9						
3a		3		18	288	649		3880	5200	2810	5010	4140				
3b		7		21	76	3290		2120	2080	1130	1250	1990				
3c		5		5	6	6		628	290	249	376	163				
3d		4		5		6		7		10						
4a" (100 m)				4												
4a' (50 m)				0		51	849	2360	4510	2910	1800	1090		369		
4a			0	40		7	838	233	428	562	828	143		1040		
3x		12		835		148		1570		291						
4b			8	6		35	1450	2930		4140	3050	3020				
4c				10		9	29	148		645	535	590		343		
4d			8	10		8	6	8	11	18	111	101		7		
5a				0		0	10	17	4	3	207	165		181	230	
5b				30		14		1070		3030		3840				
6a				10		8		4		28		752				
6b				16		14	17	58	386	1450	1660	1810		2340	2350	
6c				24		19		20		1210		2000				
6d				24		18	21	24		20	90	30		69	17	
7a						8		12		86		866		1800		
8c						7		13		16		35		742		
9a							8	8	196	424	1490	1790		1670	956	
10a								8	261	780	1210	1730		1880	2000	1690

Table 7.1: Sampled concentrations

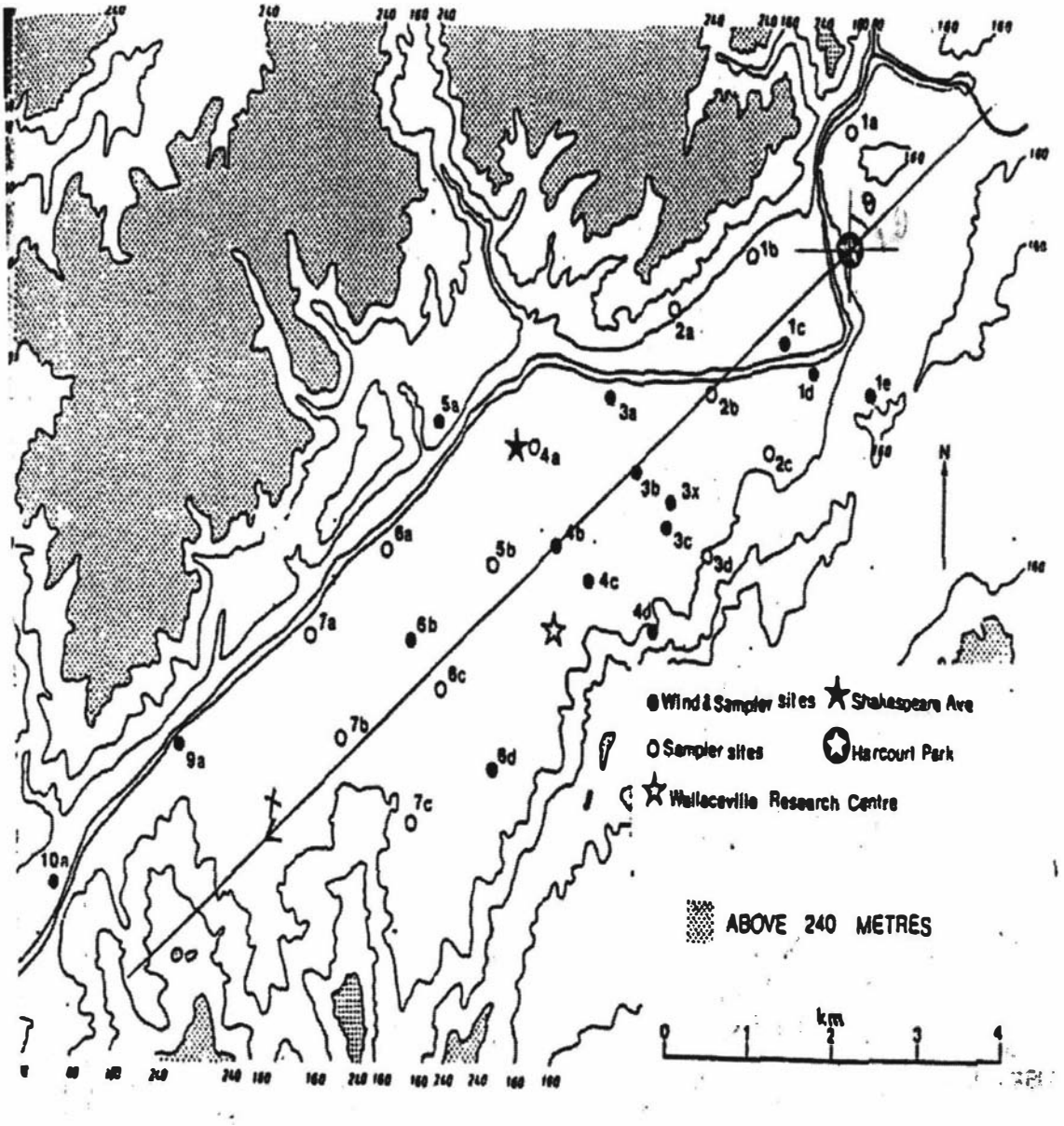

Figure 7.3: Measurement sites for the intensive survey in 1979:1a Akatarawa, 1b Tacoma Drive, 1c Kentucky Street, 1d Maoribank, 1e Mangaroa Hill, 2a Cannon Point, 2b Ebdentown Street 2c Montgomery Crescent 3a McLeod Street, 3b St Joseph's, 3c Railway Station, 3d Maidstone Park, 3x Astral Towers, 4a Shakespeare Avenue, 4b Heretaunga College, 4c Wallaceville Station, 4d Wallaceville Hill, 5a Craig's Flat, 5b Trenthan School, 6a Holdsworth Avenue, 6b Trentham Park, 6c Racecource, 6d General Motors, 7a Golf Course, 7b Heretanga Station, 7c Witako, 9a Silverstream Bridge, 10a Manor Park

graphs(Fig 42) given in [65]. It is equal to 050^o N. We then used the Cartesian co-ordinate system (X, Y, Z) with the X-axis oriented in the direction of the mean

Location	X-distance	Y-distance	Location	X-distance	Y-distance
1a	-481	-815	3x	2960	444
1b	962	-592	4b	4360	0
1c	1074	185	4c	4185	444
1d	925	592	4d	3370	2000
1e	555	1111	5a	4963	-1629
2a	1960	-777	5b	5037	-296
2b	2000	0	6a	6037	-1111
2c	1777	814	6b	6185	-333
3a	3037	-630	6c	6148	185
3b	3185	0	6d	6037	1000
3c	3074	592	7a	7222	-1074
3d	2592	1666	7b	7370	-148
4a	4185	-925	7c	7037	888
9a	9000	-1185	10a	10815	-1111

Table 7.2: Approximate distances of measurement sites from the source

wind direction, the Y-axis in the horizontal cross-wind direction, and the Z-axis in the upward vertical direction through the source location. After fixing the axis, we approximately estimated the X and Y distances of the measurement locations with respect to the origin which is 2m vertically below the source. The approximated distances are tabulated in the Table 7.2. Now we know the true release rate, and the concentration history at some known location, and therefore it is possible to approximate the values K_x, K_y, K_z and U. The approximate values of K_x, K_y, K_z and U are 10.43, 10.43, 7463 and 1.03, respectively.

7.4 Problems with inexact information

The source-term parameter estimation problem considered in this chapter is very complicated due to uncertainties in the available input data. These include:

(i) distance between the measurement locations are only an approximate values,

(ii) the parameter values such as K_x, K_y, K_z and U are not exactly known and

only an approximations are available,

(iii) the measured concentration values are 10-minutes averages, and only a few number of measurements are available ;

(iv) inaccuracies due to near by buildings and mountains. ;

(v) the main part of the Upper Hutt valley slopes roughly about 25m in 8km, and it is hard to believe the slope is steady, therefore there are likely to be variations between the height of the measurement locations.

Inaccuracies in the parameter values affect the model prediction. The effects in the model prediction due to the inaccuracy of each parameters are clearly shown in Figure 5.15 of Chapter 5. At this stage it is appropriate to simulate concentration signals at the measurements locations using the approximate parameter values and true release rates and then to compare the simulated values with the actual measurements. Figure 7.5 shows the comparison. We can see that the simulated concentrations do not agree very well with the measured values. This is because of poor standard of measured data and because the forward problem is not modelled in accordance with the topography of the experimental site. We have already discusses about the quality of the measured data. If we assume we have a good quality of data then we have to think about the forward model we use for this analysis, i.e. does the forward model include all the physical factors that contribute significantly to the data? The data we use here, is collected from the valley area where one side is surrounded by mountains and some buildings closer to the measurement locations. Therefore we have to consider all these factors when we model a forward problem. Therefore we cannot expect that the measurement data will exactly fit the forward model. As a result, $q(t)$, X_0, Y_0 and H recovered by this inverse analysis are only an estimates within the input data and regularisation errors.

7.5 The Release Rate Estimation - A Known Source Location

If we assume that the source release rate $q(t)$ at the point $(0, 0, 2)$ is unknown while all other quantities appearing in Equation (2.50) are known, our goal will be to obtain a function $q(t)$. If the location of the source is known, the release rate can be calculated using the measurement data at one location. A difficulty is that we do not have sufficient measurement data available at any one location and therefore we use concentrations $C(1c, t_j)$, $C(3b, t_j)$, $C(3c, t_j)$, $C(4b, t_j)$, $C(4c, t_j)$, $C(6b, t_j)$ at the locations 1c, 3b, 3c, 4b, 4c, 6b, respectively, along with the Equation (2.50), to estimate the release function $q(t)$. Here t is the sampling time, and $t_1 = 5.00\ hr$, $t_2 = 5.15\ hr$, ..., $t_{15} = 8.30\ hr$.

In Chapter 5 we learned that the accuracy of the release rate estimation depends on the degree of ill condition of the problem. The degree of ill condition depends on the distance between the source and receptor, number of measurements data and the source function discretisation size. In this problem the distance between the source and the receptor and the number of measurements are fixed and therefore by varying the size of the discretisation, degree of ill-conditioned can be altered. We considered the following cases to analyse the reconstructed solution:

(i) the source function is discretised into 43 points (size of 5 minutes), uniformly spaced in time from $t = 5.00\ hr$ to $t = 8.30\ hr$;

(ii) the source function is discretised into 22 points (size of 10 minutes), uniformly spaced in time from $t = 5.00\ hr$ to $t = 8.30\ hr$;

(iii) the source function is discretised into 15 points (size of 15 minutes), uniformly spaced in time from $t = 5.00\ hr$ to $t = 8.30\ hr$;

(iv) the source function is discretised into 8 points (size of 30 minutes), uniformly spaced in time from $t = 5.00\ hr$ to $t = 8.30\ hr$;

Discretisation Size	Number of source points	Relative error GCV	L-curve
5 minutes	34	0.83	0.90
10 minutes	18	0.73	0.81
15 minutes	**12**	**0.38**	**0.38**
30 minutes	6	0.42	0.42

Table 7.3: Percentage of error in release rate with number of source data points

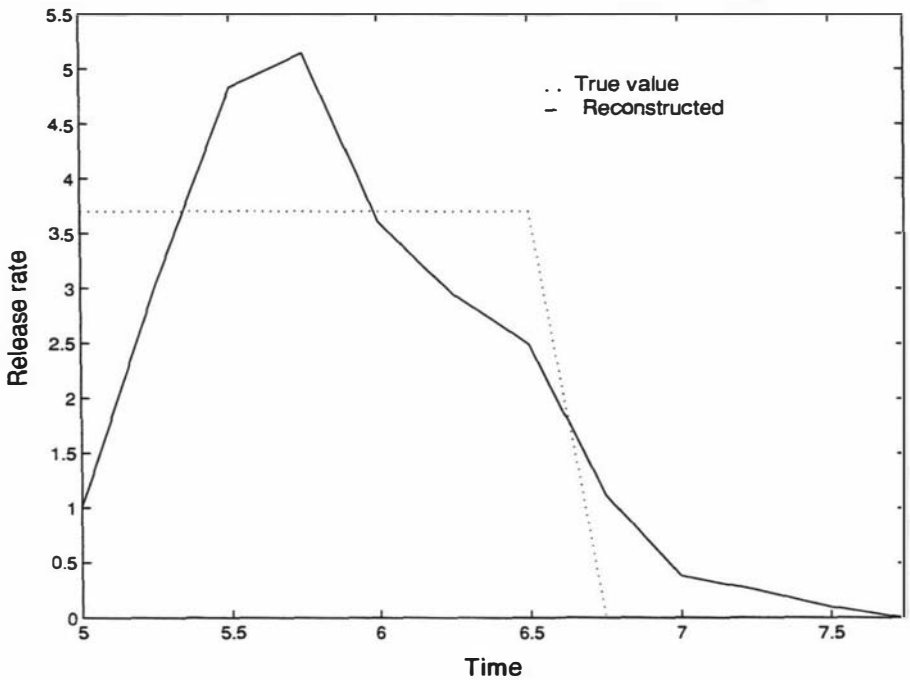

Figure 7.4: Release rate estimation

The relative error in the release rate estimation for each case is given in Table 7.3. We employ both the L-curve criterion and GCV to calculate the optimal value of the regularisation parameter. Comparisons are also given in Table 7.3. Listed in column 3 are the relative errors using the GCV criterion, and column 4, those relative errors using the L-curve criterion. Table 7.3 shows that the error in the reconstructed solution increases sharply with the increasing number of source points. This is because the condition number of the coefficient matrix increases with

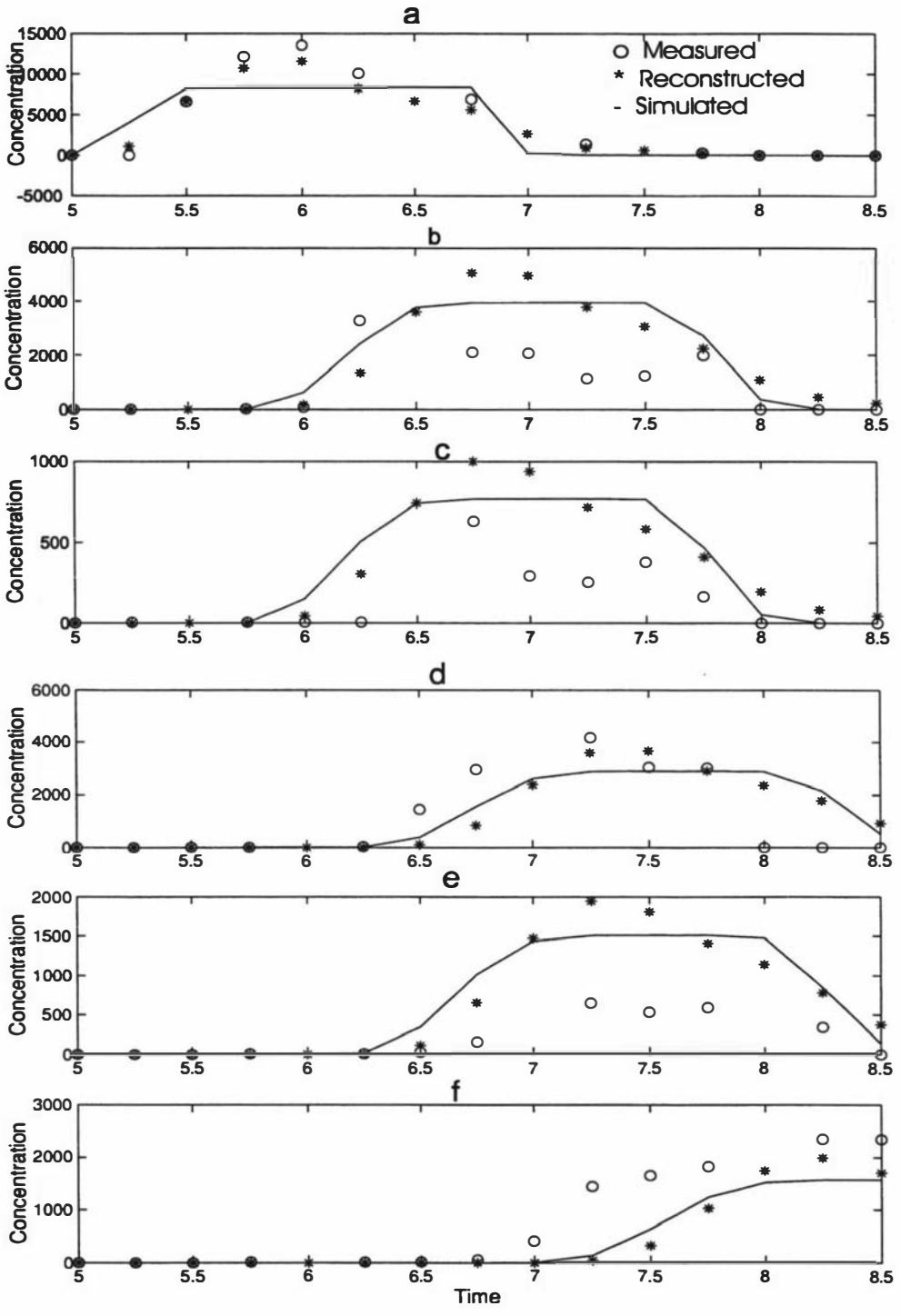

Figure 7.5: Concentration history: (a) concentration at 1c = (1074, 185), (b) concentration at 3b = (3185, 0), (c) concentration at 3c = (3074, 592), (d) concentration at 4b = (4370, 0), (e) concentration at 4c = (4185, 444), (f) concentration at 6b = (6185, -333)

the increasing number of source points and the condition number is approximately proportional to the sensitivity of the solution. But the error in the reconstructed solution also increases when the number of source points decreases. This is because of the discretisation of the integral equation (2.45). There is an optimal value of the discretisation size of the integral equation. But this is very difficult to calculate because it depends on the unavailable information on the exact data. Therefore it is recommended to take the concentration measurements as much as possible with the reasonable size of the source function discretisation. Gouveia [10] pointed out that it is possible to minimise the discretisation errors by using parameterisation of model, but in practice there will be inevitably features described by these models that are not well constrained by data.

Figure 7.4 shows the release rate estimation when the size of the source function discretisation is equal to 15 minutes, along with the true release rate. We then reconstruct the concentration history at the locations 1c, 3b, 3c, 4b, 4c and 6b using the reconstructed release rate along with (2.50). Figure 7.5 shows the comparison of measured concentrations, reconstructed concentrations, and simulated concentrations. Here we simulate concentrations using the real release rate and the forward problem (2.50).

7.6 The Release Rate Estimation - Unknown Source Location

We now consider a problem given by Equation (2.50), where the source function $q(t)$ at the point $(0, 0, H)$ and X_0, Y_0, H are unknown, while all other quantities appearing in Equation (2.50), such as U, K_x, K_y and K_z, are known. We then wish to determine the unknown source function (release rate) $q(t)$ and the location of the source (0, 0, H) relative to the monitoring point $3b = (X_0, Y_0, 0)$. To calculate the

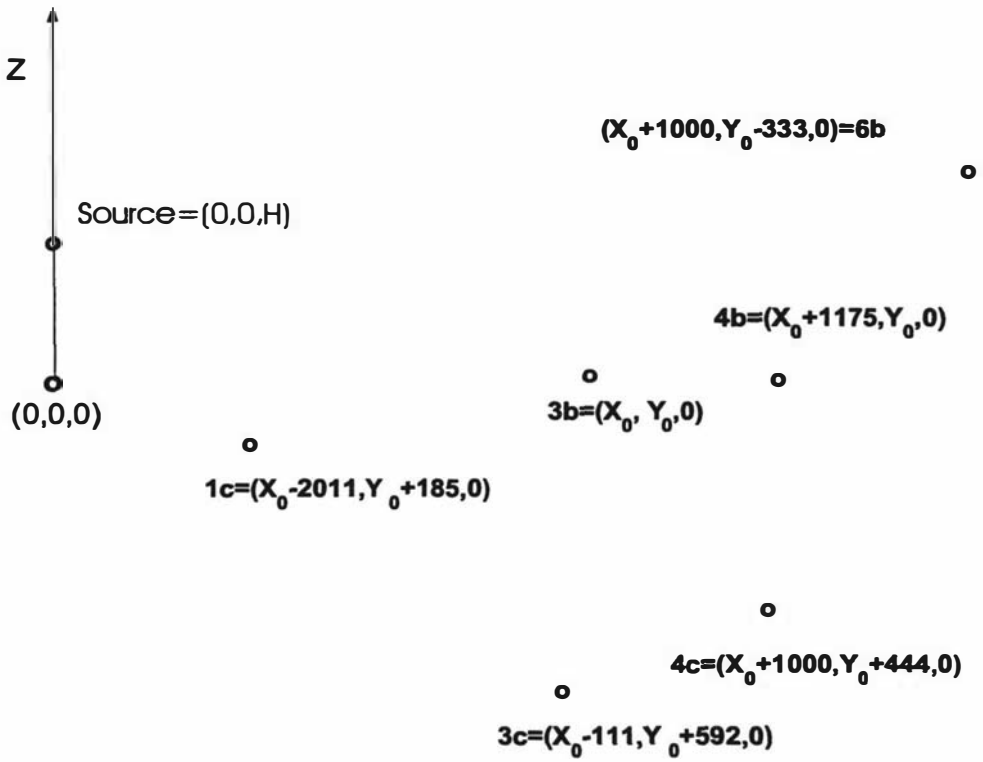

Figure 7.6: Illustration of the measurement locations

unknown values, we used the measured concentrations at the locations 1c, 3b, 3c, 4b, 4c and 6b. The coordinates of 1c, 3c, 4b, 4c and 6b with respect to the location 3b are (-2011, 185), (-111, 592), (1175,0), (1000, 444) and (3000,-333), respectively, as shown in Figure 7.6.

We employ both Method 2 and Method 3 of Chapter 6 to estimate the location parameters X_0, Y_0, H and the release rate function $q(t)$. The L-curve that has the lowest objective function value is shown in Figure 7.7. The optimal point on the L-curve is obtained by both maximum curvature and NGCV methods. Figures 7.8a and 7.8b illustrate the maximum curvature, NGCV method to estimate the regularisation parameter, where the optimal points P_1, P_2 (maximum curvature at $\lambda = 7.9433 \times 10^{-7}$ for Figure a and lowest function value at $\lambda = 6.3096 \times 10^{-8}$ for Figure b) are clearly indicated. In this case P_1 is a better choice than P_2 because of the shorter distance between the point and the origin. The graph of regularisa-

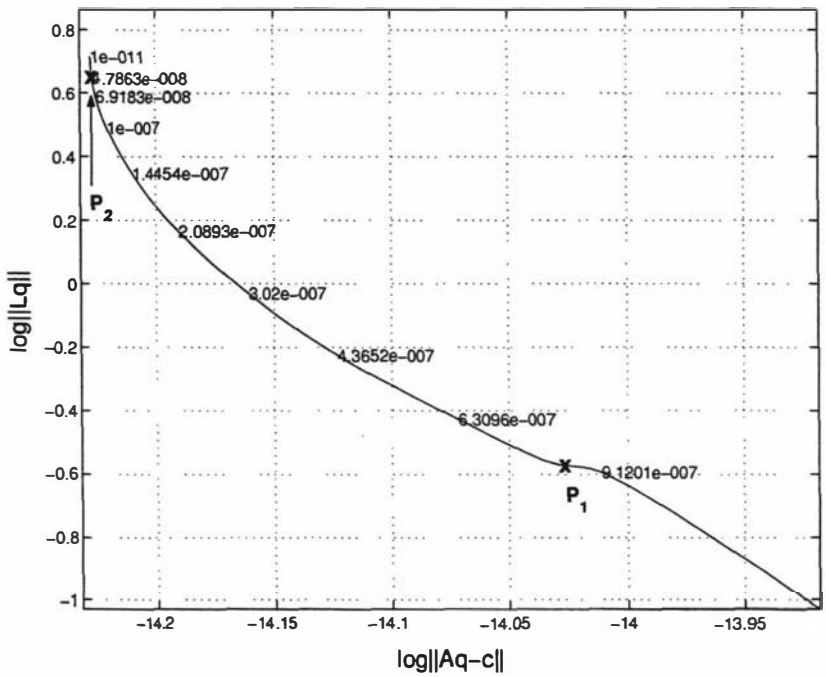

Figure 7.7: L-curve for the realdata

tion parameter *vs* number of calculations for Method 3 is shown in Figure 7.9 a. This figure clearly shows that after a few calculations the regularisation parameter converges to 1.46×10^{-7}. The estimated non-linear solutions $\mathbf{p} = [X_0, Y_0, H]$ (location) for $\lambda = 7.9433 \times 10^{-7}$ and 1.46×10^{-7} are listed in Table 7.4 along with their confidence interval estimates and their true approximates. Figures 7.9 b depicts the true linear parameters (release rates) along with their reconstructed values and confidence interval estimates for method 3.

Table 7.4: Source location estimation and confidence interval estimation

	Approximate values	Estimated values $\pm$ confidence interval	
		Method 2	Method 3
X_0	3185	2494 ± 349	2465 ± 249
Y_0	0	-95 ± 180	-61 ± 322
H	2	0.5 ± 25	0.5 ± 41

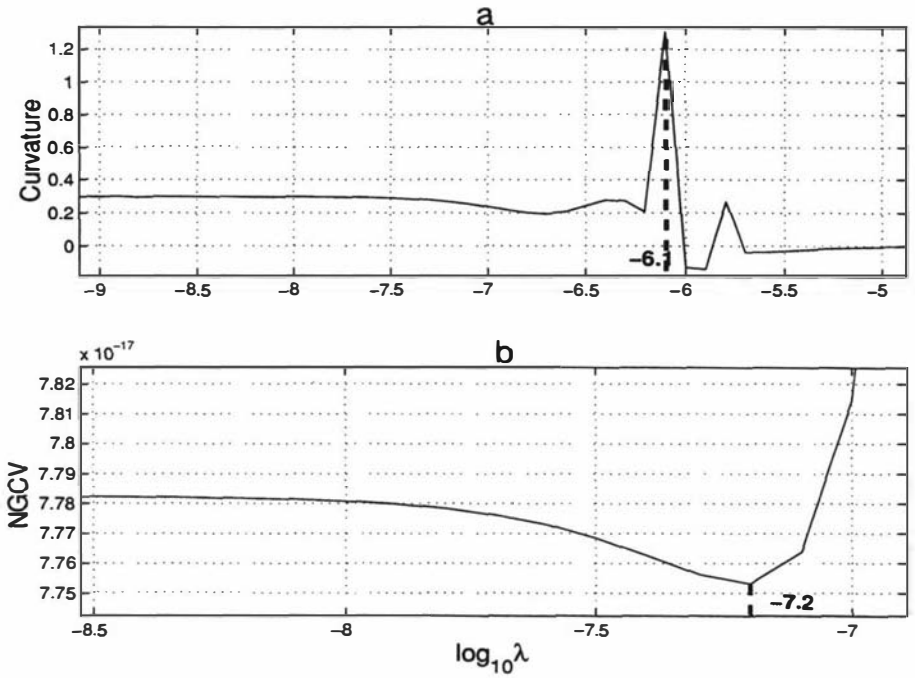

Figure 7.8: Regularisation parameter estimation- Real data (a) $\log\lambda$ *vs* curvature of the L-curve, (b) $\log\lambda$ *vs* NGCV function

7.7 Summary and Discussion

The goal of this chapter is to test an inverse model for estimating the release rate of atmospheric pollutant with real data. Two cases

(i) extended release over a period of time from a known location,

(ii) extended release over a period of time from an unknown location

are considered. The inverse model is able to recover the release rates for both cases, but is less accurate than the simulated data with noise. There may be several reasons for this inaccuracy. In this experiment samples collected at sites downstream from the release point, as nearly as possible on a circular arcs of radius 1, 2, 3,...,10. Further distance between the points on the same arc also an approximation. Secondly it has been assumed that the measurement sites are at the same height as the release point. But in reality there may be some variations. Therefore it is important to get a height of every locations with reference to the sea level and the

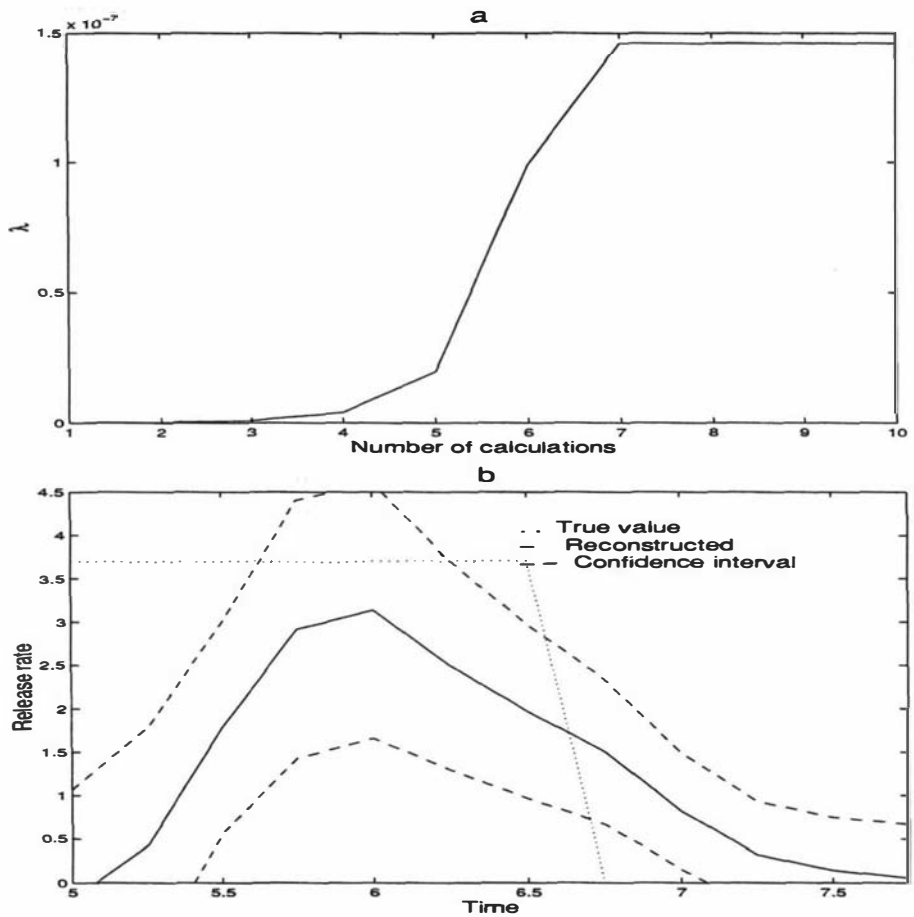

Figure 7.9: Method 2 (a)λ *vs* number of calculation, (b) release rate estimation

exact distance between all the locations have to be measured. This is one way to improve the model predictions.

Secondly the dispersion coefficients used in this study are only a rough approximation. If we know the dispersion coefficients in advance for various gases for each atmospheric stability class then these values can be used in accidental situations. These dispersion coefficients can be calculated by performing experiments using different gases at different atmospheric stability conditions. This is another way to improve the model prediction.

The concentration data used in this study are 10 minute averages. If we use the latest measurement apparatus to measure the concentration values at regular interval as much as possible then we can expect a greater accuracy in our estimation.

In this last paragraph we will give some recommendation to people trying to apply the techniques outlined in this thesis to the real word applications. These are as follows.

(a) get the height of each measurement location as accurate as possible with respect to the sea level;

(b) get the actual distance and direction between the measurement locations;

(c) measure the concentration signals at each location on a regular time interval as much as possible using the latest instruments;

(d) measure the wind speed and direction at each measurement location and use the average value for the calculation;

(e) identify the type of the gas leaked and the atmospheric stability condition and select the dispersion coefficients accordingly from your database. We mentioned earlier that it is recommended to calculate the dispersion coefficients of all the possible gases for each atmospheric stability conditions by small scale experiment in advance and store these values in the database.

Chapter 8

Final Summary and Conclusions

8.1 Summary

This thesis presents the development of an inverse model that may be used to estimate the parameters of the source term for a pollutant gas released into the atmosphere from a point source above the ground. We categorise the analysis process for the accidental release of gases from a single point source into the following:

1. The instantaneous release from a known location,

2. The instantaneous release from an unknown location,

3. The extended release over a period of time from a known location,

4. The extended release over a period of time from an unknown location.

The developments of inverse models for these cases are presented in Chapters 4, 5 and 6. The models use measured pollutant concentrations at a minimum of three observation sites on the ground, as well as meteorological data such as wind speed and percentage cloud cover. The problem where the pollutants are released instantaneously is well posed and the source term parameters are estimated with reasonable degree of accuracy. The methodology is based on nonlinear least squares,

and multiple linear regressions coupled with the solution of an advection-diffusion equation for an instantaneous point source.

But the problem where the pollutants release from the non-steady extended source is ill posed. The main difficulty with the ill-posed problem is that the measured data always have errors and consequently the solution is extremely sensitive to errors in the measurement data. Tikhonov's regularisation algorithm, which stabilises the solution process, is used to overcome the ill posedness of this problem. Here the approach taken is to develop the inverse model as a least squares minimisation problem coupled with the solution of an advection-diffusion equation for an extended point source.

For the case where the location of the source is known is a linear ill posed problem. We used the properties of L-curve and generalised cross-validation criterion from linear inverse theory to estimate the optimal value of the regularisation parameter. For the case where the location of the source is not known is a non-linear ill-posed problem. This problem is different from the linear problem in several ways. First, we can't use only linear algebra to determine the minima of the objective function. Second, the non-linear objective function may have more than one minimum for each value of the regularisation parameter. To deal with these problems, 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 often have to determine local minima for a given value of regularisation parameter. In all our algorithms exploiting the fact that some of the parameters appear in the problem linearly and hence determined using simple linear algebra for given values of nonlinear parameters and regularisation parameter. The nonlinear parameters are computed in directly in *MATLAB* as the least-squares solution of the objective function using the optimisation routine *lsqnonlin*. This is speeded up enormously because, after the elimination of linear parameters using linear algebra, only few non-linear variables remain.

To deal with the multiple minima, we found all or most of the local minimum of the objective function for a sequence regularisation parameter values and then we selected the minimum solution which has the lowest function value. We employ basically two approaches to find the optimal value of the regularisation parameter. In the first approach we solved the problem for a sequence of regularisation parameter values and then we used three different methods described in section 6.3 to pick the optimal value the regularisation parameter. In the second approach we solved the problem for a sequence of regularisation parameter values but here we start from smaller value and then slowly increase its value in conjunction with the L-curve from linear inverse theory until its steady state. In our experience, both approaches are equally useful. If we consider only the amount of computation time, then approach 2 is a better choice than approach 1.

8.2 Conclusions

The research presented in this thesis addresses the problem of finding estimates of the release rate of pollutant gas from a single point source. The main contributions towards resolving this problem are:

(i) data from at least three spatial locations are needed to estimate reliably the release rate of atmospheric pollutants where the source of the pollutant location is not known,

(ii) the model is capable of finding the release rate of atmospheric pollutants by using the measured pollutant gas concentration data at just one location on the ground where the location of the pollutant source is known,

(iii) reliable estimates are obtained even with the partial capture of the data where the pollutants are released instantaneously but accuracy of the estimates are higher for the complete sampling data,

(iv) for the cases where the pollutant is released from extended point source, the developed method is able to recover the release rate with reasonable degree of accuracy. But, if the data sampling are incomplete, then the data are unable to provide correct information about the release rate at earlier times. This is because the plume is more dispersed, and information about the plume released at the earlier times is lost. Furthermore, accuracy of the solution does not decrease much on the increasing amount of noise in the data for the known location problem (linear problem), however, if the location of the source is not known (nonlinear problem), accuracy of the solution decreases roughly at the same rate with the increasing amount of noise in the data,

(v) accuracy of the reconstructed solution depends on the following

	Instantaneous release		Extended release
(i)	size and randomness of noise in the data	(i)	size and randomness of noise in the data
(ii)	distance between the source and the receptor	(ii)	distance between the source and the receptor
(iii)	distance and angle between the receptors	(iii)	distance and angle between the receptors
(iv)	number of measurement data	(iv)	number of measurement data
		(v)	regularisation
			(a) order of regularisation
			(b) regularisation parameter selection method
		(vi)	discretisation size of the source function

Table 8.1 summarises the above in short form. Problems like the post-accident management plan, leak detection in pipelines and other installations that are used

to store gases, finding the location of gas leaks detected by monitoring stations can be addressed by means of the methods described in the thesis.

Source Type		Location		
		Known		Unknown
Instantaneous release	(i)	well-posed problem	(i)	well-posed problem
	(ii)	data at minimum of one location needed	(ii)	data at minimum of three locations needed
	(iii)	linear problem	(iii)	non-linear problem
	(iv)	source term can be estimated with partial plume	(iv)	source term can be estimated with partial plume
Extended release	(i)	ill-posed problem	(i)	ill-posed problem
	(ii)	linear problem	(ii)	non-linear problem
	(iii)	data at minimum of one location needed	(iii)	data at minimum of three locations needed
	(iv)	whole plume must be captured	(iv)	whole plume must be captured

Table 8.1: Source type characteristics

8.3 Future Research

This research establishes a foundation for future work. In particular, the following items are still requiring clarification:

(i) the accuracy of the calculated source-term parameter values depends on the distance and angle between the measurement locations. Therefore the optimal design of the locations for pollutants measurement on the ground is important. This is one area where the model could be improved.

(ii) it has been assumed throughout this thesis that dispersion coefficients K_x, K_y and K_z are known. Where these coefficients are unknown, K_x, K_y and K_z have to be treated as unknown nonlinear parameters in the minimisation objective function. This can be incorporated easily, although increasing the

number of nonlinear parameters may well increase the number of local minima a lot and it is not clear how well the current algorithm could cope with that. This could be another area to look for improving the model.

(iv) the validity of the model is tested with the simulated and some real data. The model is able to estimate the release rate and the location with reasonable degree of accuracy for the simulated data but is less accurate for the real data because of the lack of exact information and available concentration values. The validation of the model with the exact information is important and therefore the model has to be validated using a fairly good set of real data.

(iv) methods developed for the extended releases can be applied where the pollutant is from an instantaneous source. This approach is reliable if the primary aim of the problem is to estimate the concentration history of some down stream locations. However if the primary aim is to estimate the release rate of the source, it is better to use the model appropriate for the respective source releases (see Figure 5.20 for more explanation) to increase the accuracy. In order to estimate automatically the release rate of the source from the measured concentration data, it is first necessary to identify the type of source release (instantaneous or extended), before applying the methods given in this thesis. This could be yet another possibility for improving the model.

(v) In Chapter 5 we learned that the method described in this thesis unable to provide correct information about the source at earlier times if the data at earlier time is lost. Neupauer [40] pointed out in his paper that the stochastic approach using Bayesian method could identify the regions where the data give little or no information about the source history function. Therefore it is important to develop a model using the Bayesian method and its results should be compared with the model solutions described in this thesis. This could another possible research in the near future.

(vi) In our test problems we corrupted the data using multiplicative error model and assumed that the error variance are not known. Therefore, the use of least-squares with (2.32–2.33) in the thesis is a potential problem, which may be affecting some of the results. The magnitude of this effect should be studied.

Bibliography

[1] Atmadja J., Bagtzoglou A. C., Pollution source identification in heterogeneous porous media, *Water Resources Research*, 37(8), 2113–2125, 2001.

[2] Bagtzoglou A. C., Dougherty D. E., and Thompson A. F. B, Applications of particle methods to reliable identification of ground-water pollution sources, *Water Resources Research*, 26, 15–23, 1992.

[3] Beck V. J., and Arnold J. K., Parameter estimation in engineering and science, *John Wiley and Sons*, New York, 1977.

[4] Brustart W., Yeh G. T., A power wind law for turbulent transfer computations, *Water Resources Research*, 6, 1387–1391, 1970.

[5] Csanady G. T., Turbulent diffusion in the environment, *Reidel, Dordrecht*, Netherlands, 1973.

[6] 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.

[7] Engl H. W., Inverse problems and their regularisation, *Proc. of a CIME workshop on mathematics motivated by industrial problems*, 129–150, 2000.

[8] Farquharson C. G., and Oldenburg D. W., A comparison of L-curve and GCV techniques for estimating the regularisation parameter in nonlinear inverse problems, *Submitted to Geophysical Journal International*, 2002.

[9] Farquharson C. G., and Oldenburg D. W., 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.

[10] Gouveia W. P., and Scales J. A., Resolution of seismic waveform inversion: Bayes versus Occam, *Inverse Problems*, 13, 323–349, 1997.

[11] Gorelick S. M., and Remson I., Identifying sources of ground-water pollution: An optimisation approach, *Water Resources Research*, 19(3), 779–790, 1983.

[12] Groetsch C. W., Inverse problems in the mathematical sciences, *Vieweg*, Bruanschweig, Wiesbadden, 1993.

[13] Gulliksson M. E., and Wedin P. A., Analysing the nonlinear L-curve, *Submitted to SIAM Journal of Optimisation*, 1998.

[14] Gulliksson M. E., and Wedin P. A., Algorithms for using nonlinear L-curve *Submitted to SIAM Journal of Optimisation*, 1998.

[15] Hanke M., Limitations of the L-curve method in ill-posed problems, *BIT*, 36, 287–301, 1996.

[16] Hansen C. P., Rank-Deficient and discrete ill-posed problems, *SIAM, Philadelphia, PA*, 1997.

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

[18] Haber E., and Oldenburg D. W., A GCV based method for non-linear ill-posed problems, *Computational Geosciences*, 4, 41–63, 2000.

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

[20] Lukas M. A., Comparisons of parameter choice methods for regularisation with discrete noisy data, *Inverse Problems*, 14, 161–184, 1998.

[21] Kabala J. K., and Skaggs T. H., Comment on 'Minimum relative entropy inversion: Theory and application to recovering the release history of groundwater contaminant' by A. D. Woodbury and T. J. Ulrych, *Water Resources Research*, 34(8), 2077–2079, 1998.

[22] Kahane L. H., Regression basics, *Sage Publications, California*, 2001

[23] Kathirgamanathan P., McKibbin R., and McLachlan R. I., Inverse modelling for identifying the origin and release rate of atmospheric pollution-An optimisation approach, *MODSIM*, 7(1), 64–69, 2003.

[24] Kathirgamanathan P., McKibbin R., and McLachlan R. I., Source release-rate estimation of atmospheric pollution from a non-steady point source at a unknown location, Submitted to *The Journal of Environmental Modelling and Assessment*, 2002.

[25] Kathirgamanathan P., McKibbin R., and McLachlan R. I., Source release-rate estimation of atmospheric pollution from a non-steady point source at a known location, Accepted for *The Journal of Environmental Modelling and Assessment*, 2002.

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

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

[28] Kullback S., Information theory and Statistics, *John Wiley and Sons, New York*, 1989.

[29] Lee P. M ., Bayesian statistics : an introduction, *Oxford University Press, New York*, 1989.

[30] Leonard T., John S.J., Bayesian methods: an analysis for statisticians and interdisciplinary researchers , *Cambridge University Press, UK*, 1999

[31] Liu C., and Ball W. P., Application of inverse methods to contaminant source identification from aquifer diffusion profiles at Dover AFB, Delaware, *Water Resources Research*, 35(7), 1975–1985, 1999.

[32] Lukas M. A., Asymptotic optimality of generalised cross validation for choosing the regularisation parameter, *Numer. Math*, 66, 41–66, 1993.

[33] Mahar P. S., and Datta B., Identification of pollution sources in transient ground-water systems, *Water Resources Management*, 14, 209–227, 2000.

[34] Mahar P. S., and Datta B., Optimal monitoring network and ground-water pollution source identification, *Journal of Water Resources Planning and Management*, 123(4), 199–207, 1997.

[35] Morozov V. A., Methods for solving incorrectly posed problems, *Springer-Verlag, New York*, 1984

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

[37] Muniz N. B., Entropy and Tikhonov based regularisation techniques applied to the backwards heat equation, *Computers and Mathematics with Applications*, 40, 1071–1084, 2000.

[38] Murty V. V. N., and Scott V. H., Determination of transport model parameters in ground-water aquifers, *Water Resources Research*, 13(6), 941–947, 1977.

[39] Myers R. H., Classical and modern regression with applications, *Duxbury Press*, Boston : PWS-KENT, 1990.

[40] Neupauer R. M., Borchers B., Wilson J. L., Comparison of inverse methods for reconstructing the release history of a ground-water contamination source, *Water Resources Research*, 36(9), 2469–2475, 2000.

[41] Palazzi E., Faveri M. D., Fumarola G., Ferraiolo G., Diffusion from steady source of short duration, *Atmospheric Environment*, 16(12), 2785–2790, 1982.

[42] Parchevsky K. V., Using regularising algorithms for reconstruction of growth rate from experimental data, *Ecological Modelling*, 133(7), 107–115, 2000.

[43] Pasquill F., The estimation of the dispersion of wind borne material, *Meteorological Magazine*, 90, 33–49, 1961.

[44] Press S. J., Bayesian statistics: principles, models, and applications, *Wiley, New York*, 1989

[45] Rawlings J. O., Applied regression analysis: a research tool, *Wadsworth & Brooks*, 1988.

[46] Rodgers C. D., Inverse methods for atmospheric sounding; Theory and practice, *World Scientific*, 2000.

[47] Rust B. W., O'Leary D. P., Confidence intervals for discrete approximations to ill-posed problems, *Journal of Computational and Graphical Statistics*, 3(1), 67–95, 1994.

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

[49] Shum Y. S., Loveland W. D., and Hewson E. W., The use of artificial activable trace elements to monitor pollutant source strengths and dispersion patterns, *Journal of Air Pollution Control Assessment*, 25(11), 1123–1128, 1975.

[50] Skaggs T. H., and Kabala J. K., Recovering the release history of a ground water contaminant, *Hydrological Process*, 14(1), 1003–1016, 2000.

[51] Skaggs T. H., and Kabala J. K., Recovering the history of a ground water contaminant plume: method of quasi-reversibility, *Water Resources Research*, 31(11), 2669–2673, 1995.

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

[53] Snodgrass M. F., and Kitanidis P. K., A geostatistical approach to contaminant source identification. *Water Resources Research*, 33(4), 537–546, 1997.

[54] 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, 231–234, 1997.

[55] Tikhonov A. N., and Arsein V. Y., Solution of ill-posed problems, *Winston, New York*, 1977.

[56] Trefethen L. N., and Bau D., Numerical linear algebra, *SIAM*, Philadelphia, 1997.

[57] Trujillo D. M., and Busby H. R., Practical inverse analysis in engineering, *CRC Press*, Boca Raton, New York, 1997.

[58] Turner B., Workbook of atmospheric dispersion estimates: An introduction to dispersion modelling, *CRC Press - Lewis Publishers*, 1994.

[59] Vogel C. R., Non-convergence of the L-curve regularisation parameter selection method, *Inverse Problems*, 12, 535–547, 1996.

[60] Wanger B. J., Simultaneous parameter estimation and contaminant source characterisation for coupled ground-water flow and contaminant transport modelling, *Journal of Hydrology*, 135, 275–303, 1992.

[61] Whaley H., The derivation of plume dispersion parameters from measured three-dimensional data, *Atmospheric Environment*, 8, 281–290, 1974.

[62] Woodbury A. D., and Ulrych T. J., Minimum relative entropy inversion: Theory and application to recovering the release history of a ground-water contaminant. *Water Resources Research*, 32(9), 2671–2681, 1996.

[63] Woodbury A. D., and Ulrych T. J., Reply, Water Resources. Research., 34(8). *Water Resources Research*, 34(8), 2081–2084, 1998.

[64] Woodbury A. D., and Ulrych T. J., A full-Bayesian approach to the ground-water inverse problem for steady state flow. *Water Resources Research*, 36(8), 2081–2093, 2000.

[65] Wratt D. S., Salinger M. J., Clarkson T. S., Imrie B. W., Airflow and pollution dispersion in a valley, *New Zealand Meteorological Service, Scientific Report*, 4, 1984.

[66] Yeh G., and Hung C., Three-dimensional air pollutant modelling in the lower atmosphere, *Boundary-Layer Meteorology,* 9, 381–390, 1975.