

# COAL PYROLYSIS DISTRIBUTION

by

Sione Paea

A thesis  
submitted to the Victoria University of Wellington  
in fulfilment of the  
requirements for the degree of  
Master of Science  
in Mathematics.

Victoria University of Wellington  
2008

## Abstract

Coal pyrolysis is a complex process involving a large number of chemical reactions. The most accurate and up to date approach to modeling coal pyrolysis is to adopt the Distributed Activation Energy Model (DAEM) in which the reactions are assumed to consist of a set of irreversible first-order reactions that have different activation energies and a constant frequency factor. The differences in the activation energies have usually been represented by a Gaussian distribution. This thesis firstly compares the Simple First Order Reaction Model (SFOR) with the Distributed Activation Energy Model (DAEM), to explore why the DAEM may be a more appropriate approach to modeling coal pyrolysis. The second part of the thesis uses the inverse problem approach together with the smoothing function (iterative method) to provide an improved estimate of the underlying distribution in the wide distribution case of the DAEM. The present method significantly minimizes the error due to differencing and smoothes the chopped off parts on the underlying distribution curve.

# Acknowledgments

The researcher wishes to sincerely thank Associate Professor Mark McGuinness for his supervision of this study and for his vital support, guidance, inspiration and encouragement during the course of this work. Huge gratitude goes to the Arthington Davy Funds, Trinity College, Cambridge, England for their financial support. Without their help, the researcher would have not been able to undertake this study. Special thanks to Liz Richardson, Deputy Dean (Equity) and Te Ropu Awhina for the employment opportunities made available to assist the researcher financially and more importantly its contribution to developing the researcher's mentoring ability in the Faculty of Science. The researcher also wishes to thank the following people; Ruth Davidson-Toumu'a for her editorial, colleague Jonathan Crook, and Professor Matt Visser for their technical assistance with Maple. To the researcher's family, thank you for your sacrifice, support, and understanding throughout this study. Words cannot express enough gratitude for the love you have contributed over the years.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background Information on Coal Structure . . . . .	1
1.2	Pyrolysis . . . . .	3
1.3	Importance of Coal and Application . . . . .	4
1.4	Review of Previous Simplification . . . . .	5
<b>2</b>	<b>Mathematical Models of Coal Pyrolysis</b>	<b>9</b>
2.1	Existing Models . . . . .	9
2.2	Single First Order Reaction Model (SFOR) . . . . .	11
2.3	Distributed Activation Energy Model (DAEM) . . . . .	12
2.4	Comparison of the Two Models . . . . .	13
2.4.1	Evaluation of the Two Models . . . . .	17
<b>3</b>	<b>The Inverse Problem</b>	<b>19</b>
3.1	Systematic Simplifications . . . . .	19
3.1.1	The double exponential integrand . . . . .	20
3.2	The Wide Distribution Case . . . . .	26
3.2.1	Inverse Problem . . . . .	29
3.3	Numerical Analysis . . . . .	31

<b>4</b>	<b>Data Fitting</b>	<b>33</b>
4.1	Collection of data . . . . .	34
4.2	Estimation of parameters . . . . .	36
4.3	Error Description . . . . .	37
4.4	The Goodness of Fit of a Model . . . . .	38
4.5	Regression Methods . . . . .	39
4.5.1	Linear Regression Model . . . . .	39
4.5.2	Nonlinear Regression Model . . . . .	41
4.6	Determining the Least Squares Estimates . . . . .	43
4.7	Procedure of Fitting the Data . . . . .	44
<b>5</b>	<b>The Gauss-Newton Method</b>	<b>45</b>
5.1	Example 1 . . . . .	48
5.2	Step Factor . . . . .	59
<b>6</b>	<b>Method for Small-Residual Problems</b>	<b>62</b>
6.1	Levenberg-Marquardt Algorithms . . . . .	62
6.2	Data Generated with a Gaussian . . . . .	64
6.2.1	Example 2 . . . . .	64
6.3	Computer Programme . . . . .	72
6.3.1	Example 3 . . . . .	72
6.4	Digitized Data . . . . .	76
6.4.1	Example 4 . . . . .	76
6.5	Higher-Order Correction . . . . .	85
6.5.1	Example 5 . . . . .	87
6.6	Comparison of $v^{(1)}$ and $v^{(2*)}$ . . . . .	92
<b>7</b>	<b>Conclusion</b>	<b>98</b>
7.1	Recommendation . . . . .	100
7.2	Research Direction . . . . .	100
<b>8</b>	<b>Appendices</b>	<b>102</b>
8.1	Appendix 1 . . . . .	102

*CONTENTS*

iv

8.2	Appendix 2 . . . . .	103
8.3	Bibliography . . . . .	107

# List of Figures

1.1	Proposed molecular structure for bituminous coal. . . . .	2
6.1	Normalized fractional yield versus time for Example 2 . . . . .	71
6.2	Normalized fractional yield versus time for Example 3 . . . . .	74
6.3	Comparison the differences between parameter values . . . . .	75
6.4	Normalized fractional yield not yet released ( $v$ ) versus time for Example 4 . . . . .	81
6.5	Residuals of the plots in Figure 6.4 . . . . .	82
6.6	Residuals plot for digitized data for the third iteration . . . . .	84
6.7	Normalized fractional yield not yet released ( $v$ ) versus time for Example 5 . . . . .	90
6.8	Residuals of the plots in Figure 6.7 . . . . .	91
6.9	Comparison between residual results of $v^{(1)}$ and $v^{(2*)}$ . . . . .	93
6.10	Comparison between residual results versus fitted values of $v^{(1)}$ and $v^{(2*)}$ . . . . .	94
6.11	Comparisons the underlying distributions estimated inversely from values of $v$ versus time . . . . .	95

# List of Tables

6.1	The results from <i>MakeIteration</i> for Example 3 . . . . .	72
6.2	Digitized Data . . . . .	77
6.3	Data of fractional volatiles not yet released varied with time . . . . .	78
6.4	The data, fitted values and residuals for Example 4 . . . . .	79
6.5	The derivatives of $p$ , $r$ and $s$ for Example 4 . . . . .	80
6.6	The parameters $\theta^{(n)}$ when $N = 3$ for Example 4 . . . . .	80
6.7	The data, fitted values and residuals for Example 4 . . . . .	83
6.8	The data, fitted values and residuals for Example 5 . . . . .	87
6.9	The derivatives of $p_1$ , $r_1$ and $s_1$ for Example 5 . . . . .	88
6.10	The parameters $\theta^{(n)}$ when $N = 3$ for Example 5 . . . . .	89
6.11	The parameters $\theta^{(7)}$ when $N = 7$ in both fitting equations $v^{(1)}$ and $v^{(2*)}$ . . . . .	92



# Chapter 1

## Introduction

### 1.1 Background Information on Coal Structure

Coal may be described as a black, heterogeneous rock that is friable-to-hard, of variable vegetable origin with extraneous mineral additions of equally variable composition and quantity [1]. It is formed by the slow decomposition of vegetation and it originates from two distinct processes [2] about three hundred million years ago. The first process was the bacterial change in the vegetable debris before it got deeply buried (diagenesis). The other process occurs after this burial, and involves slow chemical changes due to heat and pressure (metamorphosis). The age of coal is indicated by its carbon content or rank. Coal is a complex organic polymer consisting of large polycyclic aromatic clusters of several fused rings strung together by assorted hydrocarbon chains of varying lengths and other heteroatom (O, N, S) linkages [3]. There have been substantial efforts to elucidate the molecular structure of coal, but the task is exceedingly difficult because of the variety of coal types, the heterogeneity of a single coal and the complexity of individual coal constituents [4]. The results of numerous diverse measurements have been synthesized in attempt to develop a consistent picture. Wender *et al.* [5]

proposed a structure for a highly volatile bituminous coal model molecule as shown in Figure 1.1 below.

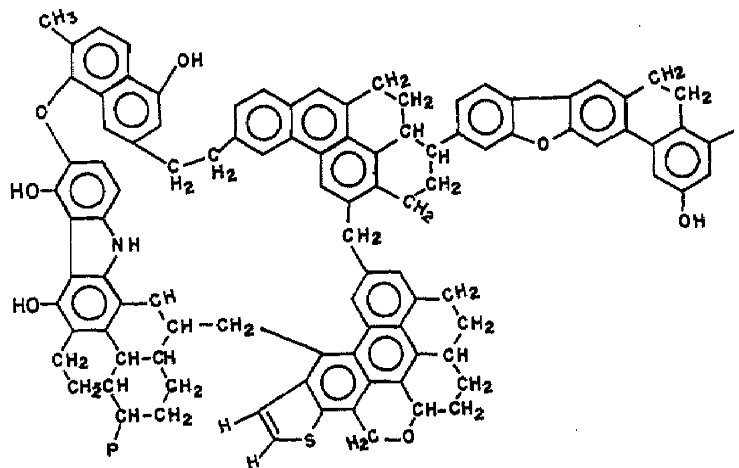


Figure 1.1: Proposed structure for a highly volatile bituminous coal model molecule following Wender *et al.* [5].

Wender *et al.* [5] discovered several characteristics of this coal molecule model. This model molecule is proposed to be part of a larger macromolecular structure and is connected by two linkages to the rest of the coal structure. The five constituent aromatic hydroaromatic structures of the model are interconnected by one aromatic ether and four aliphatic bridges. The distribution of carbon, hydrogen and oxygen atoms among different structural positions in the model molecule is discussed by Wender *et al.* [5]. The molecular structure of coal is also discussed by Van Krevelen [6], Dryden [7] and Given [8] but much work on describing the molecular structure of coal remains to be done. This model is certainly suitable for the present discussion of pyrolysis behavior. Details regarding the various functional groups present in coal have been discussed by Gavalas [9] and Howard [10].

The organic material in coal is a heterogeneous mixture of organic minerals known as macerals. Coal petrographers have identified numerous types of maceral components by microscopic analysis with reflected and transmitted light [11]. The precise chemical nature of the macerals is not well established, but their botanical origins seem to be rather well understood [6]. The various types of maceral component are often combined into three principal groups: vitrinite, exinite, and inertinite. The vitrinite is the most abundant of the three maceral groups, and usually exhibits chemical and physical properties between those of the other two [7]. The different maceral groups exhibit markedly different behaviors under pyrolysis conditions [12]. The organic matter in coal consists primarily of carbon, hydrogen, oxygen, nitrogen and sulphur, although trace quantities of other elements are found. The elemental analysis by weight of the organic constituents of different coals shows anything from 65 to 95% carbon, 2 to 7% hydrogen, up to 25% oxygen, up to 10% sulphur, and typically 1 to 2% nitrogen [1].

## 1.2 Pyrolysis

Coal can be pyrolysed by heating the solid in the absence of oxidizing compounds. Heating of coal causes its complex structure to decompose. The weaker bonds rupture at lower temperatures and the stronger ones at higher temperatures. On heating, many coals swell and subsequently contract as they pyrolyze. Internal surface area continually alters in both value and accessibility, and intrinsic reactivity changes as graphitization proceeds [1]. When coal is heated to above a certain temperature, usually between 400°C and 500°C, chemical reactions occur producing various amounts of gases, tar, and coke. These reactions in the heating process cause the coal to lose weight. The tar and gases are usually referred to as volatiles. The relative amounts of these products and their kinetics of evolution all depend on a large number of factors, such as the type of coal, the temperature-time history, the particle size and the total pressure. Thus, previous studies such as

Chermin and Van Krevelen [13] suggest the following model of three consecutive reactions during coal pyrolysis:

- i) Coal  $\xrightarrow{k_1}$  Metaplast
- ii) Metaplast  $\xrightarrow{k_2}$  Semi-coke + Primary gas
- iii) Semi-coke  $\xrightarrow{k_3}$  Coke + Secondary gas

This relatively simple model with the flexibility of three rate constants ( $k_1, k_2, k_3$ ) is used to describe the kinetics of the weight loss of coal. However this model clearly does not relate these reactions to the structure of coal or any other fundamental chemical processes. The rate constants  $k_1, k_2$  and  $k_3$  must be changed with different coal types and temperatures considered.

### 1.3 Importance of Coal and Application

It appears to be common knowledge today that petroleum and natural gas reserves have no longer kept pace with the ever-expanding energy demand in many nations, and that it will be well into the next century before nuclear power and solar energy can be expected to bridge the gap. The result is an urgent need to rely heavily on coal as a major source of energy. Indeed, as an organic fossil composed mainly of carbon and hydrogen, coal is a commodity of vast usage potential, not only in terms of its calorific value, but also for the myriad of useful chemicals that can be derived from it.

Traditionally, coal has failed to compete well with petroleum and natural gas in the energy market. This is due to solely for the economic factor of transportation and processing cost. The technology required to produce gaseous and liquid fuels and raw materials from coal existed decades ago, but nearly all commercial operations were closed down during the post-war period when large reserves of petroleum and natural gas were available. The recent surge in oil and gas prices has stimulated a thriving interest in industrial sectors developing cheaper, cleaner ways of coal utilization. Dozens of

gasification and liquefaction projects have entered the pilot plant stage, and many had reached full commercialization by 1980.

Coal has been acknowledged as the principal potential source of fuel for electrical utilities and a valuable raw material for industrial chemicals. In the United States, the Geological Survey has estimated the U.S. coal resource to be about 80% of the total fossil energy reserves in the country [4]. The energy crises of the seventies and the uncertainties associated with the supplies of petroleum have prompted the rapid development of new conversion technologies for the utilization of coal for power generation. New technology has also been developed for its conversion to synthetic gas or liquid fuels. In developing advanced technologies, three interrelated factors must be kept in view; technological feasibility, economic viability and environment constraints.

Unfortunately, coal contains many impurities like sulfur, nitrogen, sodium, potassium and other toxic impurities. To avoid environmental pollution, the emission levels of these contaminants must be kept as low as possible during combustion. Emissions must also be limited to the US Environmental Pollution Agency specified levels [4]. Most recent research on coal pyrolysis and kinetic rates is trying to improve the ways in which coal is utilized. This thesis will assist to improve the use of Distributed Activation Energy Model (DAEM) for the pyrolysis of coal and especially the use of the inverse problem in the wide distribution case.

## 1.4 Review of Previous Simplification

Considerable research on coal pyrolysis has been conducted over the years. Recent comprehensive reviews are reported predominantly in combustion and carbonization literature. Also, coal-hydrogen reactions are discussed in coal gasification and coal liquefaction literature. The more recent research,

especially that which has focused primarily on kinetic rates, includes Junggen [14], Solomon and Hamblen [15], Solomon *et al.* [16, 18] and Solomon and Serio [17]. Several past reviews of pyrolysis are available, however the best single source is probably Lowry's [19] compilation of review articles. In that work Howard [12] reviewed coal pyrolysis reactions (mostly consisting of information obtained by slow heating techniques), and several other authors reviewed subjects pertinent to pyrolysis. Jones [20] reviewed a number of pyrolysis studies but most of his work was done at low heating rates. Pyrolysis at low heating rates was also critically evaluated by Yellow [21], and work at high heating rates was summarized by Badzioch [22]. Essenhigh and Howard [23] reviewed the pyrolysis literature bearing on combustion and explosion phenomena in coal dusts. With regard to pyrolysis in the presence of hydrogen, the Institute of Gas Technology [24] published a review of its extensive coal gasification research. Anthony and Howard [3] reviewed the fundamental information on pyrolysis and hydrogasification with an emphasis on potential use of reaction models in engineering work.

Most recent literature reviews ([3], [10], [14]-[18]) of coal pyrolysis have identified numerous studies on kinetics and the amounts of total volatile yield. Some of these studies have addressed the individual volatile species and measured the kinetics of species evolution. Describing the mathematical model of coal pyrolysis is often an important part of understanding industrial processes. Modeling coal pyrolysis is important not only for improvement of combustion or gasification processes but also for the processes where coal is a part of the chemical system. Establishing a model for coal pyrolysis is also relevant to the process of thermal decomposition such as devolatilization.

Howard [10] and Solomon *et al.* [16] proposed two mathematical models of coal pyrolysis in their research. They described the two models which are Single First Order Reaction model (SFOR) and the Distributed Activation Energy Model (DAEM). From their proposed mathematical models, recent

researchers including Niksa and Lau [25], Suuberg [26], Miura and Maki [27] and McGuinness *et al.* [28] have discussed the simplification of these models. Niksa and Lau [25] claimed that holding the SFOR-base rate constant is the better way to estimate nominal rates for any given thermal history. The same devolatilization rate will be predicted as in the DAEM at every instant in the thermal history. By using this approach to explore the relationship between the DAEM and the SFOR model, the activation energy is fixed. Niksa and Lau [25] introduce an effective or *nominal* rate constant  $\langle k \rangle$  which varies with time. They also derive analytical approximations to the DAEM for testing linear or exponential temperature ramping. The resulting rapid variation of the double exponential (DExp) function is approximated by a piece-wise linear function with three regions. That is, DExp is zero, DExp is unity, and the last region is in between where it rises linearly from zero to one. This procedure can make the evaluation of the integral much easier where the initial distribution is Gaussian. Therefore it provides an accurate approximation of the full DAEM for all parameters of interest.

Niksa and Lau [25] indicated that this approximate procedure (piece-wise linear function) provides a more accurate approximation to the full DAEM for all parameters. This is a refinement of the ideas in Suuberg [26]. Suuberg used a simple step-function approximation to the double exponential (DExp) ( see also [10, 29, 30]), which jumps from zero to one at an energy that varies with time. The case of a Gaussian initial distribution in the step-function gives an error function approximation to the DAEM. This error function will be used later in this thesis as a foundation function for the development of the inverse problem in the wide distribution case. Niksa and Lau [25] note that some shortcomings remain in using their approximation at lower temperatures, particularly with the numerical solution of the equations for the position of the piece-wise linear approximation.

Miura and Maki [27] (see also [31], [32]) consider the inverse problem and present a method to estimate both  $f(E)$  and  $k_0$  from three sets of experiments performed at different heating profiles without assuming any functional forms for  $f(E)$  and  $k_0$ . They summarize a procedure of four steps for how to estimate  $f(E)$  and  $k_0$ . For a linearly-ramping temperature, they approximate DExp by a step-function because DExp changes rather steeply with  $E$  at a given temperature. Miura and Maki [27] found the rate of change of volatility with time is proportional to the distribution of volatiles ( $f$ ). They used this relationship to obtain  $f$  from the experimental measurements. McGuinness *et al.* [28] present a more accurate approximation to the double exponential (DExp), which is used in the two cases of narrow and wide distribution. A careful analysis in these two regimes is given based upon asymptotic expansions, leading to systematic methods for rapidly finding accurate approximations. The work in this thesis will build on these mathematical models.

This thesis is composed of two parts. The first compares the Simple First Order Reaction Model (SFOR) and Distributed Activation Energy Model (DAEM) in order to understand why the DAEM may be a more appropriate approach to modeling coal pyrolysis. The second part will focus on how to solve the inverse problem of finding the distribution based on the DAEM when the relative width of the initial distribution is much wider than the width of the double exponential term (DExp) of volatiles.



## Chapter 2

# Mathematical Models of Coal Pyrolysis

### 2.1 Existing Models

Mathematical models have been proposed by Howard [10], Solomon and Hamblen [15]. A model for pyrolysis using a proposed kinetic expression can almost always be applied to fit data in some limited regime of temperature and heating rate. These models suggest that for a single block of coal, the time-evolution of its constituent parts should be considered, averaged over the whole block. This approach ignores spatial variation of temperature and is appropriate for transient weight loss of pulverized coal and can be used as a component of a more complicated traveling wave model. The development here follows Howard's [10] and Solomon and Hamblen's [15] approaches. As the process of thermal decomposition of coal evolves,  $i$  denotes one particular reaction and coal's constituents are numbered with  $i= 1\dots n$ . The thermal decomposition of coal is assumed to comprise large numbers of independent chemical reactions. Large fragments of the coal molecule are present due to depolymerization and the rupture of various bonds within the coal molecule. The strength of chemical bonds depends on the coal type and rank, related to the occurrence of different reactions at various temperature intervals.  $V_i$

is the released mass fraction of volatiles corresponding to the  $i$ th constituent. Thus  $V_i^*$  is the initial mass of constituent  $i$  in the coal. The contribution to evolution by a particular reaction is described by a first order equation, so that the rate of pyrolysis is

$$\frac{dV_i}{dt} = k_i(V_i^* - V_i). \quad (2.1)$$

The proportionality constant  $k_i$  is the rate coefficient that is typically associated with temperature by an equation which is Arrhenius in form,

$$k_i = k_{0i} \exp\left(\frac{-E_i}{RT(t)}\right) \quad (2.2)$$

where  $k_{0i}$  is the pre-exponential or frequency factor in  $\text{sec}^{-1}$ ,  $E_i$  is the apparent activation energy for constituent  $i$  in J/mol,  $R$  is the ideal gas constant in (J/mol kelvins) and  $T(t)$  is the absolute temperature of the coal particle in Kelvins. Values of  $k_{0i}$ ,  $E_i$ , and  $V_i^*$  are estimated from matching with experimental data. Anthony and Howard [3] summarized a collection of experimental rate constant ( $k_i$ ) values, and the associated rate parameters and coal properties.

The solution to Equation (2.1) may be written in terms of the mass of volatiles remaining to be released at time  $t$  as

$$\frac{V_i^* - V_i}{V_i^*} = \exp\left(-\int_0^t k_i(u)du\right). \quad (2.3)$$

Then the mass of the volatiles released for one sample reaction is

$$V_i = V_i^* - V_i^* \exp\left(-\int_0^t k_i(u)du\right). \quad (2.4)$$

The model developed by Howard [10] and Solomon and Hamblen [15] has been further refined and developed specifically for SFOR and DAEM models which are briefly described in the following.

## 2.2 Single First Order Reaction Model (SFOR)

The simplest method for the description of the kinetics of the pyrolysis reactions is to use a first order reaction for overall weight loss of the volatile and for individual species evolution. The development of the mathematical models above shows that if  $i=1$  then the model is referred to as the Single First-Order Reaction Model (SFOR). Thus, the rate of pyrolysis is expressed as:

$$\frac{dV}{dt} = k(V^* - V). \quad (2.5)$$

The rate constant ( $k$ ) in Equation (2.5) is typically correlated with temperature by an Arrhenius expression:

$$k = k_0 \exp\left(\frac{-E}{RT}\right). \quad (2.6)$$

Many authors have approximated the overall process of the complex decomposition and transport phenomena involved in coal pyrolysis. They believed it to be a first order type of decomposition, occurring uniformly throughout the particle. Howard and Essenhig [33] explained their results by assuming that pyrolysis is a first order reaction with respect to the amount of undecomposed volatile matter. They assume a constant rate which is of the Arrhenius type. Badzioch and Hawksley[34], and Pitt [30] among others, have also shown that the thermal decomposition of coal occurs via first order reactions. The SFOR model is the approach based on holding the activation energy fixed and defining  $k$  in Equation (2.6) as the rate constant which varies with time,  $t \rightarrow \infty$ .

## 2.3 Distributed Activation Energy Model (DAEM)

The DAEM is one of the multi-reaction models used widely to clarify the thermal decomposition processes of coal pyrolysis. Pitt [30] assumed that the evolution of a certain substance involves an infinite number of independent chemical reactions by considering a continuous distribution of reactants. That is, many irreversible first-order parallel reactions that have different rate parameters occur simultaneously. In the DAEM model, the dependence on  $i$  is replaced by a continuous dependence on activation energy  $E$  so the values of  $k_{0i}$ ,  $E_i$  and  $V_i^*$  cannot be predicted earlier and must be estimated from the experimental data.

DAEM has been applied to represent the change in overall conversion and the change in the yield of a given component during the coal pyrolysis. The increase in the number of reactions required can cause a problem. This problem is simplified by assuming that the  $k_i$ 's differ only in activation energy so a common assumption is then to take all the pre-exponential factors,  $k_{0i}$ , to have the same value  $k_0$  for all constituents  $i$ . Then the number of reactions is large enough to permit the distribution of energy to be expressed as a distribution function  $f(E)$ , where  $f(E)$  is the distribution of activation energies, representing the differences in the activation energies of many first-order irreversible reactions. Then  $f(E)dE$  represents the fraction of the potential volatile loss  $V^*$  that has an activation energy between  $E$  and  $E+dE$ . Thus, the total amount of volatile material available for release from the coal can be written as:

$$dV^* = V^* f(E) dE \quad (2.7)$$

with the distribution function  $f(E)$  normalized to satisfy

$$\int_0^{\infty} f(E) dE = 1. \quad (2.8)$$

The solution then becomes

$$\frac{V^* - V}{V^*} = \int_0^\infty \exp\left(-\int_0^t k_0(E)e^{-E/(RT(u))} du\right) f(E)dE. \quad (2.9)$$

The SFOR and DAEM models are discussed in relation to the kinetic expressions for the pyrolysis reaction. These two models are compared in the next section.

## 2.4 Comparison of the Two Models

This section compares the two models of coal pyrolysis. From this comparison, strong evidence is created to justify which of the models is most appropriate for modeling coal pyrolysis. The SFOR model is followed in Howard [10] and Saxena's [4] work. The values of  $V^*$ ,  $E$  and  $k_0$  are determined experimentally. Details of the different experimental methods are discussed by Anthony and Howard [3]. The coal is heated so that temperature increases at a constant rate  $\frac{dT}{dt} = \text{constant} = m$  where  $m > 0$ . Then integrate this uniform heating rate with Equations (2.5) and (2.6) in the SFOR model as follows:

$$\int_0^V \frac{dV}{V^* - V} = \int_0^T \frac{k_0}{m} \exp\left(\frac{-E}{RT}\right) dT. \quad (2.10)$$

Since  $E/RT \gg 1$  is a good approximation for pyrolysis reactions, the solution to Equation (2.10) can be approximated as

$$\frac{V^* - V}{V^*} = \exp\left(\frac{-k_0 RT^2}{mE} \exp\left(\frac{-E}{RT}\right)\right). \quad (2.11)$$

This approximation applies to the solutions in Equations (2.3) and (2.4).

The distribution curve  $f(E)$  is generally assumed to be a Gaussian distribution in the DAEM and Equation (2.9). Within this distribution, the mean

activation energy  $E_0$  and standard deviation  $\sigma$  are both determined by the experimental data. Thus

$$f(E) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(E - E_0)^2}{2\sigma^2}. \quad (2.12)$$

Equations (2.9) and (2.12) provide the solution for the DAEM model as follows:

$$\frac{V^* - V}{V^*} = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \exp \left[ -k_0 \int_0^t \exp \left( \frac{-E}{RT} \right) dt - \frac{(E - E_0)^2}{2\sigma^2} \right] dE. \quad (2.13)$$

Equation (2.13) permits correlation of coal decomposition data using four parameters ( $V^*$ ,  $E_0$ ,  $\sigma$ ,  $k_0$ ) and is applicable to a non isothermal process [35] where the distribution is Gaussian.

Comparing the two models, three parameters,  $k_0$ ,  $E_0$ , and  $\sigma$  are required in addition to  $V^*$  for the DAEM model. However for the SFOR model, only two parameters, frequency factor and activation energy are required for analysis. In other word DAEM requires only one additional parameter,  $\sigma$ , from SFOR model but it is applicable to the description of thermal decomposition processes with different heating rates [10, 18]. Niksa and Lau [25] have explored the relationship between the DAEM and the SFOR models with an approach based on holding the activation energy fixed and defining an effective or nominal rate constant  $\langle k \rangle$ , which varies with time as

$$\frac{dV}{dt} = \langle k \rangle (V^* - V). \quad (2.14)$$

The nominal rates for any given thermal history can be estimated from the SFOR-based rate constant that predicts the same devolatilization rate as the DAEM at every instant in the thermal history. Niksa and Lau [25] note that there is a large variation in  $\langle k \rangle$  with time or temperature, and also a more modest variation with coal rank. They also derive analytical approximations to the DAEM for temperatures undergoing linear or exponential ramping.

Their approach is based on exploiting the rapid changes occurring in the double exponential (DExp). In Equation (2.9), the integrand consists of the product of the double exponential term

$$\text{DExp} \equiv \exp \left( - \int_0^t k_0(E) e^{-E/(RT(u))} du \right),$$

and a term representing the distribution of activation energy  $f(E)$ . Niksa and Lau [25] noted that if  $E/RT \gg 1$ , and the temperature ramps as  $T = mt$ , then

$$\int_0^t k_0(E) e^{-E/(RT(u))} du \sim \frac{k_0 RT^2}{mE} \exp \left( \frac{-E}{RT} \right). \quad (2.15)$$

Gunes and Gunes [36] discussed the influences of various parameters on the numerical solution of the nonisothermal DAEM Equation (2.13), while Brown [37] undertook a detailed review of the effect of various parameters on the SFOR model. In the SFOR model, Howard [10] plotted Equation (2.11) using various activation energies and it clearly illustrated the nature of any parameter changes. The comparison here is focused mainly on the influence of these parameters: heating rate ( $m$ ), activation energy ( $E$ ) and the pre-exponential factor ( $k_0$ ). Zsako [38] has also done this for the SFOR model. Although this is not a very realistic model, it is often assumed to apply as a first approximation. Both models used various numerical values of parameters to explore the effects of the changes in parameters.

Brown [37], Gunes and Gunes [36] and Zsako [38] examined the influence of each parameter on the numerical solution of both models. The first parameter is the influence of heating rate ( $m$ ) on both models. Their results illustrate that the DAEM and SFOR models show that remaining mass fraction curves are shifted up the temperature scale by an increase in the heating rate. The second parameter is the influence of mean activation energy ( $E_0$ ). Both models show similar influences. When  $E_0$  values increased it causes the

curves to shift towards the right. By comparing the curves of both models to Howard's [10] curve of data on the total yield of volatiles, it seems that the DAEM curve gives a more realistic result, due to the influence of standard deviation ( $\sigma$ ). The third parameter is the pre-exponential factor ( $k_0$ ). The DAEM and the SFOR models show that an increase in  $k_0$  value causes the curves to shift toward the left. The effect of each of the three parameters is to cause the curves to shift up the temperature scale.

A model such as the Single First-Order Reaction model (SFOR) is strictly applicable only to homogeneous systems in which decomposition from the source  $V^*$  is due to a single chemical process which occurs with a single activation energy  $E$ . Solomon *et al.* [18] has shown that for the devolatilization of coal, the values of  $k_0$  and  $E$  determined for one heating rate are not appropriate when used for another heating rate. Some other simple models were used also and, like the single first order model, they were applicable only under limited experimental conditions [10]. Howard [10] plotted Equation (2.9) using what might be called reasonable parameters in fitting the data to the total yield of volatile. The resulting graph showed Howard some inadequacies in the single reaction model. In an attempt to improve the SFOR model, Howard specified that the activation energy and pre-exponential factor must be very low to approximately fit the temperature dependence that results from the occurrence of different reactions at different temperature intervals. However, he still concluded that the SFOR model was inadequate for dealing with the complexities of coal pyrolysis.

When modeling industrial processes where large particles or lumps of coal are involved, heat transfer cannot be neglected and the change of temperature cannot be described with one uniform heating rate. In general, the change of temperature in solids is modelled by the transport partial differential equation, which may be nonlinear. In other words more complicated reactions like coal pyrolysis cannot be adequately modelled by a single reaction, due



to large variations in the value of  $k_0$  and  $E_0$  with the heating rate  $\frac{dT}{dt}$  [18]. Therefore, researchers realized that the SFOR model could only be applied to limited experimental conditions and that they needed a model which could be applied to more complex experimental conditions of coal pyrolysis. They then moved to a more complicated model such as the Distributed Activation Energy Model (DAEM) model. It is adapted from Vand's [29] treatment of independent parallel processes in modeling the resistance of metallic films. DAEM has proved very successful in describing the pyrolysis of various coal types under differing temperature histories.

The DAEM can also be applied to explain the thermal decomposition processes of the pyrolysis of coal and other materials, including biomass, residual oils, resin chars [39] and kerogen [40]. Anthony *et al.* [35] and Howard [10] described the DAEM as applicable over a wide range of thermal conditions. The model was originally developed to predict volatile yields during rapid pyrolysis of coal but has also been applied at the relatively low heating rate encountered during thermal decomposition of coal to coke (Merrick, [41]). DAEM is the simplest model that depicts devolatilization rates during transient heating over a broad range of heating rates. It is also the only formalism in devolatilization modeling that captures the observed density of reaction time scales for this process [10].

### 2.4.1 Evaluation of the Two Models

The two models have their shortcomings for application to the pyrolysis of coal. According to Howard [10], the most serious problem of Equation (2.5) and (2.6) in the SFOR model is the apparently asymptotic yield of volatiles that is observed after some time at the final temperature. As a result, the apparent value of  $V^*$  as a function of final temperature is mechanistically inconsistent with the equations and is mathematically unamenable. Howard also plotted Arrhenius plots for rate constants from the work of different researchers and labelled the different graphs using time zones in some of the

correlations. The relatively slow rate of weight loss observed after extended times at a given temperature requires a set of parameter values that differ markedly from those that fit the behavior of the graph over short time. Howard [10] clearly stated that coal pyrolysis is not a single reaction, but rather a multiplicity of overlapping decompositions concentrated in different time intervals for isothermal pyrolysis, or in different time and temperature intervals for the usual case of pyrolysis during heatup. He concluded that any one set of parameter values for these equations cannot be expected to represent data accurately over a wide range of conditions. The SFOR model could only be applied in limited conditions and in this way was problematic.

The problems were then solved by applying the DAEM model to these limited conditions. However, the main difficulty with the DAEM solution is a complicated double integral which can require significant computing resources, particularly when it needs evaluating many times [42]. Miura [31] also discussed in his work that the DAEM model has two major weak points. The first is the assumption of a constant  $k_0$  value for all reactions. The other is the assignment of the Gaussian distribution to  $f(E)$ . It is possible to estimate  $f(E)$  from experimental data without assuming the Gaussian distribution as performed by Vand [29]. However, in order to use the Gaussian distribution, a constant value must be assigned to  $k_0$  beforehand in order to estimate  $f(E)$ . Miura [31] presented a simple method to overcome this problem. The simple method was applied to estimate  $f(E)$  and  $k_0(E)$  from three sets of experimental data without any assumption on the functional form of  $f(E)$  or  $k_0$  for the pyrolysis of three kinds of coal.

In all of the recent research, the DAEM has been described as a more powerful model for evaluating the complex experimental conditions of coal pyrolysis. Researchers show that the shortcomings of the SFOR model can be solved by the use of the DAEM. The DAEM is generally recognized to be the most appropriate approach to model coal pyrolysis.

# Chapter 3

## The Inverse Problem

### 3.1 Systematic Simplifications

This section builds on the work of McGuinness *et al.* [28]. It follows closely their fundamental concepts, but further develops their formulae. Their work was based on an evaluation of the Distributed Activation Energy Model (DAEM) for the pyrolysis of coal. From Equation (2.9) of the DAEM, the integrand consists of the product of two major parts. First, the double exponential (DExp) term is determined by the temperature during the experiment and depends on time through the temperature history experienced by the sample. The other term representing the initial distribution  $f(E)$  is determined by the type of coal being considered and it is independent of time, and depends on the distribution of volatiles in the sample. So Equation (2.9) can be re-written as

$$v = \int_0^{\infty} (\text{DExp})(f(E))dE, \quad (3.1)$$

where  $v = 1 - (V/V^*)$  is a fraction of the volatile yield not yet released. Another way of writing the double integral in Equations (2.9) and (3.1) is in

the form

$$\begin{aligned} v &= \int_0^\infty \exp \left( \left( - \int_0^t k_0(E) e^{-E/(RT(u))} du \right) \exp(\ln(f(E))) \right) dE \\ &= \int_0^\infty \exp \left( - \int_0^t k_0(E) e^{-E/(RT(u))} du \right) + \ln(f(E)) dE. \end{aligned}$$

The effect of constant temperature  $T(u) = T_0$  on DExp is discussed first. It is followed by an investigation of ramping temperature histories.

### 3.1.1 The double exponential integrand

The equation of the double exponential term is shown as

$$\text{DExp} \equiv \exp \left( - \int_0^t k_0 e^{-E/(RT(u))} du \right). \quad (3.2)$$

McGuinness *et al.* [28] developed a more accurate approximation to DExp where  $T(u)$  is specified and  $E$  can take any positive value. Their approach was similar to the work presented by Niksa and Lau [25] but uses more systematic methods and a more accurate approximation. McGuinness *et al.* [28] discussed the typical values of the parameters (including the range of  $k_0$ , activation energy region and temperature range) and functions needed to motivate the systematic simplifications of the DExp integrand.

Most of the researchers agree that the integral of DExp is particularly simple when temperature is constant,  $T(u) = T_0$ . Equation (3.2) is integrated with respect to the dummy variable  $u$  from zero to  $t$ . Then the solution becomes

$$\text{DExp} \equiv \exp \left( -tk_0 e^{-E/RT_0} \right). \quad (3.3)$$

When both of the parameters  $E/RT_0$  and  $tk_0$  from Equation (3.3) are given large values, the DExp function changes rather steeply with  $E$ . This is further

illustrated if Equation (3.3) is re-written in the form of

$$\exp\left(-\exp\left(\frac{E_s - E}{E_w}\right)\right), \quad (3.4)$$

where

$$-tk_0e^{-E/RT_0} \approx \left(-\exp\left(\frac{E_s - E}{E_w}\right)\right).$$

Solving Equation (3.3) in the form of Equation (3.4) gives  $E_s \equiv RT_0 \ln(tk_0)$  and  $E_w \equiv RT_0$ . Evaluating Equation (3.4) in relation to the three regions of piece-wise linear function shows that when  $E_s$  is much greater than  $E$ , the DExp function is nearly zero. When  $E_s$  is much less than  $E$ , the DExp function is nearly unity. The DExp function changes rapidly from zero to one in a range of  $E$  values close to  $E_s$ . The rapid change of the DExp function can be seen by using the typical values of approximately  $E_w \approx 10$  kJ/mol and  $E_s \approx 100$  kJ/mole to show how steep the change is.

The integral in the DExp function holds the same idea for more complicated time history such as when temperature is taken to ramp linearly with  $T = mt$ . However, doing so produces an equation much more difficult to evaluate. Equation (3.2) can be re-written with  $T = mt$ , which becomes

$$\text{DExp} \equiv \exp\left(-\int_0^t k_0 e^{-E/Rmu} du\right). \quad (3.5)$$

In Equation (3.5), the integral inside the exponential function can be approximated using the conventional Laplace transform approach. The parameter  $E/(Rmt)$  is assumed to be large. Evaluating the integral of the DExp function produces the following asymptotic solution.

$$\exp\left(-\int_0^t k_0 e^{-E/Rmu} du\right) \sim \exp\left(-\frac{k_0 Rmt^2}{E} e^{-E/Rmt}\right) \quad \text{as} \quad \frac{E}{Rmt} \rightarrow \infty. \quad (3.6)$$

Equations (3.6) and (2.15) have the same form as Equation (2.11) when  $T=mt$ . These equations are exactly the same as the equation resulting from the p-function presented by Miura [31]. Equation (3.6) can be rewritten in the form of Equation (3.4):

$$\exp\left(-\exp\left(\frac{E_s - E}{E_w}\right)\right),$$

where

$$\left(-\frac{k_0 R m t^2}{E} e^{-E/Rmt}\right) \approx -\exp\left(\frac{E_s - E}{E_w}\right).$$

This function has the same nature as described in Equation (3.4) if the parameters  $E_s$  and  $E_w$  stand by themselves and are not evaluated. Therefore when  $E$  is increased over a range of size  $E_w$  around  $E_s$ , the function is changed rapidly from zero to one. It can be approximated as follows. Let

$$g(E) = \frac{E_s - E}{E_w};$$

then the solution of Equation (3.6) can be written as

$$\exp\left(-\frac{k_0 R m t^2}{E} e^{-E/Rmt}\right) = \exp(-\exp(g(E))),$$

where

$$g(E) = -\frac{E}{Rmt} + \ln\left(\frac{k_0 R m t^2}{E}\right).$$

Since only the behavior near  $E_s$  is of interest, this function is expanded in a Taylor series as follows in order to allow the parameters  $E_s$  and  $E_w$  to be identified:

$$g(E) \sim g(E_s) + (E - E_s)g'(E_s) + (E - E_s)^2 g''(E_s)/2 + (E - E_s)^3 g'''(E_s)/6 + \dots \quad (3.7)$$

Comparing Equation (3.7) and the definition of  $g(E)$  gives the values of

$g(E_s) = 0$  and  $g'(E_s) = -1/E_w$ , hence:

$$g(E_s) \equiv -\frac{E_s}{Rmt} + \ln\left(\frac{k_0 Rmt^2}{E_s}\right) = 0 \quad (3.8)$$

and

$$g'(E_s) \equiv -\frac{1}{Rmt} - \frac{1}{E_s} = -\frac{1}{E_w}. \quad (3.9)$$

Solving and simplifying Equation (3.8) and (3.9) gives the solutions  $E_s = RmtY(k_0t)$  and  $E_w = \frac{RmtE_s}{Rmt+E_s}$  where  $Y(x)$  is the LambertW function considered to be the one real root of the equation

$$Ye^Y = x. \quad (3.10)$$

Writing Equation (3.8) in the form of Equation (3.10) produces

$$\frac{E_s}{Rmt} e^{E_s/(Rmt)} = k_0t.$$

The LambertW function ( $Y(x)$ ) is used often in this thesis in order to develop the formula of the inverse problem. So it is useful to understand the approximations to  $Y(x)$  for small and large  $x$  in correspondence with the short and long time periods [43].

$$Y \sim x - x^2, \quad x \ll 1,$$

and

$$Y \sim \ln\left(\frac{x}{\ln\left(\frac{x}{\ln x}\right)}\right), \quad x \gg 1.$$

In Equation (2.9), the total integrand is the product of the double exponential function (DExp) and initial distribution ( $f(E)$ ). The DExp has been described as a smooth step-function which changes rapidly from zero to one. This rapid change is due to the large size of  $tk_0$  in a range of activation energies of width  $E_w$  around the value  $E = E_s$ , with  $E_s$  and  $E_w$  varying with time.

The  $f(E)$  is presumed to be a Gaussian distribution with  $\sigma$  as the standard deviation of the distribution and  $E_0$  as the mean activation energy. Both parameters are treated as constant values. McGuinness *et al.* [28] discussed the two cases of distribution depending on the relationship between DExp and  $f(E)$ . The difference between the two cases is in the width of DExp, with the first case having a relatively wide initial distribution compared to the width of DExp. The second case is where  $f(E)$  is relatively narrow compared to DExp. The shape of the total integrand changes with time depending on which limit applies.

This thesis focuses on the wide distribution case, where the initial distribution  $f(E)$  is much wider than DExp. When the initial distribution is wider than the width  $E_w$ , the total integrand is initially the distribution  $f(E)$ . However, the left side of the distribution progresses in sharp jumps rather than smoothly. This ragged step-like effect is due to DExp as time proceeds. The location of the maximum of the total integrand can move significantly, and the shape becomes quite skewed. The right side of the total integrand is closer to normal than the left side. This problem is evaluated as originating in the nature of DExp which is approximated by the step-function (see also [10], [29], [30], [26]).

To demonstrate the approach, the initial distribution ( $f(E)$ ) is taken to be a Gaussian distribution with mean activation energy  $E_0$  and standard deviation  $\sigma$ . Approximation is sought to the integral of Equation (2.13) as:

$$v = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \exp \left[ -k_0 \int_0^t \exp \left( \frac{-E}{RT} \right) dt - \frac{(E - E_0)^2}{2\sigma^2} \right] dE.$$

Using

$$\left( - \int_0^t k_0 e^{-E/RT} dt \right) \sim - \exp \left( \frac{E_s - E}{E_w} \right)$$

which this is then re-written as

$$v = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \exp(h(E)) dE,$$



where

$$h(E) = -\exp\left(\frac{E_s - E}{E_w}\right) - \frac{(E - E_0)^2}{2\sigma^2}.$$

The energy is rescaled as  $y = E/E_0$ , so the problem becomes

$$v = \sqrt{\frac{\alpha}{\pi}} \int_0^\infty \exp(h(y)) dy \quad (3.11)$$

where

$$h(y) = -\exp\left(\frac{y_s - y}{y_w}\right) - \alpha(y - 1)^2, \quad (3.12)$$

and the constant parameter  $\alpha = \frac{1}{2} \left(\frac{E_0}{\sigma}\right)^2$ . Note that in practice  $\alpha \gg 1$ . The rescaling of the energy helps to smooth the graph of the DAEM solution in the Maple. It also helps to minimize the time consumed during the plotting of the DAEM graph. Consider the two special cases of temperature history by using the rescaled parameters  $y$  and  $\tau = k_0 t$ .

(i) For a constant temperature where  $T = T_0$ ,

$$y_s = \frac{RT_0}{E_0} \ln \tau, \quad y_w = \frac{RT_0}{E_0}.$$

(ii) For a linear ramping temperature where  $T = mt$ ,

$$y_s = \frac{Rm\tau}{k_0 E_0} Y(\tau), \quad y_w = \frac{y_s}{1 + Y(\tau)}.$$

These equations can help to solve for  $t$  when the LambertW function is considered in some complex equations. Note that  $y_s = \frac{E_s}{E_0}$  and  $y_w = \frac{E_w}{E_0}$  and  $Y$  is the LambertW function defined above. The linear ramping temperature is used in the next section in order to improve the approach where the initial distribution is much wider than DExp. Also studied is the approximation of Equation (3.11).

## 3.2 The Wide Distribution Case

The following sections concentrate on the case where the initial distribution ( $f(E)$ ) is much wider than the double exponential (DExp). As discussed in Section 1.4, the DExp is approximated as a smoothed step-function. It rises rapidly from zero to one in a range of activation energies of width  $E_w$  around the value  $E = E_s$ , where  $E_s$  and  $E_w$  vary with time. This case considered the limit  $y_w \sqrt{\alpha} \ll 1$  and the step-function is defined to be

$$H(y - y_s) = \begin{cases} 0 & , \text{if } y < y_s \\ 1 & , \text{if } y \geq y_s. \end{cases}$$

The following approach improves upon this, and upon the linear ramp approximation discussed in the last section and in Equation (3.12). Equation (3.11) becomes:

$$v = \sqrt{\frac{\alpha}{\pi}} \int_0^\infty \exp \left[ -\exp \left( \frac{y_s - y}{y_w} \right) - \alpha(y - 1)^2 \right] dy$$

and is rewritten in the form where it includes the step-function and a term that can be written using the *complementary* error function erfc.

$$v = \sqrt{\frac{\alpha}{\pi}} \int_0^\infty \left[ \exp \left( -\exp \left( \frac{y_s - y}{y_w} \right) \right) - H(y - y_s) \right] \exp(-\alpha(y - 1)^2) dy + \sqrt{\frac{\alpha}{\pi}} \int_{y_s}^\infty \exp(-\alpha(y - 1)^2) dy. \quad (3.13)$$

The integrand of the first integral in Equation (3.13) is the product of the initial distribution and a function which is very small everywhere except in a neighborhood of size  $y_w$  around the point  $y = y_s$ . The initial distribution term can be expanded as a Taylor series about  $y = y_s$ . Hence, the first integral

gives

$$\begin{aligned} & \sqrt{\frac{\alpha}{\pi}} \int_0^\infty \left[ \exp\left(-\exp\left(\frac{y_s - y}{y_w}\right)\right) - H(y - y_s) \right] \exp(-\alpha(y - 1)^2) dy \\ &= \sqrt{\frac{\alpha}{\pi}} \int_0^\infty \left[ \exp\left(-\exp\left(\frac{y_s - y}{y_w}\right)\right) - H(y - y_s) \right] (1 - (y - y_s) \\ & \quad 2\alpha(y_s - 1) + \dots) \exp(-\alpha(y_s - 1)^2) dy. \end{aligned} \quad (3.14)$$

The right hand side of Equation (3.14) can be re-written as

$$\sqrt{\frac{\alpha}{\pi}} y_w e^{-\alpha(y_s - 1)^2} \int_{-\infty}^\infty \left[ e^{-e^{-x}} - H(x) \right] (1 - (y - y_s) 2\alpha(y_s - 1) + \dots) dx$$

and that integrating from  $-\infty$  is an approximation. Then the solution becomes

$$\begin{aligned} & \sqrt{\frac{\alpha}{\pi}} \int_0^\infty \left[ \exp\left(-\exp\left(\frac{y_s - y}{y_w}\right)\right) - H(y - y_s) \right] \exp(-\alpha(y - 1)^2) dy = \\ & \sqrt{\frac{\alpha}{\pi}} y_w e^{-\alpha(y_s - 1)^2} \left[ A_0 - 2\alpha y_w (y_s - 1) A_1 + \alpha y_w^2 \{2\alpha(y_s - 1)^2 - 1\} A_2 \right. \\ & \quad \left. + \frac{2}{3} y_w^3 \alpha^2 \{2(y_s - 1) + 2\alpha(y_s - 1)^3 + 1\} A_3 \right], \end{aligned}$$

where

$$A_i \equiv \int_{-\infty}^\infty x^i (e^{-e^{-x}} - H(x)) dx \quad i = 0, 1, 2, \dots$$

remain to be evaluated. Note that  $A_i$  are independent of any parameters and need to be evaluated only once. The first few values are evaluated as:

$$A_0 \approx -0.5772, A_1 \approx -0.98906, A_2 \approx -1.81496, A_3 \approx -5.89037.$$

Most of the previous simplifications of the step-function approximations used only the conventional error function as a dominant function and ignored the first integral. The case of a Gaussian initial distribution in the step-function gives an error function approximation to the DAEM illustrated in

the second integral of Equation 3.13. The integral can be put in the form of the conventional error function is evaluated as

$$\frac{1}{\sqrt{\pi}} \int_{\sqrt{\alpha}(y_s-1)}^{\infty} e^{-u^2} du.$$

The solution is straightforward to compute now. Where  $\text{erfc}$  is the complementary error function then

$$\sqrt{\frac{\alpha}{\pi}} \int_0^{\infty} \exp(-\alpha(y-1)^2) dy \sim \frac{1}{2} \text{erfc}(\sqrt{\alpha}(y_s-1)).$$

Combining these results gives an approximation to the solution of Equation (3.13).

$$v \sim \frac{1}{2} \text{erfc}(\sqrt{\alpha}(y_s-1)) + \sqrt{\frac{\alpha}{\pi}} y_w e^{-\alpha(y_s-1)^2} \left[ A_0 + 2\alpha y_w (y_s-1) A_1 + \alpha y_w^2 \{2\alpha(y_s-1)^2 - 1\} A_2 + \frac{2}{3} y_w^3 \alpha^2 \{2(y_s-1) + 2\alpha(y_s-1)^3 + 1\} A_3 \right].$$

This expansion is only valid when  $\alpha y_w (y_s-1) \ll 1$  and in the limit  $y_w \sqrt{\alpha} \rightarrow 0$ . This is when the initial distribution ( $f(E)$ ) is much wider than the width  $y_w$  of DExp. The asymptotic approach gives the general result

$$v \sim \int_{y_s}^{\infty} f(y) dy + y_w A_0 f(y_s) + y_w^2 A_1 f'(y_s) + y_w^3 A_2 f''(y_s)/2 + y_w^4 A_3 f'''(y_s)/6 + y_w^5 A_4 f^{iv}(y_s)/24 + y_w^6 A_5 f^v(y_s)/120. \quad (3.15)$$

Equation (3.15) is used to develop the idea of how to evaluate the wide distribution case by applying the inverse problem. The symbol  $'$  is used to indicate the derivative.

### 3.2.1 Inverse Problem

The rate of volatilization from the general result in Equation (3.15) is rewritten in non-dimensional form that gives a procedure for considering the inverse problem with greater accuracy.

$$\begin{aligned} \frac{dv}{d\tau} \sim & -f(y_s) \frac{dy_s}{d\tau} + y_w A_0 f'(y_s) \frac{dy_s}{d\tau} + A_0 f(y_s) \frac{dy_w}{d\tau} + y_w^2 A_1 f''(y_s) \frac{dy_s}{d\tau} \\ & + 2A_1 y_w f'(y_s) \frac{dy_w}{d\tau} + \frac{1}{2} y_w^3 A_2 f'''(y_s) \frac{dy_s}{d\tau} + \frac{3}{2} y_w^2 A_2 f'''(y_s) \frac{dy_w}{d\tau} \\ & + \frac{1}{6} y_w^4 A_3 f^{iv}(y_s) \frac{dy_s}{d\tau} + \frac{2}{3} y_w^3 A_3 f'''(y_s) \frac{dy_w}{d\tau} + \frac{5}{24} y_w^4 A_4 f^{iv}(y_s) \frac{dy_w}{d\tau}. \end{aligned} \quad (3.16)$$

The coefficients of  $\frac{dy_s}{d\tau}$  and  $\frac{dy_w}{d\tau}$  in Equation (3.16) are put together, then the rate of volatilization becomes:

$$\begin{aligned} \frac{dv}{d\tau} \sim & [-f(y_s) + y_w A_0 f'(y_s) + y_w^2 A_1 f''(y_s) + \frac{1}{2} y_w^3 A_2 f'''(y_s) \\ & + \frac{1}{6} y_w^4 A_3 f^{iv}(y_s)] \frac{dy_s}{d\tau} + [A_0 f(y_s) + 2A_1 y_w f'(y_s) + \frac{3}{2} y_w^2 A_2 f''(y_s) \\ & + \frac{2}{3} y_w^3 A_3 f'''(y_s) + \frac{5}{24} y_w^4 A_4 f^{iv}(y_s)] \frac{dy_w}{d\tau}. \end{aligned} \quad (3.17)$$

In this case,  $E_0$  is hard to identify beforehand, however Equation (3.16) is best to evaluate in the dimensional form in order to overcome the difficulties with greater accuracy. Then the rate of volatilization is rewritten as:

$$\begin{aligned} \frac{dv}{dt} \sim & \left( -\frac{dE_s}{dt} + A_0 \frac{dE_w}{dt} \right) f(E_s) + \left( A_0 E_w \frac{dE_s}{dt} + 2A_1 E_w \frac{dE_w}{dt} \right) f'(E_s) + \\ & \left( A_1 E_w^2 \frac{dE_s}{dt} + \frac{3}{2} A_2 E_w^2 \frac{dE_w}{dt} \right) f''(E_s) + \left( \frac{1}{2} A_2 E_w^3 \frac{dE_s}{dt} + \frac{2}{3} A_3 E_w^3 \frac{dE_w}{dt} \right) \\ & f'''(E_s) + \left( \frac{1}{6} A_3 E_w^4 \frac{dE_s}{dt} + \frac{5}{24} A_4 E_w^4 \frac{dE_w}{dt} \right) f^{iv}(E_s). \end{aligned}$$

Note that  $f(E_s)$  is still needs to be determined whereas  $E_s$  and  $E_w$  are known functions depending on  $t$  (provided that  $k_0$  is known). Niksa and Lau [25]

discussed replacing the dependence on  $t$  with dependence on  $E_s$ , by inverting  $E_s(t)$ . In order to facilitate calculation, the rate of volatilization have been expanded in terms of  $E_w = \epsilon e_w$ , where  $\epsilon \ll 1$ . The rate of volatilization then becomes

$$\begin{aligned}
\frac{dv}{dt} \sim & -\frac{dE_s}{dt} f_0 - \epsilon \frac{dE_s}{dt} f_1 - \epsilon^2 \frac{dE_s}{dt} f_2 - \epsilon^3 \frac{dE_s}{dt} f_3 - \dots + \\
& \epsilon A_0 \frac{de_w}{dt} f_0 + \epsilon^2 A_0 \frac{de_w}{dt} f_1 + \epsilon^3 A_0 \frac{de_w}{dt} f_2 + \dots + \\
& \epsilon e_w A_0 \frac{dE_s}{dt} f'_0 + \epsilon^2 e_w A_0 \frac{dE_s}{dt} f'_1 + \epsilon^3 e_w A_0 \frac{dE_s}{dt} f'_2 + \dots + \\
& 2\epsilon^2 e_w A_1 \frac{de_w}{dt} f'_0 + 2\epsilon^3 e_w A_1 \frac{de_w}{dt} f'_1 + \dots + \\
& \epsilon^2 e_w^2 A_1 \frac{dE_s}{dt} f''_0 + \epsilon^3 e_w^2 A_1 \frac{dE_s}{dt} f''_1 + \dots + \\
& \frac{3}{2} \epsilon^3 e_w^2 A_2 \frac{de_w}{dt} f''_0 + \dots + \\
& \frac{1}{2} \epsilon^3 e_w^3 A_2 \frac{dE_s}{dt} f'''_0 + \dots
\end{aligned} \tag{3.18}$$

Usually, when higher derivatives are multiplied by small parameters, the problem requires singular perturbation techniques, and consideration of boundary layers. However, the boundary conditions  $f \rightarrow 0$  when  $E_s \rightarrow 0$  or  $\infty$  are satisfied by the zeroth-order solution and regular series expansion techniques provide a simple way to approximate  $f$ . Hence  $f$  is expanded as a power series in  $\epsilon$ ,  $f \sim f_0 + \epsilon f_1 + \dots$ , and coefficients of power of  $\epsilon$  in Equation (3.18) are equated to obtain these results. The first result is the leading-order term

$$f_0 = -\frac{dv/dt}{dE_s/dt}, \tag{3.19}$$

which provides an estimate of the underlying distribution. The result obtained by adding the higher-order correction term

$$\epsilon f_1 = A_0 \left( E_w \frac{df_0}{dE_s} + \frac{\frac{dE_w}{dt} f_0}{\frac{dE_s}{dt}} \right), \tag{3.20}$$

which is non Gaussian to the leading-order term  $f_0$  does provide a more accurate estimate of the underlying distribution in the DAEM than the leading-order itself. The zeroth-order approximation  $f_0$  is the same as the result obtained by Miura and Maki [27]. In addition to Equations (3.19) and (3.20) we provide an improved formula by calculating the higher-order terms  $\epsilon^2 f_2$  and  $\epsilon^3 f_3$  shown below:

$$\epsilon^2 f_2 = \epsilon A_0 \left( \frac{\frac{dE_w}{dt} f_1}{\frac{dE_s}{dt}} + E_w f_1' \right) + A_1 E_w \left( 2 \frac{\frac{dE_w}{dt} f_0'}{\frac{dE_s}{dt}} + E_w f_0'' \right), \quad (3.21)$$

$$\begin{aligned} \epsilon^3 f_3 = & \epsilon^2 A_0 \left( \frac{\frac{dE_w}{dt} f_2}{\frac{dE_s}{dt}} + f_2' \right) + \epsilon A_1 E_w \left( 2 \frac{\frac{dE_w}{dt} f_1'}{\frac{dE_s}{dt}} + E_w f_1'' \right) \\ & + \frac{1}{2} E_w^2 A_2 \left( 3 \frac{\frac{dE_w}{dt} f_0''}{\frac{dE_s}{dt}} + E_w f_0''' \right). \end{aligned} \quad (3.22)$$

Each improvement requires higher-order derivatives (differences) to be calculated, and without some form of prior smoothing this will eventually lead to numerical errors that are too large. Therefore the more higher-order terms we add, the better the underlying distribution in the DAEM.

### 3.3 Numerical Analysis

The inverse problem of determining the distribution from measurements of  $v$  versus time shows that the results obtained by adding the higher-order correction term  $\epsilon f_1$  does upgrade the result to a more accurate estimate of the underlying distribution in the DAEM. However the errors due to differencing are becoming more significant, as the values of  $v$  were only accurate to a limited number of significant figures. McGuinness *et al.* [28] plotted and compared the underlying Gaussian distribution with distributions estimated inversely. It can be seen from this plot that there is more work to be done to smooth the distribution. The main work of this thesis is to find a method to

reduce the errors due to differencing by first fitting an appropriate smooth function to the data. This appropriate smooth function allows the calculation of higher-order terms, and gives a better fit to data, providing a more accurate estimate of the underlying distribution in the DAEM.



# Chapter 4

## Data Fitting

This chapter will provide an overview of the least squares methods of analysis of data. The first step in the analysis of data by this method is the collection of the data. Then there is the formulation of a model relating the numerical values obtained from one or more parameters and from one or more independent variables. Parameters are the unknown quantities in the model and the independent variables are the known quantities. Using the model, one then obtains estimates of the parameters and their associated errors, together with subsidiary information such as the differences between the values calculated using the model and the values obtained by observation. It is good practice to make a graph of the model which shows the experimental data, the calculated data and the residuals.

The model is assessed in terms of such criteria as goodness of fit and relation to underlying mechanisms. Indeed, one objective of the analysis may well be to try to distinguish between various possible such mechanisms. After a preliminary analysis the whole process may be repeated. More data may be postulated and compared to the previous model. The best model is selected and used to throw light on the underlying mechanisms. That model has the best procedure to minimize any errors occurring in the nature of the data collection. The reliability of any parameters obtained by any model

depends on the minimization of error. Therefore the smaller the residuals the better the fit.

## 4.1 Collection of data

Finding a particular appropriate model for data is very difficult and time consuming. There are two methods for developing data in this thesis. The first is the use of the random number generator. Equation (2.13) of DAEM assumes  $f(E)$  is taken to be a Gaussian Distribution with the mean activation energy  $E_0$  and standard deviation  $\sigma$ . The energy is rescaled and defined in Equation (3.11) and a random noise term is added as described below. These randomly generated numbers are used to test whether the model is working or not.

For the purposes of this thesis Equation (3.11) was used to generate the random numbers instead of the solution for the DAEM in Equation (2.13), since it is known that both equations are in equivalent form. The energy scaling is the only difference between the two equations. There are some reasons behind the selection of Equation (3.11) for Equation (2.13). We found out that Equation (3.11) worked very well and took less time to evaluate and graph accurately in the Maple programme compared to Equation (2.13). The procedure of developing the random numbers was done using the Maple program as follows.

The work below is performed and evaluated using Equation (3.11) together with the random number procedure in Maple programme language. Equation (3.11) is written as:

$$v(y) = \sqrt{\frac{\alpha}{\pi}} \int_0^{\infty} \exp\left(-\exp\left(\frac{y_s - y}{y_w}\right) - \alpha(y - 1)^2\right) dy$$

where the constant parameter

$$\alpha = \frac{1}{2} \frac{E_0^2}{\sigma^2}.$$

Note that in practice  $\alpha \gg 1$ . For a linear ramping temperature  $T = mt$  we have:

$$y_s = \frac{Rmk_0tY(k_0t)}{k_0E_0},$$

and

$$y_w = \frac{Rmk_0tY(k_0t)}{k_0E_0(1+Y(k_0t))}.$$

where  $Y$  is the LambertW function defined above. Therefore the equation of  $v(y)$  can be re-written as:

$$v(y) = \frac{E_0}{2\sigma} \sqrt{\frac{2}{\pi}} \int_y^\infty \left( e^{-e^{\frac{\left(\frac{RmtY(k_0t)}{E_0} - y\right) E_0(1+Y(k_0t))}{RmtY(k_0t)}}} - \frac{1}{2} \frac{E_0^2(y-1)^2}{\sigma^2} \right) dy. \quad (4.1)$$

The constant values are  $m = 650$  K/s and  $R = 8.3144$  J/mol kelvins and the parameter values used to generate the random numbers are  $E_0 = 205$  kJ/mol,  $\sigma = 40$  kJ/mol and  $k_0 = 1.07E10s^{-1}$ . The random number generator is a procedure that returns a random floating point number within a certain range. This method develops the procedure which creates the random number generator. The **rand** command generates a procedure which returns random integers. It is a procedure that returns random integers between  $-5$  and  $5\%$ , inclusive, and can be written as  $rand(-5..5\%)$ . To generate random floating-point numbers with magnitudes between  $-5\%$  and  $5\%$ , the following equation is used and written in the Maple programme as:

$$X = \text{evalf}\left(1 + \text{rand}(-1000..1000)/20000\right).$$

The **for** loop is applied to generate 20 random numbers

```
for n from 1 to 20 do X(); od
```

then the 20 random numbers are multiplied by  $v(y)$  to produce the random data in the following way:

$$S = \left[ \text{seq}([n * (3)/20, X() * \text{subs}(t = n * (3)/20, v(y))], n = 1..20) \right]$$

$$\text{evalf} \left( \text{seq}(S[i], i = 1..20) \right).$$

The evaluation of  $S$  in the last step gives make-up data which is  $v(y)$  plus a random number of  $\pm 5\%$  of  $v(y)$ , and that is the data generated from the Gaussian distribution.

The second method used in this thesis for developing data was to use a digitizer. The data may be originally non-Gaussian distributed. The idea of this data collection method is to make sure that the equations and model used in this thesis can work for data where the underlying distribution is unknown. The digitizer can be used to digitize and produce the data from published coal volatilization experiments. This method brings the nature of the data to be more closely or exactly the same as the actual data from the experiment. The next section describes the kind of errors which can occur in the data, but first discusses how to choose the estimated parameter values of a model.

## 4.2 Estimation of parameters

In all non-linear least-squares problems there is a need to furnish initial estimates of the parameters with which to begin the refinement. In some cases the parameter estimates need to be really close to the final values or the

refinement will not work at all. However in other cases, the requirements are less stringent. It is difficult in general to predict what the requirement will be, so it is prudent in all cases to use all that is known about the physics and chemistry of the system under study, and closely related systems. This will dictate the choice of model, any constraints that may apply and which parameters are to be refined.

Ideally, one would construct an expert system that could provide reliable parameter estimates by extrapolation or interpolation from known facts. Unfortunately, all too often, the known facts are too sparse to permit this and one is left with little more than intelligent guesswork. Parameter guesses will only be usable in simple systems. In more complicated systems one will need to resort to numerical techniques or simulation. In this thesis, the parameter estimates have been chosen to be close to the observed values. In this way, the parameter estimates in this model appear to be able to be evaluated in such a way as to bring the calculated value close to the observed value.

### 4.3 Error Description

All observations are subject to errors of various kinds. One of the main objectives of any fitting method is to reduce the effect of experimental errors on the calculated values of the parameters, the estimation of which is the true object of the experiments in which the observations were made. Therefore, the origin and nature of the errors in observations needs to be understood. Observational errors can be conceptually divided into two categories; systematic and random errors. To emphasize the fact that this classification is conceptual the researchers must be aware that they cannot measure the errors themselves, unless the true values of the quantities being measured are known. Errors may stem from different kinds of sources during the measurement.

Some of the most common sources of both types of errors are as follows: blunders, sampling error, bias, calibration error and electronic noise. This list is not exhaustive as there are almost as many types of errors as there are experiments [44]. In coal pyrolysis experiments, these errors also exist and cause errors in the collection of data. The presence of errors in experimental data will clearly affect the reliability of any parameters obtained by any fitting methods. Accurate results can only be achieved by reducing both systematic and random errors. The goodness of the fit of a model is also dependent on errors in experimental data.

## 4.4 The Goodness of Fit of a Model

A question that has occupied the attention of modelers is that of how well a specified model fits the data. Extensive methodology has been developed for investigating whether a proposed model provides a good description of the data [45]. The method usually involves examination of the residuals, these being the differences between the observed responses and the fitted or predicted responses. The usual modeling situation is that a model is adopted because some theory and/or empirical evidence from the use of the model over many data sets indicates that the model is appropriate. However, there are circumstances in which the use of a model is not particularly appropriate, and several competing models may appear to fit the data equally well in practice. In this thesis the choice between models must rest on this consideration; the model that comes closest to behaving as a linear model will be the preferred choice. In general, a model in which the increment is so small that there is no useful change in the elements of the parameter vector is a good model, as is one in which the relative change in the sum of squares on successive iteration is small. The next section is describing the model to apply in this thesis.

## 4.5 Regression Methods

We use a model of data fitting in order to smooth the data and then apply the inverse problem to provide a more accurate estimate of the underlying distribution in the DAEM. Any method of fitting equations to data may be called regression. Such equations are valuable for at least two purposes: making predictions and judging the strength of relationships. Since they provide a way of empirically identifying how a variable is affected by other variables, regression methods have become essential in a wide range of fields, including the social sciences, engineering, medical research and business [46].

Of the various methods of performing regression, the least squares method is the most widely used for fitting data. In fact, linear least squares regression is by far the most widely used of any statistical technique and is a powerful method for analyzing data described by models which are linear in their parameters. However, often researchers have a mathematical expression which relates the response to the predictor variables, and these models are usually nonlinear in their parameters. In such cases, linear regression techniques must be extended to a more complex method such as nonlinear regression. One of the nonlinear regression methods will be selected to calculate the residuals and sum of squares.

### 4.5.1 Linear Regression Model

This discussion begins with a brief description of linear regression because a thorough grounding in linear regression is fundamental to understanding nonlinear regression. The linear regression consists of finding those parameters that minimize a residuals function and, obviously, the smaller the residuals the better the fit. An important requirement for linear regression is that the expression should be linear in its parameters. The normal linear regression

model may be written as:

$$\begin{aligned} Y_n &= \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \varepsilon_n \\ &= (x_{n1}, \dots, x_{np})\boldsymbol{\beta} + \varepsilon_n \end{aligned} \quad (4.2)$$

where parameters  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  and superscript  $T$  denotes the transpose of the vector. In this model, the random variable  $Y_n$ , which represents the response for case  $n$ ,  $n = 1, 2, \dots, N$ , has a deterministic part and a stochastic part. The deterministic part is  $(x_{n1}, \dots, x_{np})\boldsymbol{\beta}$ , where  $(x_{n1}, \dots, x_{np})$  is a (row) vector of predictors for the  $n$  observations, usually with a 1 in the first position representing the regression constant, and  $\boldsymbol{\beta}$  is the vector of regression parameters to be estimated. The stochastic part, represented by random variable  $\varepsilon_n$ , is a shock or disturbance error. The model for the  $n$  cases can be written in matrix notation as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where  $\mathbf{Y}$  is a  $N \times 1$  vector of observable random variables,  $\mathbf{X}$  is a  $N \times p$  matrix of regressor variables,  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of unknown regression coefficients and  $\boldsymbol{\varepsilon}$  is a  $N \times 1$  vector of disturbances (i.e. a random error) whose components are assumed to be independent of the errors for other observations, with expectation 0 and constant variance. In mathematical terms, these assumptions are written as:

$$E(\boldsymbol{\varepsilon}) = 0$$

or equivalently,

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$$

and variance

$$\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2.$$



An even stronger assumption that is occasionally needed is that the errors are normally distributed. This assumption is usually needed only to obtain confidence intervals and in testing procedures. All the above assumptions can be written in the compact form of:

$$\boldsymbol{\varepsilon} \sim \text{NID}(0, \sigma^2)$$

which is read as meaning that the components of  $\boldsymbol{\varepsilon}$  are normally and independently distributed with zero mean and common variance  $\sigma^2$ .

### 4.5.2 Nonlinear Regression Model

So far it has been assumed that there is a linear relationship between the random variable and the parameters as discussed above. There are many cases in which the functional relationship cannot be written in a linear form. In that case it is necessary to move on to a more powerful technique such as nonlinear regression. A nonlinear regression model is one in which the parameters appear nonlinearly. In the more general normal nonlinear regression model, the function  $f(\cdot)$  which relates the response to the predictors, is not necessarily linear. A nonlinear regression model can be written as

$$Y_n = f(\mathbf{X}_n, \boldsymbol{\theta}) + \varepsilon_n, \quad n = 1, 2, \dots, N \quad (4.3)$$

where  $f$  is the expectation function for the parameter  $\boldsymbol{\theta}$  and  $\mathbf{X}_n$  is a vector of associated regressor variables or independent variables on which the  $i$ th calculated value depends. For example, consider an exponential relationship  $y = \theta e^{qx}$  where  $\theta$  and  $q$  are the parameters and  $x$  is the independent variable. The relationship is non-linear because the partial derivatives  $\partial y / \partial \theta$  and  $\partial y / \partial q$  both contain the parameter themselves. This model in Equation (4.3) is of exactly the same form as the linear regression in Equation (4.2), except that the expected responses are nonlinear functions of the parameters. Therefore, for nonlinear models, at least one of the derivatives of the expectation func-

tion with respect to the parameters depends on at least one of the parameters. We used the DAEM which is a nonlinear regression. Equation (4.3) can be written in the form of matrix notation

$$\mathbf{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon},$$

where  $\boldsymbol{\eta}(\boldsymbol{\theta})$  is the  $N$ -vector  $\boldsymbol{\eta}(\boldsymbol{\theta})$  with  $n$ th element

$$\eta_n(\boldsymbol{\theta}) = f(\mathbf{X}_n, \boldsymbol{\theta}) \quad n = 1, 2, \dots, N$$

with  $\boldsymbol{\varepsilon}$  assumed to have a spherical normal distribution. That is

$$E[\boldsymbol{\varepsilon}] = 0,$$

$$\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2.$$

Therefore  $\boldsymbol{\varepsilon} \sim \text{NID}(0, \sigma^2)$  as it is in the linear model.

The assumption of a spherical normal distribution for the random error term  $\boldsymbol{\varepsilon}$  leads us to consider the Euclidean geometry of the  $N$ -dimensional response space, because we will be interested in the least squares estimates  $\hat{\boldsymbol{\theta}}$  for the parameters  $\boldsymbol{\theta}$ . The  $N$ -vectors  $\boldsymbol{\eta}(\boldsymbol{\theta})$  define a  $p$ -dimensional surface called the expectation surface in the response space, and the least squares estimates correspond to the point on the expectation surface,

$$\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$$

which is closest to  $\mathbf{y}$ . That is,  $\hat{\boldsymbol{\theta}}$  is the value of  $\boldsymbol{\theta}$  which minimizes the residual sum of squares

$$\begin{aligned} \mathbf{S}(\boldsymbol{\theta}) &= \sum_{n=1}^N [\mathbf{Y}_n - f(\boldsymbol{\theta}, \mathbf{X}_n)]^2 \\ &= \|\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\|^2. \end{aligned} \tag{4.4}$$

The next section is to discuss how to obtain the least squares estimates  $\hat{\boldsymbol{\theta}}$  for the parameters  $\boldsymbol{\theta}$ .

## 4.6 Determining the Least Squares Estimates

There is a wealth of literature on how to determine the least squares estimates of the parameters once a nonlinear model has been specified and a set of data obtained (see for example [47]). The least squares estimates can be evaluated by using simple geometry. Given a data vector  $\mathbf{y}$ , an expectation function  $f(\mathbf{X}_n, \boldsymbol{\theta})$  and vector of associated regressor variables  $\mathbf{X}_n$ ,  $n=1, 2, \dots, N$  then the following steps can be taken to determine the least squares estimates:

- 1) find the  $\hat{\boldsymbol{\eta}}$  on the expectation surface which is closest to  $\mathbf{y}$ , and then
- 2) determine the parameter vector  $\hat{\boldsymbol{\theta}}$  which corresponds to the point  $\hat{\boldsymbol{\eta}}$ .

In the case of linear regression, both steps 1 and 2 are straightforward evaluations as discussed by Bate and Watts [48]. However in the case of nonlinear regression, the two steps are very difficult to evaluate. The first step is difficult because the expectation surface is curved and often of a finite extent (or, at least, has edges) so that is difficult even to determine  $\hat{\boldsymbol{\eta}}$ . The difficulty of the second step involves finding the parameter space coordinates  $\hat{\boldsymbol{\theta}}$  corresponding to that point even if we know  $\hat{\boldsymbol{\eta}}$  from mapping the parameter space to the expectation surface [48]. To overcome these difficulties, we use an iterative method to determine the least square estimates. This thesis follows this approach in order to find the appropriate iterative method to determine the least square estimates in coal pyrolysis but first, the procedure for fitting the data is summarized.

## 4.7 Procedure of Fitting the Data

The procedure of fitting the data is summarized as follows:

### Step 1

Generate the data using the Gaussian distribution and adding  $\pm 5\%$  noise.

### Step 2

Assume  $f$  is Gaussian and fit a  $v^{(1)}$  to data (is smooth), use fitted  $v^{(1)}$  to get  $\frac{dv^{(1)}}{dt}$ ,  $\frac{d^2v^{(1)}}{dt^2}$ , ... and hence  $f_0^{(1)}$ ,  $f_1^{(1)}$ , ...

### Step 3

Use improved  $f^{(1)} = f_0^{(1)} + \epsilon f_1^{(1)} + \dots$  and fit a  $v^{(2)}$  to data to get  $\frac{dv^{(2)}}{dt}$ ,  $\frac{d^2v^{(2)}}{dt^2}$ , ... and hence  $f_0^{(2)}$ ,  $f_1^{(2)}$ , ...

### Step 4

Repeat this process until negligible change in  $f$ , that is  $f^{(j)} \simeq f^{(j-1)}$ .

### Step 5

Does it always converge?

### Step 6

Get digitized data and repeat **Step 2** to **Step 5**.

# Chapter 5

## The Gauss-Newton Method

This thesis favored the Gauss-Newton iterative method to start with, and the estimates will be determined using this procedure. This is also known as the linearization method. This part of the process of reducing the residuals closely follows the work of Bate and Watts [48] which provided a comprehensive reference on nonlinear regression and nonlinear least squares estimation. The method is based upon a modification to the Newton method introduced by Gauss in 1809 and is known as the Gauss-Newton method. The Newton method uses a linear Taylor series approximation to  $f(\mathbf{X}_n, \boldsymbol{\theta})$  about  $\boldsymbol{\theta}^{(0)}$ . Gauss suggestion was to use a linear approximation to the expectation function to iteratively improve an initial guess to obtain  $\boldsymbol{\theta}^0$  for  $\boldsymbol{\theta}$  and keep improving the estimates until there is no change. That is, to expand the expectation function  $f(\mathbf{X}_n, \boldsymbol{\theta})$  in a first order Taylor series about  $\boldsymbol{\theta}^{(0)}$  as

$$f(\mathbf{X}_n, \boldsymbol{\theta}) \approx f(\mathbf{X}_n, \boldsymbol{\theta}^{(0)}) + V_{n1}(\theta_1 - \theta_1^{(0)}) + V_{n2}(\theta_2 - \theta_2^{(0)}) + \dots + V_{np}(\theta_p - \theta_p^{(0)})$$

where

$$V_{ip} = \left. \frac{\partial f(\mathbf{X}_i, \boldsymbol{\theta})}{\partial \theta_p} \right|_{\boldsymbol{\theta}^{(0)}} \quad p = 1, 2, \dots, P$$

including together all  $N$  cases and producing

$$\boldsymbol{\eta}(\boldsymbol{\theta}) \approx \boldsymbol{\eta}(\boldsymbol{\theta}^{(0)}) + \mathbf{V}^{(0)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)})$$

where  $\mathbf{V}^{(0)}$  is the  $N \times P$  derivative or Jacobian matrix with elements  $\{V_{ip}\}$ . This derivative matrix is based on the evaluation of the derivative equation with respect to each parameter. The approach is equivalent to approximating the residuals in Equation (4.4),  $\mathbf{z}(\boldsymbol{\theta}) = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})$ , by

$$\mathbf{z}(\boldsymbol{\theta}) \approx \mathbf{y} - [\boldsymbol{\eta}(\boldsymbol{\theta}^{(0)}) + \mathbf{V}^{(0)}\boldsymbol{\delta}] = \mathbf{z}^{(0)} - \mathbf{V}^{(0)}\boldsymbol{\delta},$$

where  $\mathbf{z}^{(0)} = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(0)})$ ,  $\boldsymbol{\delta} = \boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}$ . The expression of the Gauss increment  $\boldsymbol{\delta}$  involves decomposing  $\mathbf{V}^{(0)}$  into the product of an orthogonal matrix ( $N \times N$  matrix  $\mathbf{Q}$ ) and an easily inverted matrix ( $N \times P$  matrix  $\mathbf{R}$ ). Since  $\mathbf{Q}$  is orthogonal,  $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ . Note that  $\mathbf{R}$  is zero below the main diagonal,

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix},$$

where  $\mathbf{R}_1$  is a square ( $P \times P$ ) and upper triangular matrix. Writing

$$\mathbf{Q} = [\mathbf{Q}_1 \mid \mathbf{Q}_2],$$

with  $\mathbf{Q}_1$  the first  $P$  columns and  $\mathbf{Q}_2$  the last  $N - P$  columns of  $\mathbf{Q}$  then

$$\mathbf{V}^{(0)} = \mathbf{Q}\mathbf{R} = \mathbf{Q}_1\mathbf{R}_1. \quad (5.1)$$

Computing a  $\mathbf{QR}$  decomposition using Householder Transformations is shown in Appendix 1.

Geometrically, the columns of  $\mathbf{Q}$  define an orthogonal, or orthonormal, basis for the response space with the property that the first  $P$  columns span the expectation plane. Projection onto the expectation plane is accomplished by working in the coordinate system given by  $\mathbf{Q}$ . For example, transform the response vector to

$$\mathbf{w} = \mathbf{Q}^T\mathbf{z},$$

with components

$$\mathbf{w}_1 = \mathbf{Q}_1^T \mathbf{z}^{(0)}, \quad (5.2)$$

and

$$\mathbf{w}_2 = \mathbf{Q}_2^T \mathbf{z}^{(0)}.$$

The projection of  $\mathbf{w}$  onto the expectation plane is then simply

$$\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{0} \end{bmatrix}$$

with the  $\mathbf{Q}$  coordinates and

$$\hat{\boldsymbol{\eta}} = \mathbf{Q} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{0} \end{bmatrix} = \mathbf{Q}_1 \mathbf{w}_1 \quad (5.3)$$

in the original coordinates. To determine the Gauss increment, first find the value  $\boldsymbol{\delta}$  corresponding to  $\hat{\boldsymbol{\eta}}$ . Since

$$\hat{\boldsymbol{\eta}} = \mathbf{V}^{(0)} \boldsymbol{\delta},$$

combining Equation (5.1) and (5.3) gives

$$\mathbf{R}_1 \boldsymbol{\delta} = \mathbf{w}_1, \quad (5.4)$$

where  $\boldsymbol{\delta}$  is solved by back-substitution [49]. The Gauss increment  $\boldsymbol{\delta}^{(0)}$  can then be calculated to minimize the approximate residual sum of squares  $\| \mathbf{z}^{(0)} - \mathbf{V}^{(0)} \boldsymbol{\delta} \|^2$  using Equations (5.1), (5.2), (5.3) and (5.4). Therefore the point  $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}(\boldsymbol{\theta}^{(1)}) = \boldsymbol{\eta}(\boldsymbol{\theta}^{(0)} + \boldsymbol{\delta}^{(0)})$  should now be closer to  $\mathbf{y}$  than  $\boldsymbol{\eta}(\boldsymbol{\theta}^{(0)})$ , and continue on to move to this better new parameter value  $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(0)} + \boldsymbol{\delta}^{(0)}$ . Then another iteration is performed by calculating new residuals  $\mathbf{z}^{(1)} = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(1)})$ , a new derivative matrix  $\mathbf{V}^{(1)}$ , and a new increment. This iterative process is repeated until the decreasing Gauss increment is so small that continuing

will cause no useful change in the values of the parameter vector.

The asymptotic approach gives the general result defined in Equation (3.16) which arose from a term in the Taylor series which was then integrated. Equation (3.16) is then differentiated to obtain the rate of volatilization. The rate of volatilization gives a method for considering the inverse problem with greater accuracy. McGuinness *et al.* [28] used the inverse problem to develop a formula for the leading-order term  $f_0$  which is Gaussian and higher-order correction term  $\epsilon f_1$  which is non Gaussian. The leading-order term  $f_0$  given in Equation (3.19) and the result of adding the higher-order correction term from Equation (3.20) does provide a more accurate estimate of the underlying distribution in the DAEM. In this thesis, a new method is developed to upgrade the accuracy and provide a more accurate estimate of the underlying distribution in the DAEM.

## 5.1 Example 1

The leading order solution ( $v^{(1)}$ ) of Equation (3.16) provides the estimate  $f_0$  which is Gaussian. The leading-order  $v^{(1)}$  is used to fit data which is generated using Gaussian distribution, plus a small random error. We are fitting  $v^{(1)}$  to the data in order to smooth it in preparation for taking derivatives to obtain higher-order approximations to the underlying distribution. The leading-order is evaluated by using the Gauss-Newton method as shown below. All calculations use the Maple programme, which has built-in support for LambertW functions.

The leading-order solution of Equation (3.16) in the case that the underlying distribution is Gaussian, is

$$v^{(1)} = v(t, E_0, \sigma, k_0) = \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha}(y_s - 1))$$



where

$$\begin{aligned}\alpha &= \frac{1}{2} \frac{E_0^2}{\sigma^2}, \\ E_s &= Rmt Y(k_0 t), \\ y_s &= \frac{Rmt Y(k_0 t)}{E_0},\end{aligned}$$

and  $Y$  is the LambertW function. That is,

$$v^{(1)} = v(t, E_0, \sigma, k_0) = \frac{1}{2} \operatorname{erfc} \left( \frac{1}{2} \sqrt{2} \sqrt{\frac{E_0^2}{\sigma^2}} \left( \frac{Rmt Y(k_0 t)}{E_0} - 1 \right) \right). \quad (5.5)$$

In order to find the derivative or Jacobian  $N \times P$  matrix  $\mathbf{V}^{(0)}$  of variables, there is a need to find the derivative of equation  $v(t, E_0, \sigma, k_0)$  with respect to each parameter  $E_0$ ,  $\sigma$ , and  $k_0$ . The first derivative  $p$  is obtained by differentiating  $v(t, E_0, \sigma, k_0)$  with respect to parameter  $E_0$  as follows:

$$\begin{aligned}p &= -e^{-\frac{1}{2} E_0^2 \left( \frac{Rmt Y(k_0 t)}{E_0} - 1 \right)^2} \sigma^{-2} \left( \frac{1}{2} \sqrt{2} \left( \frac{Rmt Y(k_0 t)}{E_0} - 1 \right) \right. \\ &\quad \left. E_0 \frac{1}{\sqrt{\frac{E_0^2}{\sigma^2}}} \sigma^{-2} - \frac{1}{2} \sqrt{2} \sqrt{\frac{E_0^2}{\sigma^2}} Rmt Y(k_0 t) E_0^{-2} \right) \frac{1}{\sqrt{\pi}}.\end{aligned} \quad (5.6)$$

The second derivative  $r$  is obtained by differentiating  $v(t, E_0, \sigma, k_0)$  with respect to parameter  $\sigma$  as below:

$$\begin{aligned}r &= \frac{1}{2} e^{-\frac{1}{2} E_0^2 \left( \frac{Rmt Y(k_0 t)}{E_0} - 1 \right)^2} \sigma^{-2} \sqrt{2} \left( \frac{Rmt Y(k_0 t)}{E_0} - 1 \right) \\ &\quad E_0^2 \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{\frac{E_0^2}{\sigma^2}}} \sigma^{-3}.\end{aligned} \quad (5.7)$$

The third derivative  $s$  is obtained by differentiating  $v(t, E_0, \sigma, k_0)$  with respect

to parameter  $k_0$ :

$$s = -\frac{1}{2} e^{-\frac{1}{2} E_0^2 \left( \frac{RmtY(k_0 t)}{E_0} - 1 \right)^2 \sigma^{-2}} \sqrt{2} \sqrt{\frac{E_0^2}{\sigma^2}} Rmt Y(k_0 t) \\ \frac{1}{\sqrt{\pi}} (1 + Y(k_0 t))^{-1} k_0^{-1} E_0^{-1}. \quad (5.8)$$

These Equations (5.5), (5.6), 5.7) and (5.8) will be used again in Example 2, Example 3 and Example 4.

The constant values used in this example are  $m= 650\text{K/s}$  and  $R= 8.3144\text{J/}$  (mol kelvins) and the starting parameter estimates chosen are:  $E_0= 196\text{kJ/}$  mol,  $\sigma= 37\text{kJ/mol}$  and  $k_0= 1.00\text{E}10\text{s}^{-1}$ . The parameter estimates have been chosen to be close to the final values that are used to generate the random numbers as one of the criteria of parameter estimation. The data was generated with a Gaussian distribution and the parameter values of  $E_0= 205\text{kJ/mol}$ ,  $\sigma= 40\text{kJ/mol}$  and  $k_0= 1.07\text{E}10\text{s}^{-1}$ , and a  $\pm 5\%$  random error term has been added to  $v(y)$  in Equation(4.1). Using these values, the random number generator produces the following data:

$$Data = \left[ \begin{array}{l} [0.1500000000, 0.9621489806], [0.3000000000, 1.024242331], \\ [0.4500000000, 1.033051170], [0.6000000000, 0.9819960462], \\ [0.7500000000, 0.9734757992], [0.9000000000, 1.031344011], \\ [1.050000000, 0.9986545413], [1.200000000, 0.9703193539], \\ [1.350000000, 0.9130842328], [1.500000000, 0.7918876653], \\ [1.650000000, 0.6483888526], [1.800000000, 0.5047982891], \\ [1.950000000, 0.3326939449], [2.100000000, 0.1917859271], \\ [2.250000000, 0.09363834317], [2.400000000, 0.04104433104], \\ [2.550000000, 0.01602741851], [2.700000000, 0.004913205795], \\ [2.850000000, 0.001476246960], [3.0, 0.0003506119174] \end{array} \right]$$

The residuals are approximated as  $\mathbf{z}(\boldsymbol{\theta}) = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})$ , note that the parameters  $\theta$  are in this example  $E_0, \sigma, k_0$ . This equation gives the residual of the starting parameter estimates  $\theta^{(0)}$  in the form of  $\mathbf{z}^{(0)} = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(0)})$  which gives

$$\mathbf{z}^{(0)} = \begin{pmatrix} -0.0378505354, 0.0242462450, 0.0330789660, -0.0178351768, \\ -0.0256596563, 0.0350519455, 0.0119172678, 0.0098383499, \\ 0.0113429238, -0.0032702202, 0.0095844786, 0.0509029936, \\ 0.0546503886, 0.0479206872, 0.03174364522, 0.01916557094, \\ 0.9730655510\text{E-}2, 0.3448025851\text{E-}2, 0.1202123228\text{E-}2, \\ 0.3095576615\text{E-}3 \end{pmatrix}^T$$

and the residual sum of squares  $S(\boldsymbol{\theta}^{(0)}) = \|\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(0)})\|^2$  of  $\mathbf{z}^{(0)}$  is

$$S(\boldsymbol{\theta}^{(0)}) = S_{old} = 0.1514721026\text{E-}1.$$

The derivatives are collected into the derivative matrix  $\mathbf{V}^{(0)}$  by evaluating the derivative equations  $p, r$  and  $s$  for each parameter  $(E_0, \sigma, k_0)$  in the time given of  $t = (0.15, 0.30, 0.45, \dots, 3.0)$ s. The result of the derivatives  $\mathbf{V}^{(0)}$  is discussed below:

$$\mathbf{V}^{(0)} = \begin{pmatrix} 6.657472967\text{E-}11 & -3.260814262\text{E-}10 & -5.116196669\text{E-}18 \\ 4.945742232\text{E-}10 & -2.210665137\text{E-}9 & -7.615295505\text{E-}17 \\ 3.196505271\text{E-}9 & -1.288441133\text{E-}8 & -7.390246094\text{E-}16 \\ 1.747656934\text{E-}8 & -6.264706838\text{E-}8 & -5.391113343\text{E-}15 \\ 7.961230992\text{E-}8 & -2.494416517\text{E-}7 & -3.071424298\text{E-}14 \\ 2.991567349\text{E-}7 & -8.010126430\text{E-}7 & -1.385549080\text{E-}13 \\ 9.205173669\text{E-}7 & -0.000002042104341 & -4.975694290\text{E-}13 \\ 0.000002306225406 & -0.000004050412944 & -1.425099486\text{E-}12 \\ 0.000004682650784 & -0.000006047818795 & -3.256123839\text{E-}12 \\ 0.000007675553348 & -0.000006328105702 & -5.931664428\text{E-}12 \\ 0.00001012282027 & -0.000003596281322 & -8.606983433\text{E-}12 \\ 0.00001071014070 & 0.000001240510815 & -9.936068312\text{E-}12 \\ 0.000009066955560 & 0.000005337384666 & -9.114173826\text{E-}12 \\ 0.000006127541846 & 0.000006514272517 & -6.634300430\text{E-}12 \\ 0.000003298752579 & 0.000005076981573 & -3.827234243\text{E-}12 \\ 0.000001411921998 & 0.000002847008241 & -1.747565780\text{E-}12 \\ 4.796170009\text{E-}7 & 0.000001196673972 & -6.308139239\text{E-}13 \\ 1.290873580\text{E-}7 & 3.840284786\text{E-}7 & -1.797899977\text{E-}13 \\ 2.748588911\text{E-}8 & 9.499133913\text{E-}8 & -4.041301099\text{E-}14 \\ 4.623251603\text{E-}9 & 1.820710894\text{E-}8 & -7.156178286\text{E-}15 \end{pmatrix}$$

The first column elements ( $\mathbf{V}_{n1}^{(0)}$ ) are obtained from evaluation of  $p$  at each time  $t$ . The second column ( $\mathbf{V}_{n2}^{(0)}$ ) is the result of evaluation of  $r$  at each time  $t$  and the last column ( $\mathbf{V}_{n3}^{(0)}$ ) is evaluated as  $s$  varied with time. Once  $\mathbf{V}^{(0)}$  is derived then the **QR** decomposition of  $20 \times 3$  matrix  $\mathbf{V}^{(0)}$  is computed. We chose the Householder transformation  $\mathbf{H}_{\mathbf{u}_1}$  to take the first column  $\mathbf{v}_1$  of derivatives  $\mathbf{V}^{(0)}$  to the  $e_1$  using the equation below

$$\mathbf{u}_1 = \frac{\mathbf{v}_1 - n\mathbf{e}_1}{n_1}.$$

$\mathbf{e}_1 = (1, 0, \dots, 0)^T$  is the  $20 \times 1$  matrix of 1 in the first row and the rest are all 0 where  $n$  is the norm of  $\mathbf{v}_1$ , and  $n_1$  is the norm of  $(\mathbf{v}_1 - n\mathbf{e}_1)$  and hence:

$$\mathbf{u}_1 = \left( \begin{array}{l} -0.70710565542, 0.0000167261620091, 0.000108103622303, \\ 0.00059104562355, 0.0026924338778, 0.01011727619299, \\ 0.0311312678436, 0.077994966096, 0.15836404724, \\ 0.25958196523, 0.3423468589395, 0.362209634237, \\ 0.306638236511, 0.20722927485095, 0.1115615563368, \\ 0.04775023641538, 0.0162203189796, 0.00436564616988, \\ 0.000929553973200073326, 0.000156355207556618786 \end{array} \right)^T$$

The  $20 \times 3$  matrix of  $\mathbf{V}_1^0$  is calculated using this equation  $\mathbf{V}_1^0 = \mathbf{H}_{u1} \mathbf{V}^0$ , which equals:

$$\mathbf{V}_1^0 = \begin{pmatrix} 2.090840347\text{E-}5 & -0.00000008549682633 & -1.897885399\text{E-}11 \\ 1.075910928\text{E-}19 & -0.000000002208650474 & 3.727803811\text{E-}16 \\ 6.953770490\text{E-}19 & -0.00000001287139027 & 2.162496958\text{E-}15 \\ 3.801903420\text{E-}18 & -0.00000006257587704 & 1.047266243\text{E-}14 \\ 1.731909498\text{E-}17 & -0.0000002491173485 & 4.155118902\text{E-}14 \\ 6.507939207\text{E-}17 & -0.0000007997940190 & 1.329946891\text{E-}13 \\ 2.002520548\text{E-}16 & -0.000002038354586 & 3.379996588\text{E-}13 \\ 5.017029977\text{E-}16 & -0.000004041018465 & 6.683000827\text{E-}13 \\ 1.018677654\text{E-}15 & -0.000006028743875 & 9.943969931\text{E-}13 \\ 1.669761149\text{E-}15 & -0.000006296839102 & 1.035564372\text{E-}12 \\ 2.202147204\text{E-}15 & -0.000003555045706 & 5.816706908\text{E-}13 \\ 2.329914704\text{E-}15 & 0.000001284138898 & -2.142935610\text{E-}13 \\ 1.972451158\text{E-}15 & 0.000005374319185 & -8.839458485\text{E-}13 \\ 1.333002418\text{E-}15 & 0.000006539233245 & -1.072227692\text{E-}12 \\ 7.176203148\text{E-}16 & 0.000005090419142 & -8.329010750\text{E-}13 \\ 3.071534585\text{E-}16 & 0.000002852759748 & -4.659404480\text{E-}13 \\ 1.043371950\text{E-}16 & 0.000001198627706 & -1.954575098\text{E-}13 \\ 2.808203101\text{E-}17 & 0.0000003845543199 & -6.261523288\text{E-}14 \\ 5.979359287\text{E-}18 & 0.00000009510330373 & -1.546360741\text{E-}14 \\ 1.005755491\text{E-}18 & 0.00000001822594190 & -2.959575089\text{E-}15 \end{pmatrix}$$

We performed a second rotation which is orthogonal to the first rotation by choosing  $\mathbf{u}_2$  so that  $\mathbf{H}_{\mathbf{u}_2}$  zeros the rows below the diagonal of the second column of  $\mathbf{V}_1^{(0)}$  without changing the first column. To ensure that there is no change to the first column, the first element of  $\mathbf{u}_2$  must be zero. Therefore the vector  $\mathbf{u}_2$  is chosen to be

$$\mathbf{u}_2 = \frac{\mathbf{v}_2 - n_2 \mathbf{e}_2}{n_3}.$$

$\mathbf{e}_2 = (0, 1, 0, \dots, 0)^T$  is the  $20 \times 1$  matrix of unit in the second row and the rest are all 0 where  $n_2$  is the norm of  $\mathbf{v}_2$ , and  $n_3$  is the norm of  $\mathbf{v}_2 - n_2 \mathbf{e}_2$  and obtains:

$$\mathbf{u}_2 = \begin{pmatrix} -0.70710565542, 0.000016726162009, 0.00010810362230, \\ 0.00059104562355, 0.0026924338778, 0.01011727619299, \\ 0.0311312678436, 0.077994966096, 0.15836404724, \\ 0.25958196523, 0.3423468589395, 0.362209634237, \\ 0.306638236511, 0.20722927485095, 0.1115615563368, \\ 0.04775023641538, 0.0162203189796, 0.00436564616988, \\ 0.000929553973200073326, 0.000156355207556618786 \end{pmatrix}^T.$$

Then  $\mathbf{V}_2^0$  is calculated by using this equation  $\mathbf{V}_2^0 = \mathbf{H}_{\mathbf{u}_2} \mathbf{H}_{\mathbf{u}_1} \mathbf{V}^{(0)}$  as shown below:

$$\mathbf{V}_2^0 = \begin{pmatrix} 2.090840347\text{E-}5 & -8.549682633\text{E-}8 & -1.897885399\text{E-}11 \\ -1.006509781\text{E-}19 & 1.479064855\text{E-}5 & -2.431052651\text{E-}12 \\ 6.951959293\text{E-}19 & -9.824868842\text{E-}19 & 4.689310927\text{E-}17 \\ 3.801025393\text{E-}18 & -4.776485086\text{E-}18 & 1.873891538\text{E-}16 \\ 1.731559985\text{E-}17 & -1.901539908\text{E-}17 & 6.050588455\text{E-}16 \\ 6.506817272\text{E-}17 & -6.104926828\text{E-}17 & 1.536682644\text{E-}15 \\ 2.002233251\text{E-}16 & -1.555901988\text{E-}16 & 2.965857752\text{E-}15 \\ 5.016461666\text{E-}16 & -3.084555152\text{E-}16 & 4.098777332\text{E-}15 \\ 1.018592478\text{E-}15 & -4.601814814\text{E-}16 & 3.483542316\text{E-}15 \\ 1.669672999\text{E-}15 & -4.806452084\text{E-}16 & 5.854953927\text{E-}16 \\ 2.202096693\text{E-}15 & -2.713603034\text{E-}16 & -2.653785648\text{E-}15 \\ 2.329932593\text{E-}15 & 9.801987545\text{E-}17 & -3.226291726\text{E-}15 \\ 1.972527035\text{E-}15 & 4.102276073\text{E-}16 & -5.967995293\text{E-}16 \\ 1.333094630\text{E-}15 & 4.991471660\text{E-}16 & 2.592255933\text{E-}15 \\ 7.176922011\text{E-}16 & 3.885577978\text{E-}16 & 3.784792294\text{E-}15 \\ 3.071935440\text{E-}16 & 2.177547634\text{E-}16 & 2.952917661\text{E-}15 \\ 1.043540804\text{E-}16 & 9.149255317\text{E-}17 & 1.554727122\text{E-}15 \\ 2.808744344\text{E-}17 & 2.935349512\text{E-}17 & 5.919718028\text{E-}16 \\ 5.980692692\text{E-}18 & 7.259337629\text{E-}18 & 1.680307645\text{E-}16 \\ 1.006012050\text{E-}18 & 1.391208158\text{E-}18 & 3.612871787\text{E-}17 \end{pmatrix}$$



We now performed a third rotation which is orthogonal to the first and second rotations by choosing  $\mathbf{u}_3$  so that  $\mathbf{H}_{\mathbf{u}_3}$  zeros the rows below the diagonal of the third column of  $\mathbf{V}_2^0$  without changing the first and second columns. To ensure there is no change to the first and second columns, the first two elements of  $\mathbf{u}_3$  must be zero. Therefore the vector  $\mathbf{u}_3$  is chosen to be

$$\mathbf{u}_3 = \frac{\mathbf{v}_3 - n_4 \mathbf{e}_3}{n_5}.$$

$\mathbf{e}_3 = (0, 0, 1, \dots, 0)^T$  is the  $20 \times 1$  matrix of unit in the third row and the rest are all 0 where  $n_4$  is the distance of  $\mathbf{v}_3$ , and  $n_5$  is the norm of  $\mathbf{v}_3 - n_4 \mathbf{e}_3$  and hence:

$$\mathbf{u}_3 = \begin{pmatrix} 0.0, 0.0, -0.705368653379, 0.0139085505895, \\ 0.0449091710574, 0.114056912395, 0.220134312740, \\ 0.304222793655, 0.2585583185999, 0.0434571165109, \\ -0.196971442500, -0.239464455565, -0.04429614135326, \\ 0.192404533825, 0.2809179401508, 0.219173862712, \\ 0.115396224304, 0.0439378139002, 0.0124717164338, \\ 0.00268157516058 \end{pmatrix}^T$$

which gives

$$\mathbf{R} = \mathbf{H}_{\mathbf{u}_3} \mathbf{H}_{\mathbf{u}_2} \mathbf{H}_{\mathbf{u}_1} \mathbf{V}^0.$$

The matrix  $\mathbf{R}$  is in the form of

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix},$$

where  $\mathbf{R}_1$  is the  $3 \times 3$  upper triangular matrix given below

$$\begin{pmatrix} 2.09084035\text{E-}5 & -8.549682633\text{E-}8 & -1.897885399\text{E-}11 \\ 0 & 1.479064854\text{E-}5 & -2.431052651\text{E-}12 \\ 0 & 0 & 9.550287027\text{E-}15 \end{pmatrix}.$$

In order to find  $\mathbf{Q}_1 = \mathbf{V}^{(0)}\mathbf{R}_{1inv}$ , there is a need to calculate the inverse of  $\mathbf{R}_1$  and the response vector is transformed to  $\mathbf{w}_1 = \mathbf{Q}_1^T \mathbf{z}^{(0)}$ . Then the increment  $\boldsymbol{\delta}^{(0)}$  is calculated to minimize the approximate residual sum of squares, as shown below:

$$\boldsymbol{\delta}^{(0)} = \mathbf{R}_{1inv} \mathbf{w}_1$$

which gives

$$\boldsymbol{\delta}^{(0)} = \begin{pmatrix} 2.44810107473540911\text{E}6 \\ 4.45640399695003463\text{E}5 \\ 2.69085969986378856\text{E}12 \end{pmatrix}.$$

The new parameter value ( $\boldsymbol{\theta}^{(1)}$ ) is calculated in the form of  $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(0)} + \boldsymbol{\delta}^{(0)}$  to obtain

$$\boldsymbol{\theta}^{(1)} = \begin{pmatrix} 2.64410107473540911\text{E}6 \\ 4.82640399695003463\text{E}5 \\ 2.70085969986378857\text{E}12 \end{pmatrix}.$$

Another iteration is performed by calculating the new residuals  $\mathbf{z}^{(1)}$ . Using the Gauss-Newton method the residuals are written as  $\mathbf{z}(\boldsymbol{\theta}) = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})$ . This equation gives the residual of the new parameter estimates  $\boldsymbol{\theta}^{(1)}$  in the form of  $\mathbf{z}^{(1)} = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(1)})$  and the new sum of squares ( $S_{new}$ ) is

$$S(\boldsymbol{\theta}^{(1)}) = S_{new} = 7.22067597.$$

In the undertaking of this thesis a program was developed to make it easier to identify the nature of  $S_{old}$  and  $S_{new}$ . The program is read in the Maple

language as follows:

```

f := proc(sold, snew)
if snew ≤ sold then
    print("true");
else
    print("Apply Step Factor")
end if ; end proc;

```

As seen in the Example 1, the Gauss-Newton increment can produce an increase in the sum of squares, therefore the step factor is introduced.

## 5.2 Step Factor

From the Example 1,  $S_{new} = 7.220068554 > S_{old} = 0.1514721026E-1$ , a step factor is applied because the Gauss-Newton increment produces an increase in the sum of squares. The requested increment usually extends beyond the region in which the linear approximation is valid. Thus, a small step in the direction  $\boldsymbol{\theta}^{(0)}$  should produce a decrease in the sum of the squares. Therefore a step factor  $\lambda$  can be introduced and calculated as

$$\boldsymbol{\theta}^1 = \boldsymbol{\theta}^0 + \lambda \boldsymbol{\delta}^0$$

where  $\lambda$  is chosen to ensure that

$$S(\boldsymbol{\theta}^1) < S(\boldsymbol{\theta}^0). \quad (5.9)$$

A common method of selecting  $\lambda$  is to start with  $\lambda = 1$  and halve it until Equation (5.9) is satisfied.

As seen in the Example 1, the Gauss-Newton increment produced an

increase in the sum of squares which did not match with the thesis expectations. The step factor is applied to produce a decrease in the sum of squares by letting  $\lambda = 0.005$  for a start. Then the residuals are calculated in the form of  $\mathbf{z}^{(1)} = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(1)})$  which gives the new sum of squares ( $S_{new}$ ) as  $S_{new} = 7.220067027$ . The method of step factor uses different values of  $\lambda$  until the Equation (5.9) is satisfied and then the iterative process is repeated until convergence is obtained. That is, until the increment is so small that there is no useful change in the elements of the parameter vector. Also, the relative change in the sum of squares on successive iterations is small. From Example 1, the Gauss-Newton method does not agree with either of the criteria above which are used to declare convergence. This means that when the process is repeated, the requested increment extends beyond the region where the linear approximation is valid and produces an increase in the sum of squares. Note that when the step factor was applied, the  $S_{new}$  was still greater than the  $S_{old}$ . Therefore the Gauss-Newton method does not work for this type of nonlinear equation.

This erratic behavior of the Gauss-Newton method could be improved in practice by applying parameter transformations to improve constraints on parameters. For example, if  $\theta_p$  must be positive and reparametrize to  $\phi_p = \ln \theta_p$ , so throughout the iterations the value of  $\theta_p = e^{\phi_p}$  remains positive. An interval constraint on a parameter can be written as

$$a \leq \theta \leq b,$$

and can be enforced by a logistic transformation of the form

$$\theta = a + \frac{b - a}{1 + e^{-\phi}},$$

while there is an order constraint on parameters  $\theta_j, \dots, \theta_k$ , such as

$$a \leq \theta_j \leq \theta_{j+1} \leq \dots \leq \theta_k \leq b,$$

and this can be enforced by a transformation given in Jupp [50]. The recommendation is based on the fact that the range of the parameter values are expected to be all positive numbers. Example 1 illustrated that when the iterative process is repeated several times, the parameters will end up as negative values. The negative values show that this iterative method is inappropriate for use in this situation. Therefore, applying parameter transformations to improve constraints on parameters is one way to fix the problem of erratic behavior of Gauss-Newton method.

The Gauss-Newton iterative method for nonlinear least squares is a simple and useful method for finding the least square estimates  $\hat{\theta}$  for the parameters  $\theta$ . Some modifications to this method, as well as alternative methods, have been suggested primarily to deal with ill-conditioning of the derivative matrix  $V$  and to avoid having to code and to specify the derivatives. The considerations of Seber and Wild [51] regarding the Gauss-Newton algorithm lead this thesis to discuss Gauss-Newton-based algorithms under two headings: “small-residual problems” and “large-residual problems.” This thesis is primarily focused on Gauss-Newton-based algorithms for small-residual problems because the sum of squares of the starting parameter estimates is small. They are derived by taking the Gauss-Newton step as a search direction and modifying the basic Gauss-Newton algorithm in the same way as Seber and Wild [51] modified the classical N algorithm. This gives such popular least-squares algorithms as the Marquardt [53] method.

# Chapter 6

## Method for Small-Residual Problems

### 6.1 Levenberg-Marquardt Algorithms

A condition that causes erratic behavior of Gauss-Newton iterations is the singularity of the derivative matrix  $V$ , which in turn is caused by the 3rd column is close to 0. There are significant difficulties in accurately determining both  $E_0$  and  $k_0$ , because they are highly correlated. When  $V$  is nearly singular,  $\delta$  can be very large, causing the parameters to go into undesirable regions of the parameter space, as found in the Gauss-Newton method in Example 1 above. One solution to the problem of near-singularity is to perform the calculations for the increment in a numerically stable way. Bates and Watts [48] recommend using the  $QR$  decomposition rather than the normal equation ( $\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y}$ ). This solution is used in this thesis but is still unable to provide better results. Bates and Watts [48] also recommend the use of parameter transformation.

We also uses the ideas of Levenberg [54] and Marquardt [53] described by Seber and Wild [51]. Levenberg-Marquardt algorithms allow for singular or ill-conditioned matrices  $V^TV$  by modifying the Gauss-Newton step as written

in the form

$$\boldsymbol{\delta}^{(n)} = (\mathbf{V}^{(n)\text{T}}\mathbf{V}^{(n)})^{-1} \mathbf{V}^{(n)\text{T}}\mathbf{z}^{(n)}$$

to

$$\boldsymbol{\delta}^{(n)} = (\mathbf{V}^{(n)\text{T}}\mathbf{V}^{(n)} + k^{(n)}\mathbf{D}^{(n)})^{-1} \mathbf{V}^{(n)\text{T}}\mathbf{z}^{(n)}, \quad (6.1)$$

where  $k^{(n)}$  is a conditioning factor and  $\mathbf{D}^{(n)}$  is a diagonal matrix with positive diagonal elements. Often, for simplicity,  $\mathbf{D}^{(n)} = \mathbf{I}^{(n)}$ . A popular choice of  $\mathbf{D}^{(n)}$  is a diagonal matrix with entries equal to the diagonal element of  $\mathbf{V}^{(n)\text{T}}\mathbf{V}$ , so that the method is approximately invariant under rescaling of the  $\theta$ . When  $\mathbf{D} = \mathbf{I}$ , the Levenberg-Marquardt direction of  $\boldsymbol{\delta}^{(n)}$  in the Equation (6.1) interpolates between the direction of the Gauss-Newton increment ( $k \rightarrow 0$ ) and the direction of steepest-descent  $\mathbf{V}^{\text{T}}(\mathbf{y} - \mathbf{k}) / \|\mathbf{V}^{\text{T}}(\mathbf{y} - \mathbf{k})\|$  ( $k \rightarrow \infty$ ). Also, as the direction tends to steepest descent, the step length  $\|\boldsymbol{\theta}\|$  tends to zero. For  $k > 0$ ,  $\mathbf{V}^{\text{T}}\mathbf{V} + k\mathbf{D}$  is positive definite, as  $\mathbf{D}$  is positive definite. Note, as  $k \rightarrow \infty$  the length  $\delta$  tends to zero. Thus by choosing  $k$  large enough we can reduce  $S(\boldsymbol{\theta}) = \sum z^2(\boldsymbol{\theta})$ . However, if  $k$  is too large for too many iterations, the algorithm will take too many small steps and thus make little progress.

Levenberg-Marquardt algorithms differ in how they choose and update  $k$ . Originally Levenberg [54] chose  $k$  to minimize  $S(\boldsymbol{\theta})$  for  $\boldsymbol{\theta}$  of the form  $\boldsymbol{\theta}^{(n)} + \boldsymbol{\delta}^{(n)}$ , where  $\boldsymbol{\delta}$  is given by Equation (6.1). This recommendation has been disregarded because each trial value of  $k$  requires the solution of another least-squares problem. Marquardt recommends a much easier technique to use and produces an increment which is invariant under scaling transformations of the parameters. The first iteration,  $k$  is chosen as a small positive value, for example,  $k^{(1)} = 0.01$ . If the result from the first iteration agrees with the increment  $\boldsymbol{\delta}$  of Equation (6.1), which minimizes the sum of squares  $S(\boldsymbol{\theta})$ , then the second iteration,  $k$  is divided by a factor, where it is likely that a factor of 10 is preferable to use in this thesis. Therefore the new parameters are sets  $\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} + \boldsymbol{\delta}^{(n)}$  where  $k^{(n+1)} = k^{(n)}/10$  to push the

algorithm closer to the Gauss-Newton method. This process is repeated until convergence is obtained, that is, until the increment is so small that there is no useful change in the values of the parameter vectors. During the  $n$ th iteration, if the step  $\delta^{(n)}$  does not reduce the sum of squares  $S(\boldsymbol{\theta})$  then the value of  $k$  should be increased by a factor. In other words, if  $S(\boldsymbol{\theta})$  is not reduced, then try  $k_{old}^{(n)} = 10k_{new}^{(n)}$ , each time recomputing  $k^{(n)}$  until a reduction in  $S(\boldsymbol{\theta})$  is achieved. Then again  $k$  must be divided by a factor and this process repeated until convergence is obtained.

The Levenberg-Marquardt algorithm is more difficult to implement than the Gauss-Newton algorithm, since one must decide how to manipulate both the conditioning factor  $k$  and the step factor  $\lambda$ . There is general agreement that, in practice, Levenberg-Marquardt algorithms have been proved to be good general-purpose algorithms for least-squares problems [51], since they are robust and work well. Remainder, the main work of this thesis is to find a method to reduce the errors due to differencing by first fitting an appropriate smooth function to the data. This appropriate smooth function allows the calculation of higher-order terms, and gives a better fit to data, providing a more accurate estimate of the underlying distribution in the DAEM.

## 6.2 Data Generated with a Gaussian

### 6.2.1 Example 2

To illustrate calculations of the residuals, we follow the same method as in Example 1 and it will be used again in Example 3 and Example 4. The leading order is evaluated by using the Levenberg-Marquardt Algorithm to minimize the residuals as shown below. Using these values in Example 1, the



random number generator produces the following data:

$$\begin{aligned}
 \text{Data} = & \left[ [0.1500000000, 0.9621489806], [0.3000000000, 1.024242331], \right. \\
 & [0.4500000000, 1.033051170], [0.6000000000, 0.9819960462], \\
 & [0.7500000000, 0.9734757992], [0.9000000000, 1.031344011], \\
 & [1.0500000000, 0.9986545413], [1.2000000000, 0.9703193539], \\
 & [1.3500000000, 0.9130842328], [1.5000000000, 0.7918876653], \\
 & [1.6500000000, 0.6483888526], [1.8000000000, 0.5047982891], \\
 & [1.9500000000, 0.3326939449], [2.1000000000, 0.1917859271], \\
 & [2.2500000000, 0.09363834317], [2.4000000000, 0.04104433104], \\
 & [2.5500000000, 0.01602741851], [2.7000000000, 0.004913205795], \\
 & \left. [2.8500000000, 0.001476246960], [3.0, 0.0003506119174] \right].
 \end{aligned}$$

The residual sum of squares are written as  $\mathbf{z}(\boldsymbol{\theta}^{(n)}) = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(n)})$ . This equation gives the residual sum of squares of the starting parameter estimates in the form of  $\mathbf{z}^{(0)} = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(0)})$  where  $\boldsymbol{\theta}^{(0)}$  is the starting value for the parameter. Then

$$\begin{aligned}
 \mathbf{z}^{(0)} = & \left( -0.0378494324, 0.0242510975, 0.0330954475, -0.0178037418, \right. \\
 & -0.0257222098, 0.0341725915, 0.0074074423, -0.0059564746, \\
 & -0.0305957182, -0.0908213332, -0.1365770949, -0.1461597594, \\
 & -0.1614304881, -0.1459237837, -0.1113384887, -0.06822852416, \\
 & -0.03469009949, -0.01544013354, -0.005547342985, \\
 & \left. -0.001724366280 \right)^T,
 \end{aligned}$$

and the residual sum of squares of  $\mathbf{z}^{(0)}$  is  $S(\boldsymbol{\theta}^{(0)}) = \|\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(0)})\|^2$ .

$$S(\boldsymbol{\theta}^{(0)}) = s_{old} = 0.1204338066.$$

Next, the derivatives are collected into the derivative matrix  $\mathbf{V}^{(0)}$  by evaluating the derivative equations  $p$ ,  $r$ , and  $s$  for each parameter  $(E_0, \sigma, k_0)$  at time  $t = (0.15, 0.30, 0.45, \dots, 3.0)$ s. The derivative matrix  $\mathbf{V}^{(0)}$  is given below:

$$\mathbf{V}^{(0)} = \begin{pmatrix}
 1.752041770\text{E-}10 & -8.163155630\text{E-}10 & -1.013120380\text{E-}17 \\
 8.979171718\text{E-}10 & -3.855832371\text{E-}9 & -1.040276946\text{E-}16 \\
 4.175534354\text{E-}9 & -1.636803082\text{E-}8 & -7.263423812\text{E-}16 \\
 1.724075745\text{E-}8 & -6.102890225\text{E-}8 & -4.001443106\text{E-}15 \\
 6.247688786\text{E-}8 & -1.971259316\text{E-}7 & -1.813473209\text{E-}14 \\
 1.971939567\text{E-}7 & -5.456392482\text{E-}7 & -6.871384273\text{E-}14 \\
 5.391057423\text{E-}7 & -0.000001280865929 & -2.192397579\text{E-}13 \\
 0.000001271144602 & -0.000002519759620 & -5.909609280\text{E-}13 \\
 0.000002575993109 & -0.000004086652079 & -1.347632199\text{E-}12 \\
 0.000004473613741 & -0.000005317508941 & -2.601000275\text{E-}12 \\
 0.000006641297436 & -0.000005240482549 & -4.248299940\text{E-}12 \\
 0.000008409741111 & -0.000003262208299 & -5.869658057\text{E-}12 \\
 0.000009065886650 & 1.335259594\text{E-}7 & -6.856066978\text{E-}12 \\
 0.000008305875164 & 0.000003477851455 & -6.765525151\text{E-}12 \\
 0.000006456966120 & 0.000005320379789 & -5.635981026\text{E-}12 \\
 0.000004253233982 & 0.000005233197759 & -3.960466790\text{E-}12 \\
 0.000002370758312 & 0.000003883150494 & -2.345832054\text{E-}12 \\
 0.000001116875526 & 0.000002285691719 & -1.170275915\text{E-}12 \\
 4.442023518\text{E-}7 & 0.000001090985244 & -4.913520484\text{E-}13 \\
 1.489898680\text{E-}7 & 4.270842885\text{E-}7 & -1.734961120\text{E-}13
 \end{pmatrix}$$

Then the Levenberg-Marquardt Algorithm is applied to calculate the increment  $\boldsymbol{\delta}^{(0)}$  to minimize the approximate residual sum of squares as follows

$$\boldsymbol{\delta}^{(0)} = (\mathbf{V}^{(0)T}\mathbf{V}^{(0)} + k^{(0)}\mathbf{D}^{(0)})^{-1} (\mathbf{V}^{(0)T}\mathbf{z}^{(0)}) \quad (6.2)$$

where  $\mathbf{V}^{(0)T}$  is the transposed matrix of matrix  $\mathbf{V}^{(0)}$ ,  $k$  is a conditioning factor and  $\mathbf{D}$  is a diagonal matrix with entries equal to the diagonal elements of  $\mathbf{V}^{(0)T}\mathbf{V}^{(0)}$ . The calculation is started by choosing the conditioning factor  $k^{(0)}= 0.01$  and calculating the  $\mathbf{V}^{(0)T}\mathbf{V}^{(0)}$  where the diagonal elements are the same as the entries of matrix  $\mathbf{D}$ . The result is

$$\mathbf{V}^{(0)T}\mathbf{V}^{(0)} = \begin{pmatrix} 3.61477\text{E-}10 & -1.55239\text{E-}12 & -2.72267\text{E-}16 \\ 1.55239\text{E-}12 & 1.80894\text{E-}10 & -2.49901\text{E-}17 \\ -2.72267\text{E-}16 & -2.49901\text{E-}17 & 2.08856\text{E-}22 \end{pmatrix},$$

which gives

$$\mathbf{D}^{(0)} = \begin{pmatrix} 3.61477\text{E-}10 & 0 & 0 \\ 0 & 1.80894\text{E-}10 & 0 \\ 0 & 0 & 2.08856\text{E-}22 \end{pmatrix},$$

and

$$\mathbf{V}^{(0)T}\mathbf{z}^{(0)} = \begin{pmatrix} -0.000006406772516000000005 \\ 1.37678954399999998\text{E-}7 \\ 4.80986471099999980\text{E-}12 \end{pmatrix}.$$

Therefore the increment  $\boldsymbol{\delta}^{(0)}$  in Equation (6.2) is as follow:

$$\boldsymbol{\delta}^{(0)} = \begin{pmatrix} -8964.79902392720396 \\ 2249.98952546269265 \\ 1.14972112552850761\text{E}10 \end{pmatrix}.$$

Then the new parameter  $\boldsymbol{\theta}^{(1)} = \left( E_0^{(1)}, \sigma^{(1)}, k_0^{(1)} \right)$  can be calculated as

$$\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}_{Use_1}^{(0)} + \boldsymbol{\delta}^{(0)}, \quad (6.3)$$

where  $\boldsymbol{\theta}^{(0)}_{Use_1}$  is equal to the starting parameter estimates ( $\boldsymbol{\theta}^{(0)}$ ) and the increment  $\boldsymbol{\delta}^{(0)}$  from Equation (6.2). The result of Equation (6.3) gives the new parameter  $\boldsymbol{\theta}^{(1)}$  as follows:

$$\boldsymbol{\theta}^{(1)} = \begin{pmatrix} 2.11035200976072811E5 \\ 46249.9895254626899 \\ 2.47972112552850761E10 \end{pmatrix},$$

where  $E_0^{(1)} = 2.11035200976072811E5$ ,  $\sigma^{(1)} = 46249.9895254626899$  and  $k_0^{(1)} = 2.47972112552850761E10$ . Then the new residuals  $\mathbf{z}^{(1)} = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(1)})$  of the new parameters  $\boldsymbol{\theta}^{(1)}$  are calculated to be able to find the new residual sum of squares  $S(\boldsymbol{\theta}^{(1)})$ . The residual  $\mathbf{z}^{(1)}$  is given in the form of matrices as follows:

$$\mathbf{z}^{(1)} = \begin{pmatrix} -0.0378494324, 0.0242510975, 0.0330954475, -0.0178037418, \\ -0.0257222098, 0.0341725915, 0.0074074423, -0.0059564746, \\ -0.0305957182, -0.0908213332, -0.1365770949, -0.1461597594, \\ -0.1614304881, -0.1459237837, -0.1113384887, -0.06822852416, \\ -0.03469009949, -0.01544013354, -0.005547342985, \\ -0.001724366280 \end{pmatrix}^T.$$

Then the residual sum of squares  $S(\boldsymbol{\theta}^{(1)}) = \|\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^{(1)})\|^2$  of  $\mathbf{z}^{(1)}$  is

$$S(\boldsymbol{\theta}^{(1)}) = S_{new} = 0.01237329944.$$

Therefore  $S(\boldsymbol{\theta}^{(1)}) = S_{new} = 0.01237329977 < S(\boldsymbol{\theta}^{(0)}) = s_{old} = 0.1204338066$ . This process is repeated until convergence is obtained. That is, until the

increment is so small that there is no useful change in the elements of the parameter vector. Also, the relative change in the sum of squares on successive iterations is small. From Example 2, the Levenberg-Marquardt Algorithm does agree with both of the criteria above which are used to declare convergence. This means that when the iterative process is repeated, the request increment  $\delta^{(n)}$  minimizes the residual sum of squares  $S(\boldsymbol{\theta}^{(n)})$  and approaches closer to the region where the linear approximation is valid.

The plot of the leading order of the asymptotic result by using the initial parameter estimates  $\boldsymbol{\theta}^{(0)}$  and the new parameter values are illustrated in Figure 6.1. The values of  $\theta^{(0)}=(220000, 44000, 1.33\text{E}10)^T$  and  $\theta^{(1)}=(2.11035200976072811, 46249.9895254626899, 2.47972112552850761\text{E}10)^T$ .

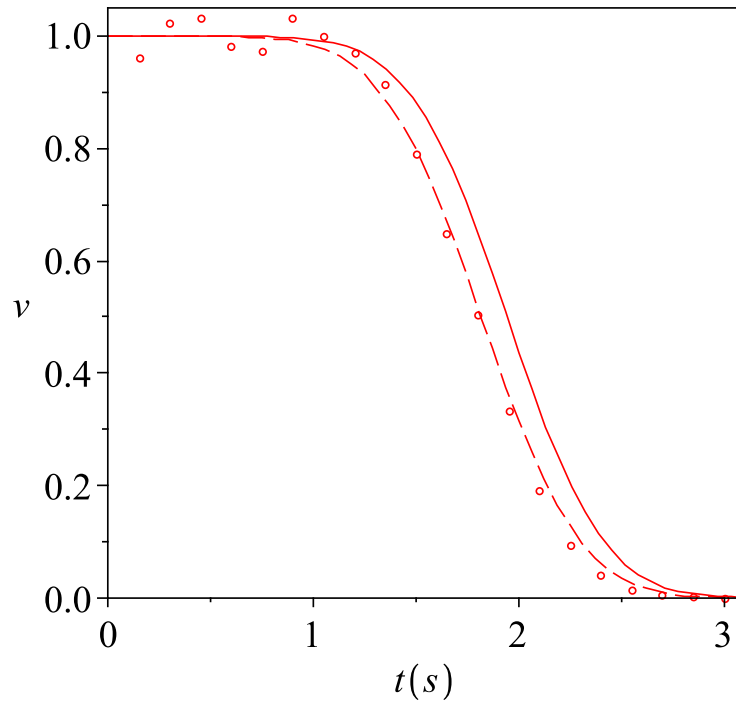


Figure 6.1: Comparison of normalized fractional yield not yet released ( $v$ ) varied with time in seconds as in Example 2. The circles are data, the solid line is the initial guessed solution ( $\theta^{(0)}$ ), and the dashed line is the result of one iteration of Levenberg-Marquardt Algorithm ( $\theta^{(1)}$ ).

The circles illustrated in Figure 6.1 indicate the data from the random number generator. This fit can provide a good approximation of whether the fitting graphs are approaching the refinement region or where it starts to move away from the converged region. The solid line shows the result of using the starting parameter estimates  $\theta^{(0)}$  and the dashed line is the result obtained by using the improved parameters from applying the Levenberg-Marquardt Algorithm. It can be seen from the graphs in Figure 6.1 that applying the Levenberg-Marquardt Algorithm does improve the fitting of the graph. The graph of the new parameters moved towards the refinement region. This result shows that the increment is becoming smaller and it minimizes the change in the elements of the parameter vector.

### 6.3 Computer Programme

This thesis has described the computer program created in order to provide guidance in implementing and speeding the above results (Example 2), and will be used in the later examples. All calculations and computing programs used the Maple programme, which has built in support for the LambertW function. Note that we still use the leading order solution ( $v^{(1)}$ ) of Equation (3.16) to provide the estimate  $f_0$  which is Gaussian in Example 3.

#### 6.3.1 Example 3

From running the programme in Appendix 2, the results from  $MakeIteration(2.20E5, 44E3, 1.33E10, 5)$  are shown in Table 6.1 below.

$\theta^{(n)}$	$E_n$	$\sigma_n$	$k_n$	$S(\theta^{(n)})$
$\theta^{(0)}$	0.220E6	0.44E5	1.33E10	0.1204338066
$\theta^{(1)}$	211035.201	46249.9895	2.47972118E10	0.01237329944
$\theta^{(2)}$	210310.815	42077.4605	2.99977767E10	0.006463076506
$\theta^{(3)}$	212992.377	44346.2851	5.45149898E10	0.009266540071
$\theta^{(4)}$	214391.234	42993.6278	4.64972779E10	0.006461018969
$\theta^{(5)}$	214344.064	42968.0639	4.70126252E10	0.006452457778

Table 6.1: The results from  $MakeIteration(2.20E5, 44E3, 1.33E10, 5)$ .

The analysis of the parameters above illustrated that the first two iterations gave better estimates for each parameter and caused the decrease of sum of squares. The conditioning factor  $k$  is divided by a factor of 10 where  $k \rightarrow 0$ . This result shows that the Levenberg-Marquardt direction [i.e. the direction of  $\delta$ ] is estimated as the direction of the Gauss-Newton increment. However the third iteration shows the sum of squares started to increase its value. To avoid the increase in the sum of squares, the conditioning factor  $k$  needs to increase by a factor of 10 to push the algorithm closer again to the Gauss-Newton direction. Then the fourth and fifth iteration shows the improvement in the sum of squares is now smaller than the previous sum of



squares. This result shows that when  $N$  is large the sum of squares becomes smaller and smaller. The Levenberg-Marquardt iterative method is repeated until the results satisfy Equation (6.4) below.

$$\frac{S_{old} - S_{new}}{S_{new}} \leq 10^{-5} \quad (6.4)$$

At this point the Levenberg-Marquardt algorithm has converged. Note that the *MakeIteration* in Example 3 uses the same equations and initial parameter values as in Example 2 but runs up to fifth iteration. The parameter values from the *MakeIteration* results are used in Equation (5.5) to calculate the leading order and the results are plotted and shown in Figure (6.2) below.

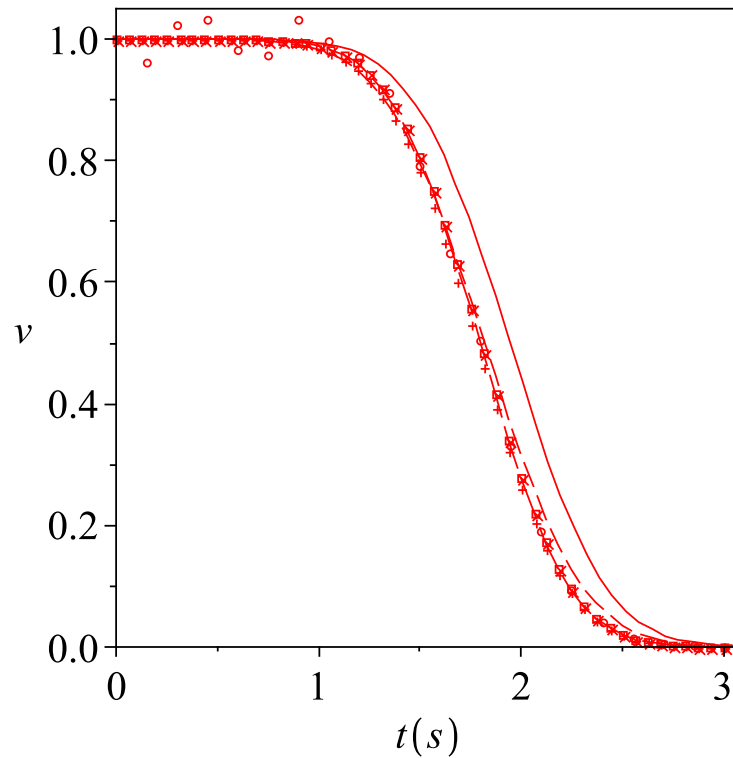


Figure 6.2: Comparison of normalized fractional yield not yet released ( $v$ ) varied with time in seconds as in Example 3. The circles are data, the solid line is the initial guessed solution ( $\theta^{(0)}$ ), the dashed line is the result of first iterative parameters ( $\theta^{(1)}$ ), the diagonal crosses indicate the second iterative parameters ( $\theta^{(2)}$ ), the crosses indicate the third iterative parameters ( $\theta^{(3)}$ ), the boxes indicate the fourth iterative parameters ( $\theta^{(4)}$ ) and the longdash line indicates the fifth iterative parameters ( $\theta^{(5)}$ ). These parameters are shown in Table 6.1.

Figure 6.2 is hard to analyse because the parameter values are close to each other. When the parameters are close to each other, limitations of the computer screen resolution make it difficult to see what is going on. It is then useful to plot the residuals, magnified as appropriate, and to base the parameter adjustment on the residuals. The residuals  $\mathbf{z}^{(n)}$ , provide information regarding assumptions about error terms and the appropriateness of the

model. Figure 6.3 shows the plot of residuals from Figure 6.2 and uses the same symbols.

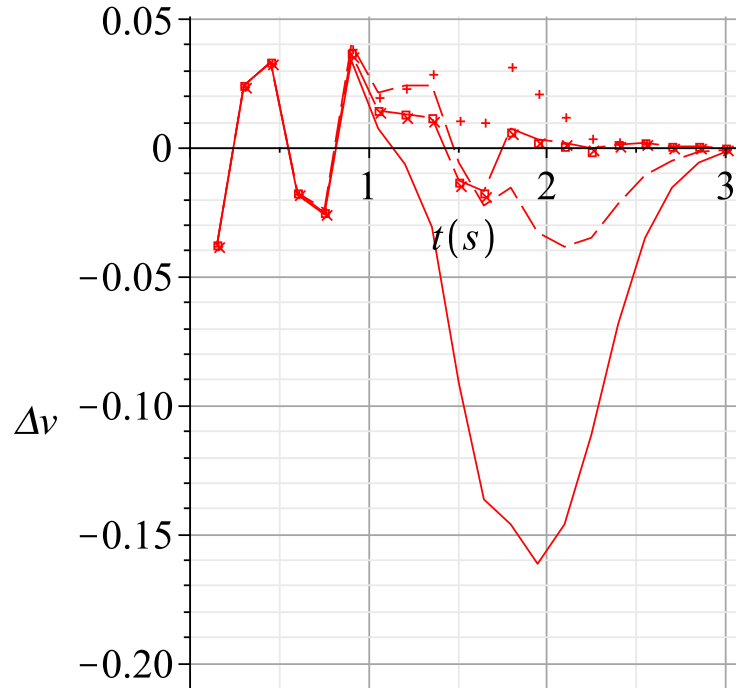


Figure 6.3: Comparison of the differences between parameter values obtained for  $v$  from the *MakeIteration* parameters result. These residuals are calculated from the parameter values that were used in Figure 6.2. The solid line is the initial guessed solution ( $\theta^{(0)}$ ), the dashed line is the result of first iterative parameters ( $\theta^{(1)}$ ), the diagonal crosses indicate the second iterative parameters ( $\theta^{(2)}$ ), the crosses indicate the third iterative parameters ( $\theta^{(3)}$ ), the boxes indicate the fourth iterative parameters ( $\theta^{(4)}$ ) and the longdash line indicates the fifth iterative parameters ( $\theta^{(5)}$ ). These parameters are shown in Table 6.1.

The starting parameter estimates graph (the solid line) in Figure 6.2 and Figure 6.3 shows the improvement achieved by applying the iterative method because all the new parameters  $\theta^{(n)}$  are moved closer to the data than the graph of the starting parameter estimates. The closest parameters to the

refinement region are illustrated by the longdash graph in both plots. Therefore we prove that the present method is working to data which has been generated with a Gaussian distribution. The next example is trying to fit data which has been digitized from published coal volatilization experiments.

## 6.4 Digitized Data

### 6.4.1 Example 4

So far we have seen that the Levenberg-Marquardt Algorithm can be successfully fitted  $v^{(1)}$  to the data which was generated using a Gaussian distribution (eg. the randomly generated data in Examples 2 and 3). However, it remains necessary to demonstrate that this algorithm can apply to data with non Gaussian distributions. Therefore, we also run the algorithm using  $v^{(1)}$  to fit data which has been digitized from published coal volatilization experiments. Note that when working with digitized actual data the distribution is unknown. Using digitized data demonstrates that this present method can apply to coal data with an unknown distribution, even if the initial guess for that  $f_0$  is Gaussian. Again the Levenberg-Marquardt Algorithm for non-linear least squares with the leading order solution ( $v^{(1)}$ ) of Equation (3.16) provides the estimate  $f_0$  which is Gaussian, is used to fit actual coal data. The results from fitting to digitized data for Total volatiles yield[wt, %] versus Peak temperature [K] at the heating rate of 1060K/s with  $V^*= 43.45\%$  of original coal is given in Table 6.2 [52].

Total Volatiles Yield [wt,%]	Peak Temperature [K]
0.284736	565.656066
2.644009	840.393492
4.353333	1029.297917
9.312743	1127.510375
21.472291	1191.954415
27.564722	1239.509385
33.829471	1307.397949
39.910990	1363.065836
41.787696	1402.774176
43.280967	1430.661257
43.352204	1476.294889

Table 6.2: Results from digitized data for Total volatiles yield[wt,%] versus Peak temperature [K] at the heating rate of 1060K/s with  $V^* = 43.45\%$  of original coal.

The original data in Table 6.2 [52] is converted to the fraction of the volatile not yet released varied with time. This is illustrated in Table 6.3, by applying the known equations of  $T = mt$  and  $v = 1 - (V/V^*)$  to calculate  $t$  and  $v$  as shown below.

Fractional volatiles yield not yet released	Time [s]
0.9934468124	0.5336377981
0.9391482394	0.7928240491
0.8998082163	0.9710357708
0.7856675949	1.063689033
0.5058160875	1.124485297
0.3655990334	1.169348476
0.2214160875	1.233394292
0.814501726E-1	1.285911166
0.382578596E-1	1.323371864
0.38902877E-2	1.349680431
0.22507710E-2	1.392731027

Table 6.3: Data of fractional volatiles yield not yet released varied with time, calculated from Table 6.2 at the heating rate of 1060K/s with  $V^* = 43.45\%$  of original coal.

To illustrate these calculations, consider the data from Table 6.3 with the starting parameter estimates  $\theta^{(0)} = (234.8E3, 29.7E3, 6.18E10)^T$ . Note that we are using the same equations and procedure as in Examples 2 and Example 3 for the calculation of residuals and sum of squares in this example. The data, along with the fitted values, and the residuals are shown in Table 6.4

$\mathbf{n}$	$\mathbf{t}_n$	$\mathbf{v}_n$	$\boldsymbol{\eta}_n^0$	$\mathbf{z}_n^0$
1	0.5336377981	0.9934468124	0.9999973670	-0.65505546E-2
2	0.7928240491	0.9391482394	0.9977228085	-0.585745691E-1
3	0.9710357708	0.8998082163	0.9496807825	-0.498725662E-1
4	1.063689033	0.7856675949	0.8453222390	-0.596546441E-1
5	1.124485297	0.5058160875	0.7274382290	-0.2216221415
6	1.169348476	0.3655990334	0.6182280855	-0.2526290521
7	1.233394292	0.2214160875	0.4465592773	-0.2251431898
8	1.285911166	0.814501726E-1	0.3113886367	-0.2299384641
9	1.323371864	0.382578596E-1	0.2274286224	-0.1891707628
10	1.349680431	0.38902877E-2	0.1769876180	-0.1730973303
11	1.392731027	0.22507710E-2	0.1110328639	-0.1087820929

Table 6.4: The data, fitted values and residuals at  $\boldsymbol{\theta}^{(0)} = (234.8\text{E}3, 29.7\text{E}3, 6.18\text{E}10)^T$ .

The derivatives evaluated at  $\boldsymbol{\theta}^0$  are shown in Table 6.5. Collecting these derivatives into the derivative matrix  $\mathbf{V}^{(0)}$ , the Levenberg-Marquardt Algorithm formulae is then applied to calculate the increment  $\boldsymbol{\delta}^{(0)}$  using Equation (6.1)

$$\boldsymbol{\delta}^0 = (\mathbf{V}^{(0)T}\mathbf{V}^{(0)} + k^{(0)}\mathbf{D}^{(0)})^{-1} (\mathbf{V}^{(0)T}\mathbf{z}^{(0)})$$

by letting  $k^{(0)} = 0.01$ . In this case,

$$\boldsymbol{\delta}^{(0)} = \begin{pmatrix} -8937.59330600000067 \\ -829.923865999999976 \\ 5.287713190000000000\text{E}10 \end{pmatrix},$$

and the sum of squares at  $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(0)} + \boldsymbol{\delta}^{(0)}$  is  $S(\boldsymbol{\theta}^{(1)}) = S_{new} = 0.02659909427\text{E}-1$ , which is much smaller than  $S(\boldsymbol{\theta}^{(0)}) = S_{old} = 0.3036006111$ . Then this result

$$\boldsymbol{\theta}^{(1)} = \begin{pmatrix} 2.25862406694067497\text{E}5 \\ 28870.0761340151948 \\ 1.14677131899214600\text{E}11 \end{pmatrix}$$

<b>n</b>	$\mathbf{V}_{n1}^0$	$\mathbf{V}_{n2}^0$	$\mathbf{V}_{n3}^0$
1	4.216347834E-10	-1.920086173E-9	-3.063962406E-17
2	2.401405379E-7	-6.812716081E-7	-2.594697586E-14
3	0.000003490242871	-0.000005730163299	-4.620692899E-13
4	0.000008012133317	-0.000008144936606	-1.162131660E-12
5	0.00001118537030	-1.715310931E-12	-1.715310931E-12
6	0.00001283813753	-0.000003862101996	-2.047467848E-12
7	0.00001331170208	0.000001788549442	-2.239497777E-12
8	0.00001190166150	0.000005854642191	-2.087699293E-12
9	0.00001015950570	0.000007592623646	-1.834118490E-12
10	0.000008741590420	0.000008102634395	-1.609572036E-12
11	0.000006373723004	0.000007782657387	-1.211086250E-12

Table 6.5: The derivatives of  $p$ ,  $r$  and  $s$  evaluated at  $\theta^{(0)} = (234.8\text{E}3, 29.7\text{E}3, 6.18\text{E}10)^T$ .

is used to perform another iteration. This process is repeated until convergence is obtained which is determined by Equation (6.4).

The *MakeIteration* programme in Appendix 2 was then run for  $N=3$  and the results are given in Table 6.6. Note the data in Table 6.3 is used in the *MakeIteration* programme.

	$\theta^{(1)}$	$\theta^{(2)}$	$\theta^{(3)}$
$E_0$	2.25862406694067497E5	225299.399463057256	225128.4315
$\sigma$	28870.0761340151948	22930.8042912352830	22935.75770
$k_0$	1.14677131899214600E11	1.23064584957803650E11	1.18018062E11
$S(\theta)$	0.02659909277	0.009649574628	0.009597161809

Table 6.6: The parameters  $\theta^{(n)}$  when  $N=3$  where  $\theta^{(0)} = (0.2348\text{E}6, 0.297\text{E}5, 0.618\text{E}11)^T$  and  $S(\theta^{(0)}) = 0.3036006111$ .

The parameter results in Table 6.6 are plotted with the normalized fractional volatiles yield not yet released versus time, as shown in Figure 6.4.



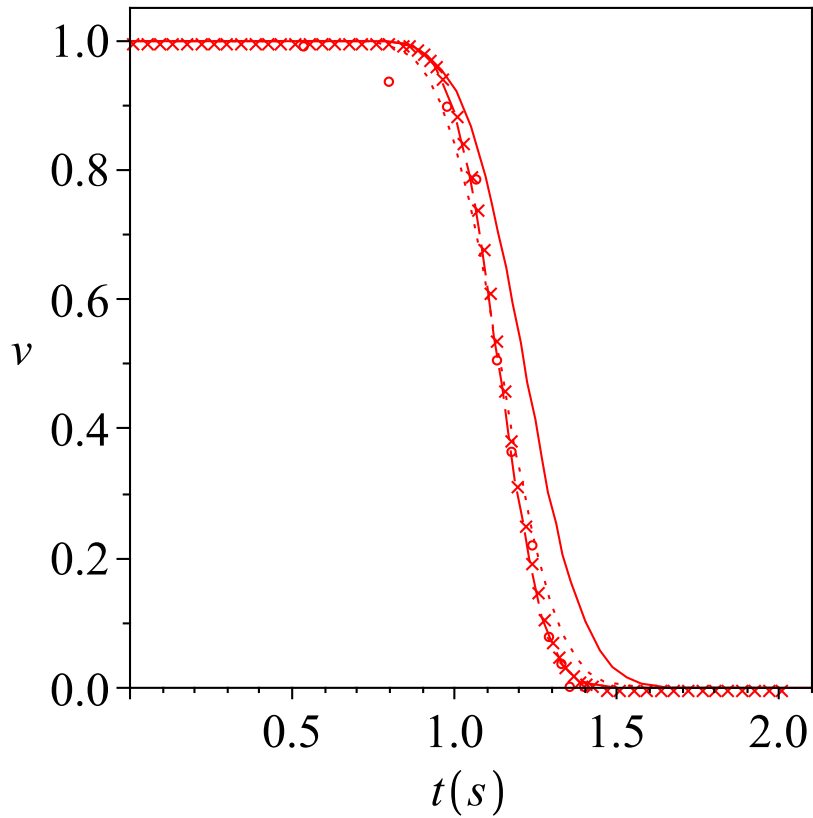


Figure 6.4: Normalized fractional yield not yet released ( $v$ ) varied with time in seconds is obtained from the result of Table 6.6. The circles are data from Table 6.3, the solid line is the initial guessed solution ( $\theta^{(0)}$ ), the dotted line is the result of first iterative parameters ( $\theta^{(1)}$ ), the crosses indicate the second iterative parameters ( $\theta^{(2)}$ ) and the dashed line indicates the third iterative parameters ( $\theta^{(3)}$ ).

Similarly to Example 3, this plot is hard to analyse because parameter values are close to each other. It is then useful to once again plot the residuals, magnified as appropriate, and to base the parameter adjustment on the residuals. The plot of residuals in Figure 6.5 is able to distinguish the closeness of the parameter values. Note that the closest parameters to the refinement region are illustrated by the dashed line graph in both plots.

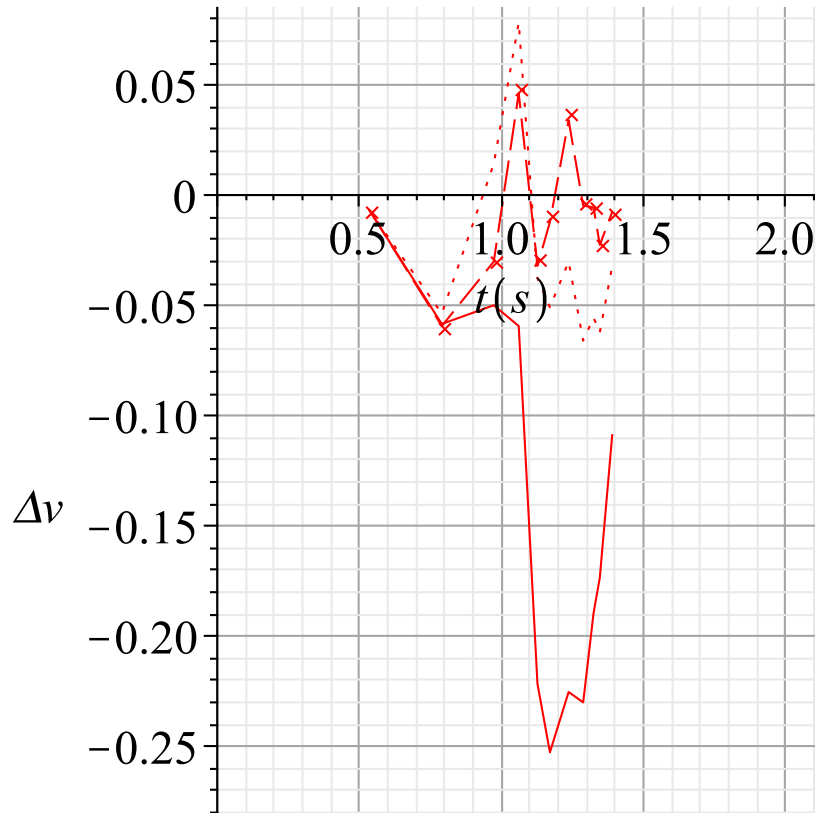


Figure 6.5: Residuals of the plots in Figure 6.4, showing differences between parameter values obtained for  $v$  versus time. The solid line is the initial guessed solution ( $\theta^{(0)}$ ), the dotted line is the result of first iterative parameters ( $\theta^{(1)}$ ), the crosses indicate the second iterative parameters ( $\theta^{(2)}$ ) and the dashed line indicates the third iterative parameters ( $\theta^{(3)}$ ). These parameters are shown in Table 6.6.

One of the most important plots in the regression is the plot of residual  $\mathbf{z}^{(n)}$  versus the fitted values  $\eta_n^{(n)}$ . Systematic features in this plot are of interest. Curvature might indicate that the fitted model is inappropriate, and might suggest a transformation of the data. Residuals that seem to increase or decrease with the fitted values might indicate nonconstant residual variance. A few relatively large residuals may be indicative of outliers-cases for which the model is somehow inappropriate. On the other hand, if the plot

of  $\mathbf{z}^{(n)}$  versus  $\eta_n^{(n)}$  shows no systematic features, then we would have little reason to suspect that the fitted model was inappropriate for the data.

For example, the fitted values  $\eta_n^{(3)}$  and residuals  $\mathbf{z}_n^{(3)}$  for the third iteration ( $N = 3$ ) are given in Table 6.7 and are plotted against each other in Figure 6.6. Notice that the residuals are generally small compared to the fitted values and that they do not suggest any distinct pattern in Figure 6.6. All residuals are typically less than 0.06 in absolute value. This suggests that the results from analyses of this plot in Figure 6.6, agree with the choice of the iterative method.

$\mathbf{n}$	$\mathbf{t}_n$	$\mathbf{v}_n$	$\eta_n^3$	$\mathbf{z}_n^3$
1	0.5336377981	0.9934468124	0.9999999555	-0.65531431E-2
2	0.7928240491	0.9391482394	0.9989062100	-0.597579706E-1
3	0.9710357708	0.8998082163	0.9296731900	-0.298649737E-1
4	1.063689033	0.7856675949	0.7394697830	0.461978119E-1
5	1.124485297	0.5058160875	0.5375941145	-0.317780270E-1
6	1.169348476	0.3655990334	0.3781659776	-0.125669442E-1
7	1.233394292	0.2214160875	0.1869741436	0.344419439E-1
8	1.285911166	0.814501726E-1	0.8618463575E-1	-0.473446315E-2
9	1.323371864	0.382578596E-1	0.4416126052E-1	-0.590340092E-2
10	1.349680431	0.38902877E-2	0.2600163112E-1	-0.2211134342E-1
11	1.392731027	0.22507710E-2	0.9790863335E-2	-0.7540092335E-2

Table 6.7: The data, fitted values and residuals at  $\boldsymbol{\theta}^{(3)} = (225128.431, 22935.7578, 1.1801806E11)^T$  for the third iteration ( $N = 3$ ).

The plot of residuals  $\mathbf{z}^{(3)}$  versus fitted values  $\eta^{(3)}$  for the third iteration ( $N = 3$ ) is shown below.

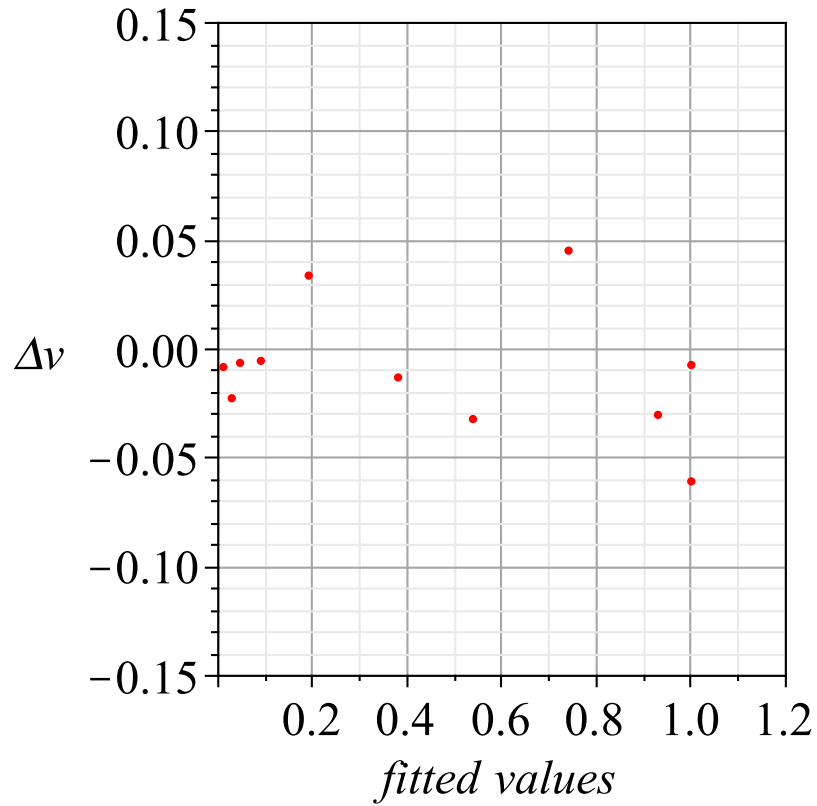


Figure 6.6: Residuals versus fitted values for digitized data from the third iteration of the Levenberg-Marquardt Algorithm.

This iterative process satisfies Equation (6.4) when  $N=7$  and the sum of squares  $S(\theta^{(7)})=0.009596682368$  with parameter values

$$\theta^{(7)} = \begin{pmatrix} 222004.3158 \\ 22623.51229 \\ 8.52305343E10 \end{pmatrix}.$$

Calculating Equation (6.4) gives  $4.173317243E-7$  which is less than  $E-5$  where  $S(\theta^{(6)})=0.009596686373$ . Therefore the iterative process converges at  $S(\theta^{(7)})$ . Now we have also proved that the present method is working to data which has been digitized from published coal volatilization experiments.

## 6.5 Higher-Order Correction

We move on to the next stage of reducing the residuals by adding the higher-order correction term  $\epsilon f_1$  which is non Gaussian to the leading-order term  $f_0$  which is Gaussian of  $v^{(1)}$  and repeating the iterative process shown in Examples 2, 3 and 4 above, by using the digitized data of Table 6.3 in Example 4. It is necessary to fit  $v^{(1)}$  to the data in order to smooth it in preparation for taking derivatives to obtain higher-order approximations to the underlying distribution. Fitting the high-order correction term ( $f_0 + \epsilon f_1$ ) to data improves the calculation of the residuals and provides the sum of squares which is smaller than the sum of squares when fitting the leading-order solution ( $f_0$ ). Therefore it provides a more accurate estimate of the underlying distribution in the DAEM.

The equation used to fit the data is defined as:

$$v^{(2)} \sim \int_{y_s}^{\infty} (f_0 + \epsilon f_1) dy$$

where

$$f_0 = -\frac{dv^{(1)}/dt}{dE_s/dt},$$

and

$$\epsilon f_1 = A_0 \left( E_w \frac{df_0}{dE_s} + \frac{\frac{dE_w}{dt} f_0}{\frac{dE_s}{dt}} \right),$$

then

$$v^{(2)} \sim \int_{y_s}^{\infty} \left( -\frac{dv^{(1)}/dt}{dE_s/dt} + A_0 \left( E_w \frac{df_0}{dE_s} + \frac{\frac{dE_w}{dt} f_0}{\frac{dE_s}{dt}} \right) \right) dy. \quad (6.5)$$

The next iterative process is based on adding the higher-order correction term  $\epsilon f_1$  to the leading-order term  $f_0$  which gives an improved formula  $v^{(2)}$  in Equation (6.5) which can be used to fit the data. Equation (6.5) improves the result of  $v^{(1)}$  to more accurately approximate parameter values since it

provides a sum of squares which will be much smaller than in the previous iterative process. Note that if we add the higher-order correction terms  $\epsilon^2 f_2$  in Equation (3.21) and  $\epsilon^3 f_3$  in Equation (3.22), the results from the iterative process will provide a more accurate estimate of the underlying distribution in the DAEM.

We change the variable of integration in Equation (6.5) to the time  $t$ . Approximations are sought to the integral

$$v^{(2)} \sim \int_{t_s}^{\infty} \left( -\frac{dv^{(1)}/dt}{dE_s/dt} + A_0 \left( E_w \frac{df_0}{dE_s} + \frac{\frac{dE_w}{dt} f_0}{\frac{dE_s}{dt}} \right) \right) \frac{K}{E_0} dt$$

where

$$K = RmY(k_0 t) + \frac{RmY(k_0 t)}{1 + Y(k_0 t)},$$

and

$$t_s = \frac{tY(k_0 t)}{2Y\left(\frac{1}{2\sqrt{\frac{1}{k_0 t Y(k_0 t)}}}\right)}.$$

Then  $v^{(2)}$  is now rewritten as  $v^{(2*)} = 1 - v^{(2)}$  to form the equation of the fraction of the volatile yield not yet released, and the problem becomes

$$v^{(2*)} \sim 1 - \int_{t_s}^{\infty} \left( -\frac{dv^{(1)}/dt}{dE_s/dt} + A_0 \left( E_w \frac{df_0}{dE_s} + \frac{\frac{dE_w}{dt} f_0}{\frac{dE_s}{dt}} \right) \right) \frac{K}{E_0} dt \quad (6.6)$$

where  $\frac{dv^{(1)}}{dt}$  is the derivative of Equation (5.5) with respect to  $t$  and can be written in the form

$$\frac{dv^{(1)}}{dt} = -\frac{1}{2\sqrt{\pi}} e^{-\frac{E_0^2 \left( \frac{RmY(k_0 t)}{E_0} - 1 \right)^2}{2\sigma^2}} \sqrt{2} \sqrt{\frac{E_0^2}{\sigma^2}} \left( \frac{RmY(k_0 t)}{E_0} + \frac{RmY(k_0 t)}{(1 + Y(k_0 t)) E_0} \right).$$

The iterative calculation of  $v^{(2*)}$  in Equation (6.6) uses the same Levenberg-Marquardt Algorithm procedure to fit the data as described in Examples 2, 3 and 4. Note that the starting parameter estimates  $\theta^{(0)}$  used in  $v^{(2*)}$  are chosen

to be the same as the last parameter values that were obtained from the last iteration of  $v^{(1)}$ . That is, when the iteration of  $v^{(1)}$  reaches convergence. Example 5 also uses Equation (6.4) to determine the convergence of the iterative process.

### 6.5.1 Example 5

Consider the data from Table 6.3 of Example 4, with the starting parameter estimates  $\theta^{(0)} = (222004.3158, 22623.51229, 8.52305343E10)^T$ . This  $\theta^{(0)}$  is the result of the last parameter values  $\theta^{(7)}$  from the fitting of  $v^{(1)}$  to data. The data, along with the fitted values, and residuals evaluated at  $\theta^{(0)}$  are shown in Table 6.8.

$\mathbf{n}$	$\mathbf{t}_n$	$\mathbf{v}_n$	$\boldsymbol{\eta}_n^{(0)}$	$\mathbf{z}_n^{(0)}$
1	0.5336377981	0.9934468124	0.99999999560	-0.65531714E-2
2	0.7928240491	0.9391482394	0.9989074760	-0.603817531E-1
3	0.9710357708	0.8998082163	0.9296810465	-0.580153853E-1
4	1.063689033	0.7856675949	0.7394189045	-0.280635149E-1
5	1.124485297	0.5058160875	0.5374672665	-0.1277459033
6	1.169348476	0.3655990334	0.3779949848	-0.1080433468
7	1.233394292	0.2214160875	0.1867994252	-0.366582035E-1
8	1.285911166	0.814501726E-1	0.8605465835E-1	-0.480675215E-1
9	1.323371864	0.382578596E-1	0.4407174470E-1	-0.323614407E-1
10	1.349680431	0.38902877E-2	0.2593793006E-1	-0.395573097E-1
11	1.392731027	0.22507710E-2	0.9759028450E-2	-0.153295219E-1

Table 6.8: The data, fitted values and residuals at  $\theta^{(0)} = (234.8E3, 29.7E3, 6.18E10)^T$ .

The derivatives of  $p_1$ ,  $r_1$  and  $s_1$  are evaluated at  $\theta^{(0)}$  where  $p_1$  is the derivative of  $v^{(2*)}$  with respect to  $E_0$ ,  $r_1$  is the derivative of  $v^{(2*)}$  with respect to  $\sigma$  and  $s_1$  is the derivative of  $v^{(2*)}$  with respect to  $k_0$ , is shown in Table 6.9.

<b>n</b>	$\mathbf{V}_{n1}^{(0)}$	$\mathbf{V}_{n2}^{(0)}$	$\mathbf{V}_{n3}^{(0)}$
1	4.178140010E-12	-2.358716647E-11	7.847513483E-16
2	7.710887433E-8	-2.637667615E-7	-6.030924213E-15
3	0.000004122810334	-0.000007318581498	-3.953867893E-13
4	0.00001224669160	-0.00001114165898	-1.287164909E-12
5	0.00001715663740	-1.906793774E-12	-1.906793774E-12
6	0.00001811656197	0.000001401951445	-2.019402747E-12
7	0.00001467294472	0.000009904810303	-1.789429679E-12
8	0.000009565241504	0.00001113978581	-1.216390203E-12
9	0.000006121034139	0.000009264585494	-8.011520630E-13
10	0.000004169006449	0.000007331267085	-5.565427790E-13
11	0.000001960655486	0.000004233369061	-2.702450672E-13

Table 6.9: The derivatives of  $p_1$ ,  $r_1$  and  $s_1$  evaluated at  $\theta^{(0)} = (222004.3158, 22623.51229, 0.852305343E11)^T$  where  $p_1$  is the derivative of  $v^{(2*)}$  with respect to  $E_0$ ,  $r_1$  is the derivative of  $v^{(2*)}$  with respect to  $\sigma$  and  $s_1$  is the derivative of  $v^{(2*)}$  with respect to  $k_0$ .

The derivatives in Table 6.9 are collected in a derivative matrix  $\mathbf{V}^{(0)}$  for which  $\mathbf{D}$  and  $\mathbf{V}^{(0)T}$  are generated. Then  $\boldsymbol{\delta}^{(0)}$  was solved using the Levenberg-Marquardt Algorithm in Equation (6.1). In this case,  $\boldsymbol{\delta}^{(0)} = (-3076.69856857, 405875, 810.148941006380483, 1.98257653671400909E10)^T$  and the sum of squares at  $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(0)} + \boldsymbol{\delta}^{(0)}$  is  $S(\boldsymbol{\theta}^{(1)}) = S_{new} = 0.9863897792E-2$ , which is much smaller than  $S(\boldsymbol{\theta}^{(0)}) = S_{old} = 0.4233597490E-1$ . Then therefore  $\boldsymbol{\theta}^{(1)} = (2.18927617231425946E5, 23433.6612310063792, 1.05056299667140091E11)^T$  and another iteration was performed.

The iterative process of the *MakeIteration* programme in Appendix 2 was then run for  $N = 3$  and obtained the parameter values shown in Table 6.10. Note the change in the *MakeIteration* programme of equation, was from  $v^{(1)}$  to  $v^{(2*)}$ . The digitized data in Table 6.3 was used.



	$\boldsymbol{\theta}^{(1)}$	$\boldsymbol{\theta}^{(2)}$	$\boldsymbol{\theta}^{(3)}$
$E_0$	218927.617231425946	218072.400530257641	218079.199449991924
$\sigma$	23433.6612310063792	22775.2147499209823	22831.5708550212548
$k_0$	1.0505629966714E11	9.9671137072299E10	9.9574500697115E10
$S(\boldsymbol{\theta})$	0.9863897792E-2	0.9582835743E-2	0.9577028426E-2

Table 6.10: The parameters  $\boldsymbol{\theta}^{(n)}$  when  $N=3$  where  $\boldsymbol{\theta}^{(0)} = (222004.3158, 22623.51229, 0.852305343E11)^T$  and  $S(\boldsymbol{\theta}^{(0)}) = 0.4233597490E-1$ .

The parameter values from Table 6.10 are indistinguishable from plots of the calculation of  $v$ , as illustrated in Figure 6.7 shown below.

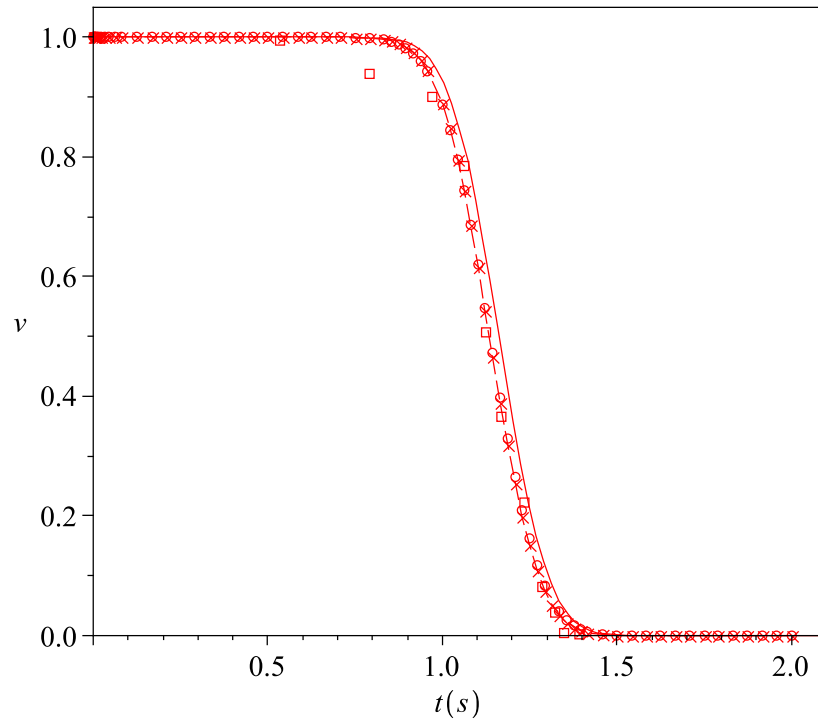


Figure 6.7: Normalized fractional yield not yet released ( $v$ ) varied with time in seconds is obtained from the result of Table 6.5. The boxes are data, the solid line is the initial guessed solution ( $\theta^{(0)}$ ), the circles are the result of first iterative parameters ( $\theta^{(1)}$ ), the crosses indicate the second iterative parameters ( $\theta^{(2)}$ ) and the dashed line indicates the third iterative parameters ( $\theta^{(3)}$ ). These parameters are shown in Table 6.10.

The plot of the residuals versus time in Figure 6.8 shows clearly which plotted residuals have the smaller sum of squares at  $\theta^{(n)}$ . The smaller the sum of squares at  $\theta^{(n)}$ , the smaller the amplitude obtained on the plot of residuals. Note that the parameter which contains the smallest sum of squares is illustrated by the dash line graph in both plots.

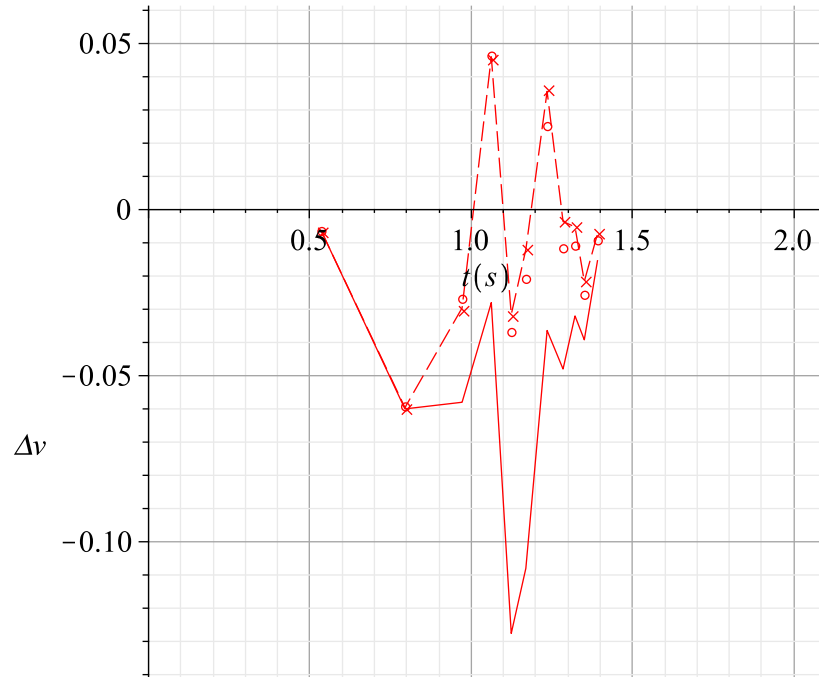


Figure 6.8: Residuals of the plots in Figure 6.7, showing differences between parameter values obtained for  $v$  versus time in Example 5. The solid line is the initial guessed solution ( $\theta^{(0)}$ ), the circles are the result of first iterative parameters ( $\theta^{(1)}$ ), the crosses indicate the second iterative parameters ( $\theta^{(2)}$ ) and the dashed line indicates the third iterative parameters ( $\theta^{(3)}$ ). These parameters are shown in Table 6.10.

The iterative process satisfies Equation (6.4) when  $N=7$  and the sum of squares  $S(\theta^{(7)})=0.009575375913$  where the parameter value of

$$\boldsymbol{\theta}^{(7)} = \begin{pmatrix} 217815.530752700288 \\ 22850.3934920922802 \\ 9.68677057029868622\text{E}10 \end{pmatrix}.$$

Calculating Equation 6.4 gives  $5.405531905\text{E}-7$  which is less than  $\text{E}-5$  where  $S(\theta^{(6)})=0.009575381089$ . Therefore the iterative process converges at  $S(\theta^{(7)})$ .

## 6.6 Comparison of $v^{(1)}$ and $v^{(2*)}$

Comparisons of the two fitted equations  $v^{(1)}$  and  $v^{(2*)}$  illustrate that the results obtained by adding the higher-order correction term  $\epsilon f_1$  to the leading-order term  $f_0$  provide an improved estimate of the underlying distribution. This can be seen from the results of the sum of squares at  $\theta^{(n)}$  and can also be illustrated clearly in the plotting of the residuals versus time graph. For example if we let  $N=7$ , the results are illustrated in Table 6.11 below.

	$\theta^{(7)}$ in $v^{(1)}$	$\theta^{(7)}$ in $v^{(2*)}$
$S(\theta^{(7)})$	0.009596682368	0.009575375913
$E_0$	222004.3158	217815.530752700288
$\sigma$	22623.51229	22850.3934920922802
$k_0$	8.52305343E10	9.68677057029868622E10

Table 6.11: The parameters  $\theta^{(7)}$  when  $N=7$  in both fitting equations  $v^{(1)}$  and  $v^{(2*)}$ .

The sum of squares in Table 6.11 illustrates that  $v^{(2*)}$  provides a more accurate estimate because the  $S(\theta^{(7)})$  in  $v^{(2*)}$  is smaller than  $S(\theta^{(7)})$  in  $v^{(1)}$ . Note that both iterative processes showed convergence when  $N=7$ . This means that  $S(\theta^{(7)})$  in  $v^{(1)}$  and  $v^{(2)}$  are both satisfied in Equation (6.4). The other alternative, which is more appropriate for comparing  $v^{(1)}$  and  $v^{(2*)}$  uses the parameter values of the smallest sum of squares in Table 6.11, that is  $\theta = \theta^{(7)}$  in  $v^{(2*)}$ .

The plot of the residual results versus time is illustrated in Figure 6.9 with the crosses being  $v^{(1)}$  calculations, and the solid circles being those of  $v^{(2*)}$  calculations. It can be seen that the solid circles provided smaller residuals than the crosses. This means that  $v^{(2*)}$  provides a more accurate estimate of parameters than  $v^{(1)}$ .

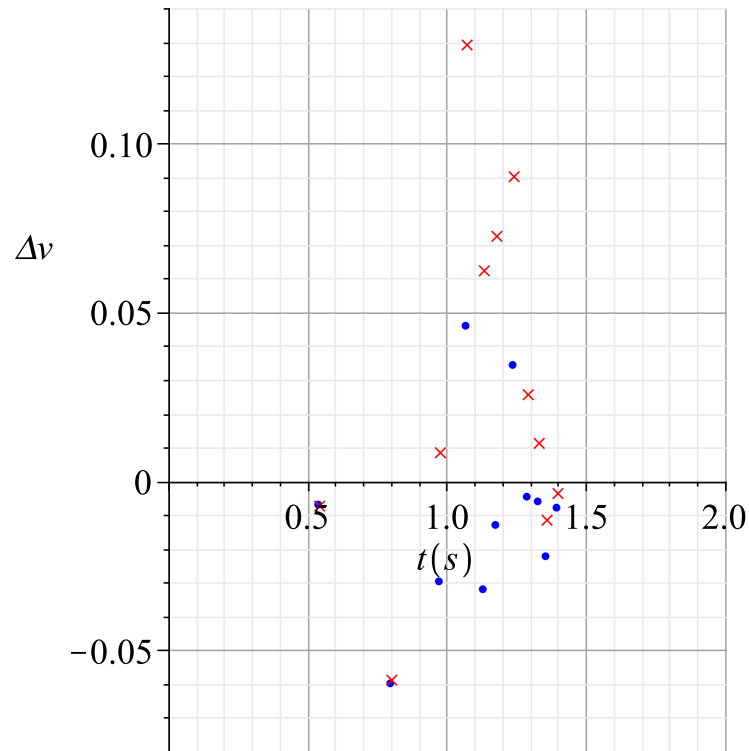


Figure 6.9: Comparison between residual results of  $v^{(1)}$  and  $v^{(2*)}$  at  $\theta = (217815.530752700288, 22850.3934920922802, 9.68677057029868622E10)^T$ . The crosses are  $v^{(1)}$  calculations, and the solid circles are of  $v^{(2*)}$  calculations.

The second plot shows the residual results versus fitted values from Figure 6.9 with the  $v^{(1)}$  results shown as crosses and  $v^{(2*)}$  results shown as solid circles. Figure 6.10 shows that the residuals are generally small compared to the fitted values and that they do not suggest any distinct pattern. All residuals in  $v^{(1)}$  are typically less than 0.13 in absolute value, whereas the residuals in  $v^{(2*)}$  are less than 0.06 in absolute value. This suggests that the results from  $v^{(2*)}$  provides a more accurate estimate parameters than  $v^{(1)}$ . The analysis of the plotting in Figure 6.10 agrees with the chosen iterative method.

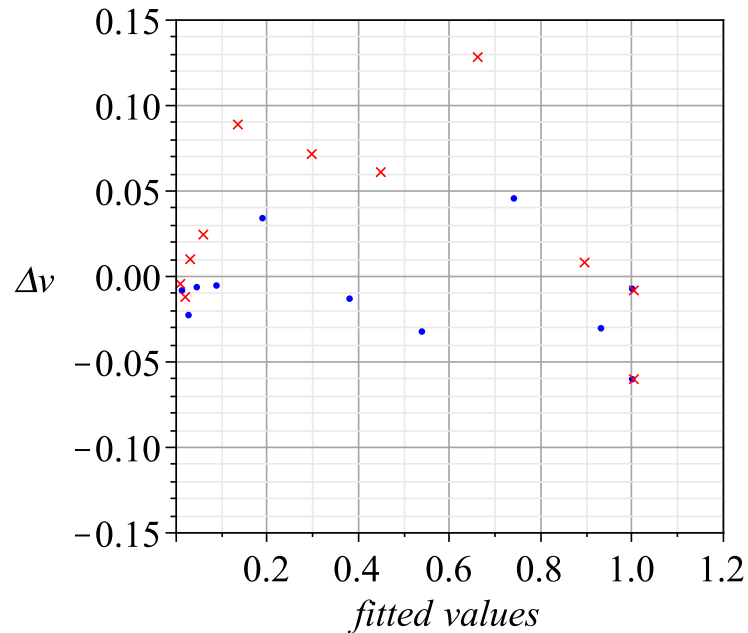


Figure 6.10: Comparison between residual results versus fitted values of  $v^{(1)}$  and  $v^{(2*)}$  at  $\theta = (217815.530752700288, 22850.3934920922802, 9.68677057029868622E10)^T$  from Figure 6.9. The crosses are  $v^{(1)}$  calculations, and the solid circles are  $v^{(2*)}$  calculations.

The inverse problem of determining the distribution from measurements of  $v$  versus time when  $\theta = (217815.530752700288, 22850.3934920922802, 9.68677057029868622E10)^T$  is illustrated in Figure 6.11. Note that when working with digitized actual data, the distribution is unknown. The crosses in Figure 6.11 show the result of calculating just the leading-order term  $f_0$  using Equation (3.19), and the solid circles are the result obtained by adding our higher-order correction term from Equation (3.20). It can be seen that adding  $\epsilon f_1$  to  $f_0$  shifts the underlying distribution a little bit toward the right which suggests that adding  $\epsilon f_1$  gives a non Gaussian total distribution, that is, the actual data distribution is non-Gaussian otherwise, the best fit of  $f_0$  would lie exactly under it. This result suggests that the method will work well for non Gaussian underlying distributions. The underlying distribution is

smooth and shown below.

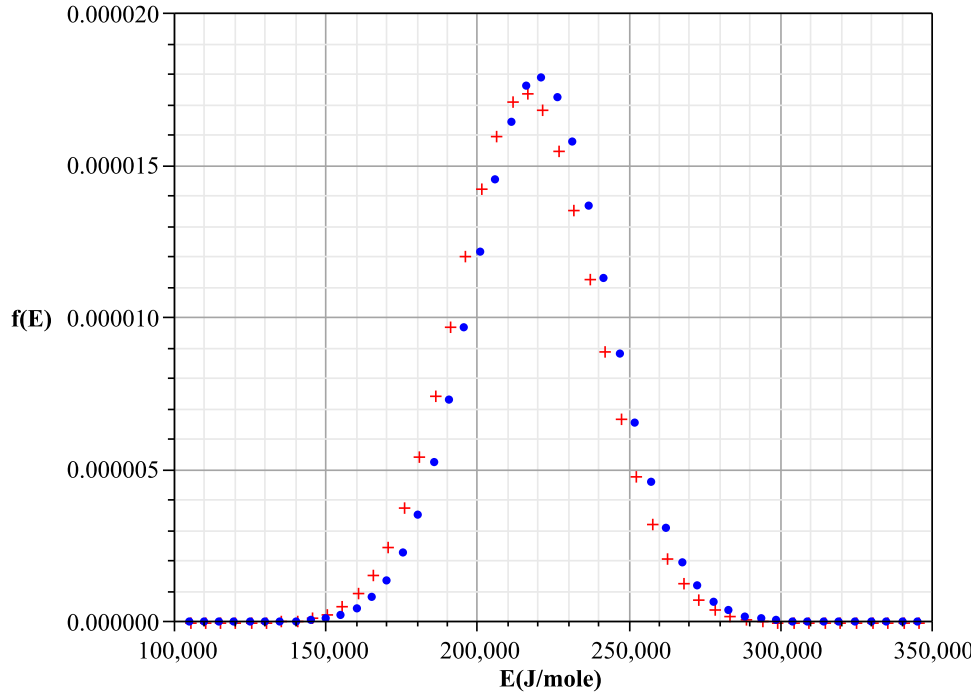


Figure 6.11: Comparisons the underlying distributions estimated inversely from values of  $v$  versus time between the leading-order result and the higher-order result. The leading-order result by a crosses, and the higher-order result by a solid circles.

Note that during the iterative process, special attention needs to be paid to how to increase and decrease the conditioning factor  $k^{(n)}$ . We find out that there are two approaches to increase  $k^{(n)}$  when the new sum of squares is greater than the previous one. The first approach is presented in Examples 2, 3 and 4 where  $k^{(n)}$  is increased by a factor and used together with the increased sum of squares parameter values to perform the next iteration. The second approach is illustrated in Example 5 where  $k^{(n)}$  is increased by a factor and used together with the last decreased sum of squares parameter values to compute the next iterative. These two methods show that the sec-

ond approach is much faster to obtain convergence than the first approach. From Example 5 if the first approach is applied, the algorithm will take too many small steps and thus make little progress.

As seen in Example 5, the converged parameter values  $\boldsymbol{\theta}^{(7)}$  from Example 4 are used as the starting parameter estimates  $\boldsymbol{\theta}^{(0)}$  to perform the new iteration in Example 5. Note that the sum of squares  $S(\boldsymbol{\theta}^{(7)})$  at  $\boldsymbol{\theta}^{(7)}$  is the smallest in Example 4. We find that the first iteration in Example 5 demonstrates that  $S(\boldsymbol{\theta}^{(0)})$  is greater than  $S(\boldsymbol{\theta}^{(7)})$ . This result illustrates that the best width  $E_w$  of  $v^{(1)}$  is not the best starting width  $E_w$  of  $v^{(2*)}$ . This is because the width  $E_w$  of  $v^{(1)}$  shifts inward and outward in  $v^{(2*)}$  as a result of adding  $\epsilon f_1$  to  $f_0$ . Therefore, the validity of the  $v^{(n)}$  depends on the numbers of higher-order terms added to improve the formula.

The Levenberg-Marquardt Algorithm showed that the direction of  $\boldsymbol{\delta}^{(n)}(k)$  is intermediate between the Gauss-Newton direction ( $k^{(n)} \rightarrow 0$ ) and the steepest-descent direction ( $k^{(n)} \rightarrow \infty$ ). These directions were both seen during the iteration process. In Examples 2, 3 and 4, the direction of  $\boldsymbol{\delta}^{(n)}(k)$  pushes the algorithm closer to the Gauss-Newton Method. While in Example 5, the direction of  $\boldsymbol{\delta}^{(n)}(k)$  pushes the algorithm closer to Gauss-Newton Method at the start of the iteration process and then diverges to push the algorithm toward the steepest descent direction. It can be seen in Example 5 when  $N = 6, 7$  and  $k = 10, 100$  that the values of  $k$  start to increase during the iterative process. The increase of  $k$  provides the calculation of  $\boldsymbol{\delta}^{(n)}$  to reduce the sum of squares.

To illustrate calculations of the residuals, the leading order solution ( $v^{(1)}$ ) of Equation (3.16) provides the estimate  $f_0$  which is Gaussian. The leading order  $v^{(1)}$  is used to fit data which is generated using a Gaussian distribution, plus a small random error or digitized data from published coal volatilization experiments. We are fitting  $v^{(1)}$  to the data in order to smooth it in prepara-



tion for taking derivatives to obtain higher-order approximations which are not Gaussian to the underlying distribution. Adding the higher-order correction term  $\epsilon f_1$  to the leading-order term  $f_0$  gives an improved formula  $v^{(2)}$  in Equation (6.5) for fitting the data. Equation (6.5) improves the result of  $v^{(1)}$  to more accurately approximate parameter values since it provides a sum of squares which is smaller than in the previous iterative process. The purpose of smoothing the data is to get the best parameter values which are closest to the final values in order to provide a more accurate estimate of the underlying distribution in the DAEM. Note that when working with digitized actual data the distribution is unknown.

# Chapter 7

## Conclusion

The validity of the two models, the Single First Order model (SFOR) and the Distributed Activation Energy Model (DAEM) were clarified through the comparison of their contribution to coal pyrolysis. Evidence is found that the DAEM is more powerful than the SFOR model in evaluating the complex experimental conditions of coal pyrolysis. The evaluation of the SFOR model and the DAEM found the SFOR model to be problematic because it can only be applied in limited conditions and cannot be expected to represent data accurately over a wide range of conditions. The shortcomings of the SFOR model can be resolved and evaluated by the use of the DAEM in coal pyrolysis. Therefore, the DAEM is generally found to be the most appropriate model for evaluating the complex experimental conditions of coal pyrolysis.

This thesis attempts to understand and describe the behavior of coal pyrolysis when the distribution of volatiles is wide compared with the double exponential term (DExp) . In this case, asymptotic approximations for the amount of volatile released in the full DAEM were derived, and have been found by numerical testing to work better than previous approximations for underlying distribution of coal pyrolysis. The iterative calculation of  $v^{(1)}$  and  $v^{(2*)}$  uses an iterative method called the Levenberg-Marquardt Algorithm to smooth the data first. This method minimizes the error due to differencing

between values obtained for  $v$  and it proved to work well with coal data. A computer programme was created to provide guidance and speed the calculation of the residuals and sum of squares.

The leading-order solution ( $v^{(1)}$ ) of Equation (3.16) provides the estimate  $f_0$  which is Gaussian, and adding the higher-order correction term  $\epsilon f_1$  gives an improved formula  $v^{(2)}$  in Equation (6.5) which is non Gaussian. It is necessary to fit  $v^{(1)}$  to the data in order to smooth it in preparation for taking derivatives to obtain higher-order approximations to the underlying distribution. Fitting the higher-order correction term to data improves the calculation of residuals and provides the sum of squares which is smaller than the sum of squares when fitting the lead-order solution. Note that if we add the higher-order correction terms  $\epsilon^2 f_2$  in Equation (3.21) and  $\epsilon^3 f_3$  in Equation (3.22), the results from the iterative process will provide a more accurate estimate of the underlying distribution in the DAEM. Therefore the more higher-order terms we add, the better the underlying distribution in the DAEM.

The iterative method result has smoothed the underlying distribution shows in Figure 6.11. Figure 6.11 illustrated the plot of the underlying distribution with the crosses showing the best Gaussian of calculating just the leading-order term  $f_0$ . The solid circles illustrate the result obtained by adding our higher-order correction term which makes the distribution non Gaussian. The analysis of Figure 6.11 shows that when adding the higher order ( $\epsilon f_1$ ), the underlying distribution curve shifted a little bit toward right from the best fit  $f_0$ . This result illustrated that the method seemed to work well for non Gaussian underlying distribution. Therefore, we anticipate that this new approach is general enough to apply to any form of energy distribution that provides a case of wide distribution in DAEM.

Substantial difficulty was encountered in finding an appropriate smooth-

ing function to deal with the complex experimental conditions of coal pyrolysis. However it is encouraging that the results from this thesis are valid for any form of energy distribution that provides a case of wide distribution. An explanation has been presented of the behavior of the double exponential time-dependent part of the double integral in the DAEM, here called DExp. This explanation follows closely from the detailed investigation done by McGuinness *et al.* [28] but further develops their formulae and applies it.

## 7.1 Recommendation

It appears at this point that the improvement that remains to be attempted is trying to solve the erratic behavior of the Gauss-Newton Method. The condition that causes this problem is the ill-conditioning of the derivative matrix  $V$ , which is caused by collinearity of the columns. This thesis recommends that this erratic behavior of the Gauss-Newton method could be improved in practice by applying parameter transformations to improve constraints on parameters. It also remains to be proved that this method (which starts by assuming  $f_0$  is Gaussian) does work for data which is not generated using a Gaussian distribution. For example, using data generated by a Gamma Distribution plus noise.

## 7.2 Research Direction

A question that arose during this work is “Can the pre-exponential or frequency factor  $k_0$  be allowed to depend on the activation energy  $E$  of Equation (2.2) in the Mathematical approach?” We understand that both  $k_0$  and  $E$  are highly correlated with each other but we have not yet reached a satisfactory conclusion on whether  $k_0$  does depend on  $E$ . Miura and Maki [27] argue in their work that  $k_0$  does depend on  $E$  by plotting the graph and showing that  $k_0$  cannot be assumed as a constant. They correlate  $k_0$  with  $E$

based on experimental data of the equation

$$k_0 = \alpha e^{\beta E}$$

where  $\alpha$  and  $\beta$  are constants dependant on the reaction system. However, it remains as an outstanding question whether  $k_0$  really does depend on  $E$  mathematically and physically, and this would be a valuable direction for future research.

# Chapter 8

## Appendices

### 8.1 Appendix 1

#### QR Decompositions Using Householder Transformations

Performing a  $QR$  decomposition of an  $N \times P$  matrix  $\mathbf{V}$  by using Householder transformations (Householder, 1958 [55]) is a generalization of reflections in the plane. These are  $N \times N$  square matrices of the form

$$\mathbf{H}_{\mathbf{u}} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^T$$

where  $\mathbf{I}$  is the  $N \times N$  identity matrix and  $\mathbf{u}$  is an  $N$ -dimensional unit vector, where the norm of  $\mathbf{u}$  is described as  $\|\mathbf{u}\| = \sqrt{\mathbf{u}^T\mathbf{u}} = 1$ .  $\mathbf{H}_{\mathbf{u}}$  is orthogonal and symmetric, since

$$\mathbf{H}_{\mathbf{u}}^T = \mathbf{I}^T - 2\mathbf{u}\mathbf{u}^T = \mathbf{H}_{\mathbf{u}}$$

and

$$\mathbf{H}_{\mathbf{u}}^T\mathbf{H}_{\mathbf{u}} = \mathbf{I} - 4\mathbf{u}\mathbf{u}^T + 4\mathbf{u}\mathbf{u}^T\mathbf{u}\mathbf{u}^T = \mathbf{I}$$

Multiplying a vector  $\mathbf{y}$  by  $\mathbf{H}_{\mathbf{u}}$ , as

$$\mathbf{H}_{\mathbf{u}}\mathbf{y} = \mathbf{y} - 2\mathbf{u}\mathbf{u}^T\mathbf{y}$$

corresponds to reflecting  $\mathbf{y}$  in the hyperplane through the origin perpendicular to  $\mathbf{u}$ . Evaluating and rearranging the equations above for  $\mathbf{u}$  gives

$$\mathbf{u} = \frac{\mathbf{y} - \|\mathbf{y}\| \mathbf{e}_1}{\|(\mathbf{y} - \|\mathbf{y}\| \mathbf{e}_1)\|}$$

or

$$\mathbf{u} = \frac{\mathbf{y} + \|\mathbf{y}\| \mathbf{e}_1}{\|(\mathbf{y} + \|\mathbf{y}\| \mathbf{e}_1)\|}$$

where  $\mathbf{e}_1 = (1, 0, \dots, 0)^T$  and  $\|\mathbf{y}\|$  is the norm of  $\mathbf{y}$ .

## 8.2 Appendix 2

### Computer Programme

*MakeIteration* := **proc**( $E_{0in}, \sigma_{in}, k_{0in}, N$ )

**local**  $v, p, r, q, s, E_0, \sigma, k_0, \theta_{0Use}, \theta_1, t, S_{old}, X, \delta, X^T, \eta, D, \delta_1, \theta_{0Use1}, S_{new}, z, z_1, y;$

$$v := (t, E_0, \sigma, k_0) \mapsto \frac{1}{2} \operatorname{erfc} \left( \frac{1}{2} \sqrt{2} \sqrt{\frac{E_0^2}{\sigma^2}} \left( \frac{\operatorname{RmtLambertW}(k_0 t)}{E_0} - 1 \right) \right);$$

$$p := \operatorname{diff}(v(t, E_0, \sigma, k_0), E_0);$$

$$r := \operatorname{diff}(v(t, E_0, \sigma, k_0), \sigma);$$

$$s := \operatorname{diff}(v(t, E_0, \sigma, k_0), k_0);$$

Let  $E_0 := E_{0in}; \sigma := \sigma_{in}; k_0 := k_{0in};$

*Data* = [0.9621489806, 1.024242331, 1.033051170, 0.9819960462,  
 0.9734757992, 1.031344011, 0.9986545413, 0.9703193539,  
 0.9130842328, 0.7918876653, 0.6483888526, 0.5047982891,  
 0.3326939449, 0.1917859271, 0.09363834317, 0.04104433104,  
 0.01602741851, 0.004913205795, 0.001476246960, 0.0003506119174];

*The constant values :*

$m := 650;$   $R := 8.3144;$

$\theta_{0Use} := E_{0in}, \sigma_{in}, k_{0in};$

$\theta_1[1] := \theta_{0Use};$

$k := 0.1E-1;$

$z := Matrix\left( evalf([seq([y[i] - (eval(v(t, E_0, \sigma, k_0), t = 0.15 \times i))], i = 1..20)]) \right);$

*Set up the iteration :*

**for**  $q$  **from** 1 **to**  $N$  **do;**

*Find the sum of squares of the starting parameter estimates ( $S(\theta^{(0)}) = S_{old}$ ) :*

$S_{old} := MatrixNorm(z, 2)^2;$

*Calculate the sum of squares of the improving parameter ( $S(\theta^{(1)}) = S_{new}$ ) :*

$X := Matrix\left( evalf([seq([eval(p, t = 0.15 \times i), eval(r, t = 0.15 \times i),  
 eval(s, t = 0.15 \times i)], i = 1..20)]) \right);$

*Apply "The Levenberg – Marquardt Compromise"*

$\delta := ((X^T X + kD)^{-1} (X^T z));$



```


$$X^T := \text{Transpose}(X);$$


$$D := \text{Matrix}((X^T, X), \text{shape} = \text{diagonal});$$


$$\theta_{0Use1} := \text{Matrix}([[E_0], [\sigma], [k_0]]);$$


$$\theta_1[q + 1] := \theta_{0Use1} + \text{delta}$$

Data from  $\theta_1[q + 1]$ :

$$E_0 := \theta_1[q + 1][1, 1];$$


$$\sigma := \theta_1[q + 1][2, 1];$$


$$k_0 := \theta_1[q + 1][3, 1];$$


$$\text{print}(E_0); \text{print}(\sigma); \text{print}(k_0);$$


$$z_1 := \text{Matrix}(\text{evalf}([\text{seq}([y[i] - (\text{eval}(v(t, E_0, \sigma, k_0), t = 0.15 \times i)]), i = 1..20)]));$$


$$S_{new} := \text{MatrixNorm}(z_1, 2)^2;$$


$$\text{print}(S_{old}); \text{print}(S_{new});$$

if  $S_{new} < S_{old}$  and  $\frac{S_{old} - S_{new}}{S_{new}} > 10 \times 10^{-7}$  then

$$z := z_1$$


$$k := \frac{1}{10}k$$

else

$$z := z_1$$


$$k := k \times 10$$

end if;
if  $0 \leq \frac{S_{old} - S_{new}}{S_{new}} \leq E - 5$  then

$$\text{print}(STOP);$$


```

```
else  
    print(CONTINUE);  
end if; od; end;  
MakeIteration( $E_0, \sigma, k_0, N$ )
```

# Bibliography

## 8.3 Bibliography

- [1] Essenhigh, R. H., Sixteenth Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 1977 p.353.
- [2] Idris Jones, W., Recent developments in the thermal treatment of coal, *J. Inst. Fuel*, **24** (1951) 69.
- [3] Anthony, D. B. and Howard, J. B., Coal devolatilization and hydrogasification, *A.I.Ch.E. J.*, **22** (1976) 625.
- [4] Saxena, S. C., Devolatilization and combustion characteristics of coal particles, *Prog. Energy Combust. Sci.* **16** (1990) 55.
- [5] Wender, I., Heredy, L. A., Neuworth, M. B. and Dryden, I. G. C., *Chemistry of Coal Utilization*, 2nd Supplementary Vol., M. A. Elliot (Ed.), John Wiley and Sons, New York, 1981.
- [6] Van Krevelen, D. W., *Coal*, Elsevier Publishing Company, Amsterdam, 1961.
- [7] Dryden, I. G. C., Chemical Constitution and Reaction of Coal, in *Chemistry of Coal Utilization*, Supplementary Vol, H. H. Lowry, Ed., John Wiley and Sons, New York, 1963 p.232-295.

- [8] Given, P. H., The Distribution of Hydrogen in Coals and Its relations to Coal Structure, *Fuel* **39** (1960) 147.
- [9] Gavalas, G. R., *Coal Pyrolysis*, Elsevier, New York 1982.
- [10] Howard, J.B., In: Elliott MA, editor. *Chemistry of Coal Utilization*, 2nd spl. vol., chap. 12, Wiley and Sons, New York, 1981.
- [11] Spackman, W., The nature of Coal and Coal Seam, in *Short Course in Coal Characteristics and Coal Conversion Processes*, Pennsylvania State University, University Park, 1975 p.19-23.
- [12] Howard, H.C., Pyrolytic reactions of Coal, in *Chemistry of Coal Utilization*, Supplementary Vol, H. H. Lowry, Ed., John Wiley and Sons, New York, 1963 p.340-394.
- [13] Chermin, H. A. G. and Van Krevelen, D. W., Chemical structure structure and properties of coal .17. A mathematical model of coal pyrolysis. *Energy and Fuel*, **36** (1957) 85.
- [14] Juntgen, H., Review of the kinetics of pyrolysis and hydrolysis in relation to the chemical constitution of coal. *Fuel*, **63** (1984) 731.
- [15] Solomon, P. R. and Hamblen, D. G., Finding order in coal pyrolysis kinetics, *Pro. Energy Combust. Sci.* **9** (1984) 323.
- [16] Solomon, P. R., Hamblen, D. G. and Carangelo, R. M., *Coal pyrolysis*, AIChE Symposium on Coal Pyrolysis 1981.
- [17] Solomon, P. R. and Serio, M. A., Evaluation of Coal Pyrolysis Kinetics, *Fundamentals of the Physical Chemistry of Pulverized Coal Combustion*, Martinus Nijhoff, J. Lahaye and G. Prado (Eds), 1987 p.126.
- [18] Solomon, P. R., Serio, M. A. and Suuberg, E. M., Coal pyrolysis: Experiments, Kinetic rates and mechanisms, *Prog. Energy Combust. Sci.* **18** (1992) 133.

- [19] Lowry, H.H., Ed., Chemistry of Coal Utilization, Supplementary Vol, John Wiley and Sons, New York, 1963.
- [20] Jones, W. I., The Thermal Decomposition of Coal, J. Inst. Fuel, **37** (1964) 3.
- [21] Yellow, P. C., Kinetics of the Thermal Decomposition of Coal, BCURA Month. Bull, **20** (1965) 285.
- [22] Badzioch, S., Thermal Decomposition, BCURA Month. Bull, **31** (1967) 193.
- [23] Essenhigh, R. H. and Howard, J. B., Combustion phenomena in Coal Dusts and the Two-Component Hypothesis of Coal Constitution (Penn State studies No. 31, Administrative committee on research, Pennsylvania State University, University Park, 1971.
- [24] Pyrcioch, E. J., Feldkirchner, H. L., Tsaros, C. L., Johnson, J. L., Blair, W.G., Lee, B. S., Schora, F. C., Huebler, J., and Linden, H. R., Production of Pipeline Gas by Hydrogasification of Coal, Res. bull. No. 39, Vol. 1, Institute of Gas technology, Chicago, 1972.
- [25] Niksa, S. and Lau, C. W., Global rates of devolatilization for various coal types, Combust. Flame, **94** (1993) 293.
- [26] Suuberg, E. M., Approximation solution technique for nonisothermal, Gaussian Distributed Activation Energy Model, Combust. Flame, **50** (1983) 243.
- [27] Miura, K. and Maki, T. , Simplified method to estimate  $f(E)$  in Distributed Activation Energy Model analyzing coal pyrolysis reaction, J Chem. Eng. of Japan, **31** (1998) 228.

- [28] Please, C. P., McGuinness, M. J. and McElwain, D. L. S., Approximations to the Distributed Activation Energy Model for the pyrolysis of coal, *Combust. Flame*, **133** (2003) 107.
- [29] Vand, V., A theory of the irreversible electrical resistance changes of metallic films evaporated in vacuum, *proc. Phys. Soc. Lond.*, **55** (1943) 222.
- [30] Pitt, G. J., The kinetics of the evolution of volatile products from coal, *Fuel*, **41** (1962) 267.
- [31] Miura, K. , A New and Simple method to Estimate  $f(E)$  and  $k_0(E)$  in the Distributed Activation Energy Model from Three sets of Experimental Data, *Energy and Fuels*, **9** (1995) 302.
- [32] Maki, T., Takatsuno, A. and Miura, K., Analysis of pyrolysis reactions of various coals including Argonna premium coals using a new Distributed Activation Energy Model, *Energy and Fuels* **11** (1997) 972.
- [33] Howard, J. B. and Essenhigh, R. H., Pyrolysis of coal particles in pulverized fuel flames, *Ind. engng Chem. Process Des. Dev.*, **6** (1967) 74.
- [34] Badzoich, S. and Hawksley, P. G. W., Kinetics of thermal decomposition of pulverized coal particles, *Ind. engng Chem. Process Des. Dev.*, **9** (1970) 521.
- [35] Anthony, D. B., Howard, J. B., Hottel, H. C. and Meissner, H. P., Rapid Devolatilization of Pulverized Coal, 15th Symposium (International)

- on Combustion, The Combustion Institute, 1975 p.1303.
- [36] Gunes, M. and Gunes, S., The influences of various parameters on the numerical solution of nonisothermal DAEM equation, *Thermochemica Acta*, **336** (1999) 93.
- [37] Brown, M. E., *Introduction to Thermal analysis: Techniques and Applications*, Chapman and Hall, New York, 1988 p.144.
- [38] Zsako, J., *Thermal Analysis*, (ed.Z.D.Ziukovic) University of Beograd, Bor, Yugoslavia, 1984 p.167.
- [39] Teng, H. and Hsieh, C. T., Activation energy for oxygen chemisorption on carbon at low temperatures, *Ind. Eng. Chem. Research*, **38** (1999) 292.
- [40] Lakshmanan, C .C., Bennett, M. L. and White, N., Implications of multiplicity in kinetic parameters to petroleum exploration: Distributed Activation Energy Models, *Energy and Fuels* **51** (1991) 110.
- [41] Merrick, D., *Mathematical Models of the Thermal Decomposition of Coal: The Evolution of Volatile Matter*, *Fuel*, **62** (1993) 534.
- [42] Donskoi, E. ,McElwain, D. L. S., Approximate modeling of coal pyrolysis, *Fuel*, **78** (1999) 825.

- [43] Armstrong, R. and Kulesza, B. L. J., An approximate solution to the equation  $x = \exp(-x/e)$ , Bull. Institute of Mathematics and its Application, **17** (1981) 56.
  
- [44] Currie, L. A., Sources of error and the approach to accuracy, in I. M. Kolthoff and P. J. Elving (editors), Treatise on Analytical Chemistry, 2nd edition, Part 1, Volume 1, John Wiley and Sons, Wiley, New York, 1978 p.95-233.
  
- [45] Daniel, C. and Wood, F. S., Fitting Equations to Data, John Wiley and Sons, New York, 1971.
  
- [46] Srivastava, M. S., Regression Analysis: Theory, Methods, and Applications, Springer-Verlag New York Inc, 1990.
  
- [47] Chambers, J. M., Computational Methods for Data Analysis, John Wiley and Sons, New York, 1977.
  
- [48] Bates, D. M. and Watts, D. G., Nonlinear regression Analysis and Its Applications. John Wiley and Sons, New York, 1988.
  
- [49] Stewart, G. W., Introduction to Matrix Computations. Academic Press, New York, 1973.
  
- [50] Jupp, david L. B., Approximation to data by splines with free knot. SIAM Journal of Numerical Analysis, **15** (1978) 328.
  
- [51] Seber, G. A., and Wild, C. J., Nonlinear Regression, Department of Mathematics and Statistics University of Auckland, New Zealand, John Wiley and Sons, 2003.



- [52] Ma, K., Hiroshi, N., Yoshida, T., Xu, G., Harada, M., Kinetics of rapid coal devolatilization measured using a spot heater apparatus. *Fuel Processing technology*, **85** (2003) 43.
- [53] Marquardt, D. W., An algorithm for the estimation of non-linear parameters. *Journal of SIAM*, **11** (1963) 431.
- [54] Levenberg, K., A method for the solution of certain nonlinear problems in least squares. *Quarterly of Applied Mathematics*, **2** (1944) 164
- [55] Householder, A. S., Unitary triangularization of a nonsymmetric matrix. *Journal of the Association for Computing Machinery*, **5** (1958) 339.