Applications of the Bayesian Approach for Experimentation and Estimation

by

Patrick A. P. de Man

M.Sc. Chemical Engineering
Eindhoven University of Technology, 1998

M.S. Chemical Engineering Practice
Massachusetts Institute of Technology, 2003

SUBMITTED TO THE DEPARTMENT OF CHEMICAL ENGINEERING
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING PRACTICE
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2006

Signature of Author.....
Department of Chemical Engineering
May 18, 2006

Certified by.....
Gregory J. McRae
Hoyt C. Hottel Professor of Chemical Engineering
Thesis Supervisor

Accepted by.....
William M. Deen
Professor of Chemical Engineering
Chairman, Committee of Graduate Students

Applications of the Bayesian Approach for Experimentation and Estimation

by

Patrick A. P. de Man

Submitted to the Department of Chemical Engineering
on May 18, 2005 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Chemical Engineering Practice

Abstract

A Bayesian framework for systematic data collection and parameter estimation is proposed to aid experimentalists in effectively generating and interpreting data. The four stages of the Bayesian framework are: system description, system analysis, experimentation, and estimation. System description consists of specifying the system under investigation and collecting available information for the parameter estimation. Subsequently, system analysis entails a more in-depth system study by implementing various mathematical tools such as an observability and sensitivity analysis. The third stage in the framework is experimentation, consisting of experimental design, system calibration, and performing actual experiments. Finally, the last stage is estimation, where all relevant information and collected data is used for estimating the desired quantities.

The Bayesian approach embedded within this framework provides a versatile, robust, and unified methodology allowing for consistent incorporation and propagation of uncertainty. To demonstrate the benefits, the Bayesian framework was applied to two different case studies of complex reaction engineering problems. The first case study involved the estimation of a kinetic rate parameter in a system of coupled chemical reactions involving the relaxation of the reactive $O(^1D)$ oxygen atom. The second case study was aimed at estimating multiple kinetic rate parameters concurrently to gain an understanding regarding the reaction mechanism of the oxygen addition to the transient cyclohexadienyl radical.

An important advantage of the proposed Bayesian framework demonstrated with these case studies is the possibility of 'real-time' updating of the state of knowledge regarding the parameter estimate allowing for exploitation of the close relationship between experimentation and estimation. This led to identifying systematic errors among experiments and devising a stopping rule for experimentation based on incremental information gain per experiment. Additional advantages were the improved

understanding of the underlying reaction mechanism, identification of experimental outliers, and more precisely estimated parameters.

A unique feature of this work is the use of Markov Chain Monte Carlo simulations to overcome the computational problems affecting previous applications of the Bayesian approach to complex engineering problems. Traditional restricting assumptions can therefore be relaxed so that the case studies could involve non-Gaussian distributions, applied to multi-dimensional, nonlinear systems.

Thesis Supervisor: Gregory J. McRae
Title: Hoyt C. Hottel Professor of Chemical Engineering

Acknowledgments

In these matters the only certainty is that nothing is certain.

Pliny the Elder

My five years at MIT have made an enormous impact on my life. I feel very fortunate to have had the opportunity to spend time in this exciting institute. Certainly, I never even imagined of my career change from engineering to finance. This journey could not have been possible without the support of a number of people to whom I would like to express my gratitude.

The person whom I am forever grateful is my wife Mika, who sacrificed a lot in order to be with me here at MIT. She has provided unwavering support during the many strenuous moments.

I extremely enjoyed working with and learning from my advisor Gregory McRae. He provided insightful guidance on both professional and personal level and gave me the push in the back I needed every now and then.

The two case studies in this thesis were based on the research of Dr. Edward Dunlea and of Dr. James Taylor. Ed and James have been very generous with their time in explaining their work and answering my endless questions.

During my research the thesis committee members always gave me thoughtful and helpful comments and questions. I would like to thank Prof. Kenneth Beers, Prof. Charles Cooney, Prof. William Green, Prof. Klavs Jensen, Prof. Dennis McLaughlin, Prof. Mario Molina, and Dr. Mark Zahniser.

Special thanks go to Sara Passone for our joint effort of learning the Bayesian approach and for proofreading my thesis, and to Ico San Martini and Jose Ortega for their advice on the Bayesian approach and Markov Chain Monte Carlo. Last but not least, thanks to my other colleagues Barthwaj Anantharaman, Nina Chen, Jeremy Johnson, Chuang-Chung Lee, Mihai Anton, Alex Lewis, Anusha Kothandaraman, and Bo Gong of the McRae research group for providing the sunshine in our windowless basement office.

Table of Contents

List of Figures

List of Tables

1
Introduction

If a man will begin with certainties, he shall end in doubts;
but if he will be content to begin with doubts, he shall end in certainties.

Francis Bacon

1.1 Thesis Statement

When mathematical models are used to describe physical systems, inevitably approximations are made. The issue is not the introduction of uncertainties, as they will always be present, but to identify those that contribute most to the uncertainties in the predicted outcomes.

A typical example is the inevitable presence of uncertainties that arise from estimation of model parameters from experimental data. Unfortunately in practice uncertainty is often regarded as being disconnected from the quantities of interest and appears as supplementary information tagged onto deterministic results of a calculation or estimation.

There is a critical need for experimental design and parameter estimation procedures that reflect the underlying complexities of real systems. In particular, methods are needed that overcome the traditional limitations of Gaussian distributions to describe and propagate uncertainties.

The key premise of this thesis is that recent advances in Markov Chain Monte Carlo simulation methodologies can radically simplify the application of Bayesian statistics to reaction engineering problems.

The main contributions of this thesis are:

- A practical description of the Bayesian approach discussing concepts, advantages, computation, and practical implementation issues for applications and examples of interest to scientists and engineers.
- Illustrations of the intuitive evaluation of uncertainty according to the Bayesian approach.
- Illustrations of wide-spread misconceptions regarding conventional statistics related to erroneous applications in parameter estimation.
- The development of a general framework addressing experimentation and estimation that has four interdependent stages: system description, system analysis, experimentation, and estimation.
- A discussion and demonstration of mathematical tools, originating from system engineering, valuable for projects in experimentation and estimation.
- Identification of three parameter estimation strategies to be employed depending on the amount of data and the size of the parameters and variables to estimate.
- The development of tools to devise a stopping rule for a series of experiments and to discriminate among data for identifying outliers.

1.2 Motivation

A typical chemical reaction mechanism consists of multiple steps and species. Prediction of species concentration and dynamics requires knowledge regarding the kinetic rate parameters, stoichiometric coefficients, and initial conditions. These are usually estimated from experimental data. The estimation involves the combining of information from various sources, such as (1) concentration measurements as a function of time for one or more species involved in the reaction, (2) experimental conditions, such as

temperature, pressure, flow rates, and initial concentrations of reactants, and (3) existing knowledge on some or all of the kinetic rate parameters. Since in addition each type of information is characterized by uncertainty, the estimation problem mounts to a complex undertaking.

1.2.1 Complex Problems

Consider for example the reaction mechanism in Figure 1-1 of the relaxation of reactive $O(^1D)$ oxygen atoms, that has been proposed as part of a larger mechanism to understand stratospheric ozone depletion. The goal is to estimate the kinetic rate parameters k_1 from noisy experimental data.

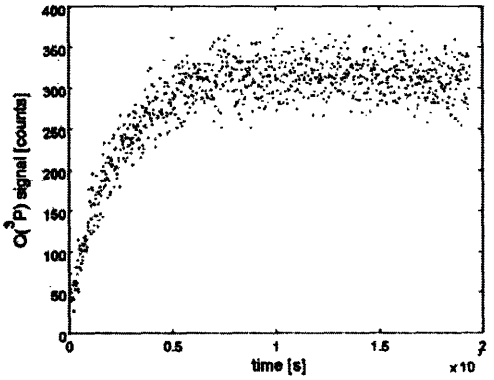
$$O(^1D) + N_2 \xrightarrow{k_1} O(^3P) + N_2$$

$$O(^1D) + O_3 \xrightarrow{k_2} 2O(^3P) + O_2$$

$$O(^1D) + O_3 \xrightarrow{k_3} 2O_2$$

$$O(^1D) + He \xrightarrow{k_4} O(^3P) + He$$

(a)



(b)

Figure 1-1. (a) First example of a reaction mechanism and (b) measurement results

With a reaction mechanism and data, such as shown in the example above, the main challenges to overcome in order to estimate the parameter k_1 are:

1. How to estimate the parameter k_1 from the data?
2. What factors contribute to the uncertainty in the value of k_1?
3. How to combine various types of uncertain measurements and information?
4. How many experiments are required to sufficiently decrease the uncertainty?
5. How to combine the results from large amounts of data?

Though the case above involves an accepted reaction mechanism and thus strictly focuses on estimating a rate parameter, the example given in Figure 1-2 is not so straightforward. To resolve contradictions in the literature regarding the reaction mechanism of the addition of oxygen to the transient cyclohexadienyl radical, quantum mechanical calculations were performed to propose the mechanism given below. Subsequently, experimental data was generated by laser techniques with the goal of confirming this mechanism.

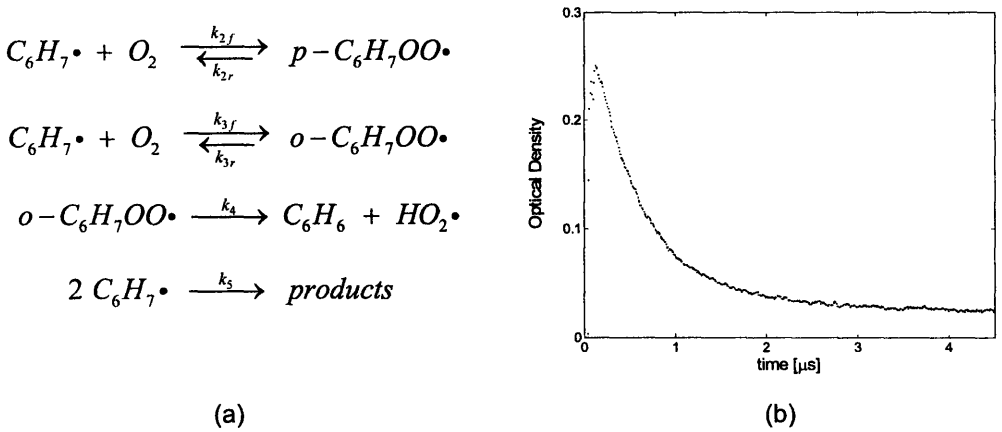


Figure 1-2. (a) Second example of a reaction mechanism and (b) measurement results

The challenges in the above example in estimating several rate parameters while attaining information regarding the proposed reaction mechanism are:

1. How to include existing uncertain estimates for particular rate parameters?
2. Which physical constraints can be applied to facilitate the estimation?
3. How reasonable is it to estimate multiple parameters from limited data?
4. How to implement different mechanism scenarios for the estimation in order to provide information regarding the applicability of the proposed mechanism?
5. How to combine the several estimates for a parameter obtained from different data sets to present one overall estimate?

The examples above illustrate key questions that arise not only for this particular problem but for estimation problems in general. The central thesis of this work is a framework for experimentation and estimation based on Bayesian statistics.

1.2.2 Description of the Bayesian Approach

The Bayesian approach can be interpreted as a learning algorithm as specified according to the relatively straightforward Bayes' theorem, which is given by

$$p(\theta \mid y) = \frac{p(y \mid \theta) p(\theta)}{p(y)} \tag{1-1}$$

where $p(\theta|y)$ is the posterior probability distribution of the parameter θ, given the data y, $p(y|\ \theta)$ is the likelihood function of the data, given the parameter θ, $p(\theta)$ is the prior distribution, and $p(y)$ is the probability distribution of the data y. Equation (1-1) effectively states that an initial parameter estimate described by $p(\theta)$ is updated with information from new data y to the posterior estimate described by $p(\theta|y)$.

The main advantages of applying the Bayesian approach are:

1. To have an intuitive estimation method based on unified underlying principles
2. To enable a consistent treatment of uncertainty incorporation and propagation
3. To evaluate uncertainty beyond 'normal' errors
4. To conveniently include multiple types of errors in the same problem
5. To robustly address multi-dimensional, nonlinear problems
6. To formally incorporate prior information
7. To have an explicit learning algorithm building on previous knowledge

These advantages of the Bayesian approach have previously been discussed in classical works arguing for Bayesian statistics [1, 2]. However, as only since recently the computational complexity resulting from the application of equation (1-1) to real life problems has been possible resolved, these advantages can now actually be attained. The

computational complexity referred to will become clear when considering that the total probability theorem to calculate $p(y)$ becomes a multi-dimensional integral, as given by

$$p(\boldsymbol{\theta} \mid y) = \frac{p(y \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int_{\theta_1} \ldots \int_{\theta_n} p(y \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\theta_1 \ldots d\theta_n} \propto p(y \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{1-2}$$

where the vector $\boldsymbol{\theta}$ contains the collection of parameters θ_1 to θ_n. The analytical solution of this multi-dimensional integral requires restricting assumptions to simplify the mathematics. Generally, in earlier works [1, 2] only Gaussian distributions were considered to characterize uncertainty, while mostly linear systems of a small number of parameters could be evaluated.

This thesis goes beyond the earlier works [1, 2] in relaxing such restricting assumptions by addressing the computational complexity through the use of Markov Chain Monte Carlo (MCMC) algorithms. MCMC simulation solves equation (1-2) by approximating the posterior $p(\boldsymbol{\theta}|y)$ through evaluating the product of the prior and likelihood function. The key realization that the denominator in equation (1-2) is merely a normalizing constant, leads to a problem formulation where non-Gaussian distributions can relatively easily be implemented in multi-dimensional nonlinear systems.

Though the MCMC algorithms require significant computational power, with the increased availability of faster and more powerful computers over the last decade practical applications of the Bayesian approach seem to become feasible in various disciplines. This is confirmed by Berger [3], who remarks the tremendous increase in activity, in the form of books, articles, or professional organizations, related to the Bayesian approach. Especially the last decade has seen an explosion of the number of publications.

1.2.3 The Relevance of Non-Gaussian Distributions

The uncertainty in model parameters will induce uncertainties in the model predictions. Even though uncertainty in the model parameters is often assumed to be characterized by the Gaussian distribution, the uncertainty in model outcomes is likely to be non-Gaussian. To illustrate this transformation of the probability distribution by uncertainty propagation through a model, consider a linear ordinary differential equation, such as a first order kinetic model, as defined by

$$\frac{dy(t)}{dt} = -ky(t) \tag{1-3}$$

where k is the first order rate parameter, y is the species concentration at time t. Suppose that rate parameter k has been estimated and can be represented by the normal probability distribution $N(\mu_k,\sigma_k^2)$. With y_0 as initial concentration, solving the model analytically leads to

$$y(t) = y_0 e^{-kt} \tag{1-4}$$

so that concentration y can be predicted at a future time. Since the uncertain rate parameter k is now an exponential term, the uncertainty in concentration y is therefore represented by a lognormal distribution, as analytically can be derived as

$$p(k) = \frac{1}{\sigma_k\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\mu_k}{\sigma_k}\right)^2} \quad \Rightarrow \quad p(y) = \frac{1}{\sigma_k t y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(y/y_0)+t\mu_k}{t\sigma_k}\right)^2} \tag{1-5}$$

Alternatively, the transformation of a normal to a lognormal distribution can be illustrated by a Monte Carlo simulation. A standard normally distributed random variable $X\sim N(0,1)$ is sampled 5000 times, and each of these values is used to calculate random variable Z according to

$$Z = e^X \tag{1-6}$$

after which the samples of Z are evaluated to approximate its probability distribution. Histograms representing the (non-normalized) probability distributions of X and Z are shown in Figure 1-3.

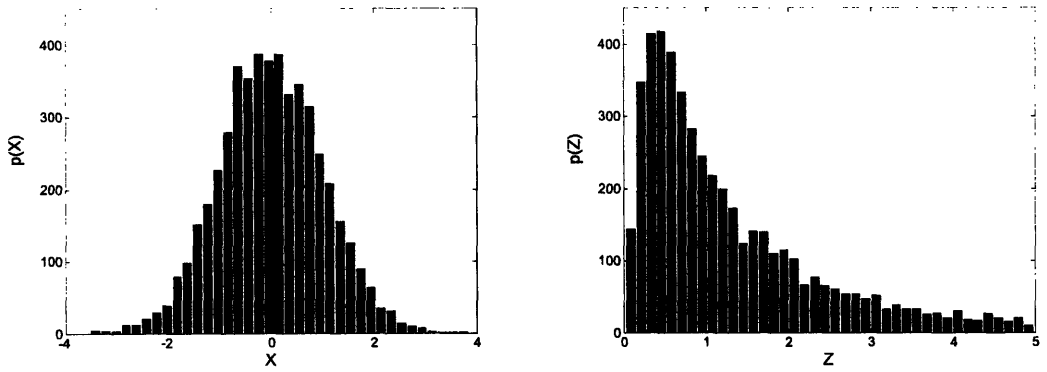


Figure 1-3. Monte Carlo results illustrating the effect on uncertainty by variable transformation

As demonstrated by the simple example above, nonlinear models likely result in predicted values with their uncertainty characterized by non-Gaussian probability distributions. This effect is only exacerbated when dealing with large complex systems, and when the uncertainty of model input parameters or variables is already of a non-Gaussian nature. Nonetheless, conventional statistical methods generally ignore the non-normality and assume Gaussian distributions. The main reason is that the available mathematical tools are inapt for consistently propagating uncertainty of observables into uncertainty of the estimates. Regrettably, the average non-statistician is almost certainly not aware of the above discussion and the established, but incorrect approach has become to accept variances as provided by standard statistical software packages as the primary measure for the parameter uncertainty.

1.2.4 Acceptance of the Bayesian Approach

At first sight the methodology of the Bayesian approach appears to be rather involved compared to the conventional statistical methods that are easily accessible via standard statistical software packages. A closer look, however, reveals a surprisingly simple and intuitive concept for which the possibilities are virtually endless. However, to explore

Bayesian interests beyond this initial barrier, the scientist or engineer should first realize the potential benefits of accepting the Bayesian approach. Unfortunately, the Bayesian literature explaining the concepts and principles of operation only illustrates applications involving relatively simple analytic or regression models. Applications of the Bayesian approach to complex engineering problems are still missing and they would be a welcome addition to the literature.

As a final note it is interesting to reiterate the opinion of Berger [3] regarding the future of statistics. According to Berger the language of statistics will be Bayesian, as this is significantly easier to understand and has been demonstrated to be the only coherent language to discuss uncertainty. On the other hand, from a methodological perspective, the Bayesian and Frequentist approaches can complement each other in particular applications, and thus at some point a unification seems inevitable.

1.3 Objectives

The primary objective of this thesis is to present tools based on Bayesian statistics that can aid scientists and engineers with experimentation and estimation. As a requirement, these tools should enable a convenient and consistent uncertainty evaluation, so that the impact of various sources of uncertain input on outcomes can be assessed. Additionally, the incorporation of information from various sources and of large amounts of data should be possible.

The secondary objective is to emphasize the importance of systems thinking regarding the estimation problem under investigation. Experimentation and estimation should be regarded as a learning process with feedback loops enabling multiple iterations to refine experiments and maximize information obtainable from the data. For example, theoretical understanding of the system contributes to the design of experiments, while also preparing the researcher for interpretation of data and identification of erratic results.

Subsequently assembled knowledge through experimentation can improve an understanding of the theory, thus shaping future experiments.

Since the acceptance and implementation of the proposed tools might be hindered by unfamiliarity with the Bayesian approach, the final objective is to describe the computational procedures in a clear and practical manner and to present the benefits of applying the Bayesian approach to case studies of complex engineering problems of interest to scientists and engineers.

1.4 Thesis Outline

This thesis has been organized around the proposed framework for experimentation and estimation and the linkages between the individual chapters are depicted in Figure 1-4.

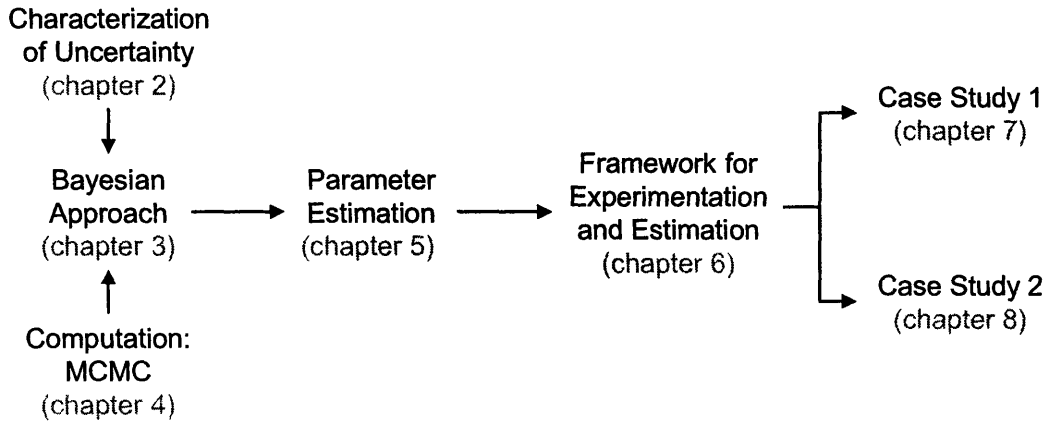


Figure 1-4. Schematic overview of the thesis structure

Chapter 2 sets the background of the thesis by discussing the interpretation of uncertainty and probability, common nomenclature, and definitions. Additionally, misconceptions and limitations resulting from conventional statistical methods are clarified before starting the discussion of the Bayesian approach.

Chapter 3 and 4 discuss the conceptual aspects, details, and computational issues regarding the Bayesian approach and Markov Chain Monte Carlo (MCMC).

Chapter 5 builds on various issues discussed in previous chapters by implementing the Bayesian approach to parameter estimation. This particular application will be described in detail and demonstrated through an example problem.

Chapter 6 introduces the proposed framework for experimentation and estimation. The discussion focuses on a structured approach to collecting essential knowledge regarding the system under investigation, and using that information towards experimentation and estimation.

The background and theory discussed in the previous chapters are implemented through two case studies elaborately discussed in Chapter 7 and 8, which both demonstrate the merits of the proposed Bayesian framework for experimentation and estimation.

Finally, Chapter 9 and 10 discuss the conclusions and directions for future research, respectively.

2

Characterization of Uncertainty

The world as we know it seems to have an incurable habit of denying us perfection.

Peter L. Bernstein [4]

For a better understanding of uncertainty existing in realistic engineering problems, it is important to specify and clarify several notions that, though seemingly evident, are often misinterpreted. First, common attributes encountered in estimation under uncertainty will be defined to clarify the confusion attached to their colloquial use. Then, issues surrounding measurement errors will be explored in-depth, after which the limitations regarding uncertainty incorporation of conventional statistical methods are demonstrated.

2.1 Introduction

This section will define several terms commonly used regarding experimentation and estimation. In particular, the distinction between accuracy and precision and the distinction between error and uncertainty will be discussed.

2.1.1 Accuracy and Precision

In analyzing measurements, there is a clear need to make a distinction between the concepts of accuracy and precision, as illustrated in Figure 2-1. Suppose the center of the concentric circles represents the true value of the measurand, and a measurement in performed four times under apparently identical conditions.

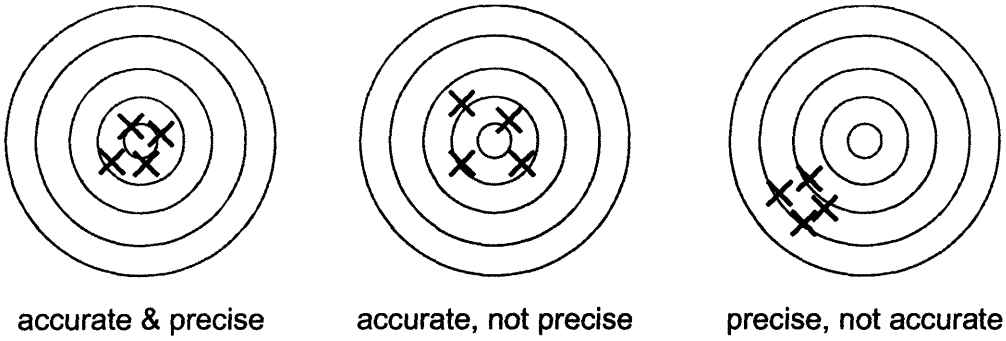


Figure 2-1. Illustration of accuracy and precision

The accuracy refers to conformity of the measurements with the true value, thus how close they are to the center. In other words, accuracy is an indication for the quality of the measurement. The precision indicates the degree of perfection in the instruments and methods the measurement is obtained with and how reproducible the measurements are. In other words, precision is an indication for the quality of the operation with which the measurement is obtained.

2.1.2 Uncertainty, Error, and True Value

A suitable point to start this discussion is to state the definitions as recommended by the International Organization for Standardization (ISO) [5, 6]:

- Uncertainty:
  "A parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurement."
- Error:
  "The result of a measurement minus a true value of the measurand."
- True value:
  "A value compatible with the definition of a given particular quantity."

Though the concepts of uncertainty and error are clearly specified, the definition for true value is not very distinctive. In addition, the operational aspects for representing and handling uncertainty and error are not specified.

Various methods exist for representation and handling of uncertainty [7]. According to the Bayesian approach, uncertainty is characterized by probability distributions and propagated using the tools of probability theory. The true value of a quantity is still considered deterministic, fixed, but unknown, even though the degree of belief is expressed by a probability distribution and the quantity is manipulated as a random variable.

Finally, it is important to realize that uncertainty arises from various sources [5-8], such as systematic and random error, approximations and assumptions during estimation, linguistic imprecision regarding the definition of quantities, etc. The ISO guide identifies the following sources:

1. incomplete definition of the measurand
2. imperfect realization of the definition of the measurand
3. non-representative sampling – the sample measured may not represent the measurand
4. inadequate knowledge of the effects of environmental conditions on the measurement, or imperfect measurement of environmental conditions
5. personal bias in reading analogue instruments
6. finite instrument resolution or discrimination threshold
7. inexact values of measurement standards and reference materials
8. inexact values of constants and other parameters obtained from external sources and used in the data-reduction algorithm
9. approximations and assumptions incorporated in the measurement method and procedure
10. variations in repeated observations of the measurand under apparently identical conditions

Ideally each source of significance should be quantified, so that its contribution to the overall uncertainty in the system can be accounted for. However, in most cases only information regarding experimental error is available and taken into account. Since error

is such an important source for uncertainty, it will be discussed extensively in the following section.

2.2 Measurement Error

Experimentation is always subject to error. The Bayesian approach explicitly includes the measurement variance in the estimation, so an understanding of what actually is considered by 'measurement error' is important. Before actually considering the concept of measurement error in detail, the Central Limit Theorem and its importance for error statistics will be explained. Subsequently, the error model generally applied in parameter estimation is described and finally manipulation of the measurement error and data averaging are discussed.

2.2.1 Central Limit Theorem

The Central Limit Theorem [9] states that when X_1, X_2,... is a sequence of independent identically distributed random variables with common mean μ and variance σ^2, and when the random variable Z_n is calculated as

$$Z_n = \frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}} \tag{2-1}$$

then the cumulative distribution function (CDF) of Z_n converges to the standard normal CDF, defined by

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx \tag{2-2}$$

in the sense that

$$\lim_{n\to\infty} p(Z_n \leq z) = \Phi(z) \tag{2-3}$$

for every value z, which is a realization of the random variable Z_n.

According to this important theorem the realizations of the sum (or average) of a large number of random variables, independently generated from an identical distribution, will eventually be normally distributed. Besides the implicit assumption that both the mean and variance are finite, there is remarkably no other requirement for the distribution of X_i, and thus any the distribution $p(X)$ can be discrete, continuous, or mixed.

Random noise in many natural or engineered systems is considered to be the sum of many small, but independent random factors. The statistics of noise have empirically been found to be represented correctly by normal distributions, and the Central Limit Theorem provides the explanation for this phenomenon.

2.2.2 Random Error: Instrument Noise and Process Variability

Not considering systematic error, measurement error in experimental data can be considered composed of two different types of random error: (1) the instrument noise, and (2) the natural variability of the system, as schematically illustrated in Figure 2-2.

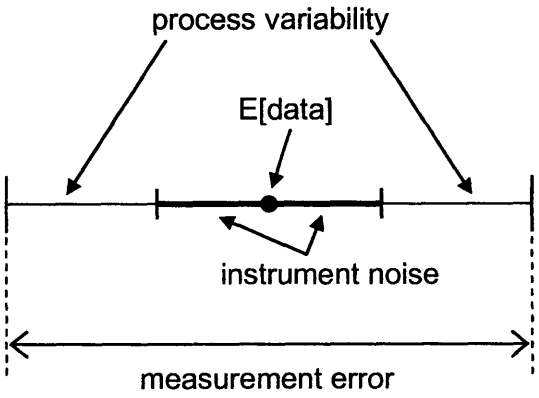


Figure 2-2. Bisection of measurement error

An example from actual data from Case Study 1 (see Chapter 7) demonstrating the difference between instrument noise and measurement error is shown in Figure 2-3. The baseline signal (at $t < 0$) can be considered representative for the instrument noise, while

during the experiment (at $t \geq 0$), the larger randomness in the signal results from a combination of instrument noise and the inherent process variability.

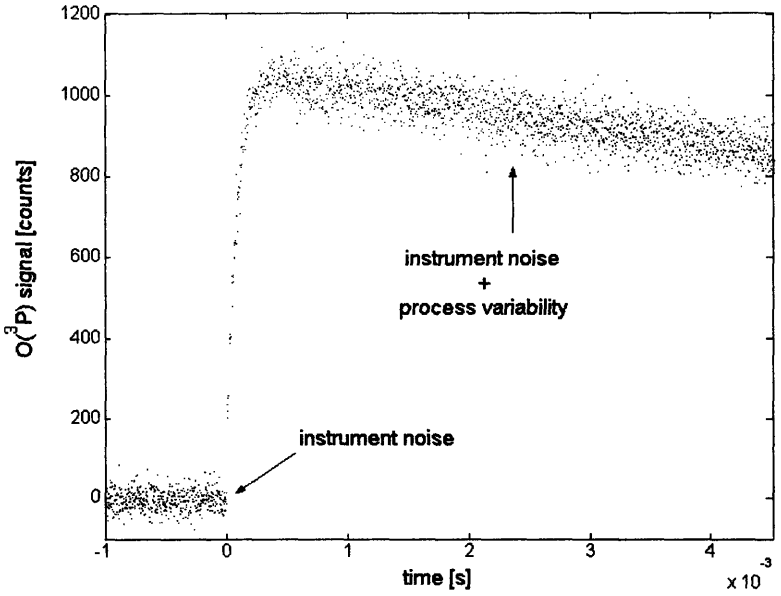


Figure 2-3. Example of instrument noise and process variability

2.2.3 Error Model

Combining the two sources of uncertainty discussed above, a measurement can be represented by

$$y_t = y_{t,true} + \varepsilon_n + \varepsilon_v \tag{2-4}$$

where y_t is the observed value of $y_{t,true}$ at time $t \geq 0$, ε_n is the instrument noise, and ε_v is the process variability. For example ε_n and ε_v can be specified as normally distributed random variables. Since normality is preserved upon linear transformations of normal random variables, the error model can be rewritten as

$$y_t = y_{t,true} + \varepsilon_0 \tag{2-5}$$

where ε_0 is the total measurement error distributed according to $\varepsilon_0 \sim N(0, \sigma_0^2)$. Assuming ε_n and ε_v are independent normal random variables, the variance σ_0^2 can be determined from [2]

$$\frac{1}{\sigma_0^2} = \frac{1}{\sigma_n^2} + \frac{1}{\sigma_v^2} \tag{2-6}$$

where σ_n^2 is the variance of the instrument noise and σ_v^2 the variance of the process variability. Though σ_n^2 can be determined from baseline measurement, σ_v^2 will for most realistic situations be unknown. Therefore, the variance of the measurement error σ_0^2 is considered unknown and will be included in the Bayesian parameter estimation.

2.2.4 Influencing Measurement Error

Of the two components of measurement error, the instrument noise can to some extent be controlled by e.g. the selection and correct tuning of the equipment, minimization of disturbances, etc. Beyond decisions regarding hardware selection and experimental setup, the effect of process variability can be decreased by measurement averaging, a feature occurring in both Case Studies 1 and 2 (see Chapter 7 and 8). For example, each data point at time t shown in Figure 2-3 is determined by summing measurements (which is equivalent to averaging in this context) at time t from several experiments.

Averaging has the useful characteristic that more accurate and precise data points are attained compared to the individual measurements. This effect is illustrated in Table 2-1, where the standard normal random variable Y_i is sampled n times to determine random variable Z as

$$Z = \frac{Y_1 + Y_2 + \ldots + Y_n}{n} \tag{2-7}$$

which is the sample mean of $\{Y_1, Y_2, \ldots, Y_n\}$.

Table 2-1. Improving accuracy and precision by averaging measurements

n	$E[Z]$	σ_Z^2
5	0.2561	0.1669
10	-0.0751	0.0975
20	-0.0666	0.0827
30	0.0312	0.0337
50	0.0123	0.0181
100	0.0018	0.0091
∞	0	0

The example shows that averaging an increasing number of samples n will improve both accuracy ($Z \rightarrow E[Y_i] = 0$) and precision ($\sigma_Z^2 \rightarrow 0$), as justified by the Strong Law of Large Numbers [9]

$$P\left(\lim_{n\to\infty}\frac{Y_1 + Y_2 + \ldots + Y_n}{n} = \mu\right) = 1 \tag{2-8}$$

where the mean of the standard normal distribution $\mu = 0$.

2.2.5 Systematic Error

Systematic error is a consistent error that is repeatable among multiple experiments. The only way to identify and quantify a systematic error is through comparisons with experimental results from different equipment or through calibration. Information on the possible existence is therefore necessary to account for systematic errors when analyzing the data, either with the proposed Bayesian approach or with any other statistical method. Nevertheless, as will be shown in Case Study 1 in Chapter 7, the Bayesian approach facilitates a convenient intercomparison among estimation results visualizing any significant differences possibly caused by a changing systematic error. Such knowledge is extremely important to decide on the validity of the data and to recognize the undesirable influence of experimental settings, environmental effects, or other factors affecting experimentation.

2.3 Limitations of Conventional Statistics

This section will highlight some limitations and widespread misinterpretations of conventional statistical estimation methods. After illustrating the effect of approximations and transformations on uncertainty, standard error propagation is discussed. Finally, the justification that the uncertainty of a parameter estimate is always normally distributed when averaging over sufficient number of data points, is shown to be based on a misconception.

2.3.1 Linear Error Statistics

A general linear model with only two parameters is specified as

$$y_i = a + bx_i + \varepsilon \tag{2-9}$$

where y_i is the dependent variable for $i = 1,\ldots,n$ data points, x_i the independent variable, a and b are the intercept and slope, respectively, of the linear model, and ε is the error term which is normally distributed according to $\varepsilon \sim N(0, \sigma^2)$ with zero mean and unknown variance σ^2. An equivalent, but more general notation is

$$y = \mathbf{X}\beta \tag{2-10}$$

where β is the parameter vector and $\mathbf{X}$ is the design matrix, in this case given by

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \tag{2-11}$$

After obtaining the estimates $\hat{a}$ and $\hat{b}$ with linear regression [10] by solving

$$\hat{\beta} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}y \tag{2-12}$$

the sample variance s^2, which is an unbiased estimator for the unknown variance σ^2, can be calculated from the residual sum of squares according to

$$s^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\left(y_i - (\hat{a}+\hat{b}x_i)\right)^2 \tag{2-13}$$

for n data points and k independent variables. The estimates $\hat{\sigma}_{\hat{a}}^{\,2}$ and $\hat{\sigma}_{\hat{b}}^{\,2}$ of the variances of $\hat{a}$ and $\hat{b}$ are the diagonal elements of covariance matrix $\mathbf{\Sigma}$, which can subsequently be determined as

$$\Sigma = \begin{bmatrix} \hat{\sigma}_{\hat{a}}^{\,2} & \hat{\sigma}_{\hat{a}\hat{b}} \\ \hat{\sigma}_{\hat{a}\hat{b}} & \hat{\sigma}_{\hat{b}}^{\,2} \end{bmatrix} = s^2(\mathbf{X}^T\mathbf{X})^{-1} \tag{2-14}$$

2.3.2 Nonlinear Approximations

Though equation (2-14) is derived for linear systems [10], it is often similarly applied for nonlinear models by approximating the nonlinearity with a Taylor series expansion, of which only the linear terms are retained. The following example [8] clearly shows that the sensitivity of the nonlinear model output to uncertain input parameters should not be underestimated. In addition, it demonstrates that the uncertainty in the model output only can be properly characterized by the full probability density function to be obtained by evaluating the complete model, and not an approximation.

We are interested in evaluating the ratio of two random variables A and B. Such calculation occurs in a variety of applications, e.g. to calculate speed when measuring distance and time, or to determine equilibrium constants in chemical reactions. To evaluate this system, the model is defined as

$$f(A,B) = \frac{A}{B} = \frac{a + \varepsilon_a}{b + \varepsilon_b} = f(\varepsilon_a, \varepsilon_b) \tag{2-15}$$

where a and b are non-negative constants, and ε_a and ε_b are independent random variables distributed according to the standard normal distribution $N(\mu,\sigma^2) = N(0,1)$, with

$$N(0,1) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{2-16}$$

The common misconception is that the probability density function for the function $f(\varepsilon_a, \varepsilon_b)$ is normally distributed as well. However, the analytical expression for this probability density function is given by

$$\Phi\left(\frac{a+\varepsilon_a}{b+\varepsilon_b}\right) = \Phi(t) = \frac{e^{-\frac{a^2+b^2}{2}}}{\pi(1+t^2)}\left[1 + \frac{q}{\phi(q)}\frac{1}{2}Erf\left(\frac{q}{\sqrt{2}}\right)\right] \quad with \quad q = \frac{b+at}{\sqrt{1+t^2}} \tag{2-17}$$

Obviously, the calculation of the mean of $f(\varepsilon_a, \varepsilon_b)$ by the ratio of the expected values of a and b is likely to be incorrect. More formally

$$E\left[\frac{a+\varepsilon_a}{b+\varepsilon_b}\right] = \int_{-\infty}^{\infty} t\Phi(t)dt \neq \frac{E[a]}{E[b]} \tag{2-18}$$

further supporting the statement that uncertainty should be incorporated at the start of a problem and propagated while solving a problem. The final solution should be obtained by evaluating the full probability distribution.

For a clear illustration that uncertain input can have a significant effect on the outcome of a problem, the probability density functions for three combinations of the constants a and b are shown below in Figure 2-4. Depending on the values of a and b the distribution can

exhibit unimodal symmetric behavior, bimodal symmetric, or even bimodal asymmetric behavior.

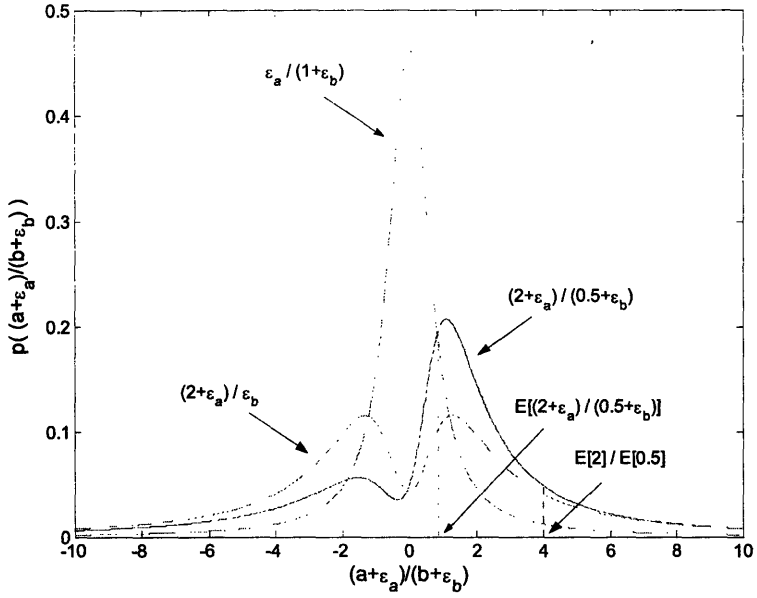


Figure 2-4. Probability density functions for $(a+\varepsilon_a)/(b+\varepsilon_b)$

2.3.3 Error Propagation

A possible representation of the uncertainty in the outcome y of a calculation as a function of the uncertainty in the independent variables can be attained by error propagation. The two basic rules in error propagation are as follows [11]

$$y = x + z - (u + w) \quad \Rightarrow \quad \delta y = \delta x + \delta z + \delta u + \delta w \tag{2-19}$$

$$y = \frac{x \cdot z}{u \cdot w} \quad \Rightarrow \quad \frac{\delta y}{|y|} = \frac{\delta x}{|x|} + \frac{\delta z}{|z|} + \frac{\delta u}{|u|} + \frac{\delta w}{|w|} \tag{2-20}$$

where δi represents the uncertainty in the corresponding variable i. In the special case that the variables are independent and normally distributed, error propagation can be calculated as

$$y = x + z - (u + w) \quad \Rightarrow \quad \delta y = \sqrt{(\delta x)^2 + (\delta z)^2 + (\delta u)^2 + (\delta w)^2} \tag{2-21}$$

$$y = \frac{x \cdot z}{u \cdot w} \quad \Rightarrow \quad \frac{\delta y}{|y|} = \sqrt{\left(\frac{\delta x}{|x|}\right)^2 + \left(\frac{\delta z}{|z|}\right)^2 + \left(\frac{\delta u}{|u|}\right)^2 + \left(\frac{\delta w}{|w|}\right)^2} \tag{2-22}$$

When the uncertainty δi is represented by the standard deviation σ_i, equation (2-22) can be derived as follows. Consider the general model denoted as

$$y(x) = f(x_1, x_2, ..., x_n) \tag{2-23}$$

and approximate $y(x)$ by a Taylor series expansion. By using the following properties of expectation operators [9] when $Z = aP + b$

$$\begin{aligned} E(Z) &= aE[P] + b \\ \sigma_Z^2 &= a^2 \sigma_P^2 \end{aligned} \tag{2-24}$$

where Z and P are random variables, and a and b are given scalars, the total variance σ_y^2 can be calculated from the individual variances of x_i as

$$\sigma_y^2 = \left(\frac{\partial y(x)}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial y(x)}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \cdots + \left(\frac{\partial y(x)}{\partial x_n}\right)^2 \sigma_{x_n}^2 + \text{covariance terms} + \cdots \tag{2-25}$$

When the covariance between the independent variables x_i can be neglected, the desired equation is obtained. This can be illustrated with the following example. Consider the system model

$$y(x) = x_1 x_2 x_3 \tag{2-26}$$

so that

$$\frac{\partial y}{\partial x_1} = x_2 x_3 \; ; \quad \frac{\partial y}{\partial x_2} = x_1 x_3 \; ; \quad \frac{\partial y}{\partial x_3} = x_1 x_2 \tag{2-27}$$

and assuming that the parameters are uncorrelated, then

$$\sigma_y^2 = (x_2x_3)^2\sigma_{x_1}^2 + (x_1x_3)^2\sigma_{x_2}^2 + (x_1x_2)^2\sigma_{x_n}^2 \tag{2-28}$$

which can be rewritten as

$$\frac{\sigma_y^2}{(x_1x_2x_3)^2} = \left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2 + \left(\frac{\sigma_{x_3}}{x_3}\right)^2 \tag{2-29}$$

leading to

$$\frac{\sigma_y}{|y|} = \sqrt{\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2 + \left(\frac{\sigma_{x_3}}{x_3}\right)^2} \tag{2-30}$$

which is the quadrature summation given in equation (2-22) that is applicable to error propagation for a system model as given in equation (2-26).

Though quadrature summation is a convenient approximation of uncertainty propagation, the mathematical manipulations include some assumptions and, more importantly, full probability distributions are not considered. Therefore, though applicable to both products and quotients, equation (2-22) can never represent the situation as discussed in Section 2.3.2, where possible probability distributions of a quotient of two normally distributed variables were considered (see Figure 2-4).

2.3.4 Abuse of the Central Limit Theorem

As inference in data analysis usually produces average estimates over series of experimental data, the common misconception is that the uncertainty of the parameter is normally distributed. This misconception has its origin in the confusing terminology of conventional statistics.

According to conventional statistics the 'true' value of a parameter θ is considered fixed, the 'statistic' $\hat{\theta}$ is introduced to account for the random effects of experimentation (more on this below in Section 3.2.1). This statistic $\hat{\theta}$, which is a random variable representing the unknown parameter θ, is merely a point estimate. When each experiment is considered as an independent random draw from the population of $\hat{\theta}$, then its expected value $E[\hat{\theta}]$ can, at a sufficiently large sample size, be considered normally distributed according to the Central Limit Theorem (see Section 2.2.1).

The misconception is then to assume that the statistic $\hat{\theta}$ is normally distributed. However, bear in mind that while the estimation result $E[\hat{\theta}]$ is likely normally distributed, the statistic $\hat{\theta}$ can be distributed according to any probability density function, while the parameter θ is fixed and thus not distributed at all.

The above discussion is illustrated by the following example. Suppose that a sequence of random variables Y_i is drawn from a particular distribution $p(Y)$, yet to be defined. The statistic $\hat{\theta}$ can be considered equivalent to the random variable Y_i. Subsequently, random variable Z is calculated by applying equation (2-7), repeated here for convenience

$$Z = \frac{Y_1 + Y_2 + \ldots + Y_n}{n} \tag{2-7}$$

so that the distribution $p(Z)$ becomes approximately normal for sufficiently large n. The expected value $E[\hat{\theta}]$ is equivalent to the normally distributed random variable Z.

Figure 2-5 shows the distributions $p(Y)$ and $p(Z)$ for the cases of sampling Y_i from a uniform and a lognormal distribution, for $n = 10$. The probability distributions $p(Y)$ are plotted exactly, while the probability distributions $p(Z)$ are obtained from a kernel density estimate of 500 samples for Z and are thus the result of $500 \cdot 10$ draws of Y_i.

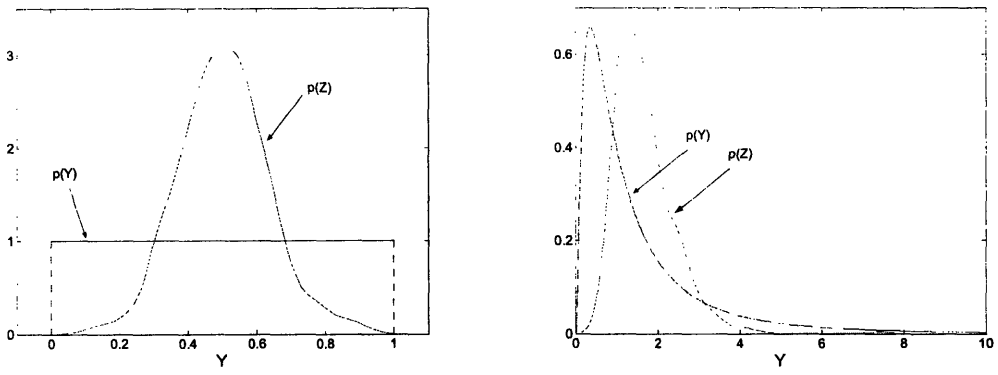


Figure 2-5. Illustrations of the Central Limit Theorem

Conventional estimation methods only provide the expected values of point estimates $E[\hat{\theta}]$, which indeed can be considered normally distributed. However, typically the normally distributed $E[\hat{\theta}]$ is mistaken for the statistic $\hat{\theta}$, which is subsequently also considered to follow a normal distribution. As seen in the example above though, this does not have to be the case at all.

3

The Bayesian Approach

Inside every Non-Bayesian, there is a Bayesian struggling to get out.

Dennis V. Lindley

When dealing with uncertain systems, incorporation of the uncertainty at the stage of problem formulation and uncertainty propagation throughout the problem-solving process is important. The Bayesian approach is a powerful method based on unified underlying principles accomplishing a consistent treatment of uncertainty. After a review of the roots of probability theory and some conceptual issues regarding plausible reasoning, three definitions of probability theory, Classical, Frequentist, and Bayesian, the characteristics of the Bayesian approach, and some application issues will be described in more detail. Finally, the relationship of the Bayesian approach with common alternative estimation methods will be illustrated.

3.1 Introduction

A variety of topics are briefly discussed to introduce the conceptual framework of the Bayesian approach that forms the core of this thesis. A note on uncertainty is followed by a discussion on plausible reasoning and induction, and an explanation how the Bayesian viewpoint fits to all of this.

3.1.1 Inescapable Uncertainty

We can never have complete knowledge regarding phenomena under investigation. Solving problems, being it in engineering or any other field, will always be plagued by

uncertainty originating from various sources [7, 8]. Our models are an approximation of reality, data is collected from inherently variable systems, the measurement equipment introduces additional error, etc. The generally accepted language to quantify uncertainty is probability theory. Probability can be interpreted as "the degree of uncertainty, which is to the certainty as the part to the whole." This very early definition by Bernoulli echoes the more recent Bayesian perspective on probability and will be applicable throughout this thesis.

3.1.2 Reasoning

The first recorded activity in the field of probability theory originates from the gambling tables of the 16th century. Realizing that though the future could not be predicted, the limited number of possible outcomes of a certain game allowed the reasoning about the best bet to place, in other words about the best decision to make with minimal risk. This type of reasoning is based on deductive logic, schematically shown in Figure 3-1 [12], and though professional gamblers can attest to its usefulness, applications to realistic problems in science and engineering are rather limited.

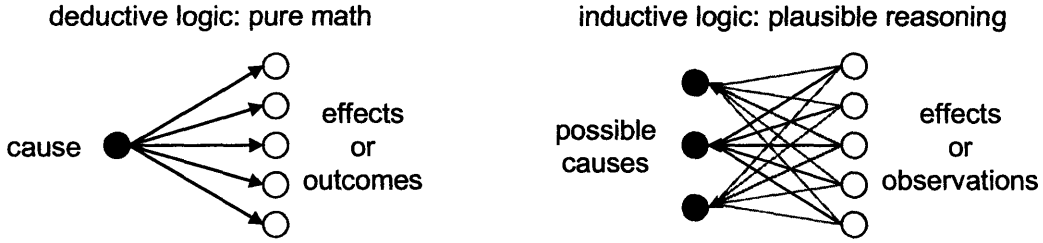


Figure 3-1. Deductive and inductive logic

Bernoulli's theorem, also known as the Law of Large Numbers, was the first formal account in calculating probabilities. Identifying the distinction between frequency and probability, the Law of Large Numbers relates the probability of occurrence in a single trial to the frequency of occurrence in a large number of independent trials. Bernoulli also posed the inverted problem of what can be said about the probability of a certain event after observing the frequency of occurrence of this event. This kind of question is

typified as inductive logic, or plausible reasoning, which is more interesting with regards to realistic applications.

3.1.3 Inductive Logic

The answer to Bernoulli's question was published posthumously by Reverend Bayes [13, 14], while Laplace [15] further developed and applied the theory with remarkable success to realistic estimation problems in astronomy.

Inductive logic, which enabled the inferences made by Bayes and Laplace, is significantly more complicated than deductive logic. For example, when the probability p that a fair coin comes up with a head is known, then the probability $p(k)$ of observing k heads in a series of n tosses can be without difficulty be calculated with the binomial probability distribution as

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{3-1}$$

However, the inverse problem of inferring the value for p when k heads are observed in a series of n tosses is not so obvious, which is unfortunate as realistic situations encountered by scientists and engineers are mostly formulated as such.

3.1.4 Degree of Belief

The fundamental idea pioneers such as Bernoulli, Laplace, and Bayes introduced, is that probability represents a *degree of belief* or *degree of plausibility* regarding the truth of a proposition, when incomplete knowledge does not suffice to make a statement with certainty. In more specific terms: probability is a real number between 0 and 1, where a probability of 1 indicates the proposition is true, and a probability of 0 it is not. A probability in between 0 and 1 measures a state of uncertainty regarding the proposition.

The degree of belief in the value of a certain parameter θ can be depicted by using probability distributions, such as the normal distribution shown in Figure 3-2.

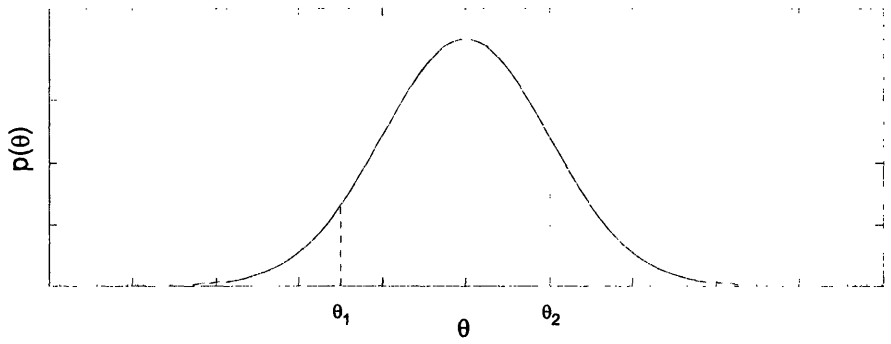


Figure 3-2. Normal probability distribution measuring the degree of belief regarding θ

It is important to clarify that the illustration above does not convey that the parameter θ is a random variable. On the contrary, the value of parameter θ is considered fixed but its value is unknown. It is the *degree of belief* or *state of knowledge* regarding the parameter θ that is distributed.

One of the many advantages of the Bayesian point of view is that particular propositions regarding parameter θ can directly be assessed using the probability distribution, for example

$$p(\theta_1 \le \theta \le \theta_2) = \int_{\theta_1}^{\theta_2} p(\theta) d\theta \tag{3-2}$$

These basic concepts will be discussed again in more detail when introducing the various definitions of probability theory.

3.1.5 Illusory Randomness

Incorrectly identifying a fixed parameter as a random variable, to account for measurement uncertainty for example, implies that some undefined randomness is the cause for the variability in the measurements. To understand this objection, consider the

output of a random number generator, which produces numbers in a sequence *appearing* random [12]. The deceptive nature of this randomness becomes clear, however, if the *algorithm* and the *seed* of the random number generator would be known, since that would allow the prediction of the future output. In other words, the randomness is the result of a deterministic process of which too little information is available to completely specify the system.

Though simplistic, this example parallels most realistic applications, where the inability to predict reflects the lack of knowledge regarding the system. The Bayesian perspective on probability is consistent with this interpretation, which on a fundamental level explains the intuitive characteristics of the Bayesian approach.

3.2 Probability Theory

The natural language to mathematically describe and deal with uncertainty is probability theory, and thus is it important to elaborate on the different perspectives regarding probability theory. This section will distinguish between the conventional definitions and the Bayesian point of view that is gaining popularity in recent years.

3.2.1 Conventional Definitions

Introductory texts on probability theory usually start with explaining the 'classical' definition of probability theory. Under the classical definition, probability is defined as *'the ratio of number of favorable outcomes to the total number of outcomes, if all outcomes are equally likely.'* Thus, the probability of event A is

$$P(A) = \frac{n_A}{N} \tag{3-3}$$

where n_A is the number of outcomes of event A, and N is the total number of possible outcomes in the sample space, where each of the elements of N is equally likely, i.e. has the same probability. The weakness in this definition of probability is the usage of the

term 'probability' in the definition itself. Indeed, the statement given in equation (3-3) is better suited to evaluate probability in cases of deductive reasoning, than to define it. On a side-note: since the "classicals" in probability theory (i.e. Bernouilli, Bayes, Laplace, Gauss, etc.) thought differently about this matter, the classification 'classical' is considered a misnomer, and the term 'combinatorial' would be more suitable [6].

The second most common definition of probability considers *the frequency of occurrence, as a percentage of success in a large number of trials*. This 'Frequentist' definition requires that the number of trials goes to infinity, either in reality or in a thought experiment. The major flaw of this definition is the assumption that the probability of future events, which are considered to be drawn from the same infinite population, is the same as the probability of past events. Such an assumption, however, can never be proven to be correct.

Difficulties arise since the Frequentist approach is only applicable to inherently repeatable events, while realistic scientific questions deal often with unique problems, for example estimating the mass of Saturn. According to the Frequentist definition, the mass of Saturn is a constant and not a random variable, so probability theory cannot be applied for the estimation. The invention of the discipline of statistics was the answer to this problem. The reasoning is that the mass of Saturn can be related to the data via a function called the 'statistic', which can be treated as a random variable as the data is subject to 'random' noise. The next question is however which statistic is suitable. The Frequentist approach does not allow for an intuitive answer and therefore led to a collection of tests and procedures seemingly lacking any unifying principles [12, 16].

A practical weakness of conventional estimation methods based on the Frequentist approach is the treatment of uncertainty as something that is detached from the system under consideration. A problem is solved by tagging an uncertainty estimate onto a deterministic solution of the problem. The confidence interval is a tool developed for the Frequentist approach representing the uncertainty as detached from the actual problem-

solving process. Though commonly applied, the interpretation of confidence intervals is often incorrect, as discussed below [17].

Suppose that the random variable $\bar{X}$, which is the statistic (an estimator) for the true value μ, can be determined from data, so that the confidence interval for known standard deviation σ can be specified as

$$P(\bar{X} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}}) = 68\% \tag{3-4}$$

or similarly for the more frequently applied 95% confidence interval

$$P(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 95\% \tag{3-5}$$

which are *pre-data* probabilistic statements regarding the random variable $\bar{X}$. This statement implies that for 68%, and 95% respectively, of the experiments to determine $\bar{X}$, the true value will be covered by the interval specified.

However, once the experiment is performed and data allows for calculation of $\bar{x}$, which is a realized value for random variable $\bar{X}$, nothing is random anymore and a probabilistic statement cannot be made. Thus the confidence interval

$$\bar{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} \tag{3-6}$$

either contains the fixed true value μ or not. The actuality, however, is unknown. Because of the statement in (3-4) there is 68% *confidence* that the true value is covered by the interval specified in (3-6), but the interpretation of a confidence interval as a probabilistic statement regarding the fixed and unknown true value μ is incorrect. For

examples and further detailed discussion on confidence intervals and similar issues of hypothesis testing, the reader is referred to the references [6, 17-19].

3.2.2 Bayesian Probability Theory

The Bayesian perspective of probability as the *degree of belief* as illustrated in Section 3.1.4 is far more general and intuitive than the conventional definitions of probability theory discussed above. Since the Bayesian point of view allows evaluation of propositions that can be anything, e.g. a model parameter, the evaluation of a data set in a unique experiment or a one-time event, there are not really any restrictions as long as the problem is well-posed.

Before evaluating the plausibility of propositions though, the above definition of probability has to allow mathematical manipulation. Fortunately, even though the notion of *degree of belief* may sound vague, a rigorous probability theory can be assured if the measure of belief answers the following qualitative desiderata [16, 20]:

I. *Degrees of plausibility are represented by real numbers*
   A theory similar to the way we reason should be associated to real numbers and the convention is that a greater plausibility corresponds to a greater number
II. *Qualitative correspondence with common sense*
   For example, evaluation of the propositions in a transitive manner
IIIa. *Internal consistency*
   If multiple ways lead to the same conclusion, each must give the same result
IIIb. *Propriety*
   All given information relevant to a question should be considered
IIIc. *Jaynes Consistency*
   Equivalent states of knowledge represent equivalent plausibility assignments

The desiderata given above allow for derivation of the mathematical content of the Bayesian definition of probability given by two axioms [21, 22], which are the sum rule

$$p(A \mid C) + p(\bar{A} \mid C) = 1 \tag{3-7}$$

stating that the probability of proposition A, given proposition C, together with the probability of "not A", given C, make up all possibilities, and the product rule

$$p(A, B \mid C) = p(A \mid B, C)p(B \mid C) \tag{3-8}$$

which specifies the calculation of the probability of both A and B, given C. The vertical bar represents conditional probability, indicating the information that is assumed for assigning the probabilities. Assumptions are always made regarding the phenomenon under consideration, and they should be explicitly stated.

With the above axioms, the probability of combinations of propositions can be calculated, given the probability of the original propositions. The assigned probability of these original propositions serves as input for application of Bayesian probability theory. The *principle of indifference* and the use of sampling distributions are common methods to assign probabilities, i.e. represent the state of knowledge regarding a proposition. Sampling distributions, also known as likelihood functions, allow calculation of a probability based on data collected and the specific functional form of a sampling distribution, such as the normal distribution, depends on the system.

The above set of rules, with the intuitive character embodied in the desiderata and the logic embodied in the axioms, forms a complete theory sufficient for the analysis of uncertainty in scientific problems. Since limitations have not been imposed, e.g. mention of relative frequencies or random variables, this theory is generally applicable.

3.2.3 The False-Positive Puzzle

Before continuing with specific details regarding the Bayesian approach, the following example [23] emphasizes the non-intuitive nature of problems related to probability and that Bayesian probability theory provides for a natural answer.

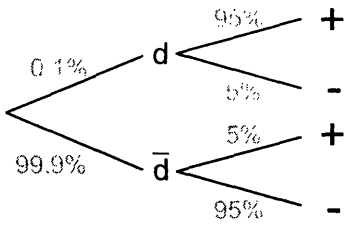
The problem is formulated as follows:
Suppose that you are selected at random from a population where on the average one person in a 1000 has a certain disease. After you have been tested positive for this disease, you are informed that the test is incorrect 5% of the time. What is the probability you actually have the disease?

The intuitive answer given by most people is a probability of 95% that you have the disease. Similarly to the probability inversion of the confidence intervals discussed above, this answer follows from the incorrect inversion

$$p(\textit{tested positive} \mid \textit{not infected}) = p(+ \mid \bar{d}) = 5\% \quad \Rightarrow \quad p(\bar{d} \mid +) = 5\% \qquad (3\text{-}9)$$

leading to the false conclusion that the probability of being infected with the disease when the test result is positive is 95%. Though this problem seems obvious, only a thorough analysis will reveal the answer.

The solution to the problem will become clear by constructing a tree diagram of the relating all possible combinations, as given in Figure 3-3. The first branch in the tree diagram represents the knowledge that approximately 0.1% of the population is infected with the disease (d), while the rest is not ($\bar{d}$). For each of these two states the information on the test accuracy is used to construct the next branch. For an infected person (d), the test result is positive (correct) 95% of the time and negative (incorrect) 5% of the time. For a healthy person the opposite is true.



$$\frac{0.001 \cdot 0.95}{0.001 \cdot 0.95 + 0.999 \cdot 0.05} = \boxed{0.0187}$$

Figure 3-3. Tree diagram and solution for the false-positive puzzle

The answer what the probability of infection is upon a positive test result can easily be derived from the tree diagram by dividing the probability for a positive test result for an infected person by the total probability that the test gives a positive result. Surprisingly, the probability for infection upon a positive test result is not even 2%. The solution is calculated with Bayes' theorem, which can for this problem be defined as

$$p(d \mid +) = \frac{p(+ \mid d)p(d)}{p(+ \mid d)p(d) + p(+ \mid \bar{d})p(\bar{d})} = \frac{p(+ \mid d)p(d)}{p(+)} \tag{3-10}$$

Clearly, the result is affected by the current knowledge of the occurrence of the disease in the general population. However, the use of this knowledge is perfectly justified when the test subject is selected at random from this population. In the different situation of selecting the test subject from a group of high-risk individuals, the doctor might suspect that the *a priori* probability of being infected is about as high as the probability of not being infected, so that

$$p(d) \approx p(\bar{d}) \tag{3-11}$$

In this case, the probability of being actually infected upon a positive test result is 95%. Thus, this example illustrates the statement made at the end of Section 3.2.1 that the probability inversion, such as in equation (3-9), is only valid for symmetric problems.

3.3 The Bayesian Approach

This section discusses the mathematical details of Bayes' theorem. After deriving Bayes' theorem is derived, several implementation issues will be discussed. Finally, the relation of the Bayesian approach to conventional estimation procedures will be illustrated.

3.3.1 Bayes' Theorem

The general derivation of Bayes' theorem follows from the product rule given in equation (3-8). Let $A = H$, the hypothesis under consideration, $B = D$, the data relevant to the hypothesis, and $C = I$, the background information that defines the problem that is addressed. With these propositions

$$p(H,D \mid I) = p(H \mid D,I)p(D \mid I) = p(D \mid H,I)p(H \mid I) \tag{3-12}$$

which states that under the assumptions I the probability of both H and data D is equal to the probability of H given that D is true, multiplied by the probability of D, and vice versa. Bayes' theorem is then easily derived by rearranging one of the terms, resulting in

$$p(H \mid D,I) = \frac{p(D \mid H,I)p(H \mid I)}{p(D \mid I)} \tag{3-13}$$

which can be considered as the updating of the prior probability $p(H|I)$ regarding proposition H. The information from data D is incorporated assuming the truth of H via the likelihood function $p(D|H,I)$ to obtain the posterior probability $p(H|D,I)$. The posterior distribution is the desired outcome as it embodies all currently available knowledge regarding the proposition under consideration.

Bayes' theorem is illustrated in Figure 3-4, where the information from D reduces the uncertainty in H. This reduction in uncertainty is represented by the decreasing width of the probability distribution when updating from $p(H|I)$ to $p(H|D,I)$.

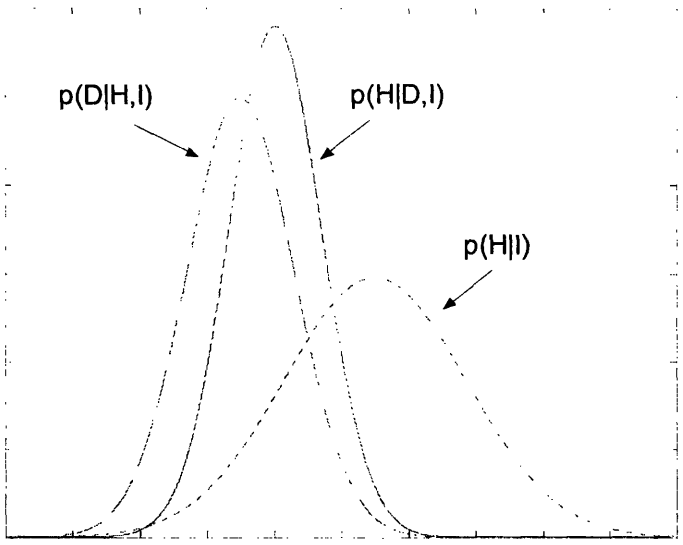


Figure 3-4. Illustration of Bayes' theorem

Throughout the rest of this thesis, Bayes' theorem will be used to evaluate the parameter θ with the information from the data y, so that

$$p(\theta \mid y) = \frac{p(y \mid \theta) p(\theta)}{p(y)} \tag{3-14}$$

where $p(\theta|y)$ is the posterior probability distribution of the parameter θ, given the data y, $p(y|\ \theta)$ is the likelihood function of the data, given the parameter θ, $p(\theta)$ is the prior distribution, and $p(y)$ is the probability distribution of the data y. Though assumptions will still be explicitly stated in the problem formulation, for notational convenience the proposition I will from now on be omitted.

3.3.2 A 'Learning' Algorithm

The procedures for evaluating data with the Bayesian approach are very flexible. Typically, Bayes' theorem can be considered as an updating or a learning algorithm.

Equation (3-14) serves from that perspective only a single step the updating sequence, which can continue indefinitely, as illustrated in Figure 3-5.

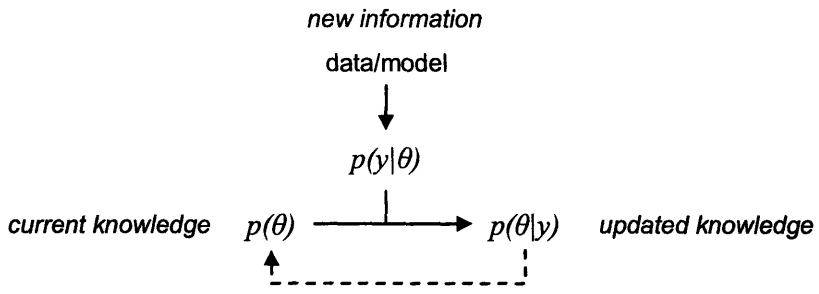


Figure 3-5. The Bayesian approach as learning or updating algorithm

The posterior distribution obtained initially can subsequently be considered as the prior distribution when additional data becomes available. The updated posterior distribution is determined by incorporating the additional data and is thus supposed to have increased information content. The updating sequence is more formally by

$$p(\theta \mid y_1) \propto p(y_1 \mid \theta)p(\theta) \tag{3-15}$$

$$p(\theta \mid y_1, y_2) \propto p(y_2 \mid \theta)p(\theta \mid y_1) = p(y_2 \mid \theta)p(y_1 \mid \theta)p(\theta) \tag{3-16}$$

$$\dots$$

$$p(\theta \mid y_1, y_2, \dots, y_n) \propto p(y_n \mid \theta)p(\theta \mid y_{n-1}) = p(y_n \mid \theta)\dots p(y_2 \mid \theta)p(y_1 \mid \theta)p(\theta) \tag{3-17}$$

This learning process appears to correspond to the natural behavior of people as is shown in an experiment where subjects had to make a decision based on prior probability and data. The experiment revealed that the decision rule most likely to be used is Bayes' Theorem [24]. Interestingly, the architecture of the nervous system seems to be very suitable for Bayesian inference. A recent publication [25] presents a Bayesian view of sensorimotor learning, where subjects were shown to be following Bayes' Theorem when combining uncertain sensory information with certain prior information. A real-life example of these results would be a tennis player studying his/her opponent to establish

prior knowledge. This prior is during the match combined with the uncertain sensory input to estimate the speed and direction of tennis ball.

3.3.3 Prior Distribution

The Bayesian approach is often criticized for the necessity of assumptions regarding available knowledge. Available knowledge enters the problem via the prior distribution, and critics of the Bayesian approach claim this to be subjective, while science should be objective. Nevertheless, any statistical method using probability is subjective when based on mathematical realizations of the world. Scientific judgment is required to select the data to evaluate and to decide on the parametric form of the distributions and model verification [26].

In the Bayesian approach scientific judgment is required to define a prior distribution based on previous experience. Depending on the actual information regarding the possible values of a phenomenon under study, the prior distribution will be either non-informative or informative, as illustrated in Figure 3-6.

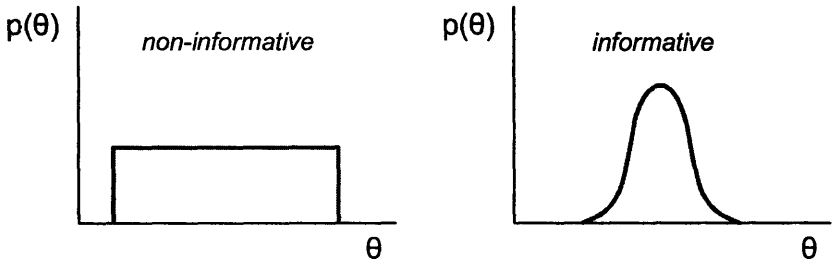


Figure 3-6. Examples of prior distributions

The non-informative prior distribution, sometimes called diffuse prior, is usually a uniform distribution only conveying a minimum amount of information, such as a physical parameter limit. Generally, a non-informative prior distribution hardly affects the posterior distribution, and inferences are primarily based on the information as provided by the data. Nevertheless, a non-informative prior distribution is certainly useful. As mentioned, the limits defined in the prior distribution are propagated into the

posterior distribution. Additionally, to carry out the desired inversion discussed above and actually obtain meaningful inferences on the probability of a proposition, the prior distribution is theoretically required [6].

If instead the prior distribution is informative, indicating a strong belief in available information, the effect on the posterior distribution can be significant. Yet, when one is very confident in prior knowledge and/or the gathered data is not trusted or only of limited quantity, then the prior dominating the conclusions is very reasonable.

The effect of an informative prior was seen in the example of Section 3.2.3, where the low prior probability significantly affected the posterior probability of actually being infected: though the test was 95% accurate, the result only implied a 2% probability of infection. However, being randomly selected from the general population, the selection of the low prior probability for the test subject was justified. Without any additional knowledge regarding the test subject there was no reason to assume a different prior probability.

3.3.4 Likelihood Function

The likelihood function, representing the probability of a proposition given the data, is equivalent to the sampling distribution, thus

$$l(\theta \mid y) = p(y \mid \theta) \tag{3-18}$$

and both notations are applied in literature when discussing the likelihood function of the Bayesian approach.

Essentially any type of probability distribution can be implemented as likelihood function. The decision on what particular distribution to use, such as a normal, lognormal, beta, or exponential distribution, should be based on the knowledge regarding the error structure of the data. To a certain extent the likelihood function is a model for the instrument performance.

A different perspective is that the likelihood function serves as the entry point for the system model into the estimation problem. Given the experimental conditions, the values of the observables calculated with the system model can be compared with the data. The likelihood function assigns a probability to these data depending on the difference between the calculated values and the data. With the data given, the model parameters can be tuned to obtain a model output of varying probability. Reasoning backwards, the uncertainty in the data represented by the likelihood function translates in an uncertainty in the parameter values. This explains many of the convenient features of the Bayesian approach, such as the relative ease of dealing with non-normal errors, multidimensional systems, and nonlinear models.

Finally, the likelihood function also offers the possibility of combining data measured with different precision, usually the result of employing different instruments. Under the assumption that the data sets are independent of each other, the overall likelihood function to be used in Bayes' theorem can be constructed as

$$p(y \mid \theta) = \prod_{i=1}^{n} p(y_i \mid \theta) \tag{3-19}$$

for n different data sets obtained with n different instruments. Compared to conventional estimation methods, the implementation of equation (3-19) is extremely convenient.

3.3.5 Exploitation of the Posterior Distribution

Though the full posterior distribution is the desired result, further manipulation is usually necessary to extract useful information. To dispose of nuisance parameters, which are often included for computation purposes but are not of interest for the solution of the problem, marginalization is performed according to

$$p(\theta \mid y) = \int p(\theta, \phi \mid y)\, d\phi \tag{3-20}$$

where $p(\theta,\phi|y)$ is the full joint posterior distribution, $p(\theta|y)$ is the marginal posterior distribution of the parameters of interest, and ϕ is the vector of nuisance parameters.

Similarly, the marginal probability distribution of only one of the parameters θ_i can be obtained as

$$p(\theta_i \mid y) = \int \cdots \int p(\theta_1,...,\theta_n \mid y)\, d\theta_1,...,\theta_{i-1},\theta_{i+1},...\theta_n \tag{3-21}$$

where $p(\theta_i|y)$ represents the belief in particular parameter values as inferred from the data.

A convenient implication of the Bayesian approach is that, instead of 'confidence intervals', the probability of an event or range of values for a proposition can be directly determined from the posterior distribution by determining 'credible intervals' [6, 17]. For example, the probability of parameter θ_i having a range of values can directly be calculated as the 'area under the curve' of the marginal probability distribution by

$$p(\theta_i = [a,b] \mid y) = \int_a^b p(\theta_i \mid y) d\theta_i \tag{3-22}$$

so that the issues discussed above regarding the confusion of the calculation and interpretation confidence intervals is thus irrelevant.

The posterior distribution is also useful in predicting future outcomes. Assuming a representative model, a perfect prediction can be made only when the model parameters are known exactly. Since this is impossible, the next best thing is the posterior distribution of the parameter estimates θ, representing the current state of knowledge. Obviously, the uncertainty in the parameter estimates is propagated into uncertainty of the prediction, as represented by the predictive distribution

$$p(y^* \mid y) = \int p(y^* \mid \theta) p(\theta \mid y)\, d\theta \tag{3-23}$$

where $p(y^*|\ \theta)$ is the likelihood function of the new measurement, given the parameters. Since the likelihood function is evaluated for all possible values, the predictive distribution can be interpreted as a weighted average of the likelihood function with the weighting determined by the posterior distribution.

3.4 Bayes' Theorem and Common Estimation Methods

With Bayes' theorem being the most general estimation method, the maximum likelihood and least squares estimation methods can be derived as shown in Figure 3-7 [6].

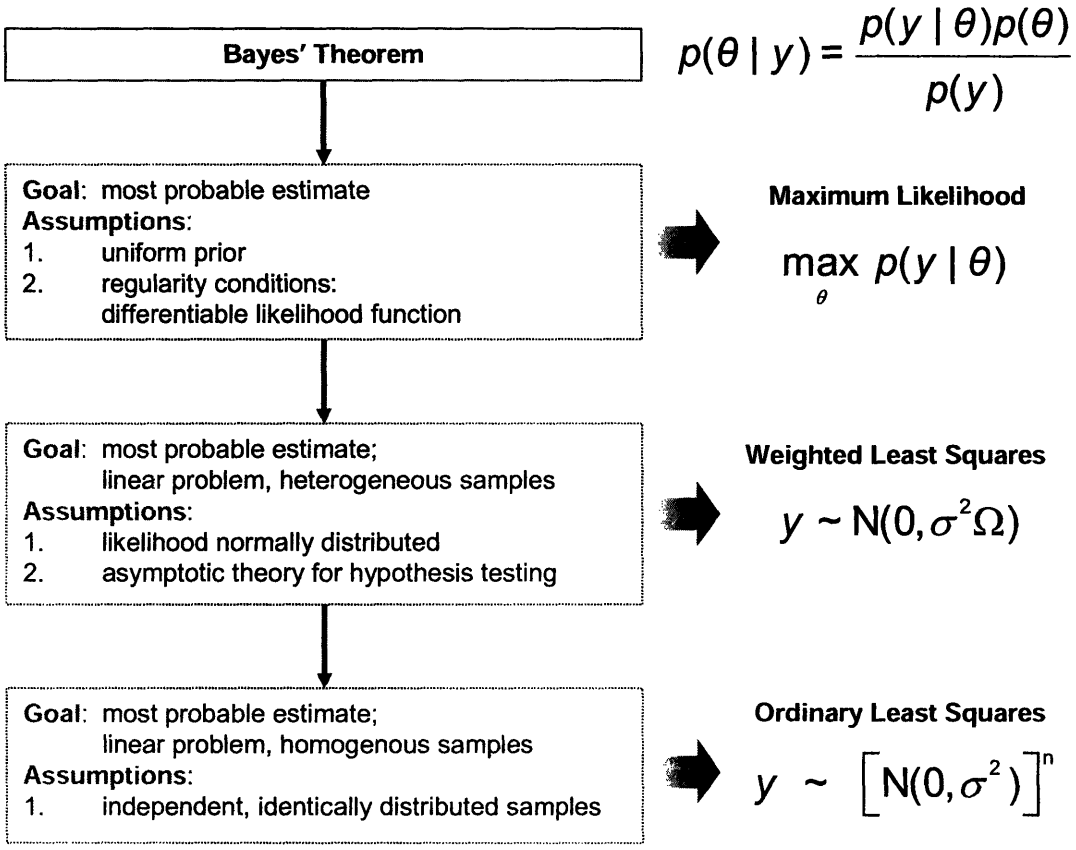


Figure 3-7. From Bayes' theorem to linear regression

The traditional estimation methods implicitly assume a uniform prior while searching for the optimum in the likelihood function. The method of maximum likelihood is the most general and does not assume any specific form for the likelihood function. The methods of least squares are based on the assumption that the likelihood function is normally distributed.

In addition, the conventional estimation methods derived above were all static. Bayes' theorem can also be used to deal with dynamic systems. As explained, Bayes' theorem essentially performs an updating of the prior distribution by incorporating information obtained from data. For a dynamic system the additional data is obtained over time, but since the principle of updating is still valid, Bayes' theorem can be applied [27].

The Kalman Filter is a very common filter that is able to update a state estimate in real time by incorporating newly acquired data. The relation to Bayes' theorem and linear regression is schematically shown in Figure 3-8. The precise mathematical derivation is given by Lorenc [28], who also derives several other estimation methods from Bayes' theorem, and more intuitively by Gamerman [29]. The relationship between linear regression and the Kalman filter is explained extensively by Strang [30].

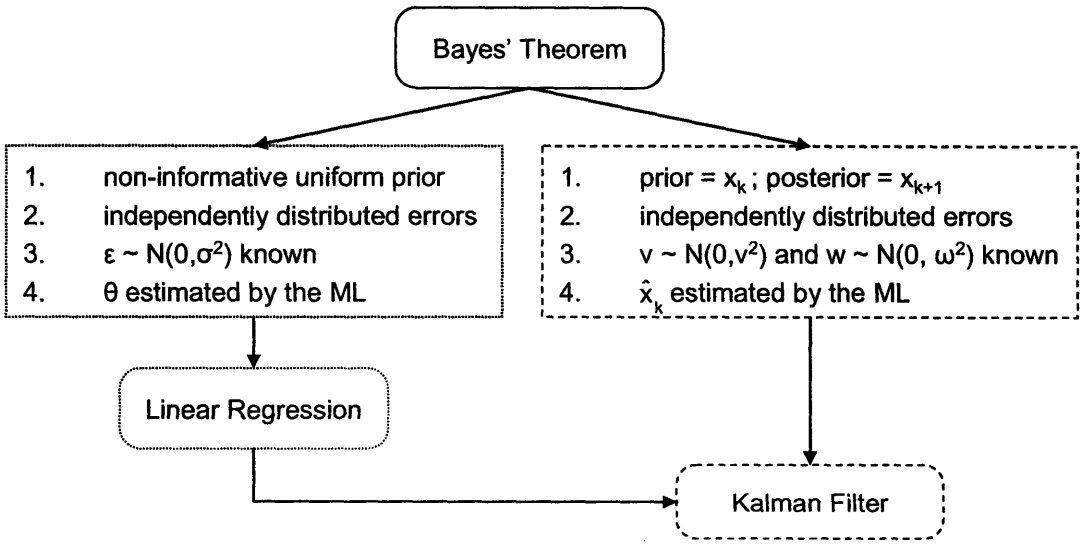


Figure 3-8. From Bayes' theorem to the Kalman filter

4

Bayesian Computation: Markov Chain Monte Carlo

Probability theory is nothing but common sense reduced to calculation.

Pierre-Simon de Laplace

Though several mathematical tools have been developed over time, Markov Chain Monte Carlo (MCMC) seems to have emerged as the most effective approach for solving the Bayesian problem formulation. The rationale for applying MCMC will be briefly discussed and in particular the Metropolis-Hastings algorithm is explained. Though the structure of this algorithm is surprisingly straightforward, implementation usually requires more thought and therefore each of the components of the Metropolis-Hastings algorithm is elaborated upon. Finally, potential sources for numerical error in the results of MCMC will be evaluated.

4.1 Introduction

Until recently, computational limitations have impeded applications of the Bayesian approach to realistic, complex systems. Though the prior and the likelihood functions in Bayes' theorem are relatively easy to obtain, the main difficulty is the denominator, which represents the probability of the data y considering all possible parameter values and can be calculated by the 'total probability theorem' as

$$p(y) = \int p(y \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{4-1}$$

where $\boldsymbol{\theta}$ is the parameter vector. Multi-dimensional parameter vectors lead to multi-dimensional integrals, which have been the major challenge throughout history in attempts to apply the Bayesian approach in a variety of estimation problems. Various mathematical techniques have been developed for computation [29, 31, 32], such as asymptotic approximation, Laplace's method, approximations by Gaussian Quadrature, and Monte Carlo integration. However, at this moment Markov Chain Monte Carlo (MCMC) techniques are considered the most effective way to circumvent the multi-dimensional integral calculations.

4.2 Markov Chain Monte Carlo

This section will provide a background on the MCMC technique, explain some elementary theory, and discuss exploiting the results obtained from MCMC simulations.

4.2.1 Rationale

MCMC methods approximate the solution to the Bayesian problem formulation by bypassing the calculation of $p(y)$, the probability function of the data. Since the posterior distribution is a function of θ at fixed y, the right hand side of Bayes' theorem must also be considered for fixed y. The key idea is then the realization that $p(y)$ is just a normalizing constant. Therefore, the posterior distribution can be approximated by only using the prior and likelihood function according to

$$p(\theta \mid y) \propto p(y \mid \theta) p(\theta) \tag{4-2}$$

after which the posterior probability can be normalized when necessary.

4.2.2 Properties

MCMC does not calculate the posterior probability precisely, but applies equation (4-2) to simulate random draws from the posterior probability. A large number of samples is

collected to obtain a representation of the characteristics of the posterior distribution. This sampling approach is similar to a Monte Carlo simulation, with the difference that the generation of samples is not completely random, but directed through the sample space by a Markov Chain [9].

In order to use a Markov Chain to sample from a distribution $\pi(\theta)$, the transition kernel $p(\theta^{i+1} | \theta^i)$ should be constructed so that $\pi(\theta)$ is the stationary distribution (also called invariant distribution) of the chain. A sufficient condition to establish this is

$$\pi(\theta^i)p(\theta^{i+1} | \theta^i) = \pi(\theta^{i+1})p(\theta^i | \theta^{i+1}) \tag{4-3}$$

which is the reversibility condition of the chain, also known as the detailed balance equation. The second requirement is that the Markov Chain is ergodic, so that there will only be one invariant distribution (the equilibrium distribution). Therefore, the chain needs to be irreducible and aperiodic. Implementation according to these relatively mild requirements causes the stationary distribution of the Markov Chain to be the limiting distribution in the sense that the distribution of the samples θ^i converges to $\pi(\theta)$ as i goes to infinity [33, 34].

Currently, two variations of MCMC are in common use: Gibbs sampling and Metropolis-Hastings sampling. The freely available software 'Bugs', or 'WinBugs', [35] is based on Gibbs sampling. Due to the convenient programming language Bugs has found widespread use in numerous applications and examples applicable to the work in this thesis are given in Appendix B. Gibbs sampling is a special case of Metropolis-Hastings sampling, which is a general method superior in coping with nonlinear and non-normal models. Therefore, only the Metropolis-Hastings algorithm, which will be discussed in Section 4.3, has been applied to the work in this thesis. As standard MCMC software applying Metropolis-Hastings sampling is not available, required programming has been performed in Matlab.

Details regarding the background and development of the algorithms are explained in [26, 31, 33], while more advanced topics regarding application are discussed in [36]. A clear and to-the-point discussion on MCMC with emphasis on practical issues is given in [37]. A comprehensive introductory tutorial on Metropolis-Hastings is provided in [38].

4.2.3 Summarizing Variables

As the samples resulting from a MCMC simulation are considered randomly drawn from the posterior distribution, they can be manipulated according to the Monte Carlo approach. And since the 'Law of Large Numbers' is valid, the precision of the results will improve with larger sample sets.

Marginal posterior distributions can be constructed by simply generating a histogram or a kernel density estimate using the samples of the parameter of interest. A Matlab toolbox available online [39] can generate kernel density estimates for up to three dimensions, i.e. for three parameters. At this moment it is not known whether computational tools are available for generating kernel density estimates of higher dimension.

Additionally, summarizing variables can be calculated with conventional sample statistics. For example, the mean and variance of a parameter can be calculated as

$$\bar{\theta} = \frac{1}{n}\sum_{i=1}^{n}\theta^i \tag{4-4}$$

$$\sigma_\theta^{\,2} = \frac{1}{n-1}\sum_{i=1}^{n}(\theta^i - \bar{\theta})^2 \tag{4-5}$$

where θ^i represents individual samples of one of the parameters and n is the total number of samples.

In general, any function of the parameters can be evaluated by using the samples so that the probability distribution of this function can be approximated. In other words, samples for y can be generated as

$$y^i = f(\theta_1^i, \theta_2^i, ..., \theta_p^i) \qquad for\ i = 1, ..., n \tag{4-6}$$

where y^i is the i^{th} sample calculated with a function of p parameters. The obtained samples y^i can subsequently be treated as the described above. A specific example of equation (4-6) is the calculation of the covariance matrix of the samples collected in order to evaluate the correlation between parameters.

The predictive distribution $p(y^*|y)$ as calculated by equation (3-23) can also be approximated by the resulting samples from MCMC. As mentioned in Section 3.3.5, the predictive distribution is a weighted average of the sampling distribution, with the weights given by the posterior distribution on the parameters. A single sample of y^* can therefore be obtained from a distribution $p(y^*|\ \theta^i)$ where θ^i is the i^{th} sample of parameters. Using the complete set of parameter samples thus approximates the predictive distribution.

4.3 Metropolis-Hastings Algorithm

This thesis applies the Metropolis-Hastings algorithm to perform the MCMC simulations. The Metropolis-Hastings algorithm is based on the work of Metropolis [40], who developed the algorithm for a special case, and Hastings [41], who is responsible for the generalization. Though the algorithm in itself is straightforward, the concept is rather difficult to grasp immediately. Therefore, this section will illustrate the algorithm extensively with the goal to facilitate understanding. The example in the next section will discuss the application and implementation issues of the Metropolis-Hastings algorithm.

4.3.1 Background

The transition kernel for the Markov Chain in the Metropolis-Hastings algorithm consists of two main elements, a probing (also called a proposal) distribution and an acceptance probability. Assume the Markov Chain started at initial design θ^0 and has reached θ^i.

From the proposal distribution $PD(\theta^*|\theta^i)$ a candidate design θ^* is generated which will be accepted with acceptance probability

$$\alpha = \min\left\{1, \frac{p(\theta^* \mid y)PD(\theta^i \mid \theta^*)}{p(\theta^i \mid y)PD(\theta^* \mid \theta^i)}\right\} \tag{4-7}$$

where α is the acceptance probability, which can indeed be calculated without needing $p(y)$, as this normalizing constant for the posterior distribution cancels out. Finally, if the proposed design is accepted, the starting point for the next step is updated to the accepted θ^*, otherwise the chain remains in the original position θ^i for another attempt to determine the next design. Each accepted design corresponds to a sample of the parameter vector θ.

4.3.2 Probing Distribution

The probing (or proposal) distribution is a tool used to generate the proposed design θ^*. Though convergence will occur for virtually any type of probing distribution, a rule of thumb is that convergence of the Markov Chain is fastest when the probing distribution is overdispersed compared to and of similar shape as the posterior distribution.

The applications in this thesis draw samples from a multi-variate normal distribution centered on the current design θ^i, defined as

$$f(\theta^* \mid \theta^i, \boldsymbol{\Sigma}) = (2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta^* - \theta^i)^T \boldsymbol{\Sigma}^{-1}(\theta^* - \theta^i)\right) \tag{4-8}$$

where d is the dimension of the parameter vector θ and $\boldsymbol{\Sigma}$ is a symmetric $d \times d$ positive-definite covariance matrix. Because of the convenient selection of a symmetrical probing distribution, as in the algorithm originally developed by Metropolis, evaluating the acceptance probability according to equation (4-7) becomes easier since

$$PD(\theta^i \mid \theta^*) = PD(\theta^* \mid \theta^i) \tag{4-9}$$

so that the ratio of regarding the probing distribution cancels out.

The selection of the covariance matrix Σ of the probing distribution is the next important consideration. A general guideline is that the probing distribution should envelop the posterior distribution. Therefore, the covariance matrix, which is a measure of the width of the probing distribution, should be sufficiently large. The effect of increasing the covariance matrix, and thus increasing the width of the probing distribution, would generate proposed designs further away from the current design, and vice versa. In view of the posterior distribution, designs located further away from the mode have a smaller probability $p(\theta^* | y)$, and thus will be accepted less often.

4.3.3 Acceptance Probability

There should be a particular balance between sampling close by and further away from the mode of the distribution so that a representative collection of samples from the posterior can be obtained.

The sample will always be accepted when the posterior probability density is higher than that of the previous sample, i.e. equivalent to moving 'uphill' of the posterior probability. When the posterior probability density of the generated sample is lower than that of the previous, the generated sample will only be accepted with a probability α. The rationale behind this is that generally moving uphill is favorable in the search for and the sampling from the posterior distribution. However, sometimes the chain should move downhill to improve the 'mixing' of the chain through the sample space. This mixing behavior of the Markov Chain is facilitated by the probing distribution, which is preferably not too wide, nor too contracted. The balance moving uphill and downhill can be assessed by evaluating the average acceptance probability $\bar{\alpha}$, which can be controlled by the width of the covariance matrix.

Draper [42] discusses an optimal average acceptance probability of $\bar{\alpha} = 0.3\text{-}0.6$. Gelman [26] mentions an optimal acceptance probability of 0.44 for a one dimensional problem,

and an acceptance probability between 0.23 and 0.44 for problems with 7 to about 50 parameters. The rather subtle balance between a covariance matrix that is too wide and one that is too contracted leading to a desirable average acceptance rate $\bar{\alpha}$ is in practice usually is a trial-and-error process in fine-tuning the covariance matrix with a constant multiplication factor.

4.3.4 Algorithm

The above discussion regarding Metropolis-Hastings sampling can be summarized in the following algorithm:

1. Initialize counter $i = 1$, specify a suitable initial design θ^0, and set $\theta^i = \theta^0$
2. With design θ^i evaluate $p(\theta^i | y)$
3. Generate the proposed design θ^* from a probing distribution $PD(\theta^* | \theta^i)$ centered on θ^i and evaluate $p(\theta^* | y)$
4. Calculate the acceptance probability:

$$\alpha = \min\left\{1, \frac{p(\theta^* | y)}{p(\theta^i | y)}\right\} = \min\left\{1, \frac{p(\theta^*)p(y | \theta^*)}{p(\theta^i)p(y | \theta^i)}\right\} \tag{4-10}$$

5. Generate u ~ Uniform(0,1) and accept or reject the proposed design:

$$\text{if } \alpha \geq u, \text{ then accept}: \quad \theta^{i+1} = \theta^*$$
$$\text{if } \alpha < u, \text{ then reject}: \quad \theta^{i+1} = \theta^i$$

6. Increase counter $i = i+1$ and repeat step 2 to 5 until convergence

4.3.5 Algorithm Illustration

Figure 4-1 and Figure 4-2 provide an overview of the Metropolis-Hastings algorithm.

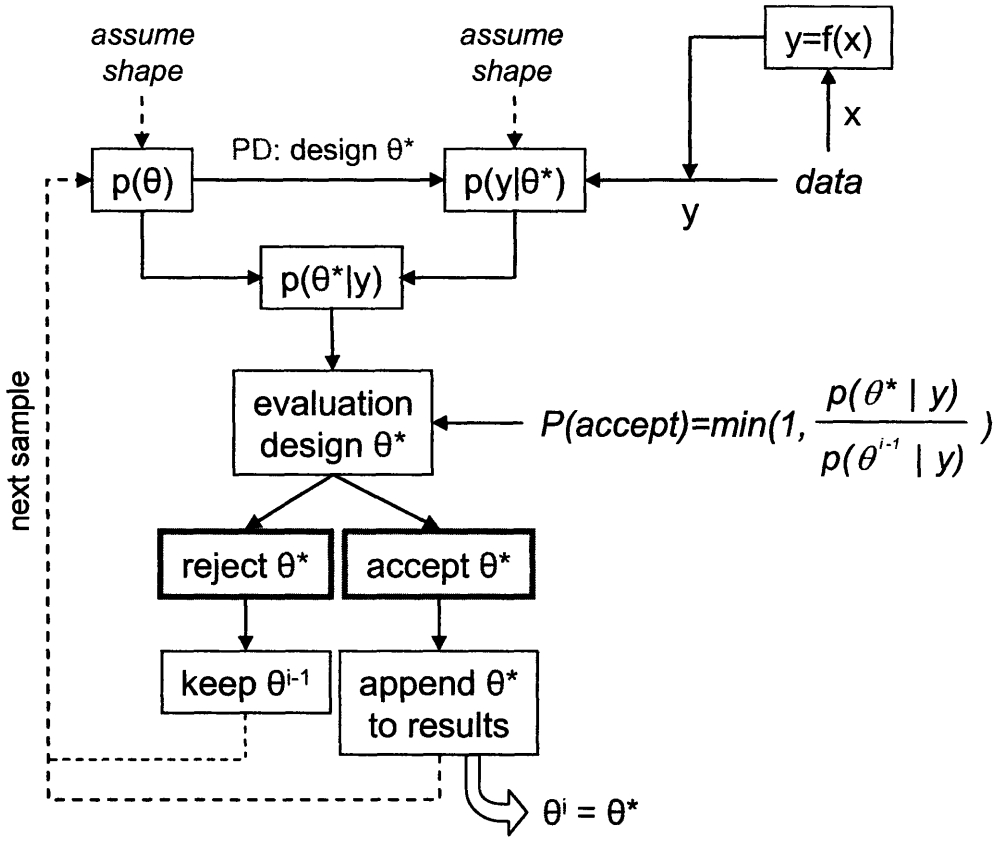


Figure 4-1. Block diagram of the Metropolis-Hastings algorithm

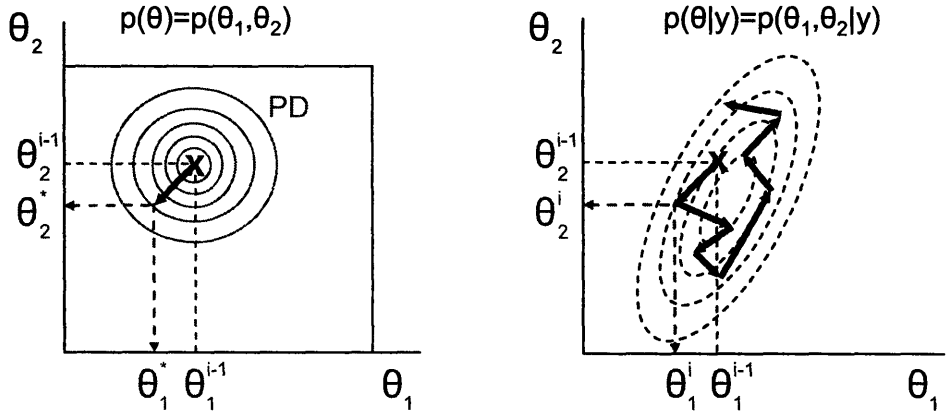


Figure 4-2. Illustration of the Metropolis-Hastings algorithm in 2D

In formulating the Bayesian problem, probability distributions for both the prior and the likelihood function first have to be defined. Using the probing distribution, a sample is

drawn from the sample space of the parameters, for which the boundaries are usually determined by the prior distribution.

Using the data and/or model output, the probability density for the sample θ^i is calculated using the likelihood function. Together with the probability density for θ^i obtained from the prior distribution, the posterior density is calculated according to equation (4-2). The sample θ^i responsible for this particular posterior probability density is then either accepted or rejected. When the sample is accepted, θ^i is appended to the vector of collected samples from the posterior distribution, and the next sample will be generated with the probing distribution centered at θ^i. Upon rejection, the Markov Chain remains at sample θ^{i-1} from where another sample will be generated. Eventually, the Markov Chain will be sampling the posterior distribution as shown in the right plot of Figure 4-2, which shows the dotted contour lines of the unknown posterior distribution.

4.4 Numerical Error

In addition to simulation error of MCMC algorithms, approximation of the posterior probability distribution introduces numerical error because of: (1) simulating a finite number of samples, and (2) assessing the posterior distribution with a histogram or kernel density estimate. This section will describe an elementary evaluation of these potential causes for numerical error.

4.4.1 Random Number Generator

The Metropolis-Hastings algorithm depends on a random number generator for both the proposal and acceptance step. If the generated numbers are not completely random, additional correlation is introduced upon sampling from the posterior distribution leading to a systematic bias in the parameter estimates [43]. Before implementing MCMC simulations, random number generators should be checked for their suitability regarding a.o. autocorrelation of the generated number series [44].

4.4.2 Number of Samples

The decision regarding the finite number of samples to simulate from the posterior distribution depends on the trade-off between accuracy and computational effort. To determine a general guideline for the minimum number of samples, a kernel density estimate of randomly generated samples of the standard normal distribution is compared to the exact probability density function. The accuracy of the approximation is measured by the residual sum of squares *RSS*, defined by

$$RSS = \sum_{i=1}^{b} \left(p_{i,KDE} - p_{i,pdf} \right)^2 \tag{4-11}$$

where $p_{i,KDE}$ is the probability density of the kernel density estimate, and $p_{i,pdf}$ is the probability density as calculated from the standard normal distribution for bin i, and b is the total number of bins. The results are shown in Figure 4-3, where the kernel density estimate is generated from n randomly generated samples, evaluated at $b = 100$ bins, and *RSS* is averaged over 25 trials of n samples to account for the randomness of the sampling.

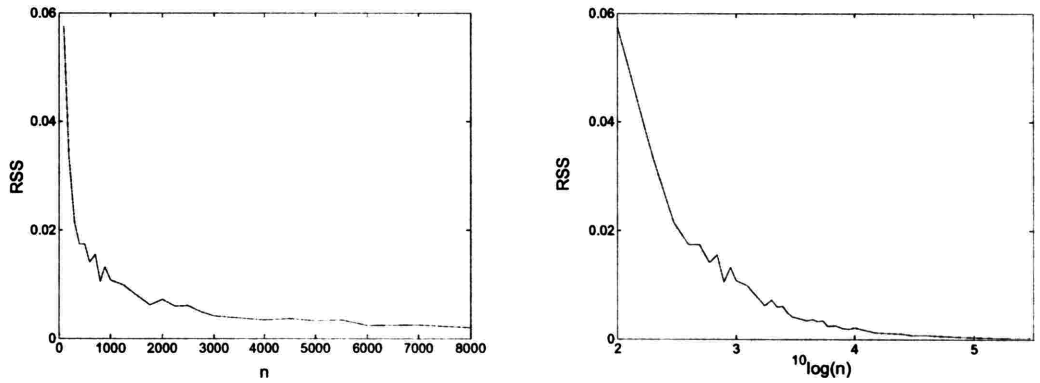


Figure 4-3. Accuracy of approximation as a function of the number of samples

As expected, the results show that the approximation by a kernel density estimate is more accurate when evaluating a larger number of samples, and will become exact for $n \to \infty$. MCMC simulations for the analyses discussed in this thesis were implemented to generate, after the burn-in period and after thinning, at least 5000 samples.

4.4.3 Number of Bins

The number of bins used for the kernel density estimate does not affect the accuracy of approximation as measured by equation (4-11), as *RSS* is only evaluated at the location of the bins, but not in between. However, the number of bins does affect the interpolation between locations where the probability density is estimated and thus the general representation of the posterior distribution. The numerical error introduced by an insufficient number of bins becomes important when the probability density of the estimate needs to be evaluated over relatively small intervals compared to the bin-width, e.g. when an approximated posterior distribution is implemented as the prior distribution in a subsequent estimation.

Figure 4-4 illustrates the approximation of the standard normal distribution by a kernel density estimate for three implementations: 10, 100, and 1000 bins. Each of these three kernel density estimates were obtained from the same set of 5000 samples, randomly generated from the standard normal distribution represented by the dashed line.

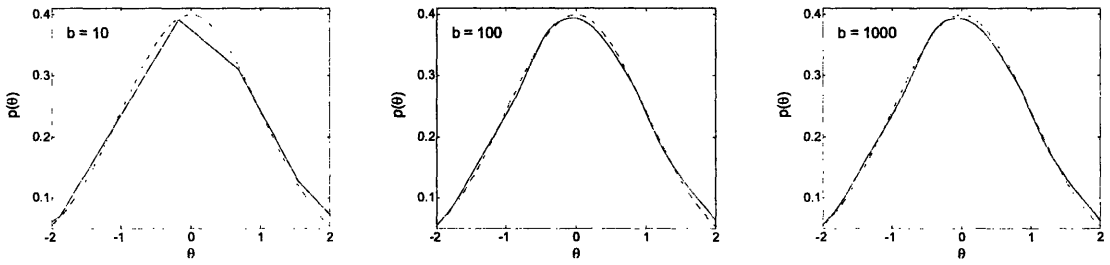


Figure 4-4. Illustration of interpolation as a function of bin-width

The general principle here is again that the larger the number of bins, the more accurate the approximation becomes. However, only for a very small number of bins the approximation is unacceptably course. Increasing the number of bins to 100 significantly improves the approximation, while the approximations resulting from using 100 and 1000 bins are virtually identical. The work in this thesis is implemented by generating kernel density estimates with 1000 bins for the range of the samples. This leads to a relatively accurate evaluation of the probability distribution in subsequent analyses.

5

Parameter Estimation

Let the data speak for themselves!

Ronald A. Fisher

The data *cannot* speak for themselves;
and they never have, in any real problem of inference.

Edwin T. Jaynes [20]

Parameter estimation is one of the most common applications of statistics. The Bayesian approach proves to be powerful in parameter estimation, providing a unified methodology for incorporating uncertainty. This chapter will start with a literature review focusing on the versatility and robustness of the Bayesian approach by citing examples of parameter estimation in the field of kinetics. Then, the general procedure of the Bayesian approach to parameter estimation will briefly be described, specific advantages will be expounded, and three different strategies to parameter estimation will be discussed. A parameter estimation involving a linear model will be formulated and resolved with the purpose of providing a detailed tutorial on the implementation of the Bayesian approach and MCMC.

5.1 Literature on Bayesian Parameter Estimation

Since kinetic rate parameters are estimated using the Bayesian approach in the case studies of Chapter 7 and 8, several articles in the field of kinetics are discussed to exemplify the specific opportunities the Bayesian approach offers with regards to parameter estimation.

5.1.1 Hierarchical Models

Poillot [45] applies the Bayesian approach for estimating the parameters and hyperparameters of a hierarchical microbial growth model using literature data. Inferences are obtained relatively easily with the Bayesian approach, as the primary parameters and hyperparameters are estimated simultaneously. Conventional methods would approach the estimation in two stages, where first the primary parameters, and subsequently the hyperparameters are estimated without a linkage between the stages, so that uncertainty is not propagated. In addition to the advantages regarding estimation, Poillot also shortly comments that the concept variability and uncertainty are a natural concept in the Bayesian approach and that different sources of information, e.g. expert opinion or previous data, can be incorporated with little effort.

5.1.2 Linear Regression with a Change Point

Sivaganesan [46] demonstrates the more accurate Bayesian approach to the parameter estimation of a first-order kinetic model with a lag phase. The general form of a linear model with a delay is defined for n data points as

$$y_i = \mu_i + \varepsilon_i \qquad for\ i = 1, ..., n \tag{5-1}$$

where y_i is the observed value, ε_i is the measurement error distributed according the normal distribution $N(0,\sigma^2)$, and μ_i is defined at time t_i as

$$\begin{aligned} \mu_i &= 0 && if\ t_i \le t_{lag} \\ \mu_i &= \beta(t_i - t_{lag}) && if\ t_i > t_{lag} \end{aligned} \tag{5-2}$$

where β is the slope of the linear relationship in effect after the lag phase t_{lag}. The parameters to estimate from the data are β, σ^2, and t_{lag}. In general, conventional regression methods such as least squares and maximum likelihood can not easily handle change point problems and they give less exact confidence bounds on the estimates for

small sample sizes. The Bayesian approach on the other hand provided the full posterior probability distribution of the model parameters, including the lag phase.

5.1.3 Non-Conventional Parameters

Using the Bayesian approach, Borsuk [47] includes the reaction order of a biochemical oxygen demand (BOD) decay model as one of the parameters to be estimated. The "mixed order" model is defined with a free parameter for the reaction order as

$$\frac{dL}{dt} = kL^n \tag{5-3}$$

where n is the pseudo-order parameter to be estimated, L is the oxygen consumed (BOD) at time t, and k is the mixed-order reaction rate constant. Mixed order models were already known to support data and correctly represent the underlying decay processes much better than the widely applied fixed order BOD decay models, but applications have been inhibited for computational reasons. Due to the high correlation between the parameters nonlinear regression showed extremely slow convergence, preventing an exact solution. The Bayesian approach however, allows for extending data analysis beyond the 'ordinary' so that reaction order n can be evaluated. Borsuk found that the Bayesian approach provided (1) better parameter estimates, as the conventional least squares method is unable to represent non-symmetrical probability distributions for the parameters, (2) better probabilistic predictions containing more information than estimates by conventional methods, and (3) an explicit consideration of the uncertainty useful for decision-making.

5.1.4 Numerical Robustness

Pillonetto [48] analyzes the numerical non-identifiability problems encountered when applying the Maximum Likelihood method to estimate parameters of the so-called "Minimal Model", which is employed to describe insulin sensitivity both in clinical and epidemiological studies. Numerical non-identifiability, implying that a small variation in the model output corresponds to large variations of some model parameters, is explained

by considering the mathematical structure of the model. Using simulation studies, the Bayesian approach subsequently is shown to be robust in parameter estimation for this model. The reasons for the superior performance are (1) the ability to include prior information and (2) the possibility to evaluate full non-symmetrical posterior probability distributions.

5.2 Formulation for Bayesian Parameter Estimation

To facilitate the formulation of a parameter estimation problem, an in-depth analysis (for which a framework will be proposed in the next chapter) should identify the characteristics of the system under investigation. This should ensure that (1) a representative system model is selected, (2) simplifying assumptions are justified and explicitly stated, (3) the output set assignments are identified, (4) the data satisfy the observability criterion, and (5) existing knowledge has been specified. Based on this foundation, though each estimation problem has its own distinctive characteristics, a general formulation regarding the likelihood function and prior distribution will be described in the following sections.

5.2.1 Likelihood Function

As discussed in Section 2.2.3, the measurement error can generally be represented by a normal distribution, which therefore is commonly implemented as likelihood function. In most cases of parameter estimation problems, the variance of the measurement error σ_0^2 will be unknown and is thus included as a parameter to be estimated from the data, so that the likelihood function is defined as

$$p(y \mid \theta, \sigma_0^2) = \prod_{i=1}^{n} p(y_i \mid \theta, \sigma_0^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{ -\frac{\left(y_i - f_i(\theta)\right)^2}{2\sigma_0^2} \right\} \qquad (5\text{-}4)$$

where individual measurements y_i are considered independent of each other, $f_i(\theta)$ is the calculated model outcome for data point i, and n represents the number of measurements.

When evaluating data involving time dependent measurements, such as discussed in the case studies of chapter 7 and 8, the likelihood can similarly be defined as

$$p(y \mid \theta, \sigma_0^2) = \prod_{t=1}^{n} p(y_t \mid \theta, \sigma_0^2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{ -\frac{\left(y_t - f_t(\theta)\right)^2}{2\sigma_0^2} \right\} \qquad (5\text{-}5)$$

where the individual data points y_t are considered independent of each other, $f_t(\theta)$ is the calculated model outcome at time t, and n represents the total number of data points.

The above formulations of the likelihood function only considered homogenous measurement errors ε_0 characterized by variance σ_0^2. Data obtained with different equipment is usually characterized by a different measurement error and can easily be included as additional terms in the multiplication of equation (5-4) or (5-5). This provides for a versatile estimation methodology, especially since the additional terms have no restrictions regarding their functional form.

A final remark concerns the numerical implementation of the likelihood function as formulated in equation (5-4) or (5-5). The probability density $p(y_i|\theta)$ of the likelihood function for measurement i, or $p(y_t|\theta)$ for the data point at time t, attains calculated values that are always positive, but not necessarily smaller than 1. Evaluation of a large number of measurements or data points rapidly approaches the upper or lower accuracy limit of the computer, causing numerical difficulties. To achieve correct computation, the likelihood function as defined in equation (5-4) can be transformed by the natural logarithm as follows

$$\ln\left[p(y \mid \theta, \sigma_0^2) \right] = \sum_{i=1}^{n} \ln\left[p(y_i \mid \theta, \sigma_0^2) \right] = \ln\left[\frac{1}{\sqrt{2\pi}\sigma_0} \right] + \sum_{i=1}^{n} \left[-\frac{\left(y_i - f_i(\theta)\right)^2}{2\sigma_0^2} \right] \qquad (5\text{-}6)$$

where robustness is improved by replacing the multiplication by a summation, likely to tolerate a larger number of measurements n. A similar transformation for Bayes' Theorem leads to

$$\ln\left[p(\theta \mid y_1, y_2, \ldots, y_n)\right] \propto \ln\left[p(\theta)\right] + \sum_n \ln\left[p(y_n \mid \theta)\right] \tag{5-7}$$

after which the posterior probability density is obtained as the exponential of the summation result. On a side-note, the latter formulation of Bayes' Theorem in equation (5-7) is appealing because of the correspondence with intuition that

$$\textit{current knowledge} = \textit{prior knowledge} + \textit{information from data}$$

which represents the underlying point of view of the Bayesian approach.

5.2.2 Prior Distribution

Existing knowledge regarding the parameters can serve to characterize the prior distribution. When sufficient information is available, an informative prior distribution can be defined according to any of the probability density functions available in probability theory [9, 26, 49]. In the majority of cases however, prior knowledge is minimal, which is best represented by a non-informative, or diffuse, probability distribution.

For the parameter θ a proper prior distribution is uniformly distributed, as defined by

$$p(\theta) = \begin{cases} \dfrac{1}{\theta_{\max} - \theta_{\min}} & \text{if } \theta_{\min} \le \theta \le \theta_{\max} \\ 0 & \text{otherwise} \end{cases} \tag{5-8}$$

where the bounds θ_{min} and θ_{max} delimit the sample space and generally represent physical limitations based on prior information. Without any prior knowledge the bounds θ_{min} and

θ_{max} should be specified wide enough not to constrain the exploration of the sample space. According to the uniform distribution the probability density for each value of parameter θ is constant throughout the sample space, not reflecting any information regarding particular parameter values. This convenient attribute can be appreciated when reviewing equation (4-10) for the acceptance probability in the Metropolis-Hastings algorithm, as the prior probability density will cancel out because of its presence in both the nominator and denominator.

Establishing a non-informative prior distribution for the variance σ_0^2 requires more sophistication than implementing a uniform distribution. As the variance σ_0^2 is a transform of the standard deviation σ_0, and both of them are implemented in the likelihood function as defined above in equation (5-4), a uniform distribution for σ_0 is likely not to be a uniform distribution for σ_0^2.

In conjunction with a normal likelihood function, the non-informative prior distribution for the standard deviation σ_0 is uniform in $\ln(\sigma_0)$, the natural logarithm of σ_0. From this it follows for a non-informative prior that [2]

$$p(\sigma_0^2) \propto \sigma_0^{-2} \tag{5-9}$$

Where the distribution is commonly defined as

$$p(\sigma_0^2) = Gamma\left(\sigma_0^{-2} \mid \alpha, \beta\right) \tag{5-10}$$

with

$$Gamma(\rho \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \rho^{\alpha-1} e^{-\beta\rho} \qquad for\ \rho > 0 \tag{5-11}$$

where the non-informative character of this prior distribution is invoked by $\alpha \to 0$ and $\beta \to 0$ [26]. The prior distribution in equation (5-10) is implemented with hyperparameters $\alpha = \beta$, which generally are assigned small values, such as 10^{-3} of 10^{-5}.

5.3 Advantages of Bayesian Parameter Estimation

Though several beneficial properties of the Bayesian approach have already been discussed, this section will list the main advantages specifically regarding parameter estimation problems. Some of these advantages have already been illustrated, while others will be discussed with examples in later sections during the case studies of chapter 7 and 8. Each of these advantages individually is desirable for a parameter estimation methodology, but the superior performance of the Bayesian approach is especially established by facilitating a combination of:

1. **Robust, multi-dimensional parameter estimation**
   The correlations among parameters are taken into account when estimating the parameters, so that multi-dimensional systems can be addressed with relative ease.

2. **Evaluation of complex models**
   The system model can be included in its original form. The transformations and simplifying assumptions leading to manageable models within conventional nonlinear estimation framework are not strictly necessary in the Bayesian approach.

3. **Straightforward incorporation of uncertainty**
   Uncertainty originating from two main sources can be included:
   a) The uncertainty in the data is reflected in the likelihood function $p(y|\theta)$, for which no restrictions apply regarding its functional form. In principle, any probability density function justified by the error model can be implemented.
   b) The uncertainty in the known, but uncertain model parameters and variables is accounted for by incorporating their probability distributions instead of point estimates.

4. **Proper propagation of uncertainty**

   As the two main sources of uncertainty above are incorporated directly into the parameter estimation, and since the model without simplifications is evaluated, the resulting posterior probability distribution correctly reflects the uncertainty in the estimates.

5. **Formal incorporation of prior knowledge**

   Whether reflecting ignorance, or a more informative state of knowledge, the prior distribution allows for a formal incorporation of existing knowledge.

6. **Obtaining the complete posterior distribution of the parameters**

   The representation of the parameter values by the posterior probability distribution, which has no restrictions regarding its shape, allows for direct assessment of the uncertainty and simple calculation of summarizing variables, credible intervals, and hypothesis tests.

7. **Appropriate combination of data obtained from different sources**

   The modular construction of the likelihood function can simultaneously consider different data sets with potentially different measurement precision.

8. **Possibility to address errors on two axis of a regression**

   Similarly as including additional terms for multiple data sets, the likelihood function can include additional terms to take the uncertainty on additional axes into account.

9. **Data discrimination on a coherent basis**

   Comparing the posterior probability distributions for the parameters, outliers can easily be spotted and the data can be scrutinized for erratic experimental conditions or observations.

10. **Possibility to identify systematic error between experiments**

    Since posterior probability distributions of a number of experiments are easily combined, groups of experiments can be compared with each other to investigate whether any systematic error has been introduced or changed during experimentation.

11. **Carry-over results to subsequent studies**

    Multiple researches have in the past gathered experimental data and determined estimates of which the informational value can be included through the prior distribution. Similarly, current results can be implemented as prior knowledge in future estimation problems.

5.4 Parameter Estimation Strategies

According to Desiderata IIIa (see Section 3.2.2) the specific order of evaluating separate data sets has no effect on the final posterior distribution. What matters is that the posterior distribution represents the state of knowledge based on existing relevant information (Desiderata IIIb) and logically, equivalent states of knowledge should be assigned equivalent probabilities (Desiderata IIIc). This convenient fact allows for either a sequential incorporation of the data, a parallel approach, or simultaneous evaluation. Each of these three strategies will be discussed and illustrated.

5.4.1 Sequential Estimation

As discussed in Section 3.3.2, Bayes' theorem has the structure of a learning algorithm. The posterior distribution obtained presently, can be implemented as prior distribution when additional data becomes available. The natural extension of this algorithm to dynamic systems was briefly discussed and the relation to the Kalman filter was illustrated in Section 3.4.

The application to static systems makes use of the premise that, though Bayes' Theorem updates the state of knowledge from *pre*-data to *post*-data, a strict temporal relationship

between these states is not required. Therefore, the posterior distribution resulting from updating the prior with the likelihood function for data set A, becomes the prior distribution to be used with the likelihood function for data set B, etc. An overview of the two different sequences $A \rightarrow B \rightarrow C$ and $A \rightarrow C \rightarrow B$ resulting in identical posterior distributions $p(\theta|A,B,C)$ and $p(\theta|A,C,B)$ is shown in Figure 5-1.

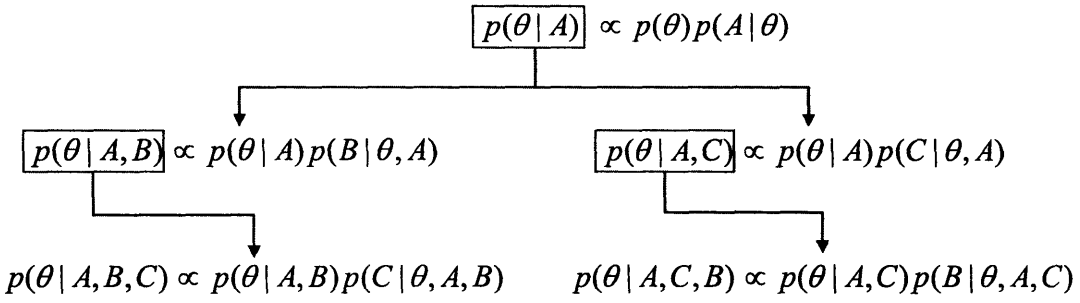


Figure 5-1. Overview of sequential estimation

Though sequential estimation is appealing intuitively, the subtle requirement that data A, B, and C are generated from the exact same population needs to be satisfied. This requirement implies that the experimental conditions must be indistinguishable between experiments, except for variables that can be accounted for in the system model. If this requirement is not satisfied, then the inferences on parameter θ from data set A, B, and C are not equivalent, a situation that will be discussed next.

5.4.2 Parallel Estimation

Suppose that generating experimental results can be imagined as drawing samples of data from a population defined by the experimental conditions specified by I (see Section 3.3.1). Identical conditions result in equivalent data sets A, B, and C as shown in Figure 5-2. However, conditions can (unintentionally) change slightly having a minor impact on the experimental results. Without accounting for these changes in the system model, each of the data sets cannot be considered to originate from the same population. This situation can be imagined as sampling from different populations defined by I_A, I_B, and I_C,

so that sequential estimation is not applicable anymore and the parallel estimation approach is recommended.

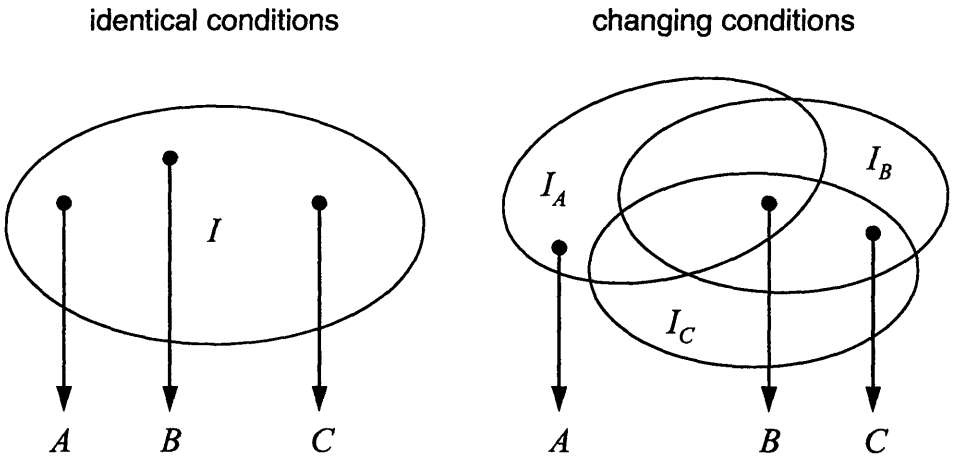


Figure 5-2. Experimentation as drawing data from a population defined by I

Parallel estimation combines multiple individual inferences on parameter θ so that the effects of the (unintentional) changing experimental conditions cancel out, assuming the changes are completely random. This combination can be interpreted as a type of weighted average over the many slightly different experimental conditions. According to the procedure of parallel estimation, as schematically shown in Figure 5-3, a separate posterior distribution is obtained for parameter θ_i by evaluating each data set i. The subsequent combining of the distinct posterior distributions for parameter θ_i results in the 'overall' posterior distribution for parameter θ.

$$
\begin{aligned}
&p(\theta_A \mid A) \propto p(\theta_A)p(A \mid \theta_A) \\
&p(\theta_B \mid B) \propto p(\theta_B)p(B \mid \theta_B) \\
&p(\theta_C \mid C) \propto p(\theta_C)p(C \mid \theta_C) \\
&\qquad\vdots \\
&\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}
\end{aligned}
$$

$$
p(\theta \mid \theta_A, \theta_B, \theta_C, \ldots, A, B, C, \ldots) \propto p(\theta)p(\theta_A \mid \theta, A)p(\theta_B \mid \theta, B)p(\theta_C \mid \theta, C) \cdots
$$

Figure 5-3. Overview of parallel estimation

The method of combining the individual posterior distributions for θ_i can be derived by a straightforward application of the laws of probability theory. For clarity, the following derivation considers the case of only two data sets, A and B, with which the posterior distributions for parameters θ_A and θ_B have been obtained. The desired posterior distribution for parameter θ is formulated as

$$p(\theta \mid \theta_A, \theta_B, A, B) = \frac{p(\theta)p(\theta_A, \theta_B, A, B \mid \theta)}{p(\theta_A, \theta_B, A, B)} = \frac{p(\theta)p(\theta_A, A \mid \theta)p(\theta_B, B \mid \theta)}{p(\theta_A, A)p(\theta_B, B)} \tag{5-12}$$

where the latter equation sign is justified by the independence of the data sets A and B, and thus also independence of the parameters θ_A and θ_B. Applying the product rule multiple twice in both the numerator and denominator gives

$$p(\theta \mid \theta_A, \theta_B, A, B) = p(\theta)\frac{p(\theta_A \mid \theta, A)p(\theta_B \mid \theta, B)}{p(\theta_A)p(\theta_B)}\frac{p(A \mid \theta)p(B \mid \theta)}{p(A \mid \theta_A)p(B \mid \theta_B)} \tag{5-13}$$

of which the last term consists of likelihood functions, that can be rewritten as

$$\frac{p(A \mid \theta)p(B \mid \theta)}{p(A \mid \theta_A)p(B \mid \theta_B)} = \frac{l(\theta \mid A)l(\theta \mid B)}{l(\theta_A \mid A)l(\theta_B \mid B)} = 1 \tag{5-14}$$

where the simplification is possible because of Desiderata IIIc (see Section 3.2.2), which requires that propositions with the same truth value are assigned the same probability. Logically, the likelihood function for parameter θ based on data set A is equal to the likelihood function for parameter θ_A based on the same data set, since the information evaluated in both cases is identical. Note that this simplification can only be made when the system is considered equivalent in both cases.

Subsequently, under the condition that

$$p(\theta_A) = p(\theta_B) = \textit{non-informative} \tag{5-15}$$

and as long as the likelihood function is completely defined within the feasible parameter space as delimited by the non-informative prior distribution, the posterior distribution can be determined as

$$p(\theta \mid \theta_A, \theta_B, A, B) \propto p(\theta) p(\theta_A \mid \theta, A) p(\theta_B \mid \theta, B) \tag{5-16}$$

since non-informative distributions only represent a proportionality constant and do not alter the information content of the posterior distribution. Equation (5-16) can be evaluated with a MCMC simulation by sampling θ and using the posterior distributions $p(\theta_A|A)$ and $p(\theta_B|B)$ of the individual parameters as likelihood functions.

5.4.3 Simultaneous Estimation

The final estimation strategy is to evaluate all the available data simultaneously, so that

$$p(\theta \mid A, B, C, ...) \propto p(\theta) p(A, B, C, ... \mid \theta) \tag{5-17}$$

which directly approximates the 'overall' estimate for parameter θ. This approach can generally be applied in place of the parallel estimation strategy. However, computationally this strategy is unattractive when a large number of data sets are analyzed. In addition, when variables such as the initial conditions of the system are estimated, the dimension of the parameter vector θ will significantly increase with an increasing number of data sets.

5.5 Parameter Estimation Example

Bayesian parameter estimation solved by MCMC simulation will be illustrated with an estimation problem involving a linear model. The results of the parameter estimation according to the Bayesian approach will be compared with the outcome of a linear

regression. First, the problem formulation according to the Bayesian approach will be described, followed by a discussion of implementation issues.

5.5.1 Problem Description

The problem involves the phenomenon that temperature decreases with increasing altitude in the troposphere, commonly called the 'Environmental Lapse Rate', as described by the model

$$T = a - bz \tag{5-18}$$

where a and b are the model parameters to be estimated from observations of the temperature T at various altitudes z. Assume the true values of the parameters are

$$a = 298K \qquad and \qquad b = 9.8K/km$$

with which data is synthetically generated by adding an error to T as calculated by equation (5-18) for z = 0.5, 1.0, 1.5,..., 5.0 km. The error is generated from a normal distribution $N(0, \sigma_T^2)$ with $\sigma_T = 3K$, but for the estimation problem σ_T^2 is considered an unknown parameter to be estimated. The calculated/true values and the synthetically generated data are shown in Figure 5-4.

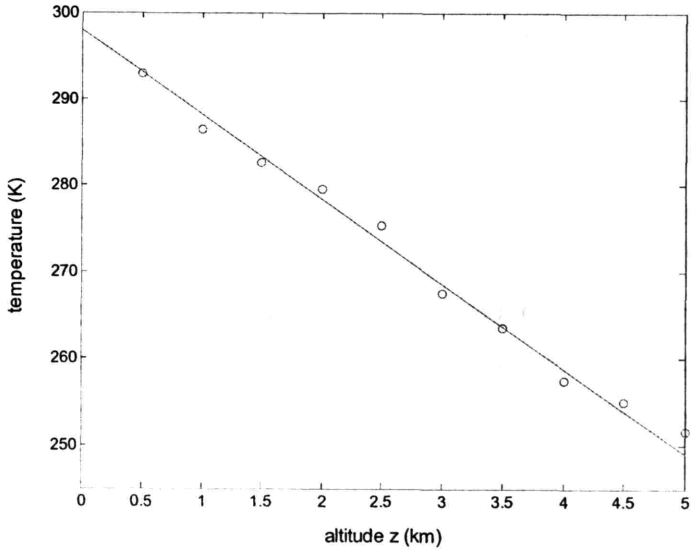


Figure 5-4. Calculated/true relationship and synthetically generated measurements (o)

5.5.2 The Bayesian Approach

For this problem, the parameter vector is defined as $\theta = \{a, b, \sigma_T^2\}$ and we are interested in finding the posterior $p(a, b, \sigma_T^2 | T)$, which can be calculated with Bayes' Theorem

$$p(a,b,\sigma_T^2 | T) \propto p(T | a,b,\sigma_T^2)p(a,b,\sigma_T^2) = p(T | a,b,\sigma_T^2)p(a)p(b)p(\sigma_T^2) \qquad (5\text{-}19)$$

where each of the prior distributions are considered independent of each other.

The likelihood function is defined according to equation (5-4) as

$$p(T | a,b,\sigma_T^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left\{ -\frac{(T_i - (a - b z_i))^2}{2\sigma_T^2} \right\} \qquad (5\text{-}20)$$

where n is the total number of measurement, the individual measurements T_i are considered independent of each other, and the measurement error is assumed to be homogenous for all observations.

As current information is not available, non-informative prior distributions are appropriate for all three parameters. For parameters a and b the uniform distributions are defined according to equation (5-8) with $[a_{min}, a_{max}] = [0, 500]$ and $[b_{min}, b_{max}] = [0, 20]$, respectively. The prior distribution for the variance σ_T^2 is defined according to equation (5-10) with $\alpha = \beta = 10^{-3}$.

5.5.3 MCMC Implementation

Building on the problem formulation above, the remaining specifications and implementation issues regarding MCMC are discussed here. For an overview, a simplified flow diagram of this estimation problem is shown in Figure 5-5. The actual code for the MCMC simulation for this estimation problem is included in Appendix A.

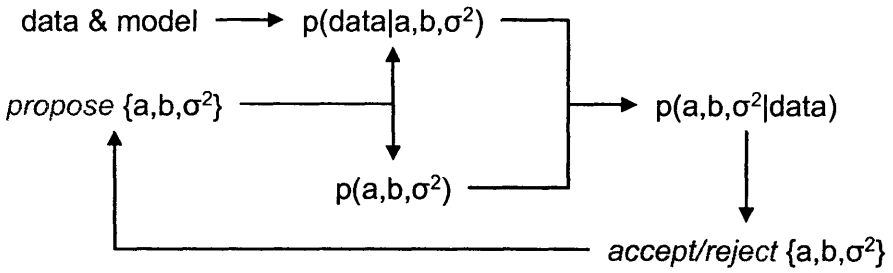


Figure 5-5. Simplified flow diagram of the Metropolis-Hastings algorithm

5.5.4 Initial Design

In principle, the selection of the initial design does not matter in terms of convergence. As the Markov Chain in the Metropolis-Hastings algorithm is aperiodic, there should always be convergence to the posterior distribution, regardless of the initial design. Nevertheless, when the initial design is close to the posterior mode, convergence is reached earlier. Thus, any available information on the estimates to select a suitable initial design would make the algorithm more efficient. In this case, the initial design is set to

$$\theta^0 = \{0.8a,\ 1.2b,\ 0.8\sigma_T^2\} \tag{5-21}$$

which was selected arbitrarily. This estimation problem is relatively simple and designs further removed from the true parameter values would still converge without difficulties.

5.5.5 Covariance Matrix of the Probing Distribution

In addition to affecting the acceptance rate, the covariance matrix also contains information on the correlations between the parameters. The method of iteratively determining a suitable covariance matrix for a system with n parameters is shown below.

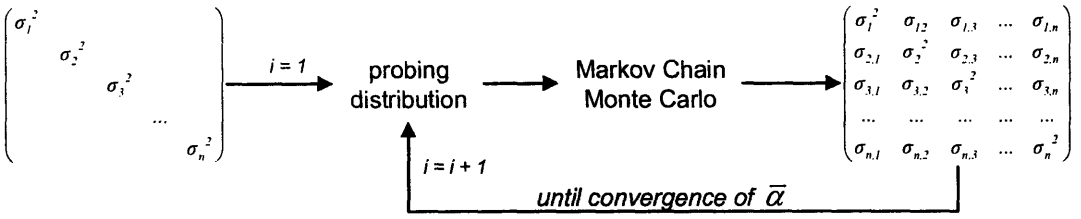


Figure 5-6. Determining the covariance matrix of the probing distribution

The initial diagonal covariance matrix is constructed from the variances of the parameters. These variances can be based on prior knowledge, or approximated from a ballpark figure for the standard deviation by multiplying the initial values of the parameters by a constant factor. The initial covariance matrix for the linear parameter estimation problem is constructed from the initial design as

$$k^2 \begin{pmatrix} (0.8a)^2 & & \\ & (1.2b)^2 & \\ & & (0.8\sigma_T{}^2)^2 \end{pmatrix} \tag{5-22}$$

where $k = 0.05$, but can be adjusted to obtain a desired average acceptance probability $\bar{\alpha}$. A diagonal matrix implies independency, while the parameters are likely correlated. Yet, due to the lack of information, the diagonal covariance matrix is the best guess, possibly leading to a relatively small average acceptance probability $\bar{\alpha}$ for iteration $i = 1$. In this case, the first MCMC simulation of 11,000 runs had an acceptance rate of $\bar{\alpha} = 0.0496$. The trace plots in Figure 5-7 show that both parameter a and b quickly converge, while parameter $\sigma_T{}^2$ still seems to be exploring the sample space.

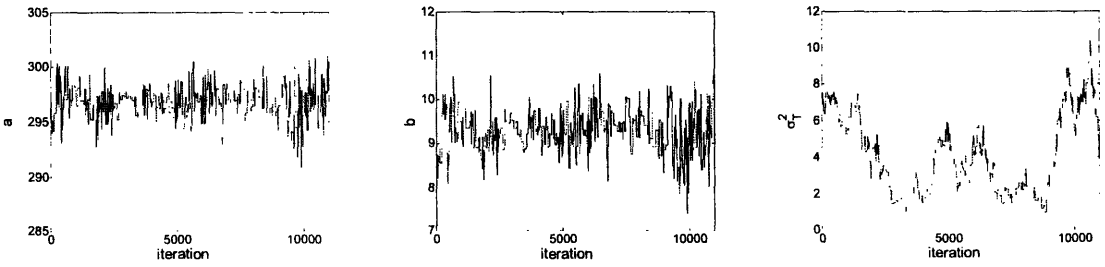


Figure 5-7. Trace and autocorrelation plots for the first MCMC simulation

As expected from the low average acceptance rate, the Markov Chain is not mixing well through the sample space. Though the mixing could be improved by reducing k, the more effective approach is to include the correlations between the parameters in the covariance matrix of the probing distribution. The desired information regarding the correlation between the parameters can be found in the accepted designs of the initial MCMC simulation. The covariance matrix of the probing distribution for the second MCMC simulation is generated from the samples obtained in the first simulation. If necessary, this second covariance matrix can be tuned with the multiplication factor k to achieve the desired average acceptance rate of $\bar{\alpha} \approx 0.2\text{-}0.5$. If necessary, additional iterative steps can improve the covariance matrix of the probing distribution. Ideally, after several steps the acceptance rate will show relatively little variation for subsequent iterations. This is what is meant by 'convergence of $\bar{\alpha}$' in Figure 5-6.

Finally, in order to keep this iterative procedure manageable regarding computation time, the number of MCMC steps can be shortened to a fraction of what would normally be required for the actual parameter estimation.

5.5.6 Burn-in Period and Convergence

To visualize the moving Markov Chain through the sample space, Figure 5-8 shows the accepted designs for the parameters a and b during the second MCMC simulation. The left plot illustrates the path traveled by the Markov Chain, starting at the initial design, exploring the sample space in search for the most likely parameter values, and converging to sampling from the posterior distribution. The exploratory phase is referred to as the 'burn-in' period.

Obviously, the samples accepted during the burn-in period are not representative for the posterior distribution and are thus discarded before analyzing the samples. To illustrate the actual shape of the posterior distribution, the right plot Figure 5-8 displays the samples after discarding the burn-in period. The scatter density of the samples is indicative of the probability and the shape of the posterior distribution reveals the

correlation between parameters a and b. As discussed in Section 4.2.3, the correlation between the parameters can be quantified by determining the (in this case 2x2) covariance matrix from the collected samples of a and b.

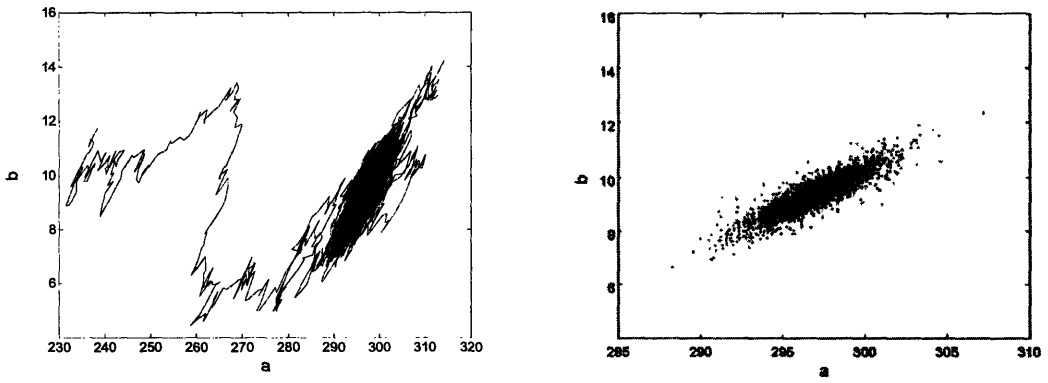


Figure 5-8. Markov Chain traveling through the 2-D sample space of a and b

Convergence should be verified for each MCMC simulation. There are several tools available in order to judge whether convergence has occurred. The procedure here closely follows the steps as described in Draper [42].

First of all, a visual inspection of a trace plot of the sample chain is important. When the Markov Chain has converged, the samples appear as stationary noise. If the consecutive samples follow a trajectory through their domain, the chain has not yet converged. Examples are shown in Figure 5-9.

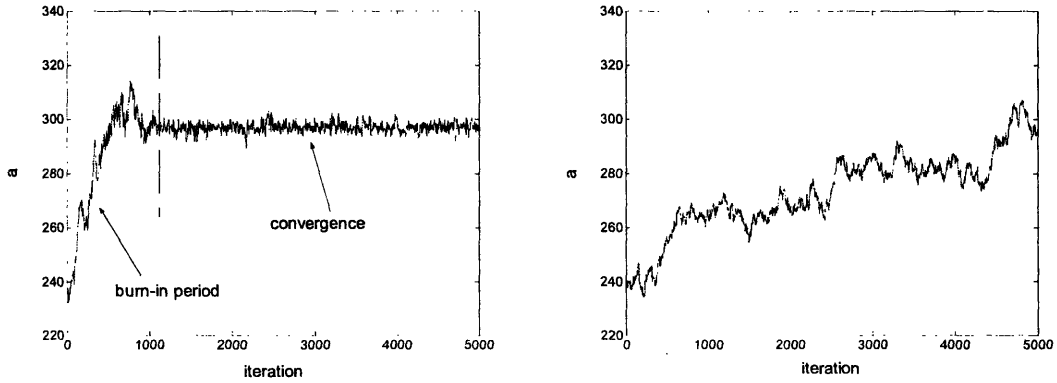


Figure 5-9. Example of a trace plot for a converged and not converged Markov Chain

A second indication of convergence can be found in the autocorrelations of the chains. Upon convergence, the Markov Chain should approximately exhibit a first order serial autocorrelation (AR1) and for each parameter, as shown in the left plot of Figure 5-10 for parameter *a*. Significantly stronger correlations occur, as shown in the right plot of Figure 5-10, when the Markov Chain moves along a trajectory through the parameter space, and the trace plot of the sample does not look like random noise, as mentioned above.

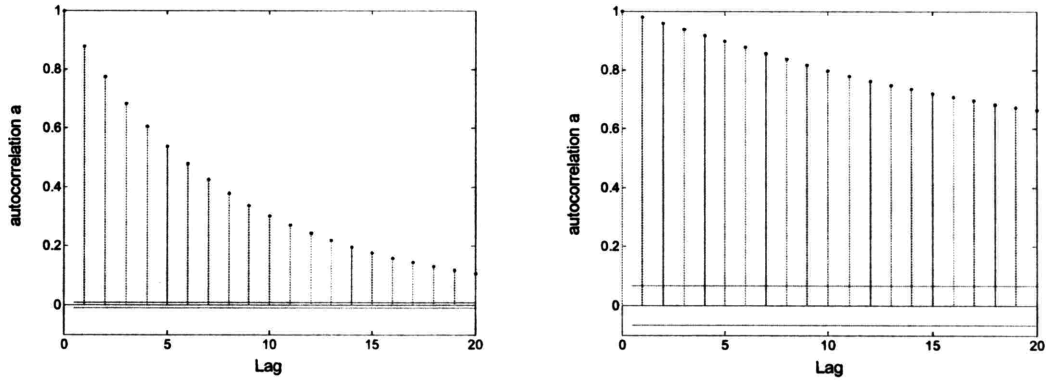


Figure 5-10. Examples of the autocorrelation for a converged and not converged Markov Chain

Finally, there are several convergence diagnostic tools for determining convergence. The common methods are based on a particular statistical analysis of the sampling chain. Cowles and Carlin [50] discuss multiple diagnostic tools and conclude that they should be used with caution. As none of them can judge convergence with absolute certainty, they recommend a mix of several tools in addition to a visual comparisons of several separate chains and a check of the auto and cross correlations. Most convenient are the diagnostic tools bundled in the package known as 'CODA' [51, 52], which is available online for S+ and R , and selected components also for Matlab [53]. A discussion on the most useful features of CODA appears in [37].

5.5.7 Thinning

Since the samples are generated by a Markov Chain, there is a serial correlation among the samples. In fact, as mentioned above, a Markov Chain exhibiting first order autocorrelation is desirable, as this is an indication of convergence. However, the autocorrelation invalidates the statement that the accepted designs represents a sample set randomly drawn from the posterior distribution. For stationary time series though, a distribution can still be approximated by simulating correlated draws from that distribution, as long as enough samples are obtained. The degree of correlation is thus an important factor in determining the simulation time [37].

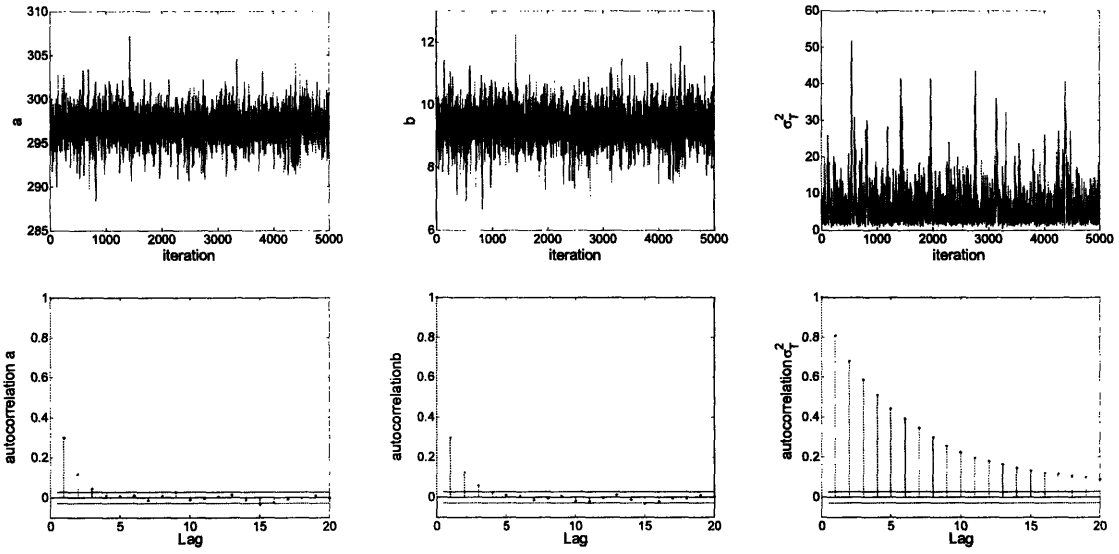


Figure 5-11. Trace and autocorrelation plots after burn-in removal and thinning

Nonetheless, the randomness of the sample set can be improved by thinning, which means that only one out of several samples is actually regarded as a sample from the posterior distribution. The trace and autocorrelation plots resulting from the second MCMC are shown in Figure 5-11. The total number of samples obtained for this simulation was 51,000, of which 1,000 are discarded as burn-in period. The remaining samples were thinned in a ratio of 1:10, so that eventually 5,000 samples were available for approximating the posterior distribution. The effect of thinning is clearly visible when comparing the autocorrelation plots for parameters *a* and *b* of Figure 5-11 with the

left plot of Figure 5-10. Because of the fact that all samples for the variance σ_T^2 are positive, the autocorrelation is still significant.

The level of autocorrelation is sufficiently low to consider the samples as randomly drawn from the posterior distribution. This favorable autocorrelation is caused in part by the implementation of a proper covariance matrix for the probing distribution, i.e. incorporating the correlations between the parameters as discussed above. The second reason for the low autocorrelation is the thinning of the samples.

5.5.8 Estimation Results

The marginal posterior probability distributions for the parameters are determined as kernel density estimates from the samples shown above in Figure 5-11. The vertical lines represent the population mean of the parameter distributions used for the data generation. The expected values for a and b are estimated reasonably accurate. Both the mode and the mean (see Table 5-1) for the estimated variance parameter σ_T^2 were lower than its population mean, probably because of the relatively little noise in the synthetic data.

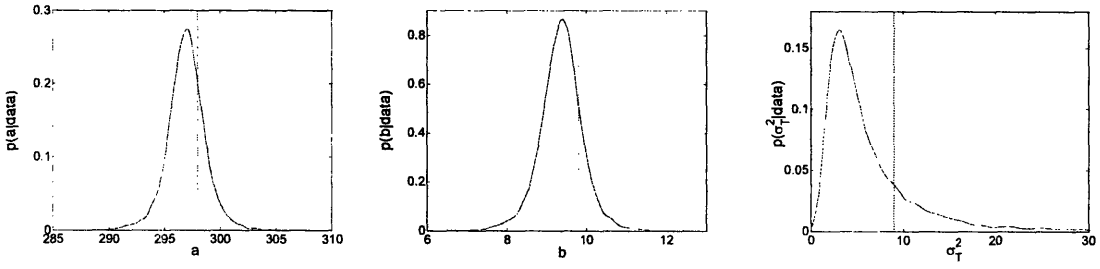


Figure 5-12. Marginal posterior distributions for the parameters

5.5.9 Comparison to Linear Regression

For comparison, a linear regression was performed on the same data. The resulting summary variables for both methods are compared in Table 5-1.

Table 5-1. Comparison of Bayesian and linear regression results

Bayesian Approach		Linear Regression	
$E[a]$	297.0	$\hat{a}$	297.0
$E[b]$	9.37	$\hat{b}$	9.38
$E[\sigma_T^2]$	6.4	s^2	1.96
$mode(\sigma_T^2)$	3.06		
σ_a^2	2.87	$\hat{\sigma}_{\hat{a}}^2$	0.91
σ_b^2	0.29	$\hat{\sigma}_{\hat{b}}^2$	0.095

The expected values for the parameters a and b obtained from both methods coincide, so that either method would perform satisfactorily for estimating the parameters of a linear system. The difference between the methods becomes obvious when considering the uncertainty estimates. As the measurement variance was considered unknown, the Bayesian approach incorporated σ_T^2 as a parameter in the estimation. This lack of information regarding the measurement variance σ_T^2 was therefore reflected in the estimates of parameters a and b, resulting in probability distributions for each of the simultaneously estimated parameters (see Figure 5-12). Linear regression on the other hand resulted in uncertainty point estimates that were calculated (according to the discussion in Section 2.3.1) after $\hat{a}$ and $\hat{b}$ were obtained. The uncertainty regarding the measurement variance σ_T^2 is thus not reflected in the uncertainty regarding the parameter point estimates $\hat{a}$ and $\hat{b}$. This comparison demonstrates that the Bayesian approach is better suited to deal with uncertainty than conventional estimation methods.

6

Framework for Experimentation and Estimation

Nothing is particularly hard if you divide it into small jobs.

Henry Ford

Each individual experiment contains information such as the state of knowledge regarding the quantities to estimate, regarding the existence of systematic error compared to previous experiments, and regarding the best experiment to perform next. The framework that will be introduced can aid experimentalists exploit the relationship between experimentation and estimation to the fullest extent. The framework consists of four components: system description, system analysis, experimentation, and estimation. The discussion on system description mainly focuses on collecting information and will lay the foundation for system analysis, for which mathematical tools are proposed for a more in-depth study of the system. The sections on experimentation and estimation will discuss methods for extracting information from the data and for deciding on future experiments.

6.1 Introduction

The framework for experimentation and estimation is shown in Figure 6-1 and can be interpreted as a guideline or checklist for scientists and engineers involved in data generation and analysis. Each of the elements will be discussed in more detail in

following sections. Some components of the proposed framework might not be familiar to certain researchers in disciplines that are less system-oriented.

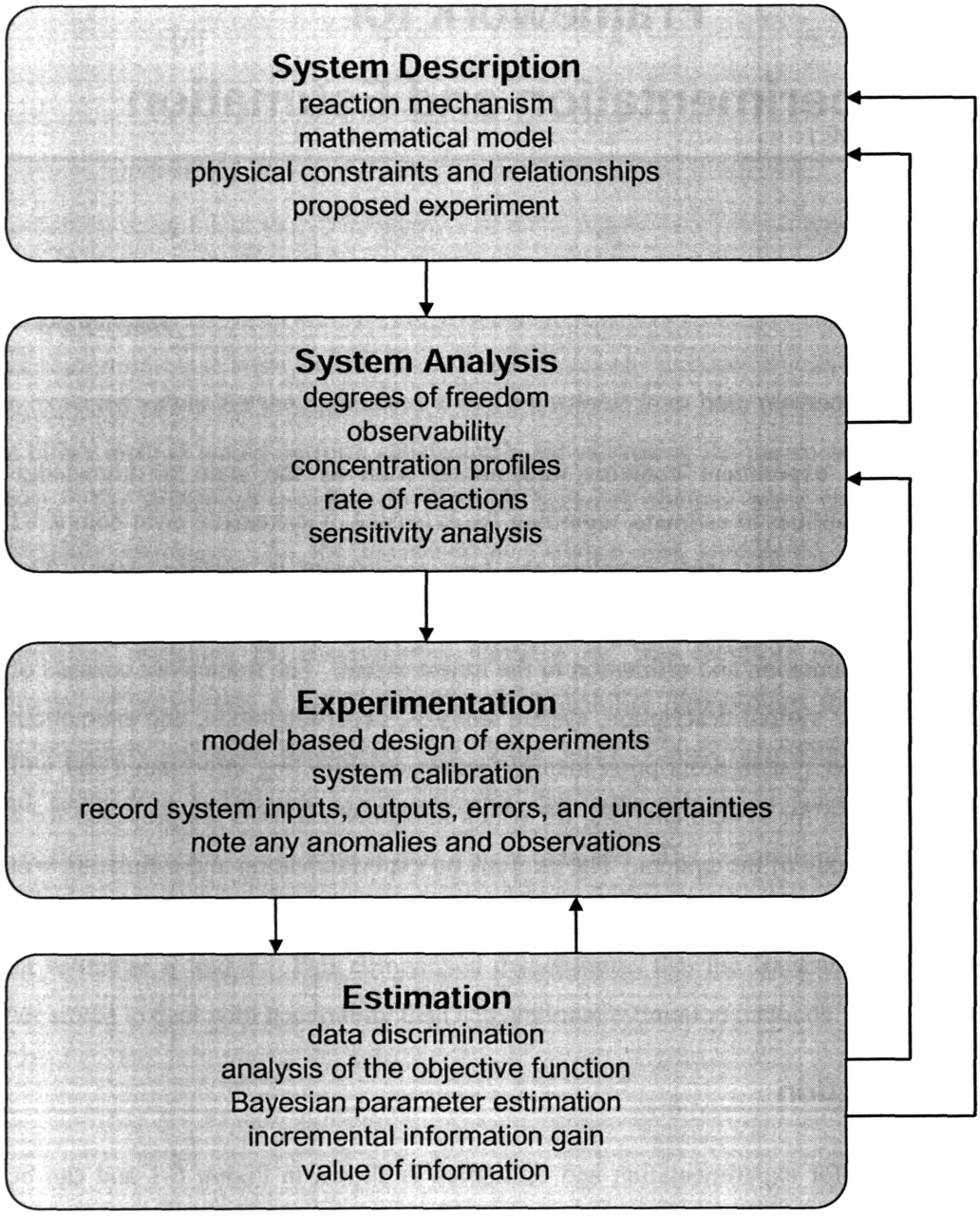


Figure 6-1. Proposed framework for experimentation and estimation

The proposed framework includes several tools from engineering that provide the opportunity to systematically investigate the feasibility of ideas regarding e.g. alternative detection methods, increasing the number of components in the experiment, including additional control variables, etc. Obviously, each experiment is unique and a strict recipe prescribing the optimal procedure does not exist. However, by following this guideline the experimentalist will perform essential analyses and thoroughly think through the system under investigation. As experimentation is a learning process, the more information gathered, the better an experimental design can be crafted. Therefore, several iterations are expected before reaching the goal of the experiment satisfactorily.

6.2 System Description

The first stage in the framework for experimentation and estimation requires a thorough description of the system. The following sections briefly discuss important elements that define a chemical reaction system.

6.2.1 Reaction Mechanism and Mathematical Model

Only elementary reactions should be considered for the reaction mechanism, as the fundamental processes in the system should be understood. Possible simplifications can be applied at a later stage when justified by information such as discussed in the next section.

The mathematical model describing the system follows directly from the mechanism in the form of kinetic rate equations, which as a set of ordinary differential equations (ODEs) require simultaneous numerical evaluation to obtain a solution. The set of kinetic rate equations describing the systems in the case studies in Chapter 7 and 8 were solved by using the Matlab ODE-solver routines.

6.2.2 Physical Constraints and Relationships

Any information regarding the physics of the system, for example feasible bounds regarding system variables and parameters or relationships between parameters, can

contribute to a better understanding. As the foundation for assumptions and system simplifications, this information should be explicitly included in the description of the system.

6.2.3 Proposed Experiment

Deciding on the actual experiment is the territory most familiar to the experimentalist. Obviously, fundamental elements of the experimental setup are the chemicals used, the equipment to carry out the experiment and the detector to measure experimental conditions and concentrations for certain species.

Nevertheless, the experiment should be approached as a chemical reaction system. For example, the time scales for diffusion, convection and chemical reaction should be considered to determine whether diffusion or convective flow contributes to the decay of reaction species so that additional decay terms are required in the kinetic rate equations. Subsequently, these additional terms require information such as flow rates, so that experimentation entails more than merely detecting the species of interest.

6.3 System Analysis

After collecting the relevant information regarding the system as described above, the tools proposed for system analysis allow for a specification of the quantities to estimate, a feasibility check of the proposed experiment, and scenario testing through simulation of the system, with the goal to develop a more thorough understanding of the system.

6.3.1 Degrees of Freedom

The difference between the number of system equations and the number of variables is known as the degrees of freedom of the system. The degrees of freedom indicate the number of variables and/or parameters that need to be specified before the system is fully determined. The appropriate selection of variables/parameters making up the degrees of freedom can be found by obtaining an output set assignment from the incidence matrix.

The incidence matrix, with the rows corresponding to the system equations, and the columns to the variables and parameters, is a representation of the system structure. Each non-zero entry in matrix element (i,j) indicates the occurrence of variable j in equation i. For example, a simple system without degrees of freedom

$$\begin{aligned} f_1(x_1, x_3) &= 0 \\ f_2(x_1, x_2, x_3) &= 0 \\ f_3(x_2) &= 0 \end{aligned} \tag{6-1}$$

can be represented by the incidence matrix

$$\begin{array}{c} \\ f_1 \\ f_2 \\ f_3 \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} \times & & \times \\ \times & \times & \times \\ & \times & \end{bmatrix} \end{array} \tag{6-2}$$

which can be used for the output set assignment by assigning each of the variables x_1, x_2, and x_3 to one of the equations f_1, f_2, and f_3. The existence of an output set assignment is a necessary condition for the existence of a solution, and thus determines the feasibility of the system. The rules for the assignment are [54]:

1) Each equation has exactly one output variable
2) Each variable appears as the output variable of exactly one equation

For the example above, an admissible output set can be determined as follows. After assigning variable x_1 to be calculated with equation f_1, both the column of x_1 and the row of f_1 can be deleted according to the rules above. Then the only possibility to complete the output set is to assign variable x_3 to f_2, and x_2 to f_3, as shown below.

$$\begin{array}{c} \\ f_1 \\ f_2 \\ f_3 \end{array}\begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} \otimes & & \times \\ \times & \times & \times \\ & \times & \end{bmatrix} \end{array} \Rightarrow \begin{array}{c} \\ f_2 \\ f_3 \end{array}\begin{array}{c} \begin{array}{cc} x_2 & x_3 \end{array} \\ \begin{bmatrix} \times & \otimes \\ \times & \end{bmatrix} \end{array} \Rightarrow \begin{array}{c} \\ f_2 \\ f_3 \end{array}\begin{array}{c} \begin{array}{cc} x_2 & x_3 \end{array} \\ \begin{bmatrix} \times & \otimes \\ \otimes & \end{bmatrix} \end{array} \tag{6-3}$$

The routine in (6-3) is easily applied manually for small problems, but for more complex systems Duff's algorithm [55, 56] is a convenient method for automating the output set assignment.

The significance of the output set assignment for estimation problems, generally dealing with underdetermined systems, is that the combinations of variables making up the degrees of freedom can easily be identified. As will be shown in the case studies in Chapter 7 and 8 this is important to decide on the approach to experimentation and estimation for obtaining the desired parameters.

6.3.2 Observability

Observability is a concept originating from Control Theory. A system is defined to be completely observable when the initial conditions of the system at t_0 can be determined from the observed output of that system at a future time $t > t_0$. The complementary but more intuitive concept of reconstructability involves the construction of the present state from past observations [57]. For linear time-invariant systems complete observability implies and is implied by complete reconstructability.

The significance of observability in the framework for experimentation and estimation is to confirm that the intended measurements are indeed suitable to completely describe the system states. An experiment that does not result complete observability will obviously be lacking the information to successfully estimate the model parameters.

The observability of a linear time-invariant system can be determined mathematically as follows. Consider the n-dimensional linear time-invariant system

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) \end{aligned} \tag{6-4}$$

where $\mathbf{x}(t)$ is the vector of the system states, $\mathbf{y}(t)$ is the output vector, $\mathbf{A}$ is the reaction matrix, and $\mathbf{C}$ is the output matrix of the linear differential system, whose elements are independent of time.

The linear time-invariant system in equation (6-4) is completely observable if the observability matrix $\mathbf{Q}$ has full rank n, where $\mathbf{Q}$ is defined as

$$\mathbf{Q} = \begin{pmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{n-1} \end{pmatrix} \tag{6-5}$$

In order to apply this observability analysis, the system equations need to be linear to obtain a linear time-invariant system as in (6-4). However, since most systems consisting of kinetic rate equations are nonlinear, the equations need to be linearized expanding the nonlinear differential equations as a Taylor series and only retaining the first-order terms.

6.3.3 Concentration Profiles

Kinetic rate equations form generally a set of ordinary differential equations, which can be generally defined as

$$\frac{d\mathbf{y}}{dt} = \mathbf{f}(\mathbf{y}, \theta); \qquad \mathbf{y}(0) = \mathbf{y}_0 \tag{6-6}$$

where $\mathbf{y}$ is the vector of species concentrations, which is a function of the parameter vector θ and time t.

Equation (6-6) can be solved using numerical computation to provide the concentration profiles $\mathbf{y}(t)$ for the reactants and products as a function of time. However, specification of model parameters θ and initial conditions $\mathbf{y}_0$ is required. Prior information, when available, could be used to specify rate parameters, experimental conditions from the specifications of the planned experiments can specify initial conditions, while for other quantities an educated guess can be the best available option. In the worst case of a total ignorance regarding a certain system input, some preliminary experiments can be performed to obtain initial approximations on these unknown parameters and variables. The important point is that at least a qualitative understanding of the system can be developed by studying the concentration profiles for various scenarios.

6.3.4 Rate of Reactions

The rate of each of the reactions in the system can easily be calculated after the concentration profiles are obtained. Comparing the reaction rates among each other and for different scenarios is important to establish that all reactions considered are significant to the system. When a certain reaction rate is orders of magnitude slower, the system could be simplified by removing this particular reaction.

6.3.5 Sensitivity Analysis

A local sensitivity analysis evaluates the sensitivity of the system output to changes in the input [8]. Differential changes of $\mathbf{y}$ around a nominal point $\bar{\theta}$ are assessed by sensitivity coefficients, defined as

$$Z_{ij} = \left.\frac{\partial y_i}{\partial \theta_j}\right|_{\bar{\theta}} \tag{6-7}$$

where Z_{ij} is the sensitivity coefficient specifying the sensitivity of species y_i to parameter θ_j. Considering the system of differential equations in (6-6), the sensitivity coefficients can be calculated according to

$$\frac{\partial}{\partial\theta}\left(\frac{d\mathbf{y}}{dt}\right) = \frac{\partial\mathbf{f}(\mathbf{y},\theta,t)}{\partial\theta} \tag{6-8}$$

$$\frac{d}{dt}\left(\frac{\partial\mathbf{y}}{\partial\theta}\right) = \frac{\partial\mathbf{f}}{\partial\mathbf{y}}\frac{\partial\mathbf{y}}{\partial\theta} + \frac{\partial\mathbf{f}}{\partial\theta}\frac{\partial\theta}{\partial\theta} \tag{6-9}$$

$$\frac{d\mathbf{Z}}{dt} = \mathbf{JZ} + \frac{\partial\mathbf{f}}{\partial\theta} \tag{6-10}$$

where **Z** is the matrix of the sensitivity coefficients, and **J** is the Jacobian. Equation (6-10) is the adjoint sensitivity equation that needs to be solved simultaneously with the model of differential equations in (6-6).

6.4 Experimentation and Estimation

Within the framework discussed, experimentation and estimation are probably the most familiar stages for experimentalists. The focus in this section is on the perspective of a more integrative approach to experimentation and estimation than what is commonly implemented. The tools introduced require a close mutual interaction for successfully generating and analyzing experimental data. Starting with experimental design, data is generated and used for estimation, of which the results again can have an impact on future experiments. The feedback loop existing between estimation and experimentation is mainly possible because of the Bayesian approach, facilitating an intuitive representation and comparison of information obtained from experimental data.

6.4.1 Model-Based Experimental Design

Having decided on the experimental setup, experimental design can aid in determining the optimal settings of the independent variables for efficiently carrying out the experiments [58, 59]. As the basics of experimental design are closely related to decision theory, the Bayesian approach can be considered as the most fundamental approach from which in fact the commonly applied methods can be derived [59]. Appendix F will give an introduction to applying the Bayesian approach to experimental design. As Bayesian

experimental design will not be discussed in this thesis outside of Appendix D, it is also briefly discussed in Chapter 10 as a topic for future research.

6.4.2 System Calibration

During a series of experiments, systematic error can gradually be introduced for various reasons, such as wear and tear of the equipment, increased leakage, or build up of impurities. For a sufficiently slow effect, changes are likely to be unobservable so that the unaware experimentalist is generating invalid data. Therefore, a regular system calibration is recommended.

The system calibration should consist of performing an experiment of which the 'true' outcome is known and well documented, so that the slightest deviations from the expected outcome will be noticed. The necessary time and effort for re-running experiments are considered a wise investment to verify system integrity.

6.4.3 Recording Information

Recording of *all* information cannot be emphasized enough. Each and every piece of information should be considered valuable and should thus be recorded in detail. Not only the data itself, but also experimental conditions, uncertainties, manufacturing specifications, anomalies and other observations during experimentation can contribute to more accurate and precise estimation results. Examples of the importance of this kind of information will be given when discussing the case studies in chapter 7 and 8.

6.4.4 Objective Function

Plotting the objective function as for one or two variables (keeping other inputs constant) can develop an understanding of the performance of the estimation procedure. As will be demonstrated in the case studies, the shape of the objective function curve depending on one system variable can indicate potential estimation difficulties. Similarly, the objective function surface depending on two system variables can be useful to inspect. Unfortunately, this tool is useful mainly for systems with a relatively small number of degrees of freedom to estimate as visualization is limited to three dimensions.

6.4.5 Bayesian Parameter Estimation Framework

All efforts finally culminate in estimating the quantities of interest (e.g. parameters, initial conditions, or state variables) and thus careful analysis of the data is important. Generally, data analysis requires significantly less time and resources than experimentation. Analysis methods should therefore be employed in such a way to get as much information out of the data as possible. Bayesian parameter estimation, which was discussed in Chapter 5, is very suitable for this purpose.

6.4.6 Data Discrimination

The posterior distribution obtained from the Bayesian approach is a very intuitive representation of the state of knowledge or information content regarding the quantity of interest. The individual posterior distributions obtained from each of the data sets can conveniently be compared. Location and shape of the posterior distribution are the two attributes that the validity can be judged from.

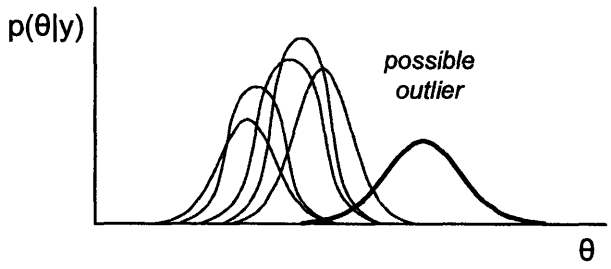


Figure 6-2. Illustration of outlier identification

The illustration in Figure 6-2 is an example where a number of posterior distributions are compared and a possible outlier is identified. The dissimilarity of the possible outlier, mainly based on its location, becomes immediately clear from the visual representation. If a posterior distribibution is not as obviously distinct as in Figure 6-2, a hypothesis test can always be carried out to determine the statistical significance of any differences. After identification of possible outliers, the suspicion should be verified from the information available regarding experimentation, such as observations and recorded anomalies.

6.4.7 Incremental Information Gain and Value of Information

Another useful tool is to assess the updating of the posterior probability distribution according to the 'learning algorithm' discussed in Section 3.3.2. Each data set obtained from a particular experiment contains a certain amount of information, which translates into the posterior distribution evolving to a more precise estimate for the rate parameter, as illustrated in Figure 6-3.

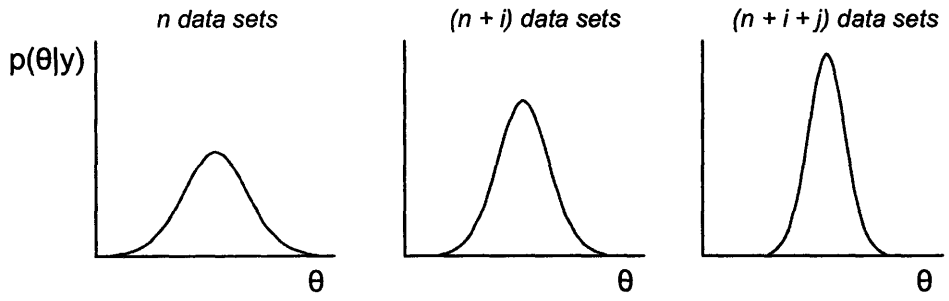


Figure 6-3. Illustration of incremental information gain

The incremental information content obtained from additional experiments can be quantified by the decrease in the variance or the increase in the precision (which equals the inverse of the variance). Experimentation can then be stopped based on a trade-off between the value of increasing precision and the expense of additional experimentation. This stopping rule requires quantification of the value of information, which is a concept of importance in decision theory [60] and will be briefly discussed in Appendix F.

7

Case Study 1

Quenching of $O(^1D)$ with Nitrogen

Example isn't another way to teach, it is the only way to teach.

Albert Einstein

This case study illustrates the approach to kinetic rate parameter estimation according to the framework developed in the previous chapter. The system that is analyzed is the quenching of the excited oxygen atom $O(^1D)$ with nitrogen. After providing a background of this estimation problem, the system description and analysis will define the problem. The Bayesian approach to parameter estimation is implemented with (1) an analytic model, in order to compare estimation results obtained with conventional statistical methods, and with (2) the original set of kinetic rate equations as system model. From the same amount of data, the Bayesian approach revealed significantly more information leading to a more precise parameter estimate, data discrimination identifying outliers, and evaluation of a suitable stopping rule for the experimentation.

7.1 Introduction

The present case study re-evaluates data obtained with the purpose to estimate the kinetic rate parameter for the atmospheric quenching of $O(^1D)$ with N_2. The experimental data used for this work originates from Dunlea [61], who applied conventional parameter estimation methods.

The research by Dunlea was part of a collective effort by three research groups [62-64]. The existing literature reported parameter estimates that conflicted with the results of recent experiments, thereby motivating the re-determination of the kinetic rate parameter. The three research groups jointly proposed an averaged value and confidence interval for the rate parameter based on a combination of the individual estimation results.

The combining of rate parameter estimates obtained from separate entities is an important issue in the atmospheric research community. Periodically, the Jet Propulsion Laboratory Panel for Data Evaluation tabulates aggregated values and confidence intervals for rate parameters of significance in atmospheric chemistry [65]. The reported parameter estimates and their uncertainties represent the subjective judgment of the panel, based on knowledge of the techniques, the difficulties of the experiments, and the potential for systematic errors. A mathematically rigorous methodology for combining results from different sources, with different experimental methods, and thus with different measurement errors, is not considered.

7.2 System Description

This section provides an overview of the atmospheric chemistry and kinetics according to the framework described in Section 6.2. The atmospheric reaction under study is the quenching of the electronically excited oxygen atom $O(^1D)$ (pronounced as *O-singlet-D*), which is considered the most important radical in the atmosphere [66]. $O(^1D)$ is responsible for the formation of two radical families, odd-hydrogen ($HO_x \equiv OH + HO_2$), and odd-nitrogen ($NO_x \equiv NO + NO_2$). These hydrogen and nitrogen radical species are intimately involved in air pollution, stratospheric ozone loss, and the overall oxidizing power of the atmosphere.

7.2.1 Atmospheric Chemistry

Dunlea [61] experimentally replicated the atmospheric photolysis process and generated $O(^1D)$ by photolysis of O_3 with a pulsed excimer laser operated at 248 nm in a

background of helium. Subsequently, the inert species N_2 quenches the electronically excited species $O(^1D)$ and the remaining O_3 reacts with $O(^1D)$, with both processes forming $O(^3P)$ ground state oxygen atoms (pronounced as *O-triplet-P*). The following set of reaction equations describes the system under investigation:

$$O(^1D) + N_2 \xrightarrow{k_1} O(^3P) + N_2 \tag{7-1}$$

$$O(^1D) + O_3 \xrightarrow{k_2} 2O(^3P) + O_2 \tag{7-2}$$

$$O(^1D) + O_3 \xrightarrow{k_3} 2O_2 \tag{7-3}$$

$$O(^1D) + He \xrightarrow{k_4} O(^3P) + He \tag{7-4}$$

where k_1, k_2, k_3, and k_4 are reaction rate parameters, of which rate parameter k_1 is of interest. Though helium was particularly selected to minimize undesirable quenching, its abundance justifies consideration of reaction (7-4). Additionally, impurities are present in the ultra-high purity *He* and also introduced via leaks in the experimental setup, leading to additional quenching of the $O(^1D)$ atoms. However, since the effect of these impurities is suspected to be minor, it is considered to be incorporated into reaction (7-4).

7.2.2 Kinetic Model

Both the bath gas *He* and the quenching agent N_2 are inert and can be assumed to remain constant throughout the reaction. Therefore, the rate equations describing the concentration of the species involved in equations (7-1) through (7-4) are given as follows

$$\frac{d[O(^3P)]}{dt} = \{k_1[N_2] + 2k_2[O_3] + k_4[He]\}[O(^1D)] - \frac{F}{V}[O(^3P)] \tag{7-5}$$

$$\frac{d[O(^1D)]}{dt} = \left\{-k_1[N_2] - (k_2 + k_3)[O_3] - k_4[He] - \frac{F}{V}\right\}[O(^1D)] \tag{7-6}$$

$$\frac{d[O_3]}{dt} = -(k_2 + k_3)[O(^1D)][O_3] - \frac{F}{V}([O_3] - [O_3]_0) \tag{7-7}$$

where F is the total flow through the reaction cell, and V the volume of the detection region. The system of kinetic rate equations is completed by the initial conditions at $t = 0$

$$[O(^3P)] = [O(^3P)]_0, \quad [O(^1D)] = [O(^1D)]_0, \quad [O_3] = [O_3]_0 \tag{7-8}$$

Each of the variables and parameters in this system of ordinary differential equations will be specified in the various stages of the system description and analysis below.

7.2.3 Physical Constraints and Relationships

The least that can be said about the rate parameters is that they must be positive values, which is information that will be used to set the lower bound for the estimation. Additional information to specify the prior distributions was obtained from literature.

Valuable information regarding rate parameters k_1, k_2, and k_3 is obtained from the JPL publication on 'Chemical Kinetics and Photochemical data for use in Atmospheric Studies' [65], according to which rate parameters k_2 and k_3 are equal. This information regarding physical properties of the system conveniently reduces the dimensionality of the estimation problem.

Additionally, estimates for the rate parameters and their uncertainty factor f at a temperature of 295 K are given follows

$$k_1 = 2.6\times10^{-11} \quad and \quad f_{k_1} = 1.2 \tag{7-9}$$

$$k_2 = k_3 = 1.2\times10^{-10} \quad and \quad f_{k_2} = 1.3 \tag{7-10}$$

where the uncertainty factors define the bounds on k_i as

$$\left[\frac{k_i}{f_{k_i}}, f_{k_i} k_i\right] \tag{7-11}$$

corresponding to a non-symmetrical distribution of the uncertainty. Since the bounds defined by the uncertainty factors are considered to approximately represent an interval of two standard deviations around the estimate k_i [65], a lognormal prior distribution capturing 68% probability within the uncertainty interval was determined. The resulting prior distributions for parameters k_1 and k_2 are shown in Figure 7-1, where the dotted lines represent the 68% probability area around the mode.

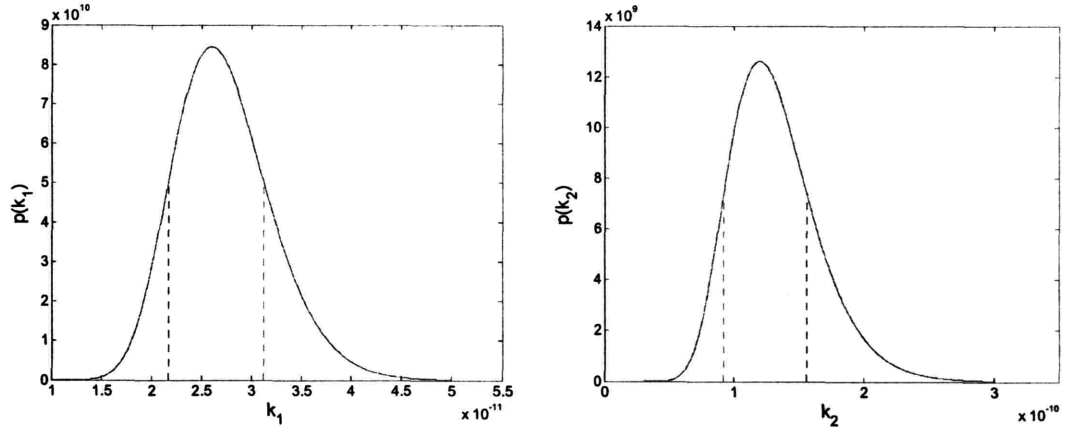


Figure 7-1. Prior distributions with uncertainty bounds

These lognormal distributions for parameter k are defined as

$$p(k) = \frac{1}{k\sigma\sqrt{2\pi}} e^{\frac{(\ln k - \mu)^2}{2\sigma^2}} \tag{7-12}$$

where μ and σ^2 are the mean and variance, respectively, of the lognormal distribution and are related to the mode by

$$\tilde{k} = e^{\mu - \sigma^2} \tag{7-13}$$

where $\tilde{k}$ is the mode of the lognormal distribution of parameter k. The mode and variance for the lognormal prior distributions, obtained with the Matlab script in Section A.3, are specified in the table below.

Table 7-1. Mode and variance for the lognormal prior distributions

	$\tilde{k}_i$	$\sigma_{k_i}^2$
k_1	$2.6 \cdot 10^{-11}$	0.0320
k_2	$1.2 \cdot 10^{-10}$	0.0651

The information regarding rate parameter k_4 is limited to specifications for the upper bound. The reported upper bounds for k_4 differ substantially [67], namely $k_4 < 7 \cdot 10^{-16}$, $k_4 < 3 \cdot 10^{-15}$, and $k_4 < 3 \cdot 10^{-13}$. Therefore, the prior distribution for parameter k_4 will be specified as a uniform distribution between 0 and the most conservative (i.e. largest) reported value as upper bound.

7.2.4 Experimental Procedure

The experimental procedures will only briefly be described in this section. For a more detailed description, the reader is referred to [61].

The experiment took place in a reaction cell with a volume of approximately 250 cm^3, kept at a constant pressure of approximately 20 Torr, varying slightly between experiments. The reactants O_3 and N_2 were mixed with the bath gas *He* before entering the reaction cell at a constant flow rate, which also varied slightly between experiments. While $[N_2]$ was controlled by the ratio of flow rates, the initial concentration $[O_3]_0$ was measured in situ for each experiment by UV absorption at 253.7 nm. The excited $O(^1D)$ oxygen atoms were generated by pulsed photolysis of O_3 using an excimer laser operated at 248 nm. The resulting $O(^3P)$ concentration profiles were detected by resonance fluorescence as a function of time, and for a specific N_2 concentration.

7.3 System Analysis

The next step of the framework described in Section 6.3 is to develop a thorough understanding of the underlying physics and chemistry. The analyses described below

are performed with the prior information for the rate parameters and the specifications of the experimental conditions as model input.

7.3.1 Degrees of Freedom

The first step in the analysis is to analyze the incidence matrix to determine the degrees of freedom in this estimation problem. The incidence matrix with the system equations located on the rows and the variables in the columns is shown below.

	$\frac{d[O(^3P)]}{dt}$	$\frac{d[O(^1D)]}{dt}$	$\frac{d[O_3]}{dt}$	$[O(^3P)]$	$[O(^1D)]$	$[O_3]$	k_1	k_2	k_3	k_4
(7-5)	×			×	×	×	×	×		×
(7-6)		×			×	×	×	×	×	×
(7-7)			×		×	×		×	×	
(7-8)a				×						
(7-8)b					×					
(7-8)c						×				
(7-10)								×	×	

From the dimensions of the incidence matrix it can be concluded that the problem is underdetermined. With 10 variables and only 7 equations, the problem has three degrees of freedom specified by the output set assignment as $\{k_1, k_2, k_4\}$ or $\{k_1, k_3, k_4\}$, of which the former is selected in this case study as the parameters to be estimated.

All of the rate parameters k_1, k_2, and k_4 are involved in depletion of $O(^1D)$. When considering $[O_3]$ constant throughout the reaction, an overall rate parameter k' can be defined as

$$k' = k_1[N_2] + 2k_2[O_3] + k_4[He] \tag{7-14}$$

which implies that in order to estimate one of the rate parameters, the other two should be known. This fact is valuable knowledge when designing experiments with the goal of estimating k_1, since additional data would be required for both k_2 and k_4.

With data generated by an experiment in the absence of N_2, the remaining linear relationship between k' and rate parameters k_2 and k_4 can be investigated visually by graphing the objective function over a range of parameter values. The objective function J is calculated as

$$J = \sum_{t=1}^{p} (y_t - [O(^3P)]_{t,model})^2 \tag{7-15}$$

for p datapoints in each data set and with $[O(^3P)]_{t,model}$ calculated using the kinetic rate equations excluding equation (7-1) as k_1 is irrelevant in the absence of N_2.

As expected, the objective function shown in Figure 7-2 does not show a unique combination of rate parameters k_2 and k_4 at the minimum, but a range of possible linear combinations. Thus, in order to estimate either rate parameter, the other should be known.

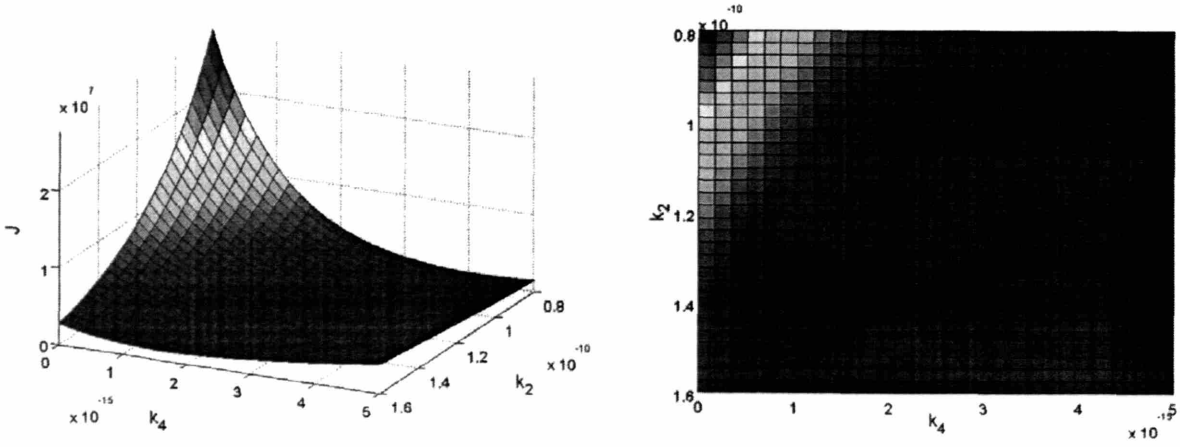


Figure 7-2. Typical objective function surface for a dataset generated in the absence of N_2

7.3.2 Observability

In order to obtain a linear system of differential equations, equation (7-5) and (7-6) are approximated by a Taylor series expansion of which only the first order terms are retained. The system matrix **A** can then be specified as

$$\mathbf{A}=\begin{bmatrix} -\frac{F}{V} & k_1[N_2]+2k_2[O_3]^*+k_4[He] & 2k_2[O(^1D)]^* \\ 0 & -k_1[N_2]-2k_2[O_3]^*-k_4[He]-\frac{F}{V} & -2k_2[O(^1D)]^* \\ 0 & -2k_2[O(^1D)]^* & -2k_2[O(^1D)]^*-\frac{F}{V} \end{bmatrix} \tag{7-16}$$

where $[O_3]^*$ and $[O(^1D)]^*$ are the concentrations around which the system is linearized.

The experiments only collect data on the changing concentration of $O(^3P)$ as a function of reaction time, so the output matrix is

$$\mathbf{C}=\begin{bmatrix}1 & 0 & 0\end{bmatrix} \tag{7-17}$$

The observability matrix was constructed with Maple according to equation (6-5) Applying Gaussian elimination resulted in

$$\mathbf{Q}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & k_1[N_2]+2k_2[O_3]^*+k_4[He] & 2k_2[O(^1D)]^* \\ 0 & 0 & -\frac{4k_2{}^2[O(^1D)]^{*2}(k_1[N_2]+k_4[He])}{k_1[N_2]+2k_2[O_3]^*+k_4[He]} \end{bmatrix} \tag{7-18}$$

which is a triangular matrix with full rank, so that the proposed system is completely observable.

7.3.3 Concentration Profiles

To solve the set kinetic rate equations, several parameters and variables need to be specified. The rate parameters will be specified according to the available information: $k_1 = 2.6 \cdot 10^{-11}$, $k_2 = k_3 = 1.2 \cdot 10^{-10}$, and $k_4 = 3 \cdot 10^{-15}$. The initial concentrations for the various components are specified according to the planned experiments: $[O_3]_0 = 10^{13}$, $[N_2] = 10^{14} - 5 \cdot 10^{14}$ depending on the experiment, and $[He]$ can be calculated according to

$$[He] = N_A \left\{ V_m \frac{760}{P} \frac{T}{273} \right\}^{-1} \tag{7-19}$$

where N_A is Avogadro's number, V_m is the molar volume (22.4 cm^3/mol at standard conditions), P is the pressure (typically 20 Torr), and T is the temperature (set at 295 K) in the reaction cell. Finally, the initial radical concentrations $[O(^1D)]_0$ and $[O(^3P)]_0$ can be estimated at constant wavelength λ and under steady state conditions according to

$$[O(^1D)]_0 = [O_3]\, \Phi\, \sigma\, I = [O_3]\, \Phi\, \sigma\, F_l \left(\frac{hc}{\lambda} \right)^{-1} \tag{7-20}$$

$$[O(^3P)]_0 = [O_3]\, (1 - \Phi)\, \sigma\, I = [O_3]\, (1 - \Phi)\, \sigma\, F_l \left(\frac{hc}{\lambda} \right)^{-1} \tag{7-21}$$

where F_l is the measured fluence of the laser in mJ·pulse^{-1}·cm^{-2}, h is Planck's constant ($6.626 \cdot 10^{-31}$ mJ s^{-1}), c is the speed of light ($2.9979 \cdot 10^{10}$ cm s^{-1}), σ is the absorption cross section of O_3 ($1.08 \cdot 10^{-17}$ cm^2 at 248 nm, [65]), and Φ is the quantum yield for the production of $O(^1D)$ (0.9 at 248 nm, [68]).

The detection volume V, being the region of overlap between the photolysis laser and the resonance fluoresence lamp, is rather difficult to determine exactly. More important is to realize that the ratio of F and V in the kinetic rate equations represent a first order decay process, which does not have a significant effect on the radical concentrations at the most interesting initial stage of the experiment. Only when the quenching process is near

completion, the flow through the reactor will become the dominating process for the $O(^3P)$ decay. Thus, since the exact value is not very important at this moment in the analysis, based on approximations for V the following will be specified

$$\frac{F}{V} = 40 s^{-1} \tag{7-22}$$

where the total flow through the reaction cell F is another variable determined by the experimental settings.

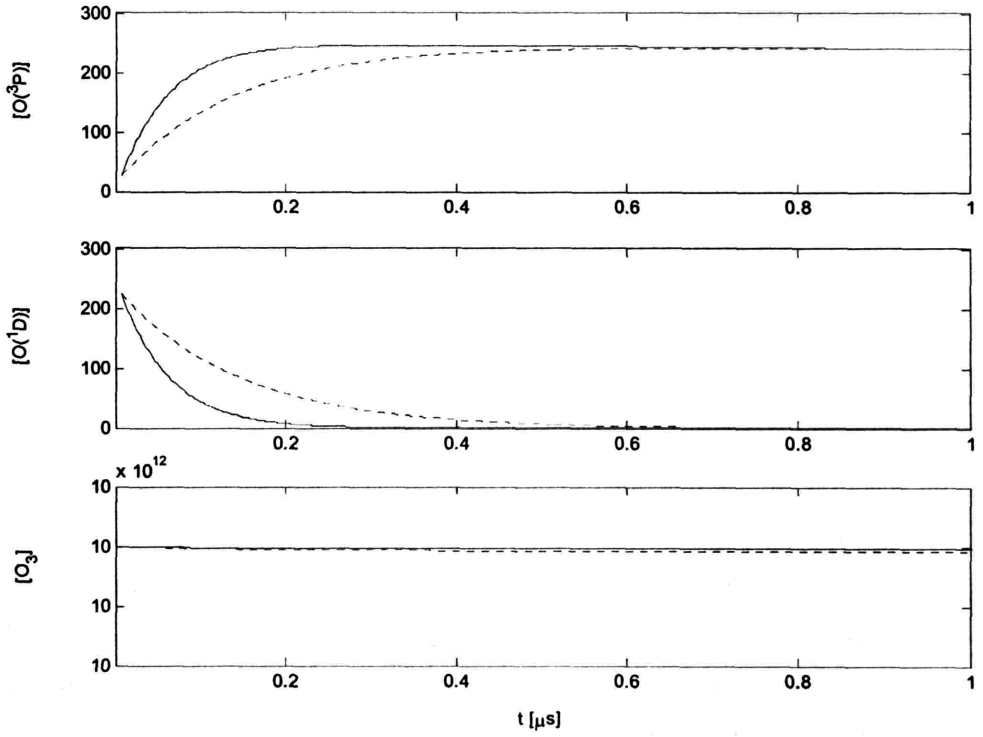


Figure 7-3. Concentration profiles (--- $[N_2]=1\cdot10^{14}$, — $[N_2]=5\cdot10^{14}$)

With equations (7-5) through (7-7) and the specified input above, the concentration profiles for O_3 and the radicals $O(^3P)$ and $O(^1D)$ are obtained by using the Matlab ODE15s solver and are shown in Figure 7-3. The concentrations in these profiles are given in the total number of molecules, which are present in the detection volume. The

dotted line represents the solution for $[N_2] = 1 \cdot 10^{14}$ and the solid line represents the solution for $[N_2] = 5 \cdot 10^{14}$, respectively the approximate lower and upper bound for the series of experiments.

The profiles show the impact of $[N_2]$, which is the independent variable in the experiments. Increasing $[N_2]$ will obviously lead to an increased depletion rate of the $O(^1D)$ radicals. The $[O_3]$ appears to remain constant throughout the reaction and therefore application of the pseudo-steady state assumption regarding $[O_3]$ is justified. Finally, the slight decay in $[O(^3P)]$ after reaching the maximum in the curve indicates that the flow through the reactor has a minor effect on the system.

7.3.4 Rate of Reactions

With the parameters and variables specified above, the following reaction rates are evaluated

$$r_1 = k_1[N_2][O(^1D)] \tag{7-23}$$

$$r_2 = (k_2 + k_3)[O_3][O(^1D)] = 2k_2[O_3][O(^1D)] \tag{7-24}$$

$$r_4 = k_4[He][O(^1D)] \tag{7-25}$$

where r_i is the reaction rate for reaction i in molecules/s. The calculated profiles for these reaction rates are shown in Figure 7-4. The results show that r_1 is approximately equal to r_2 and r_4 when $[N_2] = 1 \cdot 10^{14}$ and approximately an order of magnitude larger when $[N_2] = 5 \cdot 10^{14}$. Thus, though reaction rate r_1 is the dominating driving force for the depletion of $O(^1D)$, the difference in reaction rates is not significant to justify a simplification of the system by neglecting one of the reaction rates.

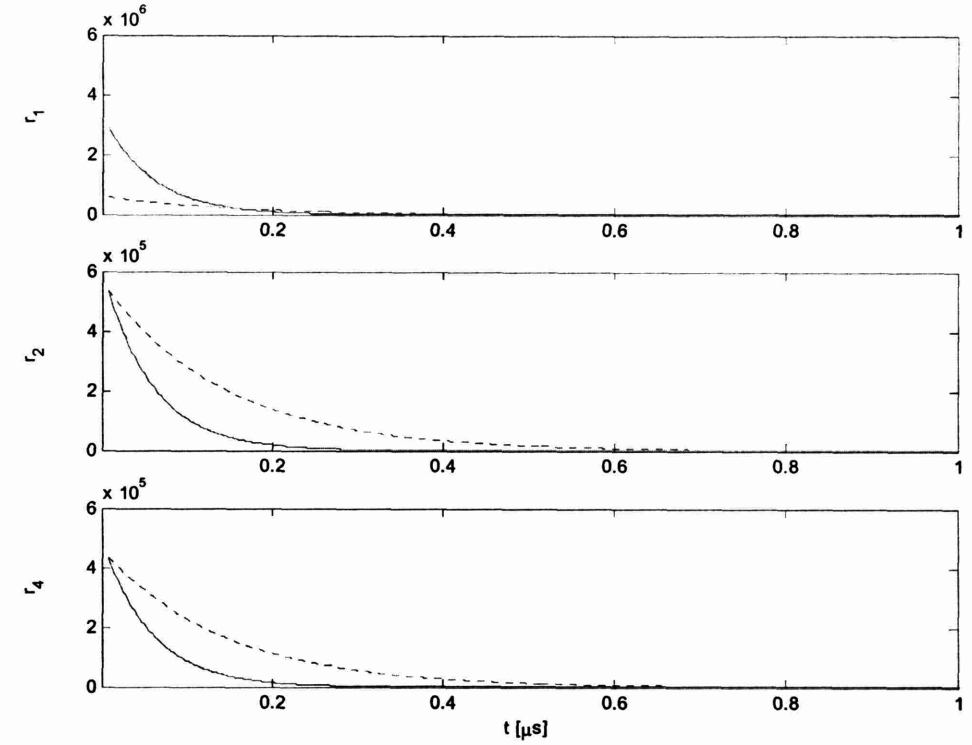


Figure 7-4. Reaction rates (--- $[N_2]=1\cdot10^{14}$, — $[N_2]=5\cdot10^{14}$)

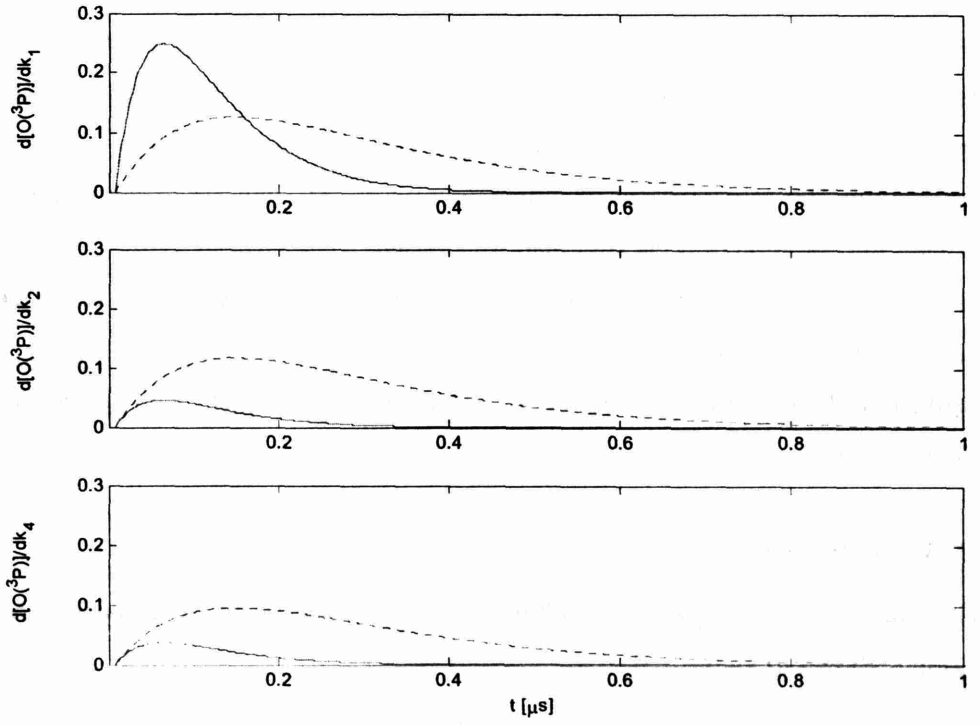


Figure 7-5. Sensitivity of $[O(^3P)]$ to the rate parameters (--- $[N_2]=1\cdot10^{14}$, — $[N_2]=5\cdot10^{14}$)

7.3.5 Sensitivity Analysis

The ordinary differential equations for the sensitivities as specified by equation (6-10) are solved simultaneously with the kinetic rate equations (7-5) through (7-7). Figure 7-5 shows the profiles of the normalized sensitivities during the reaction. The sensitivity of $[O(^3P)]$ to a change in k_1 is slightly larger than its sensitivity to the other rate parameters for $[N_2] = 1 \cdot 10^{14}$ and approximately a factor of 5 larger when $[N_2] = 5 \cdot 10^{14}$. This conveys the same message as above that all rate parameters have a significant impact on the system and that all of them should be incorporated for fitting the kinetic rate model to the data.

7.4 Experimentation

An overview and the purpose of the various data collected will be given, after which general properties of the measurements will be discussed.

7.4.1 Available Data

As discussed in Section 7.3.1, in order to estimate rate parameter k_1, the rate parameters k_2 of the reaction of $O(^1D)$ with O_3 and k_4 of the quenching process of $O(^1D)$ with He have to be known first. Therefore, three different types of data were collected:

I. k_1 estimation: 206 measurement experiments with N_2, O_3, and He in the reaction cell. The concentration of N_2 is systematically varied.

II. k_2 estimation: 35 measurement experiments with O_3 and He in the reaction cell. The concentration of O_3 is systematically varied.

III. k_4 estimation: 67 background experiments in the absence of N_2, but with O_3 and He in the reaction cell.

7.4.2 Data Characteristics

A typical example of a data series resulting from one experiment at a specific N_2 concentration is shown in Figure 7-6.

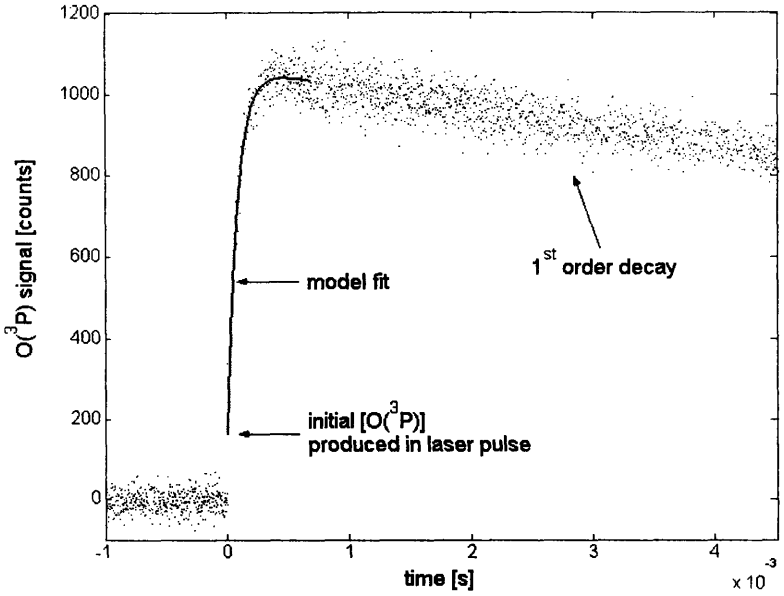


Figure 7-6. Typical temporal profile of a $O(^3P)$ measurement

The temporal profile shows a gap at $t = 0$, indicating an initial concentration of $O(^3P)$ produced by photolysis of O_3. Subsequently, the concentration of $O(^3P)$ rises as the quenching progresses and reaches a maximum at the point where $O(^1D)$ has depleted. After formation has ceased, the $O(^3P)$ concentration decreases steadily as $O(^3P)$ is flushed out from the reaction cell by the continuous laminar flow of the O_3, N_2, and He mixture.

7.4.3 Data Preparation

A complete data series consisted of several thousand data points of $O(^3P)$ concentration as a function of reaction time. However, only a few hundred of these data points were actually used for the parameter estimation, as indicated by the solid line representing the model fit in Figure 7-6. The reason for discarding most of the data points was to construct data sets that are comparable in terms of their rise times. Though the temporal profiles of $O(^3P)$ for different experiments are of similar shape, their magnitude strongly depends on the concentration of N_2. A faster rise, facilitated by a higher N_2 concentration, leads to less data points collected during the rise compared to the number of data points

in the decaying section of the plot. For the parameter estimation, however, it is desirable that the proportions of the rising and decaying sections of the temporal profile are equivalent among the various data series. Further details on preparing the data for analysis is given in [61].

7.4.4 Measurement Error

The instrument noise ε_n as observed at the baseline for $t < 0$ is likely in part caused by scattered daylight reaching the photon detector. The additional system variability ε_v during the measurement can be specified with the theory of "counting statistics", as the $O(^3P)$ atoms are detected by counting the photons released by the quenching reaction. According to counting statistics the error model for a random process, such as the quenching reaction, is given by a Poisson distribution. However, as each data point in a particular time bin is the result of the summation of counts for multiple reactions, the error model for ε_v can be approximated by a normal distribution according to the Central Limit Theorem (see Section 2.2.1). Thus, the error model discussed in Section 2.2.3 can be applied, implementing an overall error ε_0 according to equation (2-5).

7.5 Conventional Parameter Estimation

After simplifying the original model, the conventional estimation methods applied to the original estimation of rate parameter k_1 will briefly be discussed [61], followed by a critical assessment of the uncertainty assignment regarding the parameter k_1.

7.5.1 Analytic Solution of the Kinetic Model

Before the parameter estimation problem can be resolved in an attainable manner by conventional nonlinear estimation methods, several simplifications are required. The first simplification applies the knowledge that rate parameters k_2 and k_3 are equal, implying a combined rate parameter defined as

$$k_2' = (k_2 + k_3) = 2k_2 \tag{7-26}$$

so that reaction equations (7-2) and (7-3) can be substituted by the non-elementary reaction

$$O(^1D) + O_3 \xrightarrow{k_2'} O(^3P) + \tfrac{3}{2} O_2 \tag{7-27}$$

The next simplification is to exclude the background reaction with *He* and describe the radical decay as first order processes according to

$$O(^3P) \xrightarrow{k_5} loss \tag{7-28}$$

$$O(^1D) \xrightarrow{k_6} loss \tag{7-29}$$

where k_5 and k_6 are considered as loss rate coefficients for the decay processes.

The final simplification is to apply the pseudo steady state assumption, implying that the concentration of O_3 is constant. This is an acceptable assumption according to the system analysis in Section 7.3, reducing the estimation problem to two kinetic rate equations. These equations can be solved analytically, as shown in Appendix C, to give the following result

$$[O(^3P)] = Ae^{-Bt} + Ce^{-Dt} \tag{7-30}$$

where the parameters A, B, C, and D are given by

$$A = \frac{[O(^1D)]_0(k_1[N_2] + k_2'[O_3])}{k_5 - k_1[N_2] - k_2'[O_3] - k_6} \tag{7-31}$$

$$B = k_1[N_2] + k_2'[O_3] + k_6 \tag{7-32}$$

$$C = [O(^3P)]_0 - A \tag{7-33}$$

$$D = k_5 \tag{7-34}$$

and thus is the parameter estimation problem reduced to a bi-exponential equation with four unknown parameters A, B, C, and D.

7.5.2 Strategy

By deriving the analytic model, the individual rate parameters were lumped to create the new parameters A, B, C, and D, so that direct estimation of rate parameter k_1 is not possible anymore. The estimation strategy is employed in two stages, as illustrated in Figure 7-7, where the data types refer to the discussion in Section 7.4.1.

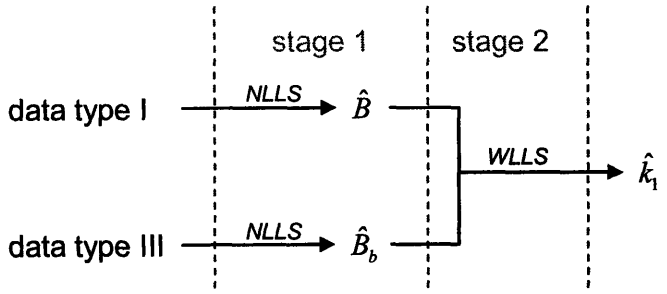


Figure 7-7. Strategy to estimate rate parameter k_1 with conventional estimation methods

In estimation stage 1, equation (7-30) is fitted to the data of type I using the method of nonlinear least squares (NLLS) to obtain $\hat{B}$, which is the point estimate of parameter B. Similarly, the point estimate of the background correction parameter $\hat{B}_b$ can be obtained by fitting to the data of type III. Since parameter B is linearly dependent on the N_2 concentration according to equation (7-32), the point estimate $\hat{k}_1$ can be obtained via the method of weighted linear least squares (WLLS). Both stages will be discussed in more detail below.

7.5.3 Estimation Stage 1: Nonlinear Least Squares

The parameters A, B, C, and D were estimated from each of the 206 data sets of type I with the statistical software package "Igor", which implements the nonlinear least squares fitting approach by minimizing the χ^2 statistic, which is defined as

$$\chi^2 = \sum_t \left(\frac{y_t - [O(^3P)]_{model,t}}{\sigma_y} \right)^2 \tag{7-35}$$

where $[O(^3P)]_{model,t}$ is the model value calculated from equation (7-30) at time t, y_t is the measured data at time t, and σ_y is an estimation of the standard deviation for the data y_t.

At first, the four parameters A, B, C, and D were fitted simultaneously, but estimation results were not acceptable as occasionally parameter D was estimated to be negative, which is physically impossible [61]. To overcome this issue, parameter D was separately estimated as the decay coefficient for the first order exponential decay $O(^3P)$. This approach is based on the fact that the quenching is essentially complete after detecting the maximum concentration of $O(^3P)$, when the first order decay becomes the dominating process. Subsequently, parameters A, B, and C were estimated according to equation (7-35) while keeping parameter D constant at its estimated value.

7.5.4 Estimation Stage 2: Weighted Linear Least Squares

After correcting for the background signal, the desired rate parameter k_1 can be determined from the method of weighted linear least squares with the following relationship

$$\hat{B} - \hat{B}_b = \hat{k}_1[N_2] \tag{7-36}$$

where $\hat{k}_1$ is the point estimate for rate parameter k_1, $[N_2]$ is the independent variable, and the weights originate from the estimates of the variance of $\hat{B}$. Each of the weights was calculated as $\hat{\sigma}_{\hat{B}}^{-2}$.

7.5.5 Original Uncertainty Calculation

The original uncertainty evaluation regarding the estimate of rate parameter k_1 [61] utilized the quadrature summation discussed in Section 2.3.

According to the calculation of equation (7-36), two separate sources affect the uncertainty in $\hat{k}_1$:

1) the uncertainty in $\hat{B}$, as represented by the standard deviation $\hat{\sigma}_{\hat{B}}$ obtained from the nonlinear least squares fitting (see Section 7.5.3) and calculated as discussed in Section 2.3.1 and 2.3.2 by the software package "Igor".
2) the uncertainty in $[N_2]$, which was calculated via quadrature summation using the manufacturing specifications on the precision of the pressure head, flow and temperature meters that were controlling the N_2 flow rate, as follows

$$\frac{\hat{\sigma}_{N_2}}{[N_2]} = \sqrt{\frac{\sigma_P^2}{P^2} + \frac{\sigma_F^2}{F^2} + \frac{\sigma_T^2}{T^2}} \tag{7-37}$$

where $\hat{\sigma}_{N_2}$ is the standard deviation of the N_2 measurement, P is the pressure, F is the flow, and T is the temperature in the reaction cell, and σ_P, σ_F, and σ_T are their standard deviations, respectively. With individual measurement errors (interpreted as a standard deviation) specified to be ±2% of the measured value, the measurement error for $[N_2]$ is approximately 3.5%.

These two sources of uncertainty were combined to obtain an overall standard deviation for the estimate $\hat{k}_1$ according to quadrature summation as

$$\frac{\hat{\sigma}_{overall}}{\hat{k}_1} = \sqrt{\frac{\hat{\sigma}_{\hat{B}}^2}{\hat{B}^2} + \frac{\hat{\sigma}_{N_2}^2}{[N_2]^2}} \tag{7-38}$$

where $\hat{\sigma}_{overall}$ is the overall standard deviation representative of the uncertainty in estimate $\hat{k}_1$. The relative uncertainty calculated with equation (7-38) was approximately 5%, leading to the reported outcome of $\hat{k}_1 = 3.01 \pm 0.16$ (10^{-11} molecule $^{-1}$ cm^{-3} s^{-1}).

7.6 The Bayesian Approach using the Analytic Model

This section describes the specifics of the Bayesian approach using the analytic equation (7-30) as system model. The goal is to find the posterior distribution of the parameters A, B, C, D, and measurement variance σ_0^2, defined as

$$p(A,B,C,D,\sigma_0^2 \mid y) \propto p(A,B,C,D,\sigma_0^2)p(y \mid A,B,C,D,\sigma_0^2) \tag{7-39}$$

which will be approximated by MCMC simulation. The estimation is performed in two stages, similarly as illustrated in Figure 7-7, where first an estimate for parameter B is obtained from each of the data series and the overall estimate for the rate parameter k_1 results from the parallel estimation method described in Section 5.4.2.

7.6.1 Estimation Stage 1: MCMC Formulation for B and B_b

The prior distributions for each of the parameters are assumed to be independent, so that the joint prior distribution can be calculated by

$$p(A,B,C,D,\sigma_0^2) = p(A)p(B)p(C)p(D)p(\sigma_0^2) \tag{7-40}$$

where each of the marginal prior distributions need to be defined separately.

Without any information on parameters A, B, and C, a uniform distribution will be implemented for the prior, as given by equation (5-8). The knowledge regarding parameter D from the separate estimation as discussed above in Section 7.5, will be incorporated as prior information, as given by

$$p(D) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_D} exp\left\{-\frac{\left(D-\hat{D}\right)^2}{2\hat{\sigma}_D^2}\right\} \tag{7-41}$$

where $\hat{D}$ is the prior estimate and $\hat{\sigma}_D^2$ is the estimated variance for $\hat{D}$. The gamma distribution with $\alpha = \beta = 10^{-3}$ as defined by equation (5-10) is implemented as prior distribution for the measurement variance σ_0^2.

The likelihood function for the data y_t is, equivalently to equation (5-5), represented by a normal distribution

$$p(y \mid A,B,C,D,\sigma_0^2) = \prod_{t=1}^{p} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{ -\frac{\left(y_t - [O(^3P)]_{model,t} \right)^2}{2\sigma_0^2} \right\} \qquad (7\text{-}42)$$

where $[O(^3P)]_{model,t}$ is the concentration at time t calculated with the bi-exponential model in equation (7-30) for a particular set of parameters A, B, C, and D, for the total of p data points.

The initial design for MCMC was assembled from the estimates A, B, C, and D, and their standard deviations as obtained from the nonlinear least squares fitting discussed in Section 7.5.3. The sample variance of the baseline measurements at $t < 0$ is used as initial value for σ_0^2. The probing distribution was constructed from the initial design as discussed in Section 5.5.5.

The information obtained from the nonlinear least squares estimation method will facilitate rapid convergence. Generally, conventional estimation methods are relatively easy to perform using standard statistical software packages and can be a convenient source for a suitable initial design. The Bayesian approach can subsequently be applied for a more rigorous parameter estimation and evaluation of the uncertainty.

7.6.2 Estimation Stage 1: Results for B and B_b

For the estimation of A, B, C, D, and σ_0^2 the MCMC algorithm was applied to generate 105,000 samples of the 5-dimensional joint posterior distribution. The first 5,000 samples were discarded as the burn-in period and the remaining samples were thinned in

a ratio of 1:10 to reduce the correlations among the parameters. The remaining 10,000 samples were used to estimate the marginal posterior distributions of the parameters.

The marginal posterior distributions $p(B|y_i)$ for each experiment i can be obtained by generating a kernel density estimate of the samples for B. Figure 7-8 compares for a particular experiment the marginal posterior probability distribution for B, obtained by the Bayesian approach, to the assumed normal distribution for the point estimate $\hat{B}$ obtained by the conventional estimation method of nonlinear least squares (see Section 7.5). The estimates are very similar, except for a slight difference in the mean of the distribution.

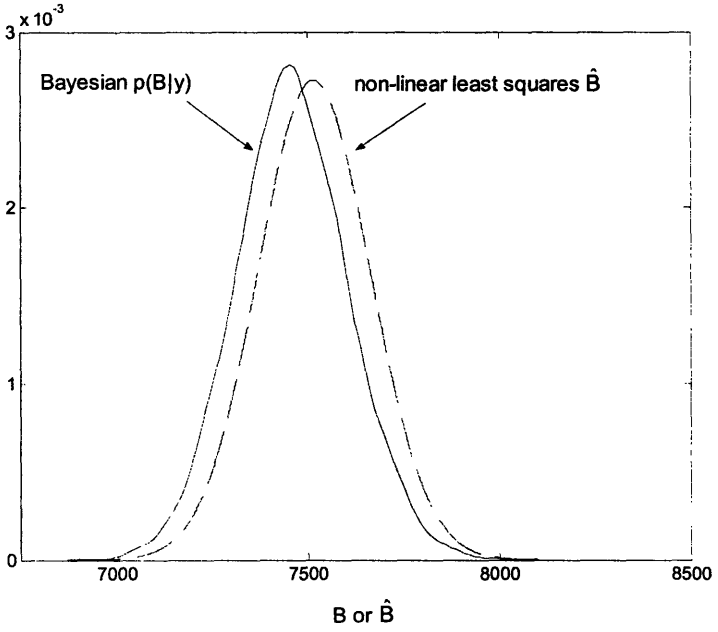


Figure 7-8. Marginal posterior distribution for B compared to $\hat{B}$ for a particular data set

7.6.3 Estimation Stage 2: Rate Parameter k_1

Similar to equation (7-36), the posterior distributions for parameter B and background parameter B_b are used to obtain the probability distribution for rate parameter k_1. The uncertain parameters B and B_b are mathematically treated as random variables and thus subtraction of random variables is performed by convolution of the probability

distributions. A Monte Carlo approach is applied to determine the samples of the probability distribution for k_1 for each of the data sets of type I as follows

$$k_1^i(n) = \frac{\left(B(n) - B_b(n)\right)_i}{[N_2]_i} \tag{7-43}$$

where $k_1^i(n)$ is the n^{th} sample from the posterior probability distribution of k_1 determined from the n^{th} sample of the B and B_b obtained from the data of experiment i. The samples for $[N_2]_j$ are obtained from a normal distribution defined by

$$[N_2]_i \sim N([N_2]_{m,i}, \hat{\sigma}_{N_2}^2) \tag{7-44}$$

where $[N_2]_{m,i}$ is the measured N_2 concentration for experiment i, and $\hat{\sigma}_{N_2}$ was approximated as 3.5% of $[N_2]_i$. Unfortunately, the available information regarding the pressure, temperature, and flow used to control $[N_2]$ is not detailed enough to specify the uncertainty more precisely for each individual experiment.

The probability distribution for each of these i sample sets for k_1 is obtained from a kernel density estimate, each resulting from the 206 data sets of type I. These probability densities were combined according to the parallel estimation method described in Section 5.4.2 to obtain an overall estimate for the rate parameter k_1. The results regarding the estimate for the rate parameter k_1 are compared in Figure 7-9. The two narrow probability distributions are the Bayesian posterior distributions. The dotted probability distribution represents the uncertainty in the estimate obtained from the conventional estimation method, assuming a normal distribution based on $\hat{k}_1$ and its 95% confidence interval [61].

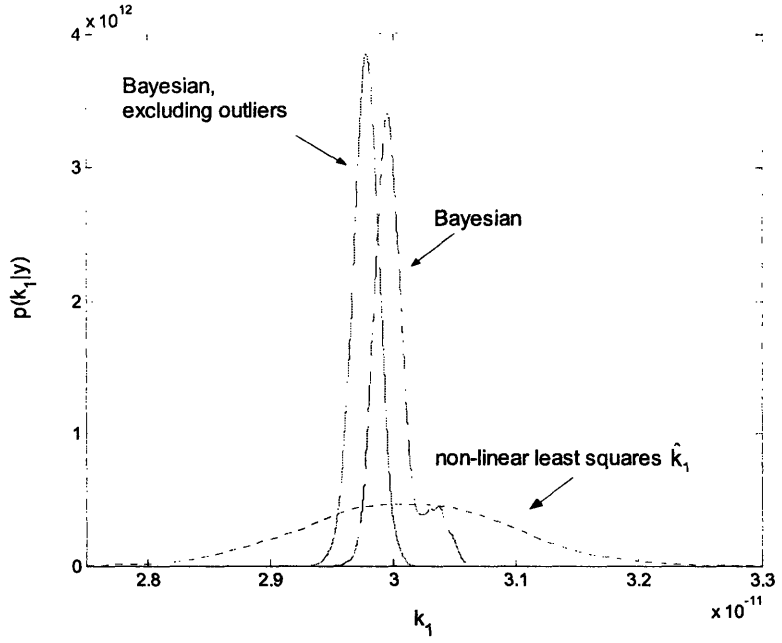


Figure 7-9. Comparisons of estimates for k_1 using the Bayesian and conventional approach

From the same amount of data, the Bayesian approach to parameter estimation resulted in a more precise estimate for rate parameter k_1. In other words, more information was obtained from the available data. The difference between the narrow Bayesian posterior distributions will be discussed in the following section.

7.7 Further Advantages of the Bayesian Approach

Besides obtaining a more precise parameter estimate, the Bayesian approach also allows for intuitive inspection of the quality of the data. Based on the discussion in Section 6.4.6, outliers and the existence of a systematic error will be identified in the data of this case study. Additionally, by updating the rate parameter estimate after each experiment, a stopping rule will be discussed.

7.7.1 Considering Individual Posterior Distributions

The Bayesian posterior distribution obtained from evaluating all 206 data sets of type I appeared to have a small shoulder (see Figure 7-9). The reason for this shoulder became

clear upon inspection of the individual posterior probability distributions for B. The majority of these distributions were of similar shape and location, but a few implied significantly higher values for the estimates of k_1, similarly as illustrated in Figure 6-2. Approximately 20 distributions were considered as outliers and removed from the data. Laboratory notes regarding the experimentation subsequently revealed that the data sets responsible for the outliers were generated under erratic experimental conditions. In fact, the data sets in question were not supposed to be included in the parameter estimation in the first place. The identification of outliers in this case study is therefore an excellent example of the intuitive nature of the Bayesian approach, and its capabilities to discriminate among data.

The removal of these outliers resulted in a smooth posterior distribution for k_1, shown as the solid curve in Figure 7-9. Though there is no significant effect on the mean, the variance of the rate parameter estimate decreases, as shown in the table below, since the shoulder is removed from the Bayesian posterior probability distribution.

Table 7-2. Comparison of summary statistics for rate parameter k_1

	$E[k_1]$ (10^{-11} cm^3 molecule^{-1} s^{-1})	σ_{k_1} (10^{-11} cm^3 molecule^{-1} s^{-1})
Bayesian estimate, complete data	3.00	0.0016
Bayesian estimate, outliers removed	2.98	0.0010
Conventional estimate	3.01	0.082

7.7.2 Updating of the Posterior Distribution

Using the parallel estimation approach (see Section 5.4.2), the evolution of the posterior distribution of the overall estimate for rate parameter k_1 was determined by accumulating the information from the individual distributions, as discussed in Section 6.4.7. The results are shown in Figure 7-10, where the data sets for updating the posterior probability distribution are incorporated both in chronological and random order.

Chronological order refers to the sequence of experiments, spanning a period of approximately two years.

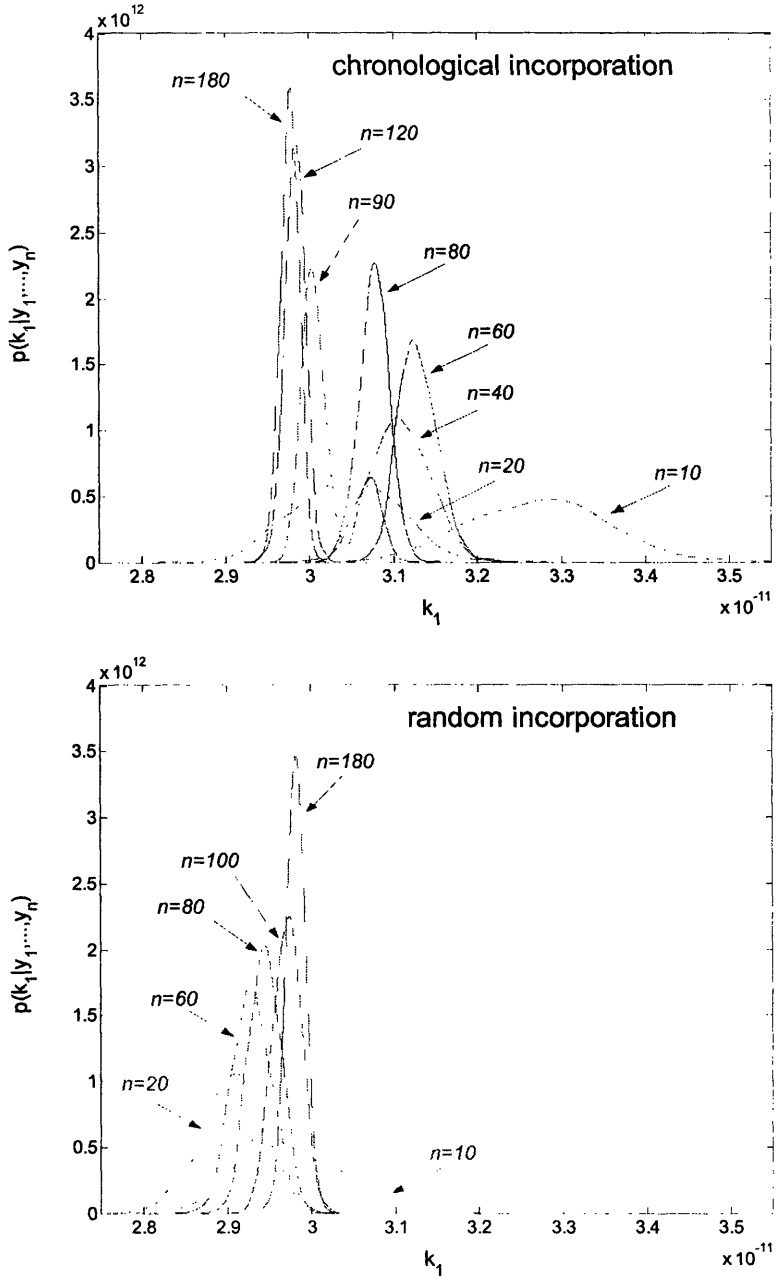


Figure 7-10. Evolving posterior probability distributions upon accumulating information

The results show that the order of updating does not matter for the final estimate: the posterior probability distributions after evaluating all the data are identical according to Desiderata IIIc (see Section 3.2.2), as the information content is equal, independent of the method of updating.

7.7.3 Systematic Error

The assessment of the data in chronological order, preferably in real-time simultaneous with experimentation, is of value as possible systematic errors among experiments can be identified. Analyzing all data at once after finishing experimentation does not easily reveal such information. For example, the updating by random incorporation of the data can be considered equivalent to applying a linear regression on the estimates for parameter B since the time element of the experimentation is not questioned at all. As expected, random incorporation results in a smooth evolution of the probability distribution data sets, because any erratic experimental result will be averaged out.

Shifting of the posterior probability distribution during updating in chronological order can reveal a change in systematic error among the experimental results. For example, in this case the initial data sets (for $n = 1,\ldots,10$) suggest a relative high estimate for the rate parameter. Additionally, the bimodal distribution obtained from incorporating 90 data sets is especially suspicious and implies a significant change occurring after collecting data set 1 through approximately data set 80.

To investigate the statistical significance of these differences, the posterior distribution was generated for data sets 1 through 80 and 81 through 180. The results in Figure 7-11 show that the estimates for rate parameter k_1 are significantly different, leading to the conclusion that a systematic error was involved depending on the different sets of experiments. For verification posterior distributions were also generated from the data sets 1, 3, 5, 7,..., 179 and for 2, 4, 6, 8,..., 180. These posterior distributions are, as expected, not significantly different.

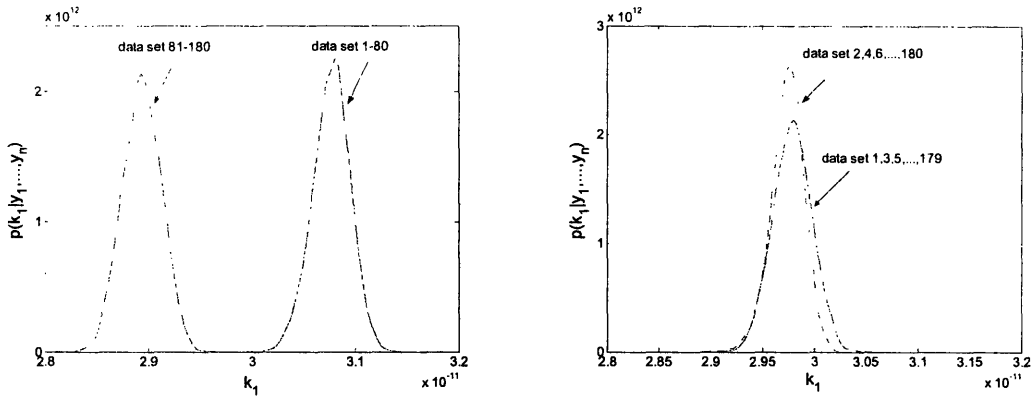


Figure 7-11. Posterior distributions for accumulated information

In an attempt to explain the reason for the significant shift of the estimated value of rate parameter k_1, the laboratory notes provided a possible answer. It appeared after collecting approximately 70 data series, the experimental equipment was dismantled to be moved and rebuilt in a new laboratory. Though it is not clear how exactly this affected the experimental conditions, such an event can obviously have a significant impact.

7.7.4 Stopping Rule

The Bayesian approach can provide insight regarding the necessity for performing future experiments by evaluating the information content gained by incorporating yet another data set. The measure of information content is usually the variance of the probability distribution. A decrease in the variance indicates a decrease in uncertainty, and thus an increase in the information content. By determining the variance of the posterior probability distribution as a function of the number of data sets incorporated when estimating rate parameter k_1, the incremental decrease in the variance can serve as a decision rule to stop experimenting. The results for the variance and mean of the posterior probability distribution of k_1 as a function of the number of data sets evaluated are shown in Figure 7-12. The solid line is obtained by incorporating the data sets in chronological order, the dashed line represents the reverse order, and the dotted line represents a random order.

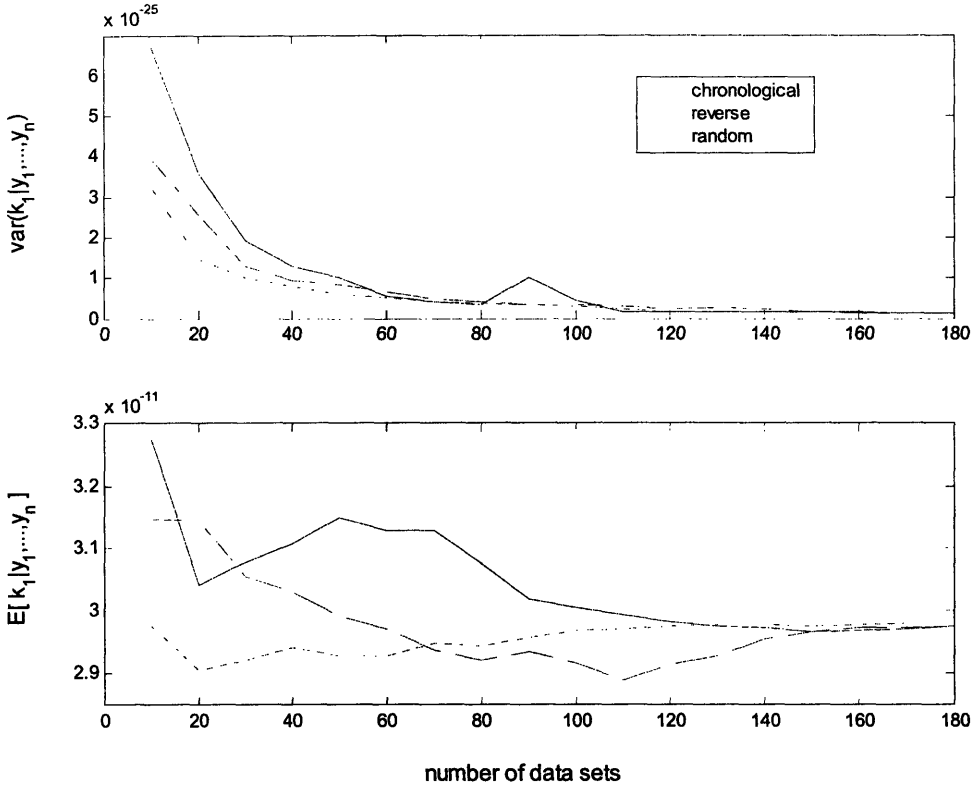


Figure 7-12. Variance and mean of $p(k_1|y_1,...,y_n)$ as a function of the number of data sets

According to these results, approximately 80 experiments would be sufficient to attain, within the framework of the analytical system model, the minimum uncertainty in the parameter estimate. As a side-note, the sudden hump in the variance after incorporating 90 data sets is caused by the bimodal distribution in Figure 7-10. However, approximately 150 experiments are required for the mean to converge to the eventually estimated value. This is still considerably less than the number of experiments carried out in the original work. By applying the Bayesian approach, the experimentalist can save time and effort by being conscious of the marginal increase in information content gained by experimental data.

7.8 The Bayesian Approach using the Kinetic Equations

The Bayesian approach is capable of directly evaluating the model of kinetic rate equations, so that model simplifications under the pseudo steady state assumption are not required. Therefore, the underlying chemistry and physical processes remain evident throughout the problem formulation, because system parameters and variables are not lumped together, as was done to obtain parameters A, B, C, and D.

Since in this case study the pseudo steady state assumption is justified, the analytic model can be considered equivalent to the original set of kinetic rate equations. Therefore, this case study provides an excellent opportunity to compare estimation results obtained by using the kinetic rate equations as the model with the results obtained by using the simplified analytic model.

The estimation of the rate parameters using the model of kinetic rate equations will be demonstrated in the following sections. After outlining the estimation strategy, the problem formulation and required information for estimating each of the parameters k_1, k_2, and k_4 will be discussed. Finally, the parameter estimates for k_1 obtained by the analytic model and the set of kinetic rate equations will be compared

7.8.1 Strategy

At the onset of the problem, the multiple unknown parameters to be estimated are the rate parameters k_1, k_2, and k_4, which were selected as the degrees of freedom of the system. According to equation (7-14) an infinite number of linear combinations of the rate parameter are possible. Therefore, in order to estimate rate parameter k_1 the remaining degrees of freedom first need to be specified by estimating rate parameters k_2 and k_4. The complete estimation strategy is schematically shown in Figure 7-13.

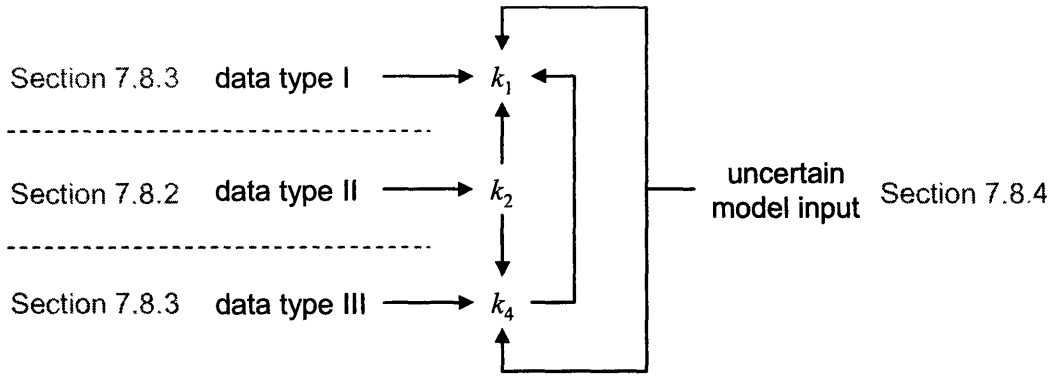


Figure 7-13. Strategy to estimate rate parameter k_1 using the kinetic rate equations as model

Rate parameter k_2 will first be estimated from the data of type II. As described below, direct estimation is unfortunately not possible for rate parameter k_2, and a similar approach as discussed in Section 7.6 was applied. The estimate of k_2 is subsequently used together with the data of type III to obtain the estimate of rate parameter k_4. Finally, the desired rate parameter k_1 is estimated from the data of type I, with the estimates of k_2 and k_4 implemented as known, but uncertain model parameters. The additional model input for $[He]$, $[O_3]$, F, and V are included as known, but uncertain quantities.

7.8.2 Estimation of Rate Parameter k_2

The rate parameter k_2 is estimated from the data of type II (see Section 7.4.1) using the bi-exponential equation (7-30) as the model. The reason the set of kinetic rate equations cannot be used, is that in the absence of N_2 a linear combination of the remaining two degrees of freedom k_2 and k_4 minimizes the objective function, as shown in Figure 7-2. In other words, the decay of $O(^3P)$ from other sources than O_3 cannot be determined separately, so that the required background correction through rate parameter k_4 is unknown. This problem actually arises due to a lack of information. As the data in this case study were originally generated to be used in conjunction with the bi-exponential equation as the model, the data are not suitable for estimating both k_2 and k_4 separately. Therefore, k_2 is estimated as the slope of the linear relationship between parameter B and $[O_3]$ according to equation (7-32) with $[N_2] = 0$. Equivalent to equation (7-43), rate parameter k_2 can be obtained from

$$k_2^{\,i}(n) = \frac{\left(B(n) - B_b(n)\right)_i}{[O_3]_i} \tag{7-45}$$

where $k_2^{\,i}$(n) is the n^{th} sample of rate parameter k_2 obtained from i^{th} data set of type II, and $[O_3]$ is normally distributed as

$$[O_3]_i \sim N\left([O_3]_{m,i}, \left(\tfrac{0.02}{1.96}[O_3]_{m,i}\right)^2\right) \tag{7-46}$$

where $[O_3]_{m,i}$ is the measured ozone concentration in molecules/cm^3, and the variance in equation (7-46) is determined by interpreting the measurement error of ±2% (according to manufacturing specifications) as the 95% confidence interval.

The background parameter B_b is determined as the y-axis intercept of the linear model between parameter B and $[O_3]$. Daily estimates of parameter B_b are obtained from data collected within one day of experiments for a range of $[O_3]$. The background decay of $O(^3P)$ is assumed to remain constant throughout that particular day [61]. As the Bayesian equivalent to linear regression is performed as described in Section 5.5, probability distributions for parameter B_b are obtained to be implemented in equation (7-45).

Finally, rate parameter k_2 is estimated according to the parallel estimation approach discussed in Section 5.4.2. The posterior probability distribution is shown in Figure 7-14. The estimate for rate parameter k_2 obtained with the Bayesian approach is more precise, but has a slightly different mean than the estimate obtained with the conventional estimation method [61]. The comparison for the rate parameter $k_2' = (k_2 + k_3)$ is shown in Table 7-3 below.

Table 7-3. Comparison of summary statistics for the overall rate parameter (k_2+k_3)

	$E[k_2 + k_3]$ (10^{-10} cm^3 molecule^{-1} s^{-1})	$\sigma_{k_2+k_3}$ (10^{-10} cm^3 molecule^{-1} s^{-1})
Bayesian estimate	2.57	0.0036
Conventional estimate	2.45	0.055

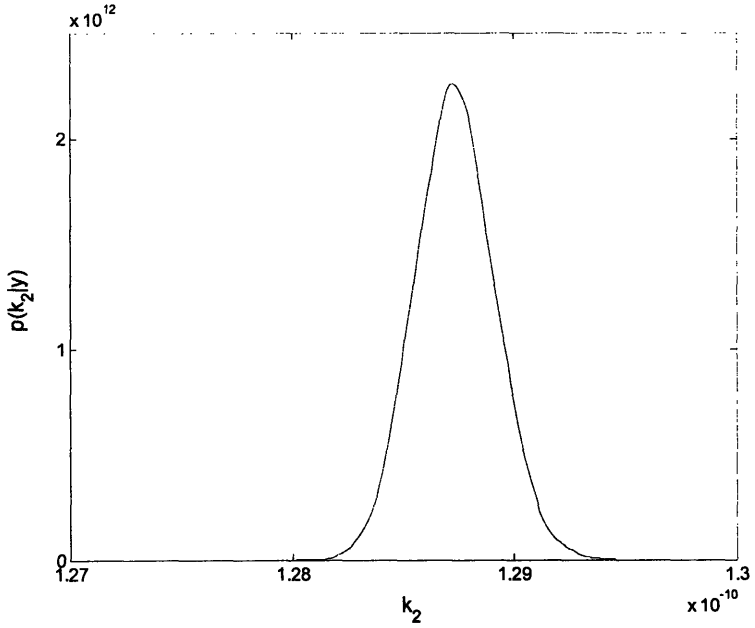


Figure 7-14. Posterior probability distribution for rate parameter k_2

7.8.3 Estimation of Rate Parameter k_1 and k_4

The formulation for estimating either rate parameter k_1 or k_4 is very similar and is summarized in Table 7-4. As the posterior distributions of parameters k_2 and k_4 are used as input for estimating k_1, rate parameter k_4 will be estimated first. The background experiments performed in absence of N_2 account for quenching by both *He* and impurities in the system. As the concentration of impurities will inherently vary among the experiments, individual estimates of k_4 for each background experiment correct the data collected at approximately the same time.

Table 7-4. Formulation of the Bayesian parameter estimation for k_1 and k_4

	Estimation of k_1	**Estimation of k_4**
data	type I	type III
model	equation (7-1) through (7-4)	equation (7-2) through (7-4), since $[N_2] = 0$
required degrees of freedom	$p(k_2), p(k_4)$	$p(k_2)$
parameter vector θ	$\left\{k_1, [O(^3P)]_0, [O(^1D)]_0, \sigma_0^{\,2}\right\}$	$\left\{k_4, [O(^3P)]_0, [O(^1D)]_0, \sigma_0^{\,2}\right\}$
prior $p(\theta)$	$p(k_1)p([O(^1D)]_0)p([O(^3P)]_0)p(\sigma_0^{\,2})$	$p(k_4)p([O(^1D)]_0)p([O(^3P)]_0)p(\sigma_0^{\,2})$
likelihood function $p(\theta\|y)$	$\prod_{t=1}^{p}\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left\{-\frac{\left(y_t - [O(^3P)]_{model,t}\right)^2}{2\sigma_0^{\,2}}\right\}$	
uncertain input	$[N_2]$, $[O_3]$, $[He]$, F, V	

7.8.4 Uncertain Model Input

In addition to $[N_2]$ given in equation (7-44) and $[O_3]$ given in equation (7-46), the remaining uncertain input consists of the helium concentration $[He]$, the total flow through the reaction cell F, and the volume of the detection region V. By including samples of their probability distributions at each step of the MCMC simulation, the uncertainty regarding these model inputs is propagated through the system and will be accounted for in the parameter estimates. As a side-note: this kind of information regarding experimental conditions is an example of what the I represents in the discussion of Bayes' Theorem in Section 3.2.2.

Equation (7-19) is used to calculate a sample for $[He]$ using the uncertain pressure P_i and temperature T_i in the reaction cell, represented by

$$P_i \sim N\left(P_{m,i}, \left(\frac{0.02}{1.96} P_{m,i}\right)^2\right) \quad and \quad T_i \sim N\left(T_{m,i}, \left(\frac{0.02}{1.96} T_{m,i}\right)^2\right) \tag{7-47}$$

where $P_{m,i}$ and $T_{m,i}$ are the measured pressure and temperature during the experiment i, and the variance is determined by considering the ±2% measurement error, as given by the manufacturing specifications, as a 95% confidence intervals.

The remaining variables required to solve the set of rate equations are the volume of the detection region V and the total flow through the reaction cell F. However, the volume V is difficult to obtain as the detection region is suspected not to be constant among the experiments with the fluence of the laser and other experimental conditions being variable. To overcome this difficulty for each experiment i, the parameter D will be applied as follows

$$\frac{F_i}{V_i} = D_i \sim N\left(\hat{D}_i, \hat{\sigma}_{D_i}^2\right) \tag{7-48}$$

where $\hat{D}_i$ and $\hat{\sigma}_{D_i}$ are the mean and variance of the first order decay parameter previously estimated by fitting a first order exponential model to the decaying portion $O(^3P)$ concentration profile, as discussed in Section 7.5.3.

7.8.5 Initial Design and Probing Distribution

The mode of the lognormal prior distribution for k_1 and the mean of the prior distribution of rate parameter k_4 are selected as the initial designs for the rate parameter. For the measurement variance σ_0^2 the initial design is obtained by determining the sample variance of the baseline data for each experiment at $t < 0$. The remaining initial designs for the radical concentrations $[O(^3P)]_0$ and $[O(^1D)]_0$ are however, due to insufficient information regarding the laser fluence, difficult to determine accurately. Therefore, the initial radical concentrations were calculated from equations (7-31) through (7-34) using the previously estimated parameters A, B, C, and D for each experiment.

A multivariate normal probing distribution is used. The covariance matrix is obtained according to the procedure outlined in Section 5.5.5, with a minor adjustment, since standard deviations are not available for the parameters. In this case, the initial design is multiplied by a constant factor of approximately 10-15% to construct an initial diagonal covariance matrix.

7.8.6 MCMC Implementation

Though computation time significantly increases when using the kinetic rate equations (7-1) through (7-4) instead of the analytic equation (7-30) as the model, each MCMC simulation generated 101,000 samples of which 1,000 were discarded as burn-in period. The remaining samples were thinned at a ratio of 1:20 before analyzing the parameter estimates from the 5,000 remaining samples. Each of the approximated posterior distributions for rate parameter k_4 will subsequently be used to correct the appropriate data sets for the estimation of rate parameter k_1.

7.8.7 Estimation Results

In keeping with the discussion on systematic error in Section 7.7.3, the initial 80 data sets were not used for estimation purposes. After removing the previously identified outliers, only 108 data sets remained to estimate the rate parameter k_1.

Combination of the individual estimates from each of these data sets according to the parallel estimation approach as discussed in Section 5.4.2 resulted in the overall estimate for rate parameter k_1. This overall estimate is in Figure 7-15 compared to the estimate resulting from the Bayesian approach using the bi-exponential equation (7-30), as applied to the same 108 data sets. The original estimate determined by nonlinear least squares [61] is also included in the figure as a reference.

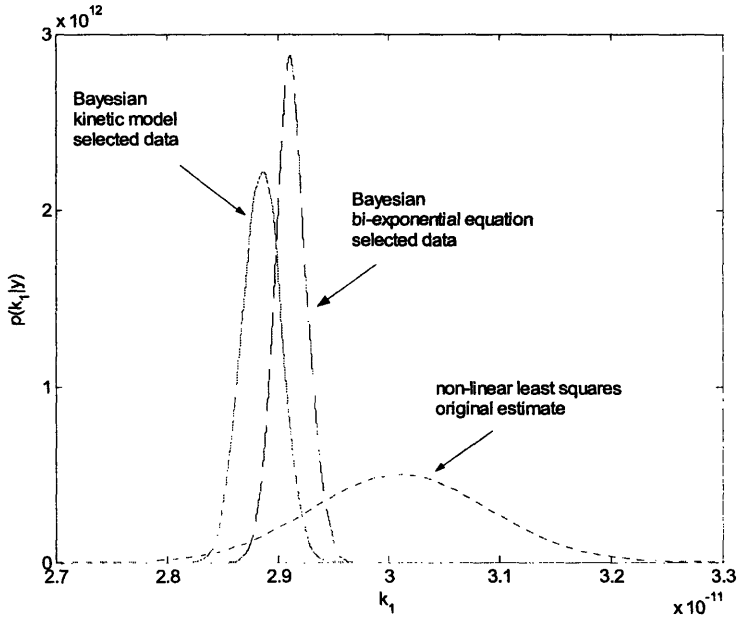


Figure 7-15. Validation of the k_1 estimate obtained with the kinetic model

Compared to the original estimate obtained by nonlinear least squares, the probability has shifted to lower values, as could be expected from Figure 7-11. The variance and mean of the probability distributions for k_1 obtained with the Bayesian approach are approximately equal, as shown in Table 7-5, and a hypothesis test would confirm the

difference is not significant. The outcome that implementation of either model (analytic and kinetic equations) leads to the same resulting parameter estimate, shows that the Bayesian approach is an effective estimation methodology able to directly incorporate the system model without the limitation of model simplification through restricting assumptions. In addition, uncertainty as introduced into the system by the various measured and estimated variables and parameters is propagated systematically, so that the probability distribution for parameter k_1 correctly represents the aggregated uncertainty. A determination of the uncertainty after the parameter estimation, as discussed in Section 7.5.5, is therefore not required.

Table 7-5. Comparison of summary statistics for the overall rate parameter k_1

Bayesian estimate	$E[k_1]$ (10^{-11} cm^3 molecule^{-1} s^{-1})	σ_{k_1} (10^{-11} cm^3 molecule^{-1} s^{-1})
kinetic equations	2.89	0.014
bi-exponential equation	2.91	0.017

Figure 7-16 compares the evolution of the variance as a function of the accumulation of information for both models by incorporating the number of data sets in a random order. The variance of the estimate behaves in a similar manner for both models and reaches a minimum value after incorporating approximately 80 data sets.

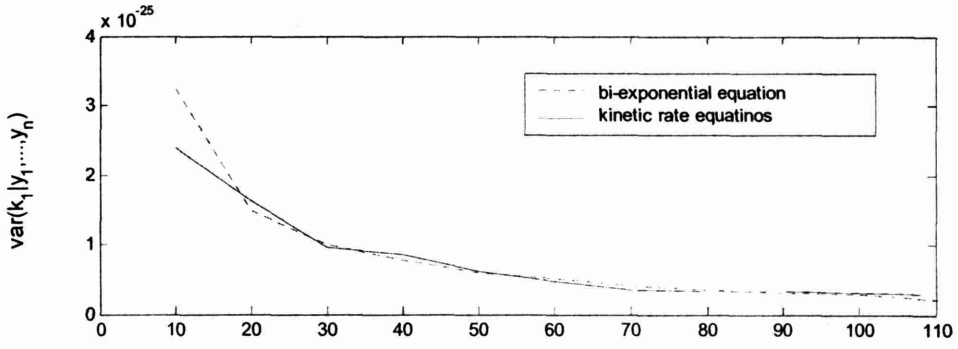


Figure 7-16. Evolving variance of $p(k_1|y_1,\ldots,y_n)$ for accumulating information

7.9 Summary of Key Points

The various features of using the Bayesian approach in parameter estimation discussed in this case study will be summarized in a few key points below.

- The Bayesian approach leads to a more precise estimate for the rate parameter k_1. In other words, more information can be obtained from the available data.
- Outliers can be identified by comparing the probability distributions obtained from each individual data set.
- Updating the available knowledge of the parameter estimate leads to an evolving posterior probability distribution that indicates the presence of a systematic error. The subsequent evaluation of subsets of the data can confirm the presence of a systematic error.
- The evolving posterior probability distribution allows for easy evaluation of the value of information as defined by the variance decrease per experiment.
- The diminishing decrease in variance of the posterior distribution resulting from incorporating more data can be the basis to formulating a stopping rule.
- The Bayesian approach is a flexible parameter estimation method that can be implemented with either the original kinetic rate equations, or the bi-exponential equation as the system model.
- The implementation of the original kinetic equations has the advantage that the underlying physics of the system remain clear. However, since the system variables and parameters are not lumped together, more information is generally needed to specify the system.

8

Case Study 2

Reversible Addition of Oxygen to Cyclohexadienyl Radicals in Cyclohexane

I've got to admit it's getting better.
It's a little better all the time.

The Beatles

The purpose of this case study is to demonstrate a second application of the Bayesian approach, illustrating the versatility and flexibility in exploring scenarios and estimating various parameters and model inputs. The system under investigation is the liquid phase reaction of resonantly stabilized cyclohexadienyl radicals with molecular oxygen. After a detailed description and analysis of this system, the Bayesian approach is compared to the method of Global Dynamic Optimization in estimating kinetic rate parameters. Providing information on parameter uncertainty, the Bayesian approach incited further investigation to explain suspicious estimation results, leading to an improved understanding of the system. The second advantage was the ability to simultaneously evaluate multiple data sets to attain more informative estimation results.

8.1 Introduction

This chapter is based on the research by Taylor *et al.* [69], who studied the title reaction in the liquid phase. Taylor *et al.* proposed an equilibrium reaction mechanism, based on experimental results and quantum mechanical calculations, consistent in both the liquid

and gas phases. The differences in equilibrium reactions in the liquid and gas phases were considered responsible for the liquid phase reaction rate being approximately two orders of magnitude larger than the gas phase reaction rate. The kinetic rate parameters as reported by Taylor *et al.* were based on literature data and quantum mechanical calculations, and verified by a comparison of model simulation results with experimental data.

Using Taylor's experimental data and proposed reaction mechanism, Singer [70] estimated the reaction rate parameters applying the Global Dynamic Optimization method, a deterministic algorithm able to optimize non-convex integral objective functions. Though this method guarantees to find the global optimum of an objective function, the results do not provide any information regarding the uncertainty of the parameter estimates.

The Bayesian approach can be useful to investigate this system to obtain information regarding the uncertainty of the parameters, and because the information of multiple data sets need to be combined. Additionally, with little effort alternative scenarios for the reaction mechanism can be compared, assumptions can easily be altered, and a variety of model parameters and variables can be included into the estimation problem.

8.2 System Description

Again according to Section 6.2, this section will describe the system in as much detail as possible in order to gain understanding regarding the system under investigation. The proposed reaction mechanism will first be discussed, followed by the kinetic model with constraints, and finally the experimental procedure will be briefly described.

8.2.1 Reaction Mechanism

The experiments are initiated by the photolysis of di-*t*-butyl peroxide, which can be considered to form instantaneously by the excimer laser pulse. The generated radicals

react with 1,4-cyclohexadiene, which is present in excess, generating the desired cyclohexadienyl radicals according to

$$(CH_3)_3CO\bullet + 1,4C_6H_8 \xrightarrow{k_1} C_6H_7\bullet + CH_3COH \tag{8-1}$$

The proposed reaction mechanism for the subsequent addition of oxygen to the cyclohexadienyl radical in non-polar solvents is as follows

$$C_6H_7\bullet + O_2 \underset{k_{2r}}{\overset{k_{2f}}{\rightleftarrows}} p-C_6H_7OO\bullet \tag{8-2}$$

$$C_6H_7\bullet + O_2 \underset{k_{3r}}{\overset{k_{3f}}{\rightleftarrows}} o-C_6H_7OO\bullet \tag{8-3}$$

$$o-C_6H_7OO\bullet \xrightarrow{k_4} C_6H_6 + HO_2\bullet \tag{8-4}$$

$$2\,C_6H_7\bullet \xrightarrow{k_5} products \tag{8-5}$$

An alternative reaction decay path for the cyclohexadienyl radicals is a direct hydrogen abstraction by oxygen

$$C_6H_7\bullet + O_2 \xrightarrow{k_6} C_6H_6 + HO_2\bullet \tag{8-6}$$

though this reaction is not considered in this parameter estimation case study.

8.2.2 Kinetic Model

For notational convenience, the following substitutions are introduced

$$\begin{aligned} C_6H_7\bullet &\Leftrightarrow A \\ (CH_3)_3CO\bullet &\Leftrightarrow Z \\ 1,4C_6H_8 &\Leftrightarrow Y \\ o-C_6H_7OO\bullet &\Leftrightarrow B \\ p-C_6H_7OO\bullet &\Leftrightarrow D \end{aligned}$$

so that the proposed mechanism as given in equations (8-1) through (8-5) can be described by the following set of kinetic rate equations

$$\frac{dx_A}{dt} = k_1 x_Z x_Y - (k_{2f} + k_{3f}) x_{O_2} x_A + k_{2r} x_D + k_{3r} x_B - k_5 x_A^2 \tag{8-7}$$

$$\frac{dx_Z}{dt} = -k_1 x_Z x_Y \tag{8-8}$$

$$\frac{dx_Y}{dt} = -k_1 x_Z x_Y \tag{8-9}$$

$$\frac{dx_B}{dt} = k_{3f} x_{O_2} x_A - (k_{3r} + k_4) x_B \tag{8-10}$$

$$\frac{dx_D}{dt} = k_{2f} x_{O_2} x_A - k_{2r} x_D \tag{8-11}$$

The system of kinetic rate equations is completed by the initial conditions

$$x_A(t=0) = 0,\ x_Z(t=0) = x_{Z,0},\ x_Y(t=0) = x_{Y,0},\ x_B(t=0) = 0,\ x_D(t=0) = 0 \tag{8-12}$$

where $x_{z,0}$ and $x_{Y,0}$ are determined for each experiment.

8.2.3 Physical Constraints and Relationships

Because of the chemical equilibrium, the forward and reverse reaction rate parameters for equations (8-2) and (8-3) are related according to

$$k_{2r} = \frac{k_{2f}}{K_2} \qquad \text{and} \qquad k_{3r} = \frac{k_{3f}}{K_3} \tag{8-13}$$

where K_2 and K_3 are equilibrium constants, which can be approximated from thermodynamic properties. Taylor [69] calculates K_2 = 2100, though the uncertainty on K_2 is suspected to be about one order of magnitude. Taking into account this significant

uncertainty would surely hinder a successful outcome of the parameter estimation. To complicate matters, thermodynamic properties provide arguments for

$$K_3 = K_2 \qquad \text{or} \qquad K_3 = 2K_2 \tag{8-14}$$

without indicating a clear preference for either relationship. However, as the system appeared to be relatively insensitive to K_2, as demonstrated later in Figure 8-3, implementing the calculated value for K_2 and either of the equalities in (8-14) is considered acceptable.

Relationships between the forward reaction rate parameters are based on physical arguments. As the reaction of cyclohexadienyl with oxygen in the liquid phase is found to be diffusion limited, the overall reaction rate parameter k_f' should be equal to the diffusion parameter, thus

$$k_f' = k_{2f} + k_{3f} \approx D = 1200 M^{-1}\mu s^{-1} \tag{8-15}$$

where the diffusion limit D was experimentally obtained [69].

An alternative assumption regarding the forward rate constants is

$$k_{3f} = 2k_{2f} \tag{8-16}$$

which is based on the statistical argument that two enantiomers exist for $o\text{-}C_6H_7OO\cdot$, so that its formation is considered to occur twice as likely as the formation of $p\text{-}C_6H_7OO\cdot$. However, it is unknown whether kinetic effects of the molecular structures of the ortho- and para-species affect the 1:2 ratio as presented in equation (8-16).

Finally, the parameters k_1 and k_5 are obtained from literature. Rate parameter k_1 is actually determined in benzene [71] as

$$k_1 = 53\ M^{-1}\mu s^{-1} \tag{8-17}$$

but since formation of $C_6H_7\cdot$ according to equation (8-1) is not a diffusion limited reaction, the effect of using the structurally similar cyclohexane as solvent is considered negligible.

Rate parameter k_5 was originally also determined in benzene [72]. However, the recombination reaction in equation (8-5) is diffusion limited and thus the rate parameter can be adjusted for the viscosity of cyclohexane according to the Stokes-Einstein equation. This leads to

$$k_5 = 1200\ M^{-1}\mu s^{-1} \tag{8-18}$$

which confirms the diffusion limit for the reaction of $C_6H_7\cdot$ discussed above. Due to the lack of information on the uncertainty in both of these parameters, their values will be implemented as being exact. This should be kept in mind, as the uncertainties of the resulting parameter estimates will appear smaller than they actually are.

8.2.4 Experimental Procedure

This section will only discuss the key issues, and for further experimental details the reader is referred to Taylor *et al.* [69].

The reaction took place in a cuvet with the reaction mixture continuously flowing through in order to flush out the reaction products and prepare for the next excimer laser pulse. Contrary to the gas phase experiments discussed in the previous chapter, the outflow of reactants during the reaction was negligible in this case as the reaction rate timescale is several orders of magnitude larger than the residence time of the reactant mixture in the cuvet.

The radical $(CH_3)_3CO\cdot$ is generated instantaneously by the excimer laser pulse and initiates the reaction mechanism. The cyclohexadienyl radical $C_6H_7\cdot$ generated is detected with UV absorption spectrometry measuring the absorption band at 316 nm.

The measured optical density $I_{M,t}$ at time t can be related to the $C_6H_7\cdot$ radical concentration as

$$I_{M,t} = 2100[C_6H_7\cdot]_t + 200([o\text{-}C_6H_7OO\cdot]_t + [p\text{-}C_6H_7OO\cdot]_t) \qquad (8\text{-}19)$$

assuming that the *o*- and *p*-cyclohexadienylperoxyl radicals (species *B* and *D*, respectively) are partially responsible for the measured absorption. The recorded data was the result of averaging 30 successive reactions, in between which the reaction mixture was completely refreshed.

8.3 System Analysis

Similar to Section 7.3, this section will thoroughly analyze the system to gain an understanding of the chemistry and physics. The rate parameters k_{2f}, k_{3f}, and k_4 used for these analyses are specified by Taylor *et al.* as

$$k_{2f} = 400\ M^{-1}\mu s^{-1},\ k_{3f} = 800\ M^{-1}\mu s^{-1},\ k_4 = 0.8\ M^{-1}\mu s^{-1} \qquad (8\text{-}20)$$

calculated from the overall rate parameter $k_f' = (k_{2f} + k_{3f})$ and the assumption in equation (8-16). The estimate for rate parameter k_4 is a lower bound for the rate parameter obtained from the literature.

Table 8-1. Incidence matrix for the system of case study 2

	$\frac{dx_1}{dt}$	$\frac{dx_2}{dt}$	$\frac{dx_3}{dt}$	$\frac{dx_B}{dt}$	$\frac{dx_D}{dt}$	x_A	x_Z	x_Y	x_B	x_D	k_{2f}	k_{3f}	k_4	k_{2r}	k_{3r}	I
(8-7)	×					×	×	×	×	×	×	×		×	×	
(8-8)		×					×	×								
(8-9)			×				×	×								
(8-10)				×		×			×			×	×		×	
(8-11)					×	×				×	×			×		
(8-12)a						×										
(8-12)b							×									
(8-12)c								×								
(8-12)d									×							
(8-12)e										×						
(8-13)a											×			×		
(8-13)b												×			×	
(8-19)						×			×	×						×

8.3.1 Degrees of Freedom

According to the dimensions of the incidence matrix for the system as shown in Table 8-1, 13 equations in the rows and 16 variables in the columns, the system has three degrees of freedom. The output set assignment identifies the set of parameters requiring further specification, namely $\{k_{2f}, k_{3f}, k_4\}$.

8.3.2 Observability

After linearization the time invariant matrix can be constructed as

$$\mathbf{A} = \begin{bmatrix} -(k_{2f}+k_{3f})x_{O_2} - 2k_5 x_A^* & k_1 x_Y^* & k_1 x_Z^* & \frac{k_{3f}}{K_3} & \frac{k_{2f}}{K_2} \\ 0 & k_1 x_Y^* & k_1 x_Z^* & 0 & 0 \\ 0 & k_1 x_Y^* & k_1 x_Z^* & 0 & 0 \\ k_{3f} x_{O_2} & 0 & 0 & -\left(\frac{k_{3f}}{K_3} + k_4\right) & 0 \\ k_{2f} x_{O_2} & 0 & 0 & 0 & -\left(\frac{k_{2f}}{K_2}\right) \end{bmatrix} \quad (8\text{-}21)$$

where x_A^*, x_Y^*, and x_Z^* are the concentrations around which the system is linearized. The output matrix follows from equation (8-19) and is defined as

$$\mathbf{C} = [2100 \quad 0 \quad 0 \quad 200 \quad 200] \quad (8\text{-}22)$$

The observability matrix **Q** was constructed with Maple according to equation (6-5) (the observability matrix is too complex to display here). Gaussian elimination performed on this matrix **Q** led to a full rank matrix, meaning that this system is completely observable.

8.3.3 Concentration Profiles

In addition to the rate parameters specified above, the initial concentrations of the radical $(CH_3)_3CO\cdot$ and the species $1,4C_6H_8$ are required to solve the set of kinetic rate equations. These initial concentrations are specified according to the planned experiments as $[1,4C_6H_8]_0 = 0.4\ M$ and $[(CH_3)_3CO\cdot] = 1.4\cdot10^{-4}\ M$, where the latter is determined by a similar calculation as equation (7-20) based on the laser fluence. With the system completely specified, the concentration profiles are determined for a reaction time of $4.5\mu s$ as shown in Figure 8-1.

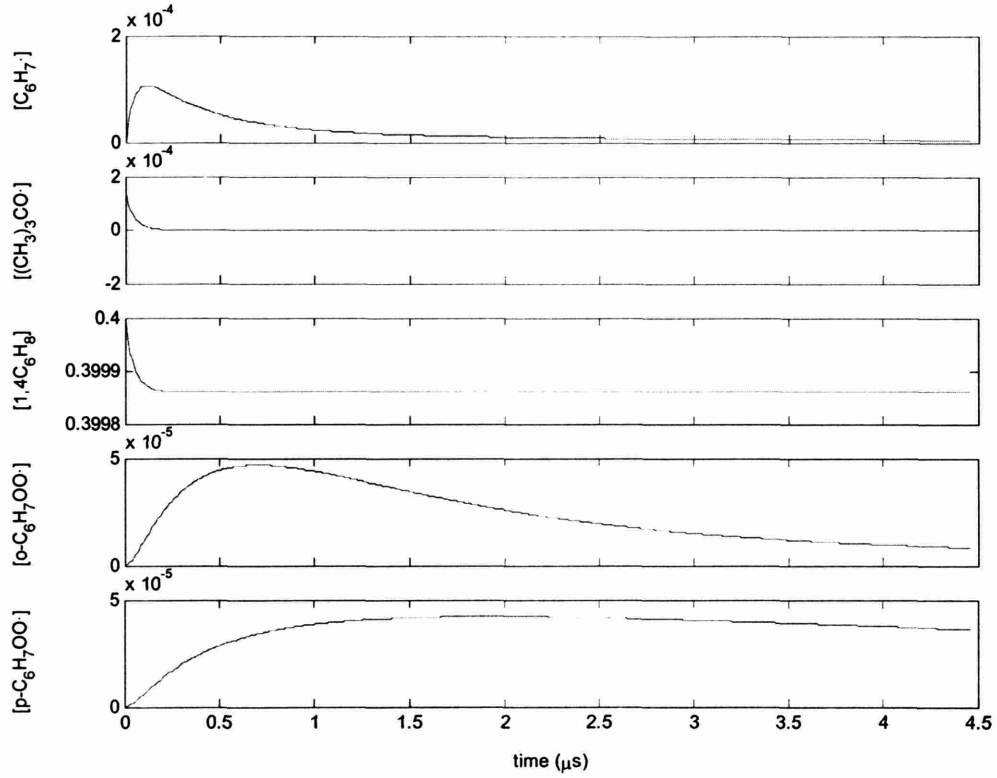


Figure 8-1. Concentration profiles

The initial raise in $[C_6H_7\cdot]$ is clearly dependent on the initial concentration of $(CH_3)_3CO\cdot$, so $[(CH_3)_3CO\cdot]_0$ should be known with a specified uncertainty. The concentration of $1,4C_6H_8$ does not change significantly during the experiment and could be considered constant. Finally, the effect of the reaction equation (8-4) is visible because of the faster decay of $o\text{-}C_6H_7OO\cdot$ compared to $p\text{-}C_6H_7OO\cdot$.

8.3.4 Rate of Reactions

The reaction rates of the individual forward reactions are shown in Figure 8-2. The forward reactions are of similar order of magnitude, where reaction 3 is about twice as fast as reaction 2, which can be expected from the specified rate parameters in (8-20). The reverse reactions rates (not shown) are several orders of magnitude lower than the forward reactions, as is understandable when $K_2 = 2100$ and implementing equation (8-13) and (8-14).

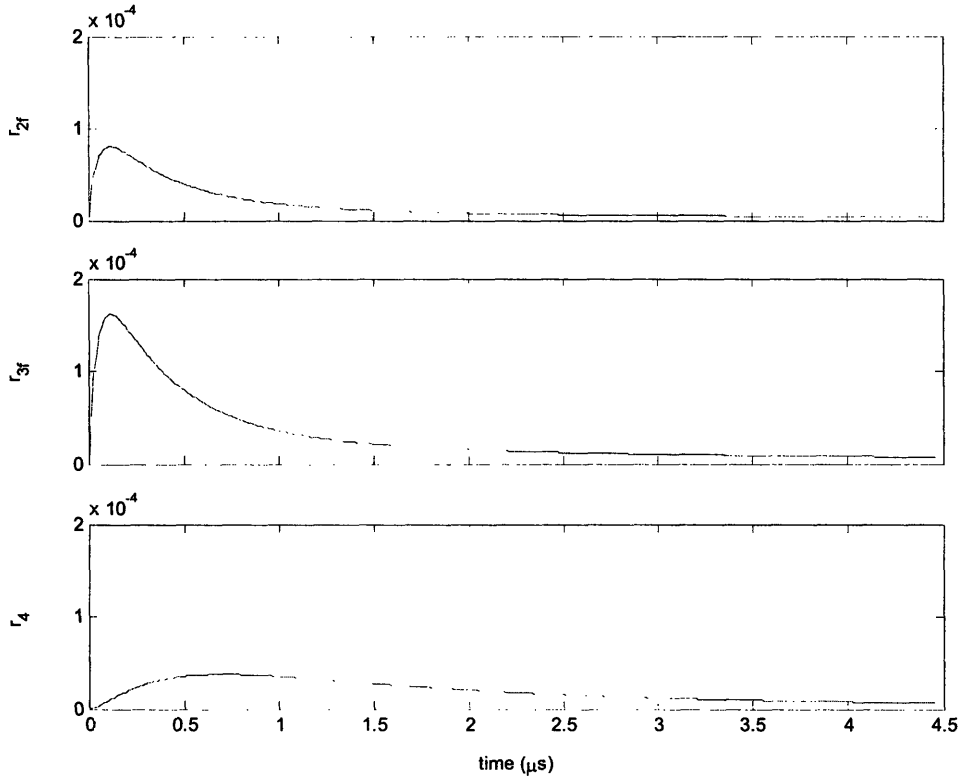


Figure 8-2. Forward reaction rates

8.3.5 Sensitivity Analysis

The sensitivity of the $C_6H_7\cdot$ radical concentration to the rate parameters k_{2f}, k_{3f}, and k_4 and equilibrium constant K_2 is shown in Figure 8-3. This sensitivity analysis incorporates the

assumption that K_2 and K_3 are equal and that the reverse rate parameters have been substituted according to equation (8-13).

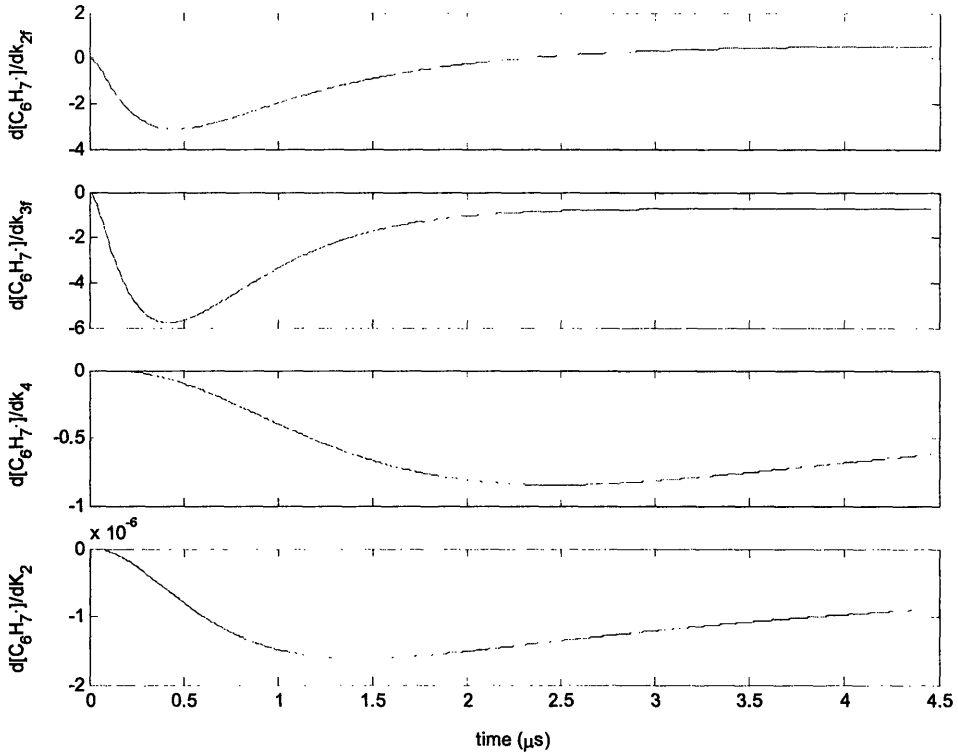


Figure 8-3. Sensitivity of $[C_6H_7\cdot]$ to the rate parameters and equilibrium constant K_2

The negative dependence of $[C_6H_7\cdot]$ on each of the rate parameters and equilibrium constant is understandable from the reaction mechanism. Only near the end of the reaction, the dependence on rate parameter k_{2f} becomes positive. This can be explained by the fact that by then the species $p\text{-}C_6H_7OO\cdot$ becomes a significant source for the radical $C_6H_7\cdot$. A final observation from the sensitivity analysis is that, compared to the rate parameters, equilibrium constant K_2 has a negligible effect, so that disregarding its uncertainty for the parameter estimation is reasonably acceptable.

8.4 Experimentation

The experimental data were collected at 298*K* for five different oxygen concentrations. Typical profiles of the measured optical density I_E show remarkably little noise compared to the data discussed in Chapter 7, partially because each data point is the average of 30 measurements. The data shown in Figure 8-4 originate from two experiments performed under identical conditions at $[O_2] = 0.0019M$. The reason for the significantly different profiles is probably a different $[(CH_3)_3CO\cdot]_0$ caused by the variation in the laser fluence. The laser fluence was observed to vary approximately ±10% among experiments. This knowledge would justify considering $[(CH_3)_3CO\cdot]_0$ as one of the variables in the parameter estimation.

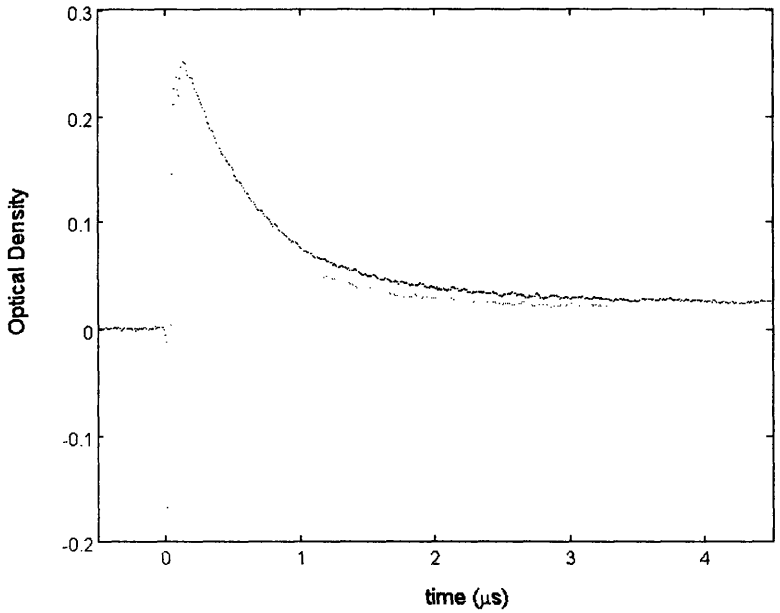


Figure 8-4. Measurements of optical density during reaction time

For one of the data sets displayed, the baseline measurements before the start of the experiment at $t < 0$ are also included. This data set shows that immediately after the excimer laser pulse at $t = 0$, the measured optical density exhibits a sharp negative spike, before initiating the profile resembling the model simulation shown in Figure 8-1. Since the optical density I_E is a measure for the absorbance, these negative values would imply

the creation of energy. The explanation of this phenomenon is that fluorescence of various excited species in the reaction mixture interferes with the actual signal of interest. The first positive optical density measurement occurs therefore at $t = 0.05\mu s$, though the actual system has been reacting since $t = 0$, after which a sharp rise only allows for a few data points until the maximum in the curve. The effect of the fluorescence is expected to vanish around the maximum in the measured optical density, these initial data points are therefore considered suspicious. Though intuitively the complete data set from $t = 0.05\mu s$ would be included for parameter estimation, the knowledge regarding the initial stages of the reaction justifies an alternative approach of only using the data points after reaching the maximum optical density.

Towards the end of the experiment the optical density does not return to the baseline, even though, according to Figure 8-1, $[C_6H_7\cdot]$ appears to have reacted almost completely. The baseline offset exists because of the minor absorption of $o\text{-}C_6H_7OO\cdot$ and $p\text{-}C_6H_7OO\cdot$, which has been taken into account in equation (8-19).

8.5 Bayesian Parameter Estimation

The Bayesian approach to parameter estimation for this system will be described in this section. Starting with a brief overview of the general Bayesian formulation, the Bayesian approach will be compared with the method of Global Dynamic Optimization [70]. The remainder of this section will focus on the flexibility of the Bayesian approach in parameter estimation, and its application in this case study in implementing and analyzing several scenarios.

8.5.1 Problem Formulation

The Bayesian formulation is similar to the specifications in Chapter 7 and will therefore only be discussed briefly. For this case study, Bayes' theorem is implemented as

$$p(k_{2f}, k_{3f}, k_4, \sigma_0^2 \mid I_E) \propto p(k_{2f}, k_{3f}, k_4, \sigma_0^2)\, p(I_E \mid k_{2f}, k_{3f}, k_4, \sigma_0^2) \qquad (8\text{-}23)$$

where I_E represents the data, and σ_0^2 the unknown variance of the measurement error, as before. The parameters are assumed to be independent of each other, so that the prior distribution can be obtained by multiplying the individual prior distributions of the parameters. Uniform distributions, as defined in equation (5-8), are used as prior distribution for the rate parameters k_{2f}, k_{3f}, and k_4 and the prior distribution for the inverse of the variance σ_0^2 is again the Gamma distribution of equation (5-10). Because the data points are averages of 30 separate measurements, the likelihood function can be considered to be a normal distribution defined as

$$p(I_E \mid k_{2f}, k_{3f}, k_4, {\sigma_0}^2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{ \frac{-\left(I_{E,t} - I_{M,t} \right)^2}{2{\sigma_0}^2} \right\} \tag{8-24}$$

where the calculation of $I_{M,t}$ will depend on the particular scenario under investigation. The implementation of MCMC to solve the Bayesian formulation for this case study is analogous to the procedure described before and will therefore not be repeated.

8.5.2 Comparison with Global Dynamic Optimization

Global Dynamic Optimization [70] finds the optimal values for the rate parameters by minimizing the objective function J defined as

$$J = \sum_{t=1}^{n} (I_{E,t} - I_{M,t})^2 \tag{8-25}$$

where $I_{E,t}$ is the experimentally determined optical density, and $I_{M,t}$ is calculated with equation (8-19) after solving the set of kinetic rate equations (8-7) through (8-11) for measurement time t, and n is the total number of data points for an experiment.

Based on the data set obtained at $T = 298K$ under the following conditions

$$[O_2] = 0.0019M$$
$$[(CH_3)_3CO\bullet]_0 = 1.4 \cdot 10^{-4}M \quad (8\text{-}26)$$
$$[1,4C_6H_8]_0 = 0.4M$$

and implementing $K_3 = 2K_2 = 4200$, the expected values of the resulting Bayesian parameter estimates for k_{2f}, k_{3f}, and k_4 are approximately equal to the estimates obtained from Global Dynamic Optimization, as shown in Table 8-2.

Table 8-2. Estimation results of Bayesian approach compared to Global Dynamic Optimization

	Bayesian		Global Dynamic Optimization
	$E[k_i]$	σ_{k_i}	k_i
k_{2f}	530.9 $M^{-1}\mu s^{-1}$	13.7	529.5 $M^{-1}\mu s^{-1}$
k_{3f}	403.4 $M^{-1}\mu s^{-1}$	8.1	401.1 $M^{-1}\mu s^{-1}$
k_4	22.2 μs^{-1}	9.7	24.5 μs^{-1}

Global Dynamic Optimization has the advantage that, if convergence occurs, the global optimum can be guaranteed. However, as Global Dynamic Optimization provides point estimates for a particular data set, information regarding the uncertainty of the estimate is not available. Therefore, accumulating knowledge obtained from different experiments becomes difficult.

8.5.3 Detailed Bayesian Estimation Results

The first remarkable insight provided by the Bayesian approach is that the standard deviation for the estimate of rate parameter k_4 seems disproportionally large. This observation is further inspected by displaying the simulation trace plots generated by MCMC. The trace plots shown in Figure 8-5 display the rate parameters k_{2f} and k_{3f} converging to a relatively tightly bounded area, while the sampling of k_4 spans mostly the range of $5\ \mu s^{-1} < k_4 < 40\ \mu s^{-1}$.

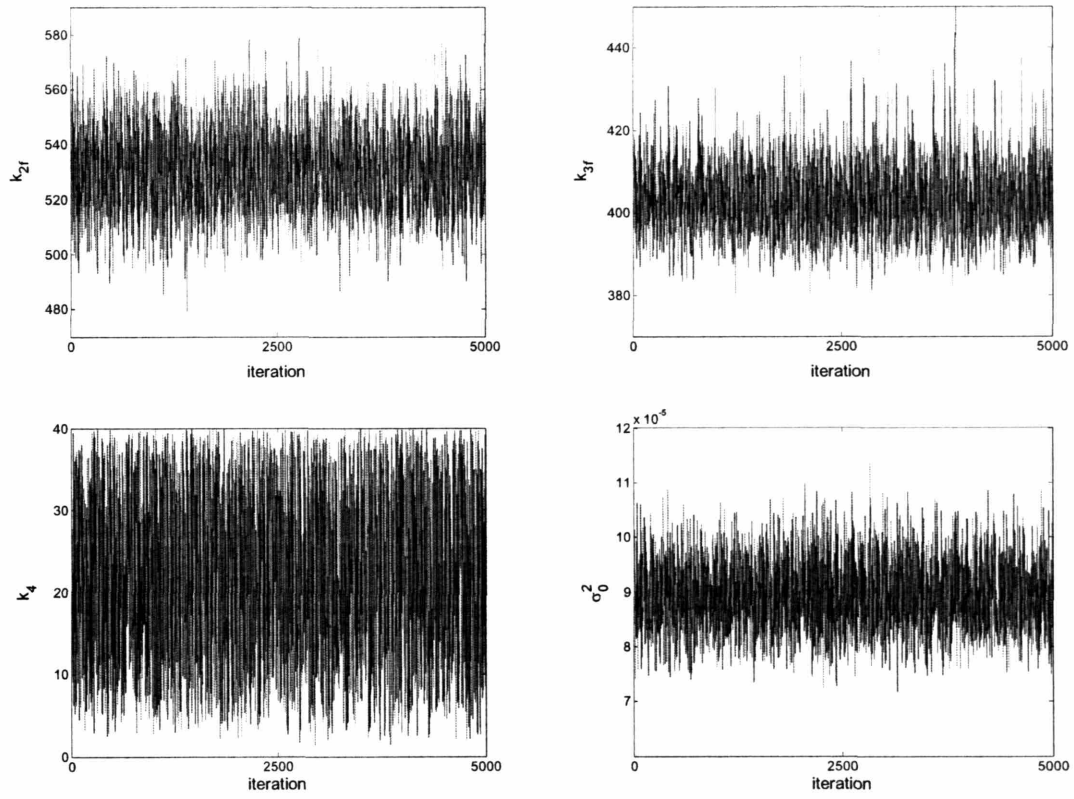


Figure 8-5. MCMC trace plots for the Global Dynamic Optimization scenario

The marginal posterior distributions for the rate parameters, determined as a kernel density estimate of the sample sets, are shown in Figure 8-6. The marginal posterior distribution for rate parameter k_4 is somewhat skewed to larger parameter values, but is too wide to provide a precise estimate.

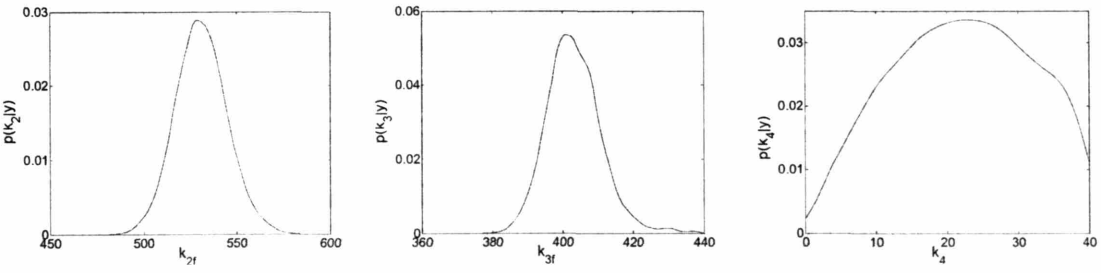


Figure 8-6. Marginal posterior distributions for the Global Dynamic Optimization scenario

Expanding the bounds for rate parameter k_4 only led to the sampling of k_4 spanning the wider range, while the rate parameters k_{2f} and k_{3f} were not significantly affected. The interpretation of this behavior is that this particular system, as defined in Section 8.5.2, is relatively insensitive to the rate parameter k_4. This insensitivity was also noted by Singer [70] when applying Global Dynamic Optimization to estimate the model parameters. To verify this statement, the following sections will analyze this system in more detail.

8.5.4 Parameter Estimates Explained: Objective Function

The objective function J is calculated as a function of rate parameter k_4, as shown in Figure 8-8, keeping the k_{2f} and k_{3f} constant at the values estimated by Global Dynamic Optimization as given Table 8-2 above.

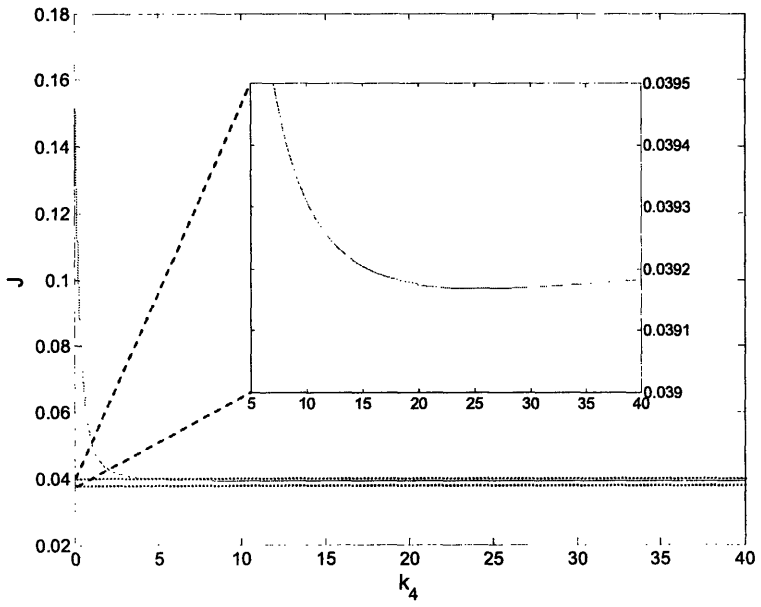


Figure 8-7. Objective function as a function of k_4 for k_{2f}=401.0 and k_{3f}=529.5

As the objective function is virtually horizontal for approximately $k_4 > 5\ \mu s^{-1}$, the estimation of k_4 is not straightforward. Only after significant enlargement, a slight minimum can be observed. To place the objective function of k_4 in perspective, its range was expanded beyond the defined bounds of $[10^{-3}\ \mu s^{-1}, 40\ \mu s^{-1}]$ and plotted in conjunction with the objective functions dependent on rate parameter k_{2f} and k_{3f} respectively, in

Figure 8-8. While the objective functions for rate parameters k_{2f} and k_{3f} exhibit a sharp, clearly defined minimum, the objective function for k_4 decreases sharply in the range $0 < k_4 < 5\ \mu s^{-1}$, after which an almost unnoticeable well is followed by a virtually flat plateau at a higher level. The difference between the minimum of the well and the plateau level is approximately 10% of the uncertainty in the objective function, which was determined by the Global Dynamic Optimization method as $J = 0.039 \pm 0.001$.

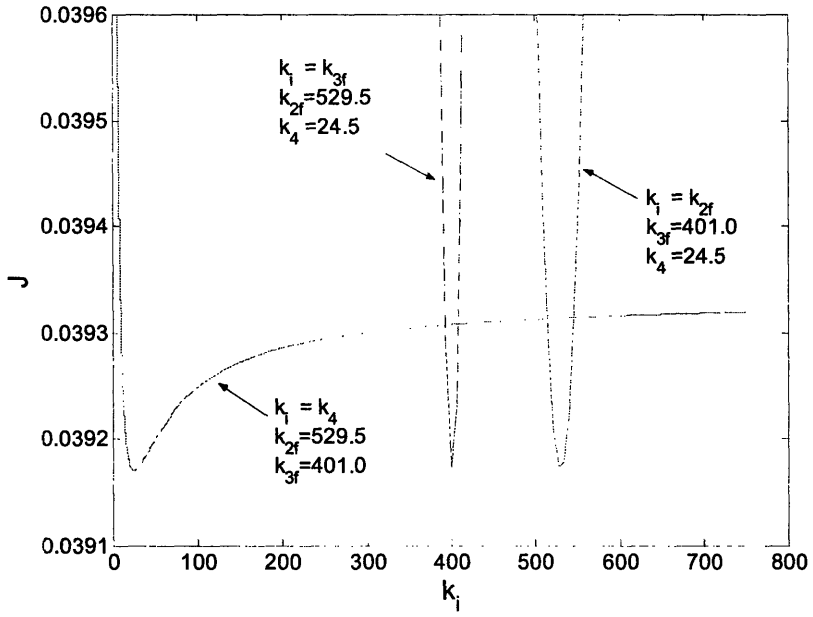


Figure 8-8. Objective functions for the scenario of Global Dynamic Optimization

The point estimate for k_4 obtained by Global Dynamic Optimization is equal to the minimum of the well, though the uncertainty in the minimum value for J would justify a larger range of possible parameter values. The Bayesian approach on the other hand conveyed the message that any value for rate parameter k_4 would suffice as long as approximately $k_4 > 5\ \mu s^{-1}$. Such insights are valuable because it motivates an investigation whether the estimated value for k_4 is relevant in relation to the physical behavior of the system. This issue will be addressed in the following section.

8.5.5 Parameter Estimates Explained: Model Fit

The effect of rate parameter k_4 on the system behavior can be further analyzed from the fit of the model to the data. Figure 8-9 shows the model outcomes for various values of rate parameter k_4, while fixing rate parameters k_{2f} and k_{3f} to their estimated values given in Table 8-2. Consistent with the interpretation of the MCMC trace plots, the system appears insensitive to the actual value of k_4 as long as $k_4 > 5\ \mu s^{-1}$.

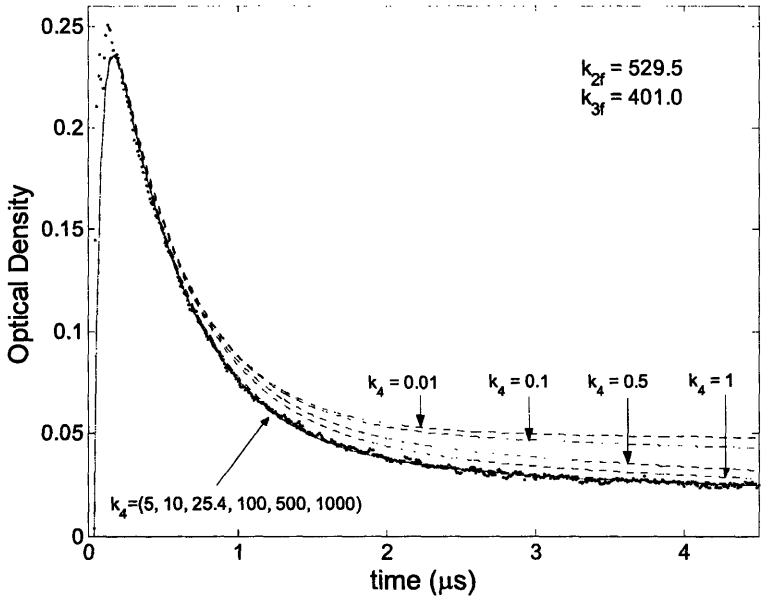


Figure 8-9. Model fit for various values of k_4 with k_{2f} and k_{3f} fixed

Thus, though a particular value has been estimated for rate parameter k_4, further investigation motivated by the Bayesian estimation results lead to a doubt regarding the physical significance of this point-estimate. Since the point-estimate for k_4 was obtained under specific model assumptions, the next step will be to relax these assumptions and consider various scenarios for the system model used in the parameter estimation.

8.6 Advantages of the Bayesian Approach

The parameter estimation above was performed according to particular assumptions regarding the system model and constraints. In many cases there is however no clarity

about either model or constraints, or both. The Bayesian approach allows for relatively easy testing of different scenarios, as described and demonstrated in this section. The implementation of different scenarios will first be described, followed by an example involving the adjustment of a constraint.

8.6.1 Convenient Scenario Testing

Various scenarios, for example to consider different constraints, are easy to implement in the Bayesian problem formulation. Within the MCMC framework, the model and constraints are incorporated when calculating the posterior probability density. To illustrate the relative ease of evaluating different constraints or variables, consider for example the parameter vector

$$\boldsymbol{\theta} = \left\{k_{2f},\ k_4,\ [(CH_3)_3CO\bullet]_0,\ \sigma_0^{\,2}\right\} \tag{8-27}$$

which includes an initial concentration as a variable to estimate and does not consider k_{3f} by implementing equation (8-16) in the Matlab function 'getposterior' as presented in Section A.1. Within the backbone of MCMC, this function would substitute the straightforward calculation of posterior probability density as shown in the Matlab m-file 'lapserate', which was discussed in Section 5.5.

The modularity of the function 'getposterior' makes the Bayesian approach such a versatile and flexible tool. Via this function, virtually any model can be incorporated, any parameter can be estimated, or any variable can be included as uncertain model input. In this particular case study the model is determined by the kinetic rate equations that are solved as a set of ordinary differential equations. The standard model parameters to estimate are k_{2f}, k_{3f}, and k_4, but also the equilibrium constants or other rate parameters could be estimated by relaxing the appropriate assumptions. More creative options could be e.g. initial concentrations, parameters of equation (8-19) to calculate the optical density, the ratio between k_{2f} and k_{3f}, and/or a bias in the oxygen concentration. Obviously, the number of parameters and variables that can be evaluated at once is

limited by the available data. Nevertheless, relaxing an assumption or a fixed parameter generally leads to further insights into the system behavior.

8.6.2 A Possible Scenario

The parameter estimation described here combines various adjustments inspired by the discussion above regarding the observed data and the model constraints. This particular scenario implemented consists of four components:

1. Start with the Original Problem Formulation

   The rate parameters to be estimated are k_{2f}, k_{3f}, and k_4, based on the degrees of freedom analysis in Section 8.3.1.

2. Introducing Constraints

   A possible constraint could be based equation (8-15) when information is available regarding the diffusion limit D. The constraint implemented is then

   $$k_{3f} = D - k_{2f} \tag{8-28}$$

   to establish a relationship between k_{2f} and k_{3f}, without rigidly fixing a 2:1 ratio for k_{3f} : k_{2f} according to equation (8-16). Parameter D will be estimated simultaneously with rate parameters k_{2f} and k_4.

3. Simultaneous Data Evaluation

   An important advantage of Bayesian approach is the ability to combine various sources of information. In this case, five data sets obtained at different oxygen concentrations will be evaluated simultaneously, according to the explaination in Section 5.4.3. Contrary to the above parameter estimation that only evaluated a single data set, the additional data will lead to more precise estimates.

4. Cleaning of the Data

   As discussed before, the data collected at the initial stage of each experiment might be distorted as fluorescence interferes with the absorption signal. Therefore,

the data up to the maximum in the concentration profile is removed and only the remainder of the data used for parameter estimation. Secondly, the reaction is assumed to initiate after the excimer laser pulse (lasting 12-20 *ns*), so the starting time of the reaction is changed to 0.02 μs instead of the time at which the first positive optical density is detected.

To summarize, five data sets (of which the first few datapoints are removed) measured at different oxygen concentrations are evaluated simultaneously to obtain the parameter vector

$$\boldsymbol{\theta} = \left\{k_{2f},\ k_4,\ D,\ \sigma_0^{\,2}\right\} \tag{8-29}$$

where the parameter D is included instead of the original parameter k_{3f}. An informative prior distribution for parameter D could be defined based on existing information, but in this case a uniform prior is implemented and the knowledge that $D = 1200\ M^{-1}\mu s^{-1}$ [69] is used to verify the estimated parameter D.

The marginal posterior probability distributions obtained from the simultaneous evaluation of the five data sets are shown in Figure 8-10.

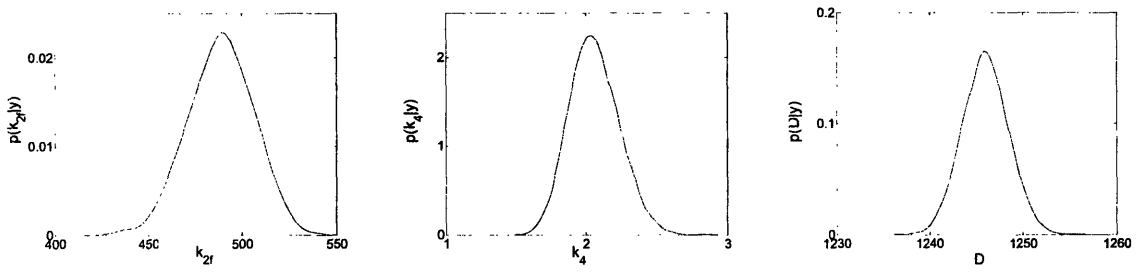


Figure 8-10. Marginal posterior distributions for rate parameters and diffusion limit

When examining the probability distributions above, the most significant result is that the rate parameter k_4 is now actually confined to a relatively narrow range. Moreover, the estimated value is very near the value of $k_4 = 0.8\ \mu s^{-1}$ referenced by Taylor *et al.* [69]. The increased information content attained by inclusing more data clearly had an impact.

A comparison of the mean and standard deviation for the estimates of k_4 obtained with one and five datasets are given in Table 8-3, emphasizing the importance of the ability to incorporate all available information.

Table 8-3. Impact of evaluating multiple datasets on the estimate for k_4

# datasets	$E[k_4]$	σ_{k_4}
1	$22.2\ \mu s^{-1}$	9.7
5	$2.06\ \mu s^{-1}$	0.17

Corroborating the improvement in system behavior is the fact that the estimated parameter D is approximately equal to value of 1200 $M^{-1}\mu s^{-1}$ initially assumed in equation (8-15). In addition, the model output (using the mean of the parameter estimates), fits the data reasonably well, as shown in Figure 8-11.

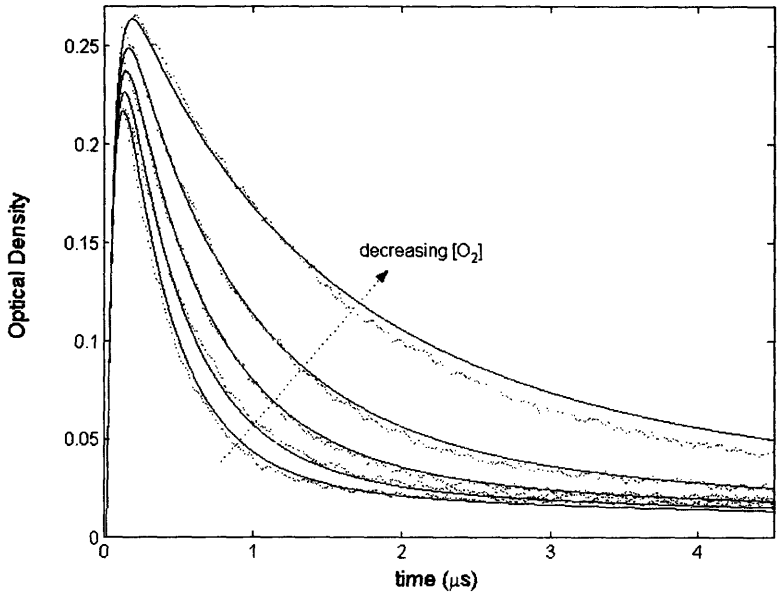


Figure 8-11. Model and data fit obtained from simultaneous data evaluation

Obviously, the model output does not fit each data set to the same degree. In particular, the data set obtained at lowest oxygen concentration deviates more from the model than any other data set.

Based on the estimated probability distributions of k_{2f} and D, rate parameter k_{3f} can be determined according to equation (8-28). Subsequently, the ratio of k_{3f} over k_{2f} can be obtained. The resulting probability distributions are shown in Figure 8-12.

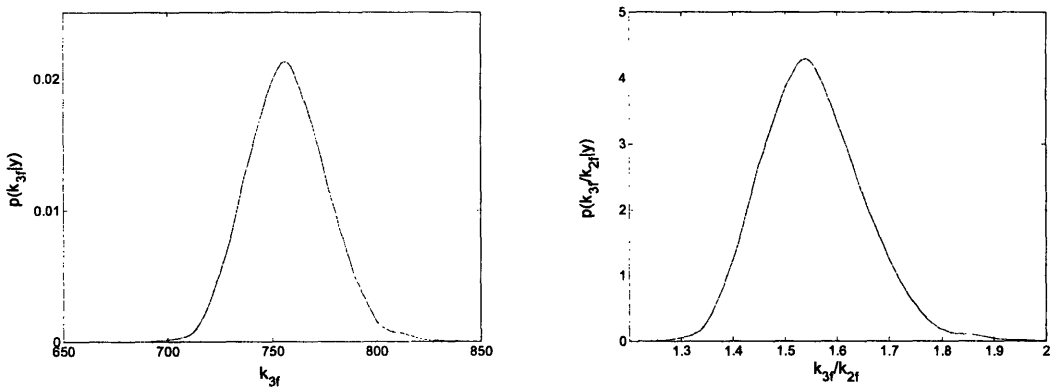


Figure 8-12. Derived rate parameter k_{3f} and ratio of forward rate parameters

Compared to the parameter estimates initially proposed by Taylor *et al.* [69], rate parameter k_{3f} is estimated slightly lower (and rate parameter k_{2f} slightly higher according to Figure 8-10). These results imply that the ratio of rate parameters k_{3f} : k_{2f} is lower than the 2:1 ratio that was initially assumed purely based on statistical arguments. The estimation results obtained in this particular scenario therefore provoke a critical evaluation of one of the original intuitions. To further analyze this ratio of rate parameters k_{3f} : k_{2f}, it is recommended to perform further experimentation as the currently available data alone are insufficient to provide a conclusive answer.

8.7 Summary of Key Points

The various features of using the Bayesian approach in parameter estimation discussed in this case study will be summarized in a few key points below.

- The marginal probability distributions for the parameter estimates and the MCMC simulation results can reveal information regarding the system behavior. Erratic

observations invoke further, more detailed investigation, eventually leading to a better understanding of the system.

- The Bayesian approach to parameter estimation is very flexible. The model and constraints can easily be adjusted, so that scenario analyses are easy to implement.
- Multiple data sets can be evaluated simultaneously to directly obtain parameter estimates. The increased amount of information leads to a relatively precise estimate for the rate parameter k_4, which was previously difficult to estimate upon evaluation of only a single data set.

9

Conclusions

終わり良ければ全て良し

Japanese Proverb

Eind goed, al goed

Dutch Proverb

Uncertainties will always be present, so the key is to identify those that contribute most to uncertainties in predicted model outcomes. Therefore, proper characterization of uncertainty in model parameters and consistent propagation of uncertainty are important. With this in mind, Bayesian statistics, for which probability represents a degree of belief, have been discussed as being the most appropriate method for dealing with uncertainties. Until recently, however, computational complexities have hindered the implementation of the Bayesian approach for multi-dimensional nonlinear problems. The work in this thesis has elaborately discussed and demonstrated that the use of Markov Chain Monte Carlo simulation techniques can overcome these complexities and solve the Bayesian formulation of parameter estimations in complex reaction engineering problems.

To formalize the application of the Bayesian approach to parameter estimation for reaction engineering problems, this thesis proposed a framework that systematically investigates an engineering system with the goal to estimate unknown parameters. The four stages of system description, system analysis, experimentation, and estimation represent the natural progress of a parameter estimation project. Nevertheless, by applying this Bayesian framework, the experimentalist is encouraged to consciously apply the suggested tools at each step to efficiently and effectively generate and interpret data. As is shown in the case studies, a detailed analysis at the outset of the parameter

estimation problem leads to a thorough understanding of the system and generally benefs the estimation stage of a project.

The parameter estimation results obtained in the two case studies developed in this thesis demonstrated that the Bayesian approach to parameter estimation is superior to conventional estimation methods. Several advantages of the Bayesian approach have elaborately been discussed in Chapter 5, but several benefits specifically relevant to the case studies are worth reiterating:

1. Uncertain information from various sources can easily be combined.
2. Compared to conventional estimation methods, significantly more information can be obtained from a particular dataset.
3. Evaluating incremental information gain per experiment becomes an intuitive step in the learning cycle of the Bayesian approach, thereby providing a straightforward measure to determine the value of information.
4. Systematic errors between different sets of experiments can be identified.
5. The algorithm for the estimation is modular, so that model and constraints are easily adjusted, thereby facilitating the evaluation of various scenarios.
6. There are no restrictions regarding the system model so that both a simplified analytic model and a model of kinetic rate equations can be implemented.
7. Conventional estimation methods can be useful to obtain suitable initial parameter values required for MCMC.

Clearly, the benefits of applying the Bayesian framework proposed in this thesis are numerous. As the recent increase in computational power has made the Bayesian approach feasible to be implemented for parameter estimation in complex reaction engineering problems, scientists should expand their toolkit of statistical methods and include the Bayesian framework as their main approach in dealing with uncertain data and information.

10

Future Research Opportunities

Always listen to experts.
They'll tell you what can't be done and why.
Then go do it.

Lazarus Long

As demonstrated in this thesis, applying the Bayesian approach to the analysis of engineering problems has compelling advantages. In the setting of experimentation and estimation, the case studies have focused mainly on parameter estimation and have merely scratched the surface of potential applications in the area of experimental design. Following the research presented in this thesis, several potential research projects can be suggested.

- *Developing Procedures for Evaluation and Propagation of Simulation Results*
  The applications of Bayesian parameter estimation reported in literature generally provide model parameters based on a particular dataset. The work in this thesis actually went a step further by combining individual parameter estimates (as represented by kernel density estimates) in order to provide an overall parameter estimate representing all information from a large number of datasets. Therefore, the procedures by which the sampling results are represented and implemented in further analyses will have an impact on the final outcomes.

  The ad-hoc approach described in this thesis is by no means optimal and there are several issues that should be further studied. As discussed, diagnostic methods have

been developed to determine the required number of samples to be generated with an MCMC simulation, but these diagnostic methods were developed for the situation that the parameter estimate would be the final product, not an intermediairy. Secondly, a kernel density estimate might not be the optimal representation for a probability distribution. Histograms are an alternative approach. Finally, using either kernel density estimates or histograms, the number of bins does likely have an impact on the accuracy of the parameter estimates.

The questions to be dealt with are then:

- How many samples need to be generated?
- Which method is preferable to represent the probability distribution?
- What should be the number of bins for a kernel density estimate or histogram?

These issues are expected to have an impact on the numerical error in final result of a parameter estimation project. Research should be directed at finding whether this impact is significant, and if so, at developing procedures to minimize this impact.

- *Application of the Bayesian Approach to Large and Complex System*

  The case studies in this thesis described relatively straightforward system models, involving only a single output variable as a function of time. The demonstrated advantages (e.g. dealing with nonlinear models, non-gaussian errors, flexibility of parameter estimation, sequential estimation, etc.) would be very welcome in larger and more complex systems.

  One possible application considered when starting research for this thesis is the application of an air pollution monitoring system. Such a system would involve a 3D air pollution model of a spatial area, where multiple species change their concentration over time. Different sensors (with different accuracy) placed at several locations would measure air quality and the collective data can be assimilated with the model to construct the current state of air pollution and predict future system behavior, while accounting for uncertainty.

As the Bayesian parameter estimations for relatively straightforward system models in this thesis already proved to be computationally intensive, more advanced applications require intelligent approaches in reducing the computational burden. One suggested approach to increase computational efficiency is to implement the Bayesian approach with the Deterministic Equivalent Modeling Method (DEMM [7]). Several applications of DEMM with complex engineering systems [8, 73] have shown the power of this method in reducing model complexity and would be an excellent lead in the research towards applying the Bayesian approach to large and complex systems.

- *Model Based Design of Experiments*

  The application of experimental design in this thesis has been limited to evaluating a stopping rule regarding the number of experiments to perform. However, the Bayesian approach to experimental design can be extended to more general applications. An introduction to Bayesian Experimental Design, supported with an example problem, is given in Appendix D.

  The Bayesian framework for experimental design has been applied to relatively simple systems [1, 74] as computational restrictions have inhibited the evaluation of large and complex systems. Where parameter estimation requires the evaluation of a single integral, the general formulation of Bayesian experimental design, given in equation (D-1), requires the evaluation of an integral within an integral. Though MCMC simulation is a powerful method to solving Bayesian problems, the computation of the optimal experimental design for complex systems will be challenging. Other challenging research opportunities in this area are to couple the experimental design decisions with the concepts of value of information, real options, and portfolio theory.

Appendix A.
Matlab Scripts and Functions

A.1. Metropolis-Hastings

Script 'lapserate'

The following script is used in Section 5.5.3. A linear regression with unknown variance is formulated according to the Bayesian approach and solved by applying the Metropolis-Hastings algorithm. The samples are thinned and the kernel density estimates are determined. The auxiliary functions 'getdesign' and 'acceptreject' are given below.

```
% lapserate.m
% Patrick de Man - MIT (June 2006)

%======== SPECIFY THE PROBLEM ========
% Initialize the random number generator
seed = 0;
randn('seed',seed);
rand('seed',seed);

% Population means / true parameters
a = 298 % K
b = 9.8 % K/km

% Height of the measurements (assume no error)
z = [0.5; 1; 1.5; 2; 2.5; 3; 3.5; 4; 4.5; 5];

% Define the design matrix
dim = length(z);
X = [ones(dim(1),1)  -z ];

% Generate the true values
trueT = a - b*z;

% Add random, normally distributed noise
sigT = 3;   % K
```

```
varT = sigT^2;
noise = sigT*randn(dim,1);
T = trueT + noise;   % (only for the 1st run)

%======== SPECIFY THE MCMC SIMULATION ========
% Specify the # of MCMC simulation runs, burn-in period, and thinning rate
nMCMC = 51000 ;
burnIn = 1000;
thinRate = 10;

% The matrix 'design' will collect all the accepted designs
design = zeros(nMCMC+1,3);

% Define the initial design for [a  b  varT]
design(1,:)= [0.8*a   1.2*b   0.8*varT];

% Specify the covariance matrix of the probing distribution
covarFactor = 0.05;
covarPD = diag( (covarFactor * design(1,:)).^2 );   % (only for the 1st run)

% Specify the tuning of the covariance matrix to control the acceptance rate
tuningFactor = 1;

% Define the domain of the prior / sample space
priorDomain.min = [0     0   0];
priorDomain.max = [500  20  200];

%======== RUN THE MCMC SIMULATION ========
for iRun = 1 : nMCMC

    if iRun > 1
        if accept == 1
           lnCurrentPosterior = lnProposedPosterior;
       end
    end

    % Propose a new design
    [proposedDesign] = getdesign(design(iRun,:), ...
            tuningFactor*covarPD, priorDomain.min, priorDomain.max);

    % Reinitiate the auxiliary variable signaling acceptance at previous step
    accept = 0;
```

```
    % determine prior and posterior of current and proposed design
    if iRun == 1
        lnCurrentPrior = log( gampdf(1/(design(iRun,3)),1e-3,1e3) );

        lnCurrentLikelihood = log( design(iRun,3)^(-dim/2)) - ...
             (T - X*design(iRun,1:2)')'*(T - X*design(iRun,1:2)') / ...
                  (2*design(iRun,3)) ;

        lnCurrentPosterior = lnCurrentPrior + lnCurrentLikelihood;
    end

    % Calculate the natural logarithm of the posterior of the proposed design
    lnProposedPrior = log( gampdf(1/(proposedDesign(3)),1e-3,1e3) );

    lnProposedLikelihood = log( proposedDesign(3)^(-dim/2)) - ...
             (T - X*proposedDesign(1:2)')'*(T - X*proposedDesign(1:2)') / ...
                  (2*proposedDesign(3)) ;

    lnProposedPosterior = lnProposedPrior + lnProposedLikelihood;

    % Determine the acceptance probability alpha
    alpha(iRun+1,1) = min(1, exp(lnProposedPosterior - lnCurrentPosterior) );

    %   Determine   the   acceptance/rejection   and   store   next   design
    [design(iRun+1,:), accept] = acceptreject(alpha(iRun+1, 1), ...
                  design(iRun,:), proposedDesign);

end % end of the MCMC loop

%======== PROCESSING OF THE ACCEPTED DESIGNS ========
% Discard the burn-in period
cleanDesign = design(burnIn : nMCMC, :);
cleanAlpha = alpha(burnIn : nMCMC);

% Perform thinning
iCounter = 1;
for iThin = 1 : thinRate : length(cleanDesign(:,1))
    thinnedDesign(iCounter, :) = cleanDesign(iThin, :);
    iCounter = iCounter + 1;
end
```

```
% Determine the estimated a and b from the thinned samples
aEst = mean(thinnedDesign(:,1))
bEst = mean(thinnedDesign(:,2))
avgAlpha = mean(cleanAlpha)

% Determine marginal probability distributions for each of the parameters
for i = 1 : 3
    [prob, values] = ksdensity(thinnedDesign(:,i), 'npoints', 1000);
    p(:, i) = prob';
    y(:, i) = values';
end
```

Function 'getdesign'

```
function proposedDesign = getdesign(currentDesign,covarPD,priorMin,priorMax);

% Patrick de Man - MIT (June 2006)
% This function generates a new design within the prior / sample domain.
% The design proposition is based on a random draw from a multi-variate
% normal probing distribution (PD), with mean located at the current design.

% Determine the dimension of the current design
dimension = length(currentDesign);

% Initialize
proposedDesign = zeros(1,dimension);
stopProposing = 0;
stopCheck = zeros(dimension,1);

% Propose a sample
while stopProposing ~= 1
    proposedDesign = mvnrnd(currentDesign , covarPD);

    % Check whether the sample is located within the domain
    for i = 1 : dimension
        if proposedDesign(i) <= priorMax(i)
            stopCheck(i) = 1;
            if proposedDesign(i) < priorMin(i)
                stopCheck(i) = 0;
            end % if
        elseif proposedDesign(i) > priorMax(i)
            stopCheck(i) = 0;
        end % elseif
    end % for
    stopProposing = all(stopCheck);
end % while

return; % function getdesign
```

Function 'acceptreject'

```
function [newDesign, accept] = ...
                        acceptreject(alpha, currentDesign, proposedDesign);

% Patrick de Man - MIT (June 2006)
% This function determines the newDesign by accepting or rejecting the
% proposedDesign.
% The proposedDesign is accepted with acceptance probability alpha

% Generate uniformly distributed random number
u = rand(1);

% Determine newDesign
if alpha >= u
    newDesign = proposedDesign;
    accept = 1;
elseif alpha < u
    newDesign = currentDesign;
    accept = 0;
end

return; % function acceptreject
```

Function 'getposterior'

The following function is discussed in Section 8.6.1. The natural logarithm of the posterior probability density is calculated as part of the MCMC algorithm for the parameter estimation in Chapter 6. This function would replace the straightforward calculation of the posterior probability density in the function 'lapserate' above.

```
function [lnPosterior] = getposterior(design, data)

% This function calculates the natural logarithm of the posterior probability
% density for a design.
%
% Patrick de Man - MIT (June 2006)

% INPUT:
% design            1x4 vector of design values:
%                   [k2f, k4, [(CH3)3CO.]0, sigma2]
% data              structure: data.data and data.O2
% data.data         nx2 vector of time values and measurements in columns
% data.O2           scalar of oxygen concentration of the experiment

% OUTPUT:
% lnPosterior       scalar of natural log of posterior probability density

%======== INITIALIZE THE VARIABLES ========
lnLikelihood = 0;
lnPrior = 0;

%======== SPECIFY THE MODEL ===============
% Split up the design vector
k(1) = 53;
k(2) = design(1);
k(3) = 2*k(2);
k(4) = design(2);
k(5) = 1.2e3;

% Specify the equilibrium constants
K2 = 2100;
K3 = 2100;
```

```
% Specify the time vector
t = data.data(1 : length(data.data(:,1)), 1));

% Specify the initial conditions for the ODE solver
%          [ [C6H7.]   [(CH3)3CO.]   [C6H8]   [o-C6H7OO.]   [p-C6H7OO.]   ]
initialC = [ 0          design(3)      0.4         0              0       ];

%======== SOLVE THE MODEL ================
% Solve the set of ODEs
options = odeset('RelTol', 1e-6, 'AbsTol', 1e-8);
[t, y] = ode15s(@kinetics, t, initialC, options, k, K2, K3, data.O2 );

%======== DATA AND MODEL COMPARISON =======
% Calculate the posterior probability density using data and model output
for i = 1 : length(data.data(:,1))

    % Determine the data and model input
    dataY = data.data(i, 2);
    modelY = 2100*y(i,1) + 200*(y(i,4) + y(i,5));

    % Calculate the likelihood function
    lnLikelihood = lnLikelihood  + log(1/(sqrt(2*pi*design(4)))) ...
                                    - (dataY - modelY)^2 / (2*design(4));

end % i

% Calculate the prior
lnPrior = log(gampdf(1/design(4),1e-5,1e5)) ;

% Calculate the posterior probability density
lnPosterior = lnLikelihood + lnPrior;

return; % function getposterior
```

A.2. Processing Samples

The following collection of scripts and functions serve various purposes in processing the samples obtained from MCMC. They are included as their content is likely to be useful when working with Matlab to solve problems according to the Bayesian approach using MCMC.

Function 'getpdf'

```
function [PDF] = getpdf(sampleSet)
% Patrick de Man (June 2006) - MIT
% This function determines the PDF and its mean & variance from a sample set.
% The PDF is determined with binwidth appr. 0.1% of the range of the samples.
% On both sides of the PDF, a significant number of bins is appended in
% order to guarantee overlap with other PDFs to apply e.g. Bayes' theorem.

% Define the number of bins
nBins = 1000;

% Initialize the vector specifying the bins
y = zeros(nBins + 20000, 1);

% Determine the range of the sample set
binWidth = (0.001)*(max(sampleSet) - min(sampleSet));

% Specify the lowest bin
yMin = min(sampleSet) - 10000*binWidth;

% Specify the remainder of the bins
for iBin = 1 : (nBins + 20000)
    y(iBin) = yMin + binWidth*(iBin - 1);
end

% Remove the bins that are located below 0
% (only necessary when the parameter is >0)
if yMin < 0
    iBin = 1;
    while y(iBin) < 0
        iBin = iBin + 1;
    end
    y = y( iBin : length(y));
end % if yMin < 0
% Determine the marginal posterior distribution
```

```
p = ksdensity(sampleSet, y);
PDF.PDF = normalizePDF(y, p);

% Determine the mean and variance of the sample set
PDF.mean = mean(sampleSet);
PDF.var = var(sampleSet);

return; % function getpdf
```

Function 'normalizePDF'

```
function [normalizedPDF] = normalizePDF(yInput, pInput);
% Patrick de Man - MIT (June 2006)
% This function normalizes a PDF, which should be given as a kernel density
% estimate or histogram with constant bin widths.
% INPUT: y and p have to be column vectors where:
% y = the random variable
% p = p(y) the probability values
% OUTPUT: is a (nBins x 2) matrix holding the y and the normalized p.

% Ensure that p is a column vector
p = zeros(length(pInput), 1);
for i = 1 : length(pInput)
    p(i) = pInput(i);
end

% Ensure that y is a column vector
y = zeros(length(yInput), 1);
for i = 1 : length(yInput)
    y(i) = yInput(i);
end

% Determine the dimension of the PDF
nBins = length(y);

% Determine the total area under the curve of the PDF
areaPDF = 0;
for iBin = 2 : nBins
    binWidth = y(iBin) - y(iBin - 1);
    areaPDF = areaPDF + (binWidth * p(iBin));
end

% Normalize the PDF
normalizedPDF(:, 1) = y;
normalizedPDF(:, 2) = p / areaPDF;

return; % function normalizePDF
```

Function 'getprobability'

```
function [probability] = getprobability(variable, PDF);
% Patrick de Man - MIT (June 2006)
% This function determines the probability density of a value
% for a random variable using its PDF.
% The function scans through the bins of the PDF and selects the proper one.

% Determine the dimension of the PDF
nBins = length(PDF(:,1));

% Initialize search
iBin = 1;

% Assign a small probability density when outside the range of the PDF
if variable < PDF(iBin, 1)
    probability = 1e-100;

% Search through the bins until the variable fits in
elseif variable == PDF(iBin, 1)
    probability = PDF(iBin, 2);
elseif (variable > PDF(iBin,1)) & (variable <= PDF(nBins,1))
    while (variable > PDF(iBin, 1)) & (iBin <= nBins)
        iBin = iBin + 1;
    end % while

    % Determine the probability density
    probability = PDF(iBin, 2);

% Assign a small probability density when outside the range of the PDF
elseif (variable > PDF(nBins))
    probability = 1e-100;

end % elseif

return; % function getprobability
```

Script 'compilesamples'

```
% compilesamples.m
% Patrick de Man - MIT (June 2006)
% This function compiles a number of PDF of different length into one array.
% The advantage of the compiled array is easy retrieval of the PDFs for
% further processing.
% To create PDFs of equal dimensions lengths of the PDFs, columns
% holding 'NaN' (=not a number) entries are attached when required

% Define the number of PDFs to compile
nSeries = 100;

% Initialize the counter for the array of compiled PDFs
% Use this variable if intermittent PDFs should be skipped
counter = 1;

% Cycle through the different PDFs
for iSeries = 1 : nSeries

    % Load the structure holding the PDF
    load(strcat(pwd, '\results\MCMCoutput_', num2str(iSeries) ));

    % Extract the PDF from the loaded structure
    PDF = MCMCoutput.PDF;

    % Specify the number of rows in the PDF
    rowMax(counter) = length(MCMCoutput.PDF(:,1));

    if counter > 1

        % Add 'NaN' columns of the compiled array when necessary
        if length(PDF(:,1)) > length(compiledPDF(:,1,1))
            addLength = length(PDF(:,1)) - length(compiledPDF(:,1,1));
            addNaN = zeros(addLength, 2, (counter - 1));
            for i = 1 : (counter-1)
                for iNaN = 1 : addLength
                    addNaN(iNaN, :, i) = [NaN  NaN];
                end % for iNaN
            end % for i

            compiledPDF( (length(k2PDF(:,1,1)) + 1) : length(PDF(:,1)), ...
```

```
                                                :, (1 : (counter - 1))) = addNaN;

        % Add 'NaN' columns of the imported PDF when necessary
        elseif length(PDF(:,1)) < length(compiledPDF(:,1,1))
            addLength = length(compiledPDF(:,1,1)) - length(PDF(:,1));
            addNaN = zeros(addLength, 2);
            for iNaN = 1 : addLength
                addNaN(iNaN, :) = [NaN  NaN];
            end % for iNaN

            PDF( (length(PDF(:,1)) + 1) : ...
                        length(compiledPDF(:,1,1)),:) = addNaN;

        end % elseif
    end %if counter > 1

    % Assign the PDF to the compiled array now dimensions are equal
    compiledPDF(:, :, counter) = PDF;

    % Update the counter
    counter = counter + 1;

end % for iSeries

% Create a structure as output
compiled.PDF = compiledPDF;
compiled.RowMax = RowMax;
```

A.3. JPL Prior: Lognormal Fitting

The following script is used to find a lognormal distribution for the JPL rate parameter estimates, as discussed in Section 7.2.3.

Script 'findsigmalognormal'

```
% findsigmalognormal
% Patrick de Man - MIT (June 2006)
% This script finds sigma that fits a lognormal distribution fitted through
% the rate parameter estimate with given uncertainty obtained from JPL report

% Specify the mode of the lognormal distribution
kmode = 2.6e-11;

% Specify the range calculated according to the uncertainty factors
k = linspace(2.17e-11, 3.12e-11, 1000);

% Minimize f by manipulating sigma
[sigma, f] = fminbnd(@lognormalfit, 0, 1, [], k, kmode)

% Specify the lognormal distribution with the determined sigma over range k
for i = 1 : length(k)
    pdf(i) = lognpdf(k(i), log(kmode) + sigma^2, sigma);
end

% plot the lognormal distribution over range k
figure;
plot(k, pdf);
```

Function 'lognormalfit'

```
function f = lognormalfit(sigma, k, kmode)

% Patrick de Man - MIT (June 2006)
% This function specifies the calculation of f, the variable to minimize.
% Cumulative lognormal distributions, specified by mode and sigma, are used
% to calculate f as the area under the curve captured by the range k

% Specify the probability captured by the distribution between the range of k
probability = 0.6827;

% Calculate f
f = abs(logncdf(max(k), log(kmode) + sigma^2, sigma) ...
            - logncdf(min(k), log(kmode) + sigma^2, sigma)- probability);

return; % function lognormalfit
```

Appendix B.

WinBugs code for Parameter Estimation

While the Matlab scripts given in Appendix A above were based on the Metropolis-Hastings algorithm, WinBugs [35] applies Gibbs sampling for obtaining the posterior distribution. As shown below, WinBugs provides a convenient programming environment specifically designed for the Bayesian approach. Excel add-ins, Matlab functions, and SAS macros are available to transfer data between these applications and WinBugs.

Linear Regression

The following WinBugs code implements the linear regression as discussed in Section 5.5.3.

```
# specify the linear model with parameters
model{
        for(i in 1:N){
                        y[i] ~ dnorm(T[i], tau)
                        T[i] <- a - b*z[i]
                        }

        # specify the prior distributions
        a ~ dunif(0.0, 500)
        b ~ dunif(0.0, 20)
        tau ~ dgamma(0.001, 0.001)
        variance <- 1/tau
        }

# specify the data and other constants
list(y = c(300.85, 288.88, 282.33, 268.67, 266.13, 252.45, 242.54, 229.26),
z = c(0,1,2,3,4,5,6,7), N=8)

# specify the initial values
list(a = 238.4, b = 11.76,  tau = 0.139 )
```

Parameter Estimation with the Bi-Exponential Model

The advantage of WinBugs is the ease of use and simplicity with which MCMC can be implemented. However, when more complex systems are under investigation, Matlab is still preferred because of its flexibility and computational capabilities. Regarding the parameter estimation of Case Study 1 in Chapter 7, Matlab allows for automating the large number of estimations, the possibility to include models consisting of differential equations, and for subsequent sampling from the obtained posterior distributions.

For illustrative purposes, the parameter estimation according to the analytic equation in Case Study 1 (Section 7.5.3) can also be performed by implementation in WinBugs. An example of a particular data set is shown below. Note that the number of data points has been abridged.

```
# specify the bi-exponential model with parameters

model{
        for(i in 1:N){
                        y[i] ~ dnorm(O[i], tau)
                        O[i] <- A*exp(-B*t[i]) + C*exp(-D*t[i])
                        }

        # specify the prior distributions
        A ~ dunif(-2000, 0)
        B ~ dunif(0.0, 25000)
        C ~ dunif(0.0, 1500)
        D ~ dnorm(61.7,0.0088)
        tau ~ dgamma(0.001, 0.001)
        sigma <- sqrt(1/tau)
        }

# specify the data and other constants
list( t = c( 6.00E-06,8.00E-06,....,0.0014,0.001402,0.001404,0.001406,0.001408),
        y = c(117,33,....,426,422,409,408,439), N=701 )

# specify the initial values
list(A = -434.4, B = 5676.36,  C=484, D=61.7, tau = 0.0013 )
```

Appendix C.

Derivation of the Bi-Exponential Equation

The derivation of the analytic equation obtained from the kinetic model discussed in Section 7.2.2 is shown below.

The reaction mechanism for which the bi-exponential equation is derived as follows

$$O(^1D) + N_2 \xrightarrow{k_1} O(^3P) + N_2 \quad \text{(C-1)}$$

$$O(^1D) + O_3 \xrightarrow{k_2} 2O(^3P) + O_2 \quad \text{(C-2)}$$

$$O(^1D) + O_3 \xrightarrow{k_3} 2O_2 \quad \text{(C-3)}$$

$$O(^3P) \xrightarrow{k_5} loss \quad \text{(C-4)}$$

$$O(^1D) \xrightarrow{k_6} loss \quad \text{(C-5)}$$

By applying the pseudo steady state assumption implying that $[O_3]$ is constant, the set of ordinary differential equations becomes

$$\frac{d[O(^3P)]}{dt} = k_1[O(^1D)][N_2] + 2k_2[O(^1D)][O_3] - k_5[O(^3P)] \quad \text{(C-6)}$$

$$\frac{d[O(^1D)]}{dt} = -k_1[O(^1D)][N_2] - (k_2 + k_3)[O(^1D)][O_3] - k_6[O(^1D)] \quad \text{(C-7)}$$

The kinetic rate equation for $[O(^1D)]$ can be rewritten as

$$\frac{d[O(^1D)]}{dt} = -\{k_1[N_2] + (k_2 + k_3)[O_3] + k_6\}[O(^1D)] \quad \text{(C-8)}$$

which can be integrated to give

$$[O(^1D)] = [O(^1D)]_0 \, e^{-k_a t} \tag{C-9}$$

where

$$k_a = k_1[N_2] + (k_2 + k_3)[O_3] + k_6 \tag{C-10}$$

The above result can be applied to the kinetic rate equation for $[O(^3P)]$ as follows

$$\frac{d[O(^3P)]}{dt} = [O(^1D)]\{k_1[N_2] + 2k_2[O_3]\} - k_5[O(^3P)] \tag{C-11}$$

$$\frac{d[O(^3P)]}{dt} = [O(^1D)]_0 \, e^{-k_a t}\{k_1[N_2] + 2k_2[O_3]\} - k_5[O(^3P)] \tag{C-12}$$

Multiplication by the factor $e^{k_5 t}$ facilitates integration as follows

$$\left\{\frac{d[O(^3P)]}{dt} + k_5[O(^3P)]\right\} e^{k_5 t} = [O(^1D)]_0 \, e^{-k_a t}\{k_1[N_2] + 2k_2[O_3]\} e^{k_5 t} \tag{C-13}$$

$$\frac{d}{dt}\{[O(^3P)]e^{k_5 t}\} = [O(^1D)]_0\{k_1[N_2] + 2k_2[O_3]\} e^{(k_5 - k_a)t} \tag{C-14}$$

Integration gives

$$[O(^3P)]e^{k_5 t} = \frac{[O(^1D)]_0\{k_1[N_2] + 2k_2[O_3]\}}{k_5 - k_a} e^{(k_5 - k_a)t} + \Omega \tag{C-15}$$

where Ω is the integration constant, which is found by evaluating (C-15) at $t = 0$, resulting in

$$\Omega = [O(^3P)]_0 - \frac{[O(^1D)]_0 \{k_1[N_2] + 2k_2[O_3]\}}{k_5 - k_a} \quad \text{(C-16)}$$

so that equation (C-15) becomes

$$[O(^3P)] = Ae^{-Bt} + Ce^{-Dt} \quad \text{(C-17)}$$

with the parameters A, B, C, and D defined as

$$A = \frac{[O(^1D)]_0 \{k_1[N_2] + 2k_2[O_3]\}}{k_5 - k_a} \quad \text{(C-18)}$$

$$B = k_1[N_2] + (k_2 + k_3)[O_3] + k_6 \quad \text{(C-19)}$$

$$C = [O(^3P)]_0 - A \quad \text{(C-20)}$$

$$D = k_5 \quad \text{(C-21)}$$

Appendix D.
Basics of Bayesian Experimental Design

Introduction

The goal of experimental design is optimizing the information content to be gained by an intelligent selection of the control variables for the next experiment. Equivalent to the situation in parameter estimation displayed in Figure 3-7, Bayesian Experimental Design can be considered as the fundamental approach from which conventional experimental design methods, such as D-optimality [58], can be derived. Therefore, with the recent surge in attention for Bayesian methods in general, Bayesian Experimental Design [59] is also becoming more popular.

This appendix will first discuss the general formulation of Bayesian Experimental Design, after which an example problem will illustrate the basic principles. This example problem generalizes the approach as described by Bard [1], who was forced to implement several restricting assumptions in order to avoid mathematical difficulties when solving the experimental design problem. As will be shown, the Bayesian approach is able to relax these assumptions and treat the problem in a generalized manner.

General Formulation

The basic premise of experimental design is to evaluate a utility or loss function concurrent with performing a parameter estimation based on hypothetical data to be obtained from a potential future experiment. Among all possible future experiments, the selection that will maximize the expected utility within the resource constraints is considered the optimal experimental design. Mathematically this can be represented by

$$U(\eta^*) = \max_{\eta \in \mathcal{H}} \int_Y \max_{d \in \mathcal{D}} \int_\Theta U(d,\theta,\eta,y)\, p(\theta \mid y,\eta)\, p(y \mid \eta)\, d\theta dy \qquad \text{(D-1)}$$

where $U(d, \theta, \eta, y)$ is the utility function, $p(\theta|y, \eta)$ is the posterior distribution function, $p(y|\eta)$ is the predictive distribution, η is the design to be selected from design space $\mathcal{H}$, η^* is the optimal design, data y will be observed from the sample space Y, a decision d will be chosen from the set $\mathcal{D}$, and the unknown parameters θ are from parameter space Θ. The key in experimental design is the choice of utility function, for which the quadratic loss function or functions based on the Shannon information theory are common [59, 75].

Shannon Information Theory and Information Gain

In order to evaluate future possible experiments, the impact of their information gain needs to be determined. A unique measure for information content is given by

$$H\left(p(\zeta)\right) \equiv -E[\log p(\zeta)] = -\int p(\zeta) \log p(\zeta) d\zeta \tag{D-2}$$

where $H\left(p(\zeta)\right)$ is the information measure for the probability distribution $p(\zeta)$, which characterizes the uncertainty in parameter ζ. The information gain achieved with experimental data can then be determined by

$$I(\theta) = H\left(p(\theta)\right) - H\left(p(\theta \mid y)\right) \tag{D-3}$$

where the information gain $I(\theta)$ regarding parameter θ will increase by the updating of prior $p(\theta)$ with data y to obtain the posterior distribution $p(\theta|y)$.

As discussed above, the goal of experimental design is to maximize the information gain. Since $H\left(p(\theta)\right)$ is constant for a given prior, the information measure $H\left(p(\theta \mid y)\right)$ for the posterior distribution must be minimized. To make this discussion more concrete, consider that in the case of $p(\theta|y) \sim N(\mu, \sigma^2)$, then $H\left(p(\theta \mid y)\right) \sim \sigma^2$, so that the variance must be minimized. The increase in information content with a decreasing variance has previously been illustrated in the discussion of the estimation results of Case Study 1 (Chapter 7).

Experimental Design for a Linear Model

The implementation of the general formulation of experimental design according to equation (D-1) has been too complex for practical applications. The first simplification that can be implemented is to consider the situation without the decision d involved, reducing the formulation to

$$U(\eta^*) = \max_{\eta \in \mathcal{H}} \int_{Y} \int_{\Theta} U(\theta, \eta, y)\, p(\theta \mid y, \eta)\, p(y \mid \eta)\, d\theta dy \tag{D-4}$$

which would still require significant computational effort as all possible future data y have to be considered. Bard [1] summarizes the issues of conventional solution methods as follows: "...while the procedures are conceptually simple and appealing, their implementation is difficult in most practical situations. ...severe computational difficulties which arise from the need to evaluate multiple infinite integrals for all possible values of $\mathcal{H}$, Y, and Θ."

By further assuming that the posterior distribution $p(\theta|y)$ and distribution of the data y are Gaussian, the maximization of equation (D-4) is equivalent to minimizing the information measure, leading to

$$\min_{\eta} H\left(p(\theta \mid y)\right) \triangleq \min_{\eta} \left\{\det \boldsymbol{\Sigma}\right\} \triangleq \max_{\eta} \left\{\det \boldsymbol{\Sigma}^{-1}\right\} \tag{D-5}$$

where det $\boldsymbol{\Sigma}$ is the determinant of the covariance matrix of the Gaussian posterior distribution $p(\theta|y)$, defined by equation (4-8). The optimal design η^* is then obtained by satisfying the objective function of equation (D-5). Because of the simplifying assumptions only the magnitude of the covariance matrix of the posterior distribution needs to be evaluated [1].

In order to determine the optimal experimental design analytically, the restricting assumption of linearity is required. Implementing the linear system model

$$y = \mathbf{X}\theta \tag{D-6}$$

reduces the objective function to

$$\max_{\mathbf{x}} \left\{ \det\left({\sigma_m}^2 + \mathbf{X}\mathbf{V}_0\mathbf{X}^T \right) \right\} \tag{D-7}$$

where x is the experimental design (equivalent to η), ${\sigma_m}^2$ is the measurement variance, and $\mathbf{V_0}$ is the covariance matrix of the prior distribution.

An Example Problem: the Traditional Approach

Experimental design for a linear system with a Gaussian error model for data y will be implemented using the Bayesian approach. For an illustration of the model, the linear regression results including the 95% confidence interval bounds are shown below [10].

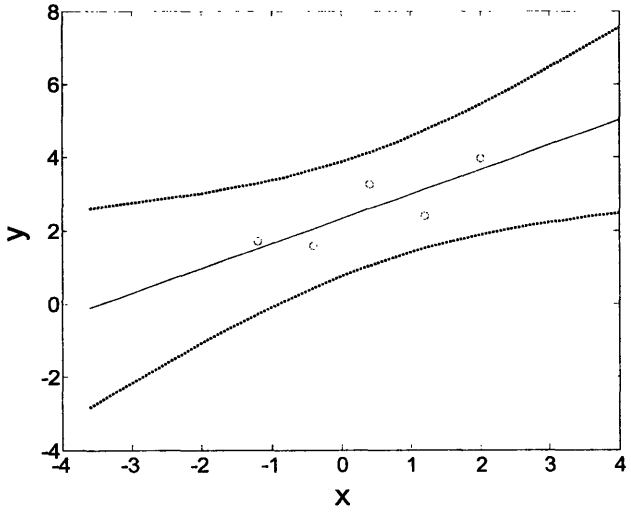


Figure D-1. Linear model including the 95% confidence interval bounds

As the confidence interval is wider further removed from the data, the obvious future measurement location increasing det $\mathbf{\Sigma}^{-1}$ (thus increasing the information content) is on

either extremes of the range of x. This is also clear from rewriting equation (D-7) for the two-dimensional system with parameters a and b, to attain the objective function

$$\max_{x}\left\{\sigma^2+\begin{pmatrix}1\\ x\end{pmatrix}\begin{pmatrix}(\sigma_{0,a})^2 & \sigma_{0,ab}\\ \sigma_{0,ab} & (\sigma_{0,b})^2\end{pmatrix}\begin{pmatrix}1 & x\end{pmatrix}\right\} \tag{D-8}$$

which is indeed maximized by increasing the absolute value of the decision variable $|x|$.

An Example Problem: the Bayesian Approach

As discussed in this thesis, one of the advantages of the Bayesian approach is that restricting assumptions can be relaxed. Bayesian Experimental Design can therefore generalize the problem statement described above. Figure D-10-2 illustrates the algorithm implemented to determine the optimal design for the future measurement regarding a linear system model. The algorithm requires as input 1) the linear model, 2) the measurement error distribution, and 3) the current knowledge regarding parameter θ in the form of posterior distribution $p(\theta|y_0)$ based on existing data y_0. As shown, the computation of the Bayesian formulation relies on two nested MCMC simulations.

Though the particular algorithm shown in Figure D-10-2 maintains the linearity and normality assumptions, it treats the system model and probability distributions in a generalized manner. Therefore, the objective function will not anymore solely depend on decision variable x, as is the case in equation (D-8) above. Instead, the future experiment will have to be considered with the future data evaluated according to the predictive distribution given by equation (3-23)

$$p(y|y_0)=\int p(y|\theta)p(\theta|y_0)\,d\theta$$

where y is the future data and y_0 is the existing data. Notice that the integral to obtain the predictive distribution is also included in the general formulation of Bayesian Experimental Design as given by equation (D-1) above.

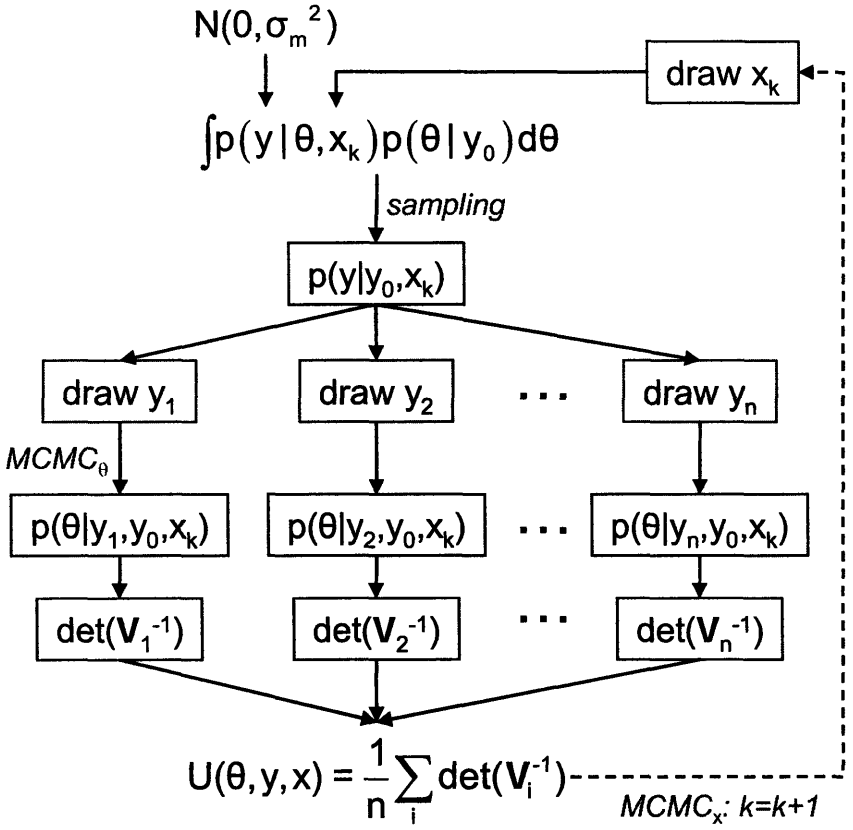


Figure D-10-2. Diagram describing the experimental design algorithm

The predictive distribution $p(y|y_0)$ is sampled n times and for each sample y_i the updated posterior distribution (incorporating the fictive data y_i) is determined using MCMC simulation. This 'inner' $MCMC_\theta$ simulation is formulated similarly to the parameter estimation problem described in Section 5.5.

After determining the updated posterior distributions $p(\theta|y_i,y_0)$, the utility function $u_i(\theta,y_i,x)$ of equation (D-4) is evaluated as

$$u_i(\theta, y_i, x) = \det\{\mathbf{V}(\theta, y_i, x)\}^{-1} \tag{D-9}$$

Since in this case the probability distributions are Gaussian, the inverse of the determinant of the covariance matrix is calculated. The integration over y according to equation (D-4) is subsequently performed by averaging the utility function values obtained for $y_1,\ldots, y_n$. The overall utility function $U(\theta,y,x)$ is subsequently applied to calculate the acceptance probability of the Metropolis-Hastings algorithm of Section 4.3.4 as follows

$$\alpha = \min\left\{1, \frac{U^*(\theta,y,x)}{U^i(\theta,y,x)}\right\} \tag{D-10}$$

which is implemented for the 'outer' $MCMC_x$ that is sampling x from space $\mathcal{H}$.

The result of a search for the optimal experimental design is shown in Figure D-10-3. As expected, the sampling by the MCMC simulation moves away from the initial value of $x = 1$ and towards the extremes of the range $x = [-2, 2]$.

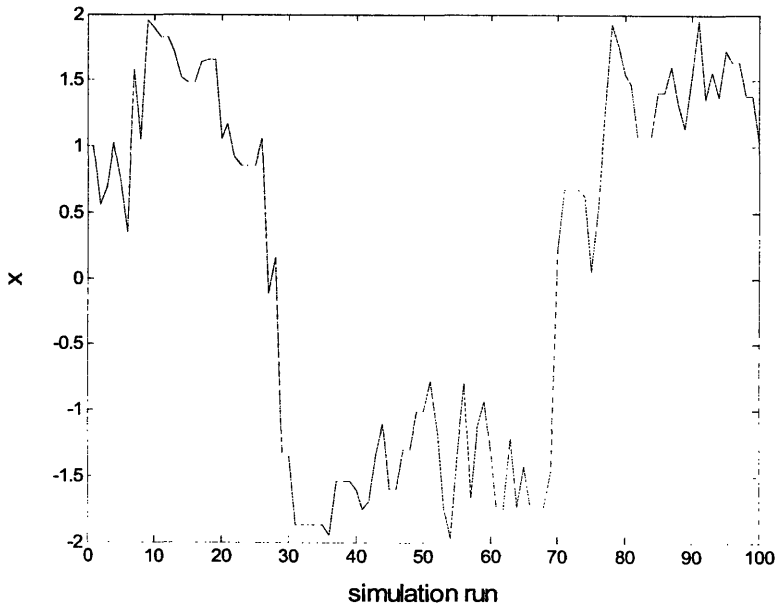


Figure D-10-3. Example of a search for the optimal experimental design using MCMC

These results indicate that this generalized approach to experimental design performs as desired. However, the algorithm encounters difficulties when the implemented

probability distributions are too wide. These difficulties are caused by relatively uncertain knowledge regarding parameter θ and/or by large measument errors regarding data y. These two factors reinforce each other when determining the predictive distribution, leading to a large forecasted uncertainty causing difficulties for the optimization of x.

The Matlab script and accompanying functions for the Bayesian experimental design algorithm are provided in this appendix below.

Script 'ExperimentalDesign'

```
% % This script determining the best next measurement location.
% The data to be collected is used to for estimation of the model parameters.
% This problem is based on a linear example from Bard (1974) p. 264
%
% Functions 'getdesign' and 'acceptreject', which were specified in Appendix
A
% Function 'getalphanetworkdesign1D' is given below

% Patrick de Man - MIT (June 2006)

%======== THE LINEAR MODEL ========

% Specify the current estimates for parameters a and b
estParameters = [2   1];
estCovar = [0.1  0  ;  0  0.5];

% Specify the measurement variance~N(y,varMeasurement)
varMeasurement = 0.1;

% Specify the domain of variable x
xMin = -2;   xMax = 2;

%======== MCMC SIMULATION ========

% Specify various variables for the MCMC simulation sampling x
nRunX = 100;                      % maximum number of MCMC runs
designX = zeros(nRunX+1, 1);      % storing vector for designs of x
runCounterX = zeros(nRunX+1, 1);  % tracking vector for # of MCMC runs
varPDX = 0.5;                     % variance of the probing distribution
acceptXtracker = 0;               % counter for # of accepted designs for x

% Specify the initial design
designX(1) = 1;

% Start the 'outer' MCMC simulation sampling x
for iRunX = 1 : nRunX

    runCounterX(iRunX) = iRunX;
    iRunX
```

```
% Propose a new design for x
[proposedDesignX] = getdesign(designX(iRunX), varPDX, xMin, xMax);


%======== DETERMINE THE PREDICTIVE DISTRIBUTION P(Y*|Y) ========

% Specify the number of sample to take from p(y*|y)
nSampleY = 200;     % at least >100 for a good estimate

% Generate random samples from the parameters: [a b]
parameterSample = mvnrnd(estParameters, estCovar, nSampleY);

% Generate random samples from the predictive distribution
predictiveY = parameterSample(:,1) + ...
                parameterSample(:,2)*designX(iRunX) ...
                                + sqrt(varMeasurement)*randn(nSampleY,1);

%======== INTEGRATION OVER DATA Y ========

for iSampleY = 1 : nSampleY

    % Generate the future measurement data y by sampling p(y|y0)
    y = predictiveY(iSampleY);

    %======== INTEGRATION OVER PARAMETER THETA ========

    % Specify the MCMC simulation
    n = 10;              % samples to determine PD covariance
    burnInAB = 2*n;      % burn in period
    nRunAB = 2*n + 200;  % maximum number of MCMC runs

    % Initialize storing vectors
    designAB = zeros(nRunAB+1, 2);
    alphaAB = zeros(nRunAB,1);

    % Specify the initial design
    designAB(1, :) = estParameters;

    % Specify the initial covariance matrix of the probing distribution
    covarPD = estCovar;
    covarTuning = 1;
```

```
% Specify the domain for sampling
ABMax = [( estParameters(1) + 6*sqrt(estCovar(1,1)) ) ;
                    ( estParameters(2) + 6*sqrt(estCovar(2,2)) ) ];
ABMin = [( estParameters(1) - 6*sqrt(estCovar(1,1)) ) ;
                    ( estParameters(2) - 6*sqrt(estCovar(2,2)) ) ];

for iRunAB = 1 : nRunAB

    runCounterAB(iRunAB) = iRunAB;

    % Update covarPD
    % Note: for this problem updating did not
    %       lead to a significant improvement
    %if iRunAB==n || iRunAB==2*n || iRunAB==3*n
    %    covarPD = covarTuning * cov(designAB(iRunAB-n+1 : iRunAB,:));
    %end

    % Propose a new design for a and b
    [proposedDesignAB] = getdesign(designAB(iRunAB, :), ...
                                          covarPD, ABMin, ABMax);

    % Determine the acceptance probability alpha
    [alphaAB(iRunAB)] = getalphanetworkdesign1D(designAB(iRunAB,:),...
        proposedDesignAB, estParameters, estCovar, y, ...
                                    proposedDesignX, varMeasurement);

    % Determine the acceptance/rejection and store next design
    [designAB(iRunAB+1,:), accept] = acceptreject(alphaAB(iRunAB), ...
                                designAB(iRunAB, :), proposedDesignAB);

end % for iRunAB


%======== PROCESS MCMC RESULTS FOR SAMPLING A AND B ========

% Correct the final entry of the runCounter vector
runCounterAB(iRunAB+1) = nRunAB + 1;

% Remove the burn-in period
cleanRunCounterAB = runCounterAB(burnInAB : nRunAB);
cleanDesignAB = designAB(burnInAB : nRunAB, :);
alphaABavg = mean( alphaAB(burnInAB : nRunAB, 1) );
```

```
    % Append sampling results regarding a and b to storing vector
    nResults = length(cleanDesignAB(:,1));
    resultsIntegrationAB( ((iSampleY-1)*nResults + 1 ) : ...
                                    iSampleY*nResults, :) = cleanDesignAB;
    resultsAlphaAB(iSampleY) = alphaABavg;

    % Optional: Track several simulation results
    yTracker(iSampleY,1) = y;
    meanABTracker(((iRunX - 1)*nSampleY + iSampleY), :) = ...
                                                        mean(cleanDesignAB);
    detCovarABTracker(((iRunX - 1)*nSampleY + iSampleY), 1) = ...
                                              det(inv(cov(cleanDesignAB)));

end % for iSampleY

% Optional: determine the variance of the samples y of p(y|y0)
TrackerVarY(iRunX,1) = var(yTracker);

%======== CONTINUE THE MCMC SIMULATION SAMPLING X ========

% Average the det(inv(cov)) over the samples y
% (equivalent to integrating over y in the general formulation)
resultsDetCovar(iRunX,1) = ...
   mean(detCovarABTracker(((iRunX - 1)*nSampleY + 1) : iRunX*nSampleY,1));
resultsMean = mean(resultsIntegrationAB);

% Determine acceptance probability based on the ratio of utility
proposedCriterion = resultsDetCovar(iRunX);
if iRunX == 1
    alphaX(iRunX,1) = min(1, (proposedCriterion / det(inv(estCovar))));
else
    alphaX(iRunX,1) = min(1, (proposedCriterion / criterion(iRunX-1)));
end

% Determine the acceptance/rejection and store next design
[designX(iRunX+1), accept] = acceptreject(alphaX(iRunX), ...
                                        designX(iRunX),  proposedDesignX);

% Track the utility function value during sampling
if accept == 1
    criterion(iRunX) = resultsDetCovar(iRunX);
```

```
        acceptXtracker = acceptXtracker + 1;
    elseif accept == 0
        if iRunX == 1
            criterion(iRunX) = det(inv(estCovar));
        else
            criterion(iRunX) = criterion(iRunX - 1);
        end
    end

end % for iRunX

% Determine the average acceptance rate of the simulation
avgAlphaX = mean(alphaX)
acceptRate = acceptXtracker/nRunX

% Correct the counter as the final count nRun+1 is not yet included
runCounterX(nRunX + 1) = nRunX + 1;

% Plot the sampling of x
figure;
plot(runCounterX, designX);
xlabel('simulation run');
ylabel('x');
```

Function 'getalphanetworkdesign1D'

```
function [alpha] = getalphanetworkdesign1D(currentDesign, proposedDesign, ...
                                        mean, covar, y, x, varMeasurement);
% This function calculates the acceptance probability alpha for the
% proposed design [a, b] when the mean and covar of the prior,
% the measurement y, and the measurement location x are given.

% Patrick de Man - MIT (June 2006)

[dummy, nParameters] = size(currentDesign);

% Calculate the prior probability from the multivariate normal distribution
currentPrior = ((2*pi)^(-nParameters/2)) * ( (det(covar))^(-0.5) ) * ...
        exp( -0.5 * (currentDesign - mean) * inv(covar) * ...
                                            (currentDesign - mean)' );

proposedPrior = ((2*pi)^(-nParameters/2)) * (det(covar))^(-0.5) * ...
        exp( -0.5 * (proposedDesign - mean) * inv(covar) * ...
                                            (proposedDesign - mean)' );

% Calculate the model prediction for the current and proposed designs
currentModelY = currentDesign(1) + currentDesign(2)*x;
proposedModelY = proposedDesign(1) + proposedDesign(2)*x;

% Calculate the likelihood probability
currentLikelihood = ( 1/(sqrt(2*pi*varMeasurement)) ) *  ...
        exp( -(y - currentModelY)^2 / (2*varMeasurement))   ;

proposedLikelihood = ( 1/(sqrt(2*pi*varMeasurement)) ) * ...
        exp( -(y - proposedModelY)^2 / (2*varMeasurement))   ;

% Calculate the acceptance probability
alpha = min(1, (proposedPrior * proposedLikelihood) / ...
                                    (currentPrior * currentLikelihood));

return; % function getalphanetworkdesign1D
```

Appendix E.
Survey of Chemical Microsensors

Motivation for this Survey

At the outset of the thesis research, the goal was to investigate the feasibility of devising an extensive monitoring network to measure and assimilate air pollution data on a large scale. Air pollution data, such as the temporal and spatial variations of several air pollution species, is generally obtained through major measurement campaigns requiring significant commitments of equipment and researchers. Therefore, the large amounts of data delivered over a prolonged period of time by an extensive monitoring network would be extremely valuable for air pollution research.

Conventional sensor technologies were not considered suitable for realizing this vision of an extensive monitoring network. Though technological sophistication enabled selective and sensitive measurement of air pollution species, the high cost and often large equipment size were two important factors hindering large-scale implementation of these conventional measurement techniques.

Recent advances in microelectronics, low cost solid-state sensors, telecommunications networks, and computing power seemed to have created an opportunity for deployment of large-scale environmental monitoring networks. Several small and inexpensive sensors were reviewed to identify potential candidates for building a monitoring network. The most promising candidate for implementation was the cermet sensor technology.

Unfortunately, laboratory experiments proved this cermet sensor unsuitable for the desired application. Therefore, this project involving microsensors was concluded and the thesis research shifted to the question of handling and processing large amounts of uncertain data generated from different measurement equipment.

Introduction to Chemical Microsensors

Microsensors have been under investigation now for many years. Their main advantages are that they are small (scale of micro- or millimeters) and inexpensive, allowing the deployment of an economically feasible dense measurement network. The discipline of microsensors started in the '70s when research into silicon sensors was initiated. Since then, various research groups around the world have developed many different kinds of physical, optical, and chemical microsensors. The performance of gas sensors compared to conventional analytical equipment is summarized in Table E-1. Literature offers a wealth of information on the various types of sensors and several reviews offer an insight into the latest developments [76-78]. Prevalent types of chemical microsensors, categorized by their principle of operation, are briefly described below.

Table E-1. Comparison between conventional analytical instruments and gas sensors [78]

	Analytical Instrument	Gas Sensor
Resolution	Excellent	Comparable
Cost	Very High	Fair
Size	Bulky (Factory)	Compact
Rigidity	Fragile	Rigid (replaceable)
Process control	Difficult	Easy
Mass Production	Difficult	Easy
Measurement	Instantaneous	Continuous

Conductivity

Conductivity is the most common principle of operation of a chemical gas sensor and is applied in commercial gas sensors since the '70s (e.g. www.figarosensor.com and www.citytech.co.uk). These sensors, also called chemiresistors, are devices with a semiconductor thick or thin film and rely on the change in conductivity upon interaction of the analyte gas with the semiconductor film. An example of a conductivity sensor is shown in Figure E-10. The material of the sensitive film is often a semiconductor like tin oxide or other types of metal oxides [79], while recently conducting polymers have been shown to be effective [77].

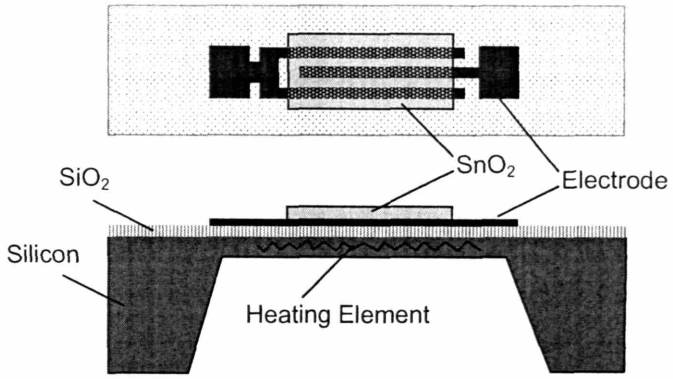


Figure E-10. Typical design of a tin oxide conductivity sensor (micrometer order of magnitude)

As semiconductors only become conductive after the electrons pass the band gap to the conduction band, the device is heated to supply the necessary energy. Chemiresistors are usually operated at temperature ranges of about 300-500°C. The mechanism is schematically illustrated in Figure E-.

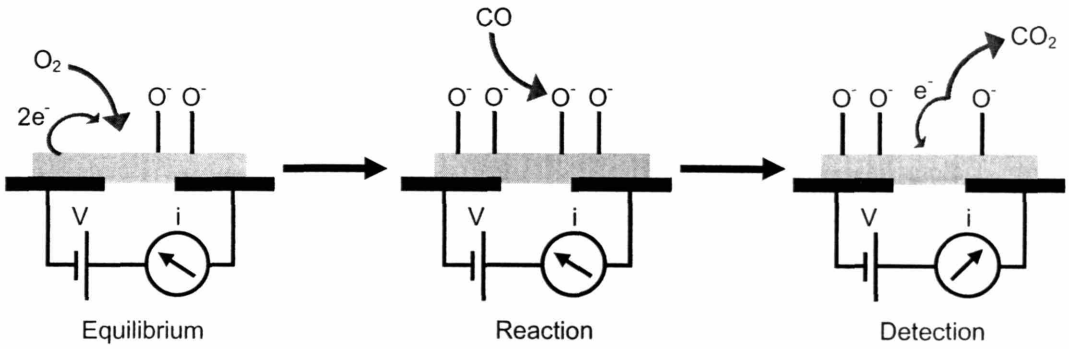


Figure E-2. Detection mechanism of chemiresistor sensors

In the equilibrium situation in absence of the analyte, oxygen from the air adsorbs to the film. These oxygen molecules thereby 'consume' electrons from the semiconductor film. When the analyte gas interacts with the film, these oxygen molecules are removed and the consumed electrons return to the semiconductor film, thereby increasing conductivity of the film. The change in conductivity depends on the gas concentration and calibration of the sensor with standardized gasses is necessary before application.

Potentiometric

CHEMFET

A common type of potentiometric sensor is based on the Field Effect Transistor (FET). When used to detect chemicals, the name is often CHEMFET [77, 80], of which a schematic is shown in Figure E-. When a positive potential is applied to the gate, electrons will be attracted to the semiconductor surface and form a conductive layer directly underneath the oxide. The n^+ source and n^+ drain are then connected by a conducting surface layer (or channel) through which a current can flow. The conductance of this channel can be modulated by varying the gate voltages.

The principle of a CHEMFET is based upon the interaction of the analyte with the sensitive layer deposited on the gate of the FET. This interaction, which is proportional to the concentration of the analyte in the surrounding environment of the active layer, changes the gate potential of the transistor, so that the current in the conductive channel is affected. The altering current is determined as a measure for the analyte concentration.

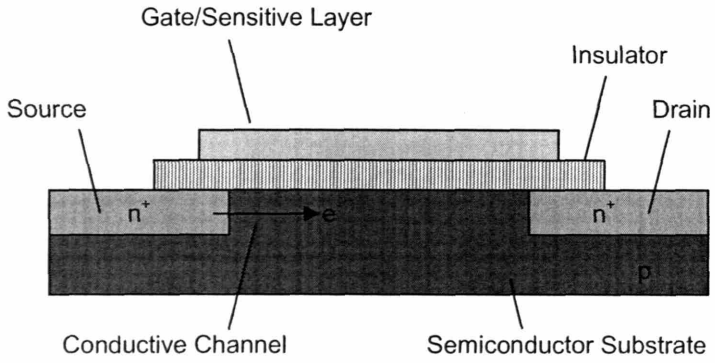


Figure E-3. Schematic of a CHEMFET sensor

Solid Electrolyte

The other main type of potentiometric sensor is the type of gas sensor used in the engine exhaust of every car to measure the oxygen concentration. Solid electrolyte, sandwiched between two electrodes, functions as ion conductor for ions traveling from one electrode to the other. The difference in chemical potential at the electrodes, due to concentration

differences of the analyte oxygen, is the driving force for the ion conduction. As there is no current flowing in the sensor, ion conduction builds up a potential difference between the two electrodes. This potential difference functions as a measure for the analyte concentration.

Amperometric

The design of amperometric sensors is very similar to the solid electrolyte potentiometric sensors. Again, solid electrolyte is sandwiched between catalytically active electrodes. In this case however, a constant voltage sufficiently high for the species of interest to react, is applied over the electrodes. The surface reaction of the analyte generates ions that are conducted through the solid electrolyte generating a specific current that is a measure for the concentration of the analyte. In addition to the sensor materials, the voltage is an important control of the selectivity of the species reacting on the electrode. Examples of an amperometric oxygen sensor are shown in Figure E-4.

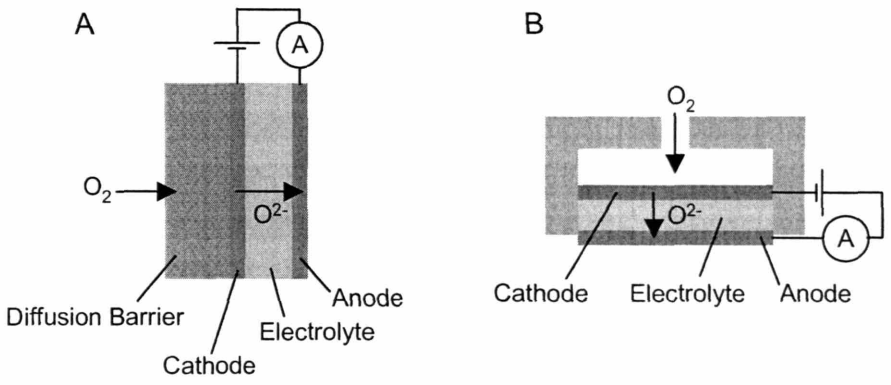


Figure E-4. Schematic representation of amperometric oxygen sensors

Amperometric sensors always have a diffusion barrier installed before the electrode, either a permeable material (Figure E-A) or a narrow opening (Figure E-B). The reason is to obtain a 'limiting current' sensor operation, meaning that the generated current is limited by the diffusion of the analyte towards the electrode. Diffusion limited operation leads to a linear relation between the bulk concentration of the analyte and the generated current, while the concentration of the analyte at the electrode surface is negligible.

Mass-Sensitive

The micromachined *cantilever beam* is a common type of mass-sensitive device that can be applied as chemical sensor. To apply this device an active layer is deposited on top of the cantilever beam. When the cantilever beam is exposed to the analyte, adsorption or absorption will result in a mass increase of the cantilever beam. The mass difference is measured by determining the change in frequency of vibration of the cantilever and can be related to the bulk concentration of the analyte [81, 82].

Another common mass-sensitive sensor is the *surface acoustic wave* sensor, in which the difference in wave behavior of the active layer is measured upon interaction with the analyte. An acoustic wave propagates through the active layer, where interaction with the analyte changes the wave's amplitude, frequency, and phase. These properties are measured as an indicator for the analyte's concentration [83].

The Cermet Microsensor

The cermet sensor is based on a patented ceramic-metallic technology developed by Argonne National Laboratory (ANL). After an extensive review of literature, the cermet sensor appeared to be the most promising on several performance criteria, such as a potentially high selectivity and negligible drift (the sensor cleans itself by periodically burning off all contaminants), an insignificant response to humidity, low in cost ($0.25 for the sensor, $10 for the accompanying electronics), and a long lifetime of approximately five years [84-86].

The sensor is a thick film device consisting of the following layers (shown in Figure E-):

1. Platinum (Pt) sensing electrode exposed to the analytes for reaction to occur,
2. Yttria stabilized zirconia (YSZ) solid electrolyte for conduction of the oxide ion involved in the reaction on the surface platinum electrode,
3. Platinum reference electrode,
4. Non-stoichiometric metal oxide that serves as a reference source for the oxide ions, in this case nickel/nickel oxide (Ni/NiO),

5. Ceramic substrate, in this case Alumina (Al_2O_3), and a heating element to raise the sensor temperature sufficiently to facilitate the ion conduction of the YSZ.

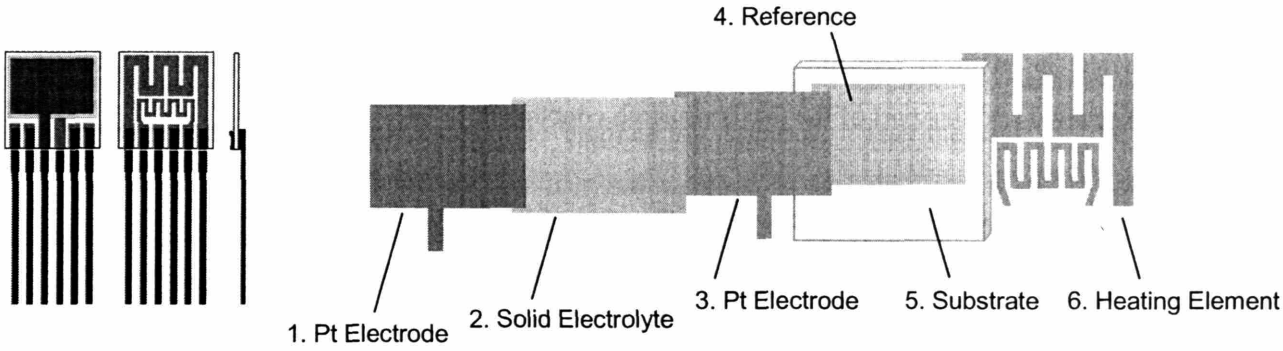


Figure E-5. Cermet sensor front, back, and sideviews and blown up schematic

This design of the sensor is very similar to the amperometric sensor as explained above. The main difference between these two types of sensor is the mode of operation. Whereas the voltage over the amperometric sensor is constant, the voltage over the cermet sensor changes according to a cyclic function, and thus the sensor is called voltammetric. Voltammetry, a rather uncommon principle for gas sensors, is a powerful electrochemical technique usually applied in chemical analysis of liquid electrolyte systems [87, 88]. Voltammetry creates fingerprint or signature signals for each component and for each mixture and is therefore a selective detection method.

As the cermet sensor seemed a suitable candidate, its detection and measurement capabilities were tested. Initial experiments subjected the sensor to a range of CO_2 concentrations in a background of nitrogen. Unfortunately, even under these idealized conditions (a constant temperature, an inert medium of nitrogen, and the presence of only one component at a constant concentration) the cermet sensor performed unsatisfactorily.

Conclusion of the Microsensor Survey

The cermet sensor was considered unsuitable for deployment in a large-scale monitoring network. Additionally, alternative sensor technologies were not sufficiently advanced to operate in practical, real-life conditions. Therefore, the search for a suitable sensor was suspended and this part of the thesis research concluded.

Instead of researching the hardware component of the envisioned monitoring network, it was decided to investigate the data handling and processing aspect of the large amounts of data to be generated by the network. The Bayesian approach was selected as the preferred methodology and further research conducted led to the results discussed in this thesis.

Appendix F. Ph.D.CEP Capstone
Probabilistic Evaluation of Petroleum Reserves

This thesis has demonstrated that the Bayesian approach is very valuable for parameter estimation involving complex chemical engineering systems. The message of the capstone is again that full probability distributions need to be evaluated in order to properly and consistently deal with uncertainty. This will be demonstrated with an example originating from petroleum exploration and production. After an extensive discussion on uncertain reserves arguing in favor of a probabilistic approach, the perspective of a petroleum company with regards to information on uncertainty will be used to optimize the production schedule of multiple reserves under the constraints imposed by a Volumetric Production Payment contract, a financial security on the company's reserves. The methodology and algorithms implemented to solve this optimal control problem were developed during the thesis research.

F.1. Problem Statement

The business of petroleum exploration and production (E&P) is extremely complex and fraught with numerous uncertainties. Characterizing uncertainty is important for making successful business decisions and petroleum companies have established internal procedures to probabilistically evaluate their petroleum assets. On the other hand, the Securities and Exchange Committee (SEC) merely requires, for the sake of standardization among the industry, the public reporting of deterministic estimates of petroleum reserves. As demonstrated in this thesis, by focusing on point estimates without considering uncertainty, important information will be ignored.

Obviously, the information available and what is required to make decisions depends on the stakeholder. A petroleum company would have access to and need more detailed information regarding their reserves in order to decide on the E&P projects than a

financial institution underwriting a security for the petroleum company. An investor would have the least information available and will most likely depend on the company's public reporting as required by the SEC. In order to judge the importance of these different degrees of information availability, the impact of uncertainty regarding petroleum reserves needs to be assessed.

Knowledge regarding the uncertainty of reserves and their production schedule is also essential to structuring a Volumetric Production Payment (VPP), a financial product commonly offered to monetize petroleum reserves. The current challenge is to design a product aimed at monetizing the reserves with a relatively high degree of uncertainty.

F.2. Introduction

This section gives a brief description of the dynamics of the E&P industry to explain the rationale for the increased interest in structured financial products. Secondly, the technical and economic assessments of a prospective reservoir are discussed as these are important background knowledge to structuring the financial products.

F.2.1. Petroleum Industry Dynamics

Research has shown that oil field sizes generally follow a lognormal distribution [89] so that relatively few large fields and many small fields occur. Obviously, large and readily observable oil fields have the highest probability of being found so that over time smaller fields remain undiscovered, making the discovery of new oil fields more difficult. The discovery of a large oil field in the US is very unlikely, as already approximately 3 million out of the estimated worldwide total of 5 million oil wells have been drilled.

Exploiting the small oil fields is unprofitable for the major oil companies, such as ExxonMobil and Shell, which have subsequently implemented more advanced technology in search for highly profitable oil fields. With the large oil companies' move to capital intensive projects such as deep-water drilling in the Gulf of Mexico, the second and third tier companies with relatively low overhead costs have been exploiting the

remaining smaller fields in the US. Even with the recent high oil prices, the balance sheets of these smaller oil companies' are generally not as strong, creating a demand for financial products facilitating risk management.

F.2.2. Technical Assessment of Reserves

The assessment of a prospective oil reservoir moves through three stages to determine: 1) an estimate of the total amount of oil in place, 2) an estimate of the recoverable amount of oil, and 3) a likely production rate schedule. Each of these stages together with the typical estimation methods [90] will be discussed in more detail below.

Estimating the Total Amount of Oil in Place

Before drilling an actual well to confirm the presence of oil, information about the reservoir is obtained from analogy and seismic data. Analogy compares the prospective reservoir with reservoirs of similar geology and/or geographic proximity. Seismic data would be able to coarsely identify a subsurface volume potentially containing petroleum, though the presence can not be guaranteed without drilling. When exploratory wells have actually confirmed the existence of petroleum, multiple wells will be drilled to delineate the reservoir and calculate a more accurate estimate for the total oil-in-place.

Estimating the Recoverable Amount of Oil

The second goal of exploratory drilling is to gain information regarding physical properties of the oil and subsurface characteristics of the reservoir. Well-logging and core sample analyses are used to estimate parameters such as porosity and permeability of the reservoir. The pressure observed in the well and performance data obtained from a short-term production test run also contribute to a better understanding of the reservoir. The information from these analyses is used to calculate the physically recoverable amount of oil from the reservoir using volumetric methods.

Determining a Production Schedule

Determining possible production schedules requires forecasting of the reservoir performance. Because of the extensive experience in the petroleum industry with

exploiting reservoirs, the methodologies to forecast future production are well established. Obviously, the more production data is available, the better the required model parameters can be estimated. Three types of performance methods are available:

1. Material-balance calculations for the reservoir volume based on the physical reservoir characteristics and petroleum properties.
2. Reservoir simulation based on reservoir engineering models implementing a material balance for each element in the 3D grid representing the reservoir. Significant expertise is required for detailed simulators and over time a variety of advanced reservoir engineering software packages have become commercially available (e.g. Eclipse from Schlumberger). An interesting comparison of methods for production forecasting including uncertainty has been done in the PUNQ project [91].
3. Production decline curves are based on historical production data and propagate the reservoir production rate or well-pressure as a function of time according to a fitted function.

F.2.3. Economic Assessment of Reserves

Assessing the economic viability of an exploration and production project builds on the technical assessment discussed above.

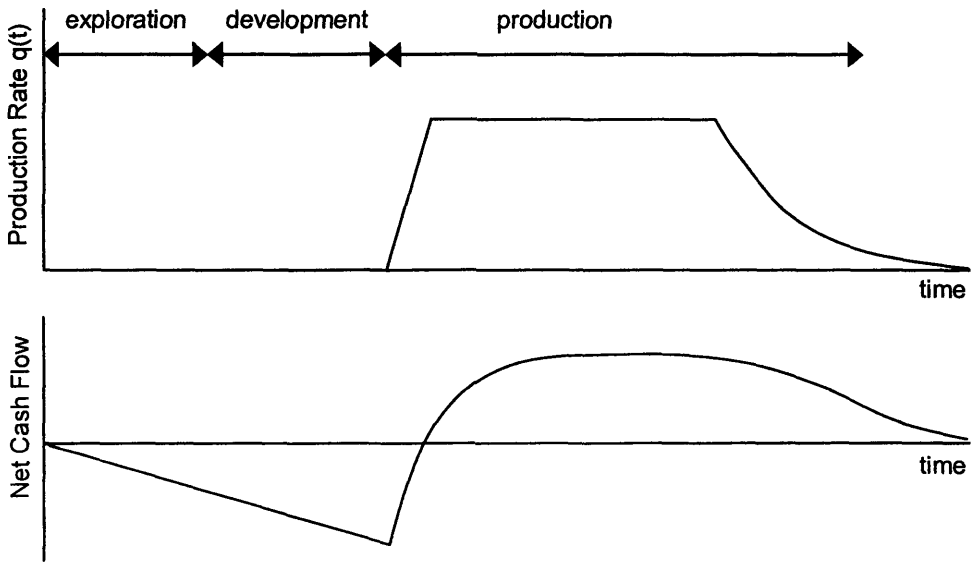


Figure F-1. Simplified production rate curve and accompanying net cash flow

After forecasting a production schedule, revenues and costs for exploitation of the reservoir can be estimated [92]. A simplified production and net cash flow schedule is shown in Figure F-1.

Revenues depend on the forecasted production schedule and estimated future prices of petroleum. Besides market price fluctuations, the quality of the petroleum is an important determinant. Attributes such as whether the petroleum is sweet/sour, light/heavy, or conventional/unconventional significantly impact the attractiveness and thus the price.

Costs need to be estimated for each of the exploration, development, and production stages of the project. Both fixed and variable costs can change over time and depend on issues such as the geographical location (transportation cost), size, depth, and other physical characteristics of the reservoir.

When future revenues and costs have been estimated, the Discounted Cash Flow (DCF) method is generally applied for valuation of properties for which a reasonable production forecast is possible. Figure F-2 illustrates the general steps involved in performing a DCF valuation.

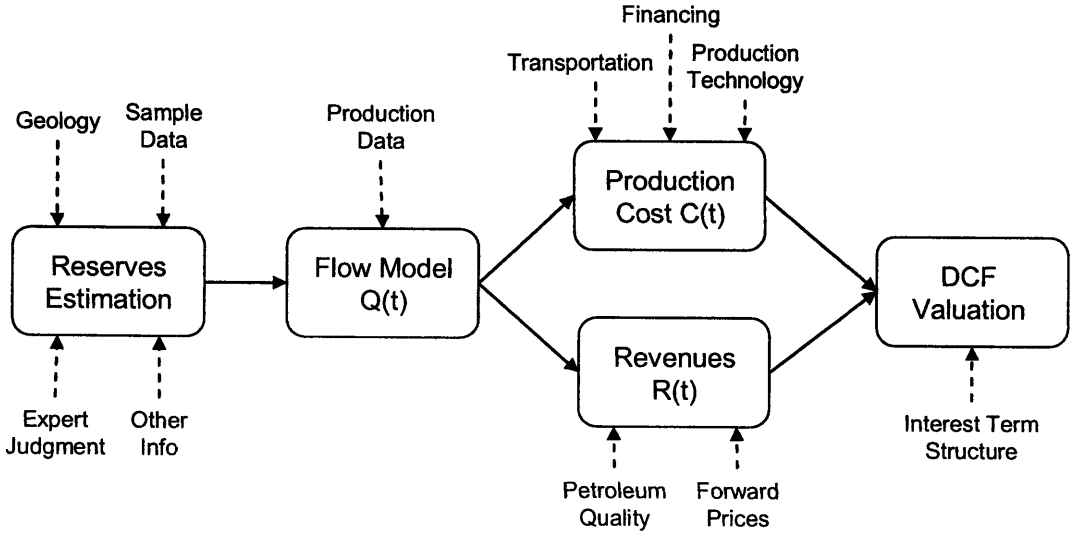


Figure F-2. Overview of a reserves valuation by the DCF method

The DCF valuation method is acceptable when uncertainty in the cash flows is absent or relatively small. Obviously, though companies apply the DCF as the base case to estimate the value of proven reserves for purposes of reporting to the public, internal more advanced valuation methods are often implemented, such as real option analysis and portfolio theory [93, 94].

F.3. Resources and Reserves

As mentioned above, uncertainty is inherent to the E&P industry and this section will give an overview of how uncertainty has traditionally been characterized. After a broad discussion regarding resource classifications and reserves definitions, the SEC definition for proven reserves will be discussed in more detail. Finally, the practical aspects of reserves estimation will be illustrated, followed by the general rationale of why companies engage in reserves estimation.

F.3.1. Resource Classification

Petroleum companies are always undertaking exploration projects to maintain their inventory of oil fields for future production. Exploration occurs in various stages and resources under consideration will be categorized accordingly. The Society of Petroleum Engineers (SPE) provides independent guidelines to promote international consistency in total resource assessment and most companies have included these in their internal systems. Indeed, a recent comparison found that these guidelines are very similar throughout the world. The main resource categories are schematically shown below [95].

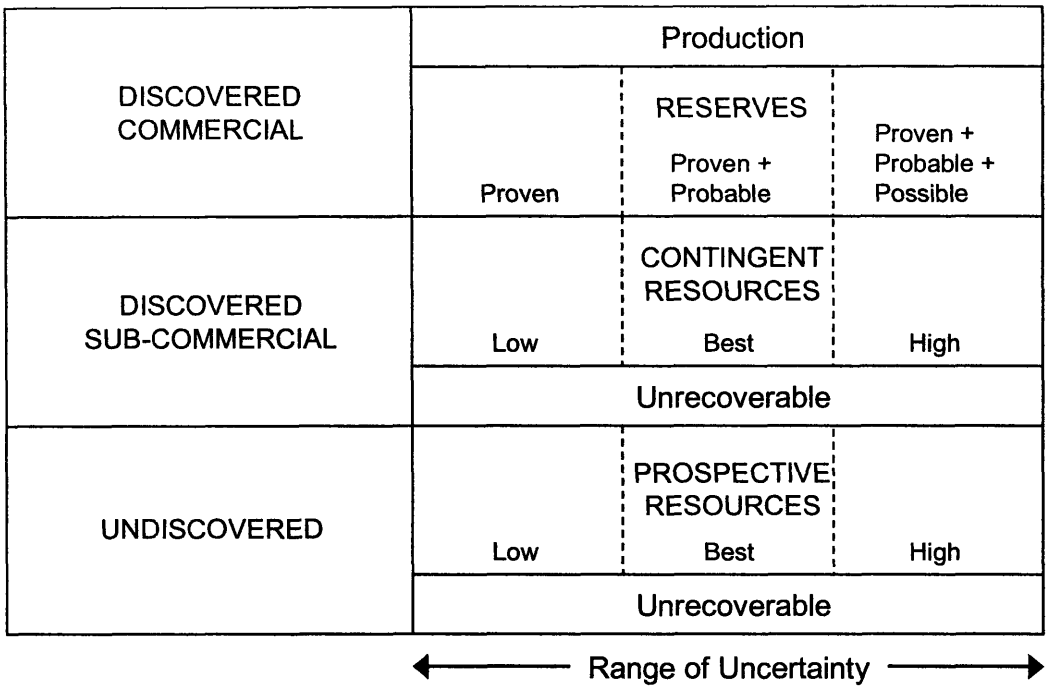


Figure F-3. Schematic overview of the main resource categories

Obviously, there is uncertainty with regards to the resource quality or quantity and categorization can change based on new information. Unfortunately, all parties categorizing according to Figure F-3 have different subjective interpretations of what each category represents. As full probability distributions are not evaluated, there is no unified approach to characterizing uncertainty.

F.3.2. Proven Reserves Subcategories

Among the categories shown above, the discovered commercial resources have been investigated most thoroughly by data collection and modeling, in particular the proven reserves. The proven reserves category is important to forecasting near term production and is thus of most interest to the stakeholders of oil companies.

As shown above, the category of reserves is generally subdivided into deterministic estimates: a low estimate, a best estimate, and a high estimate. The SPE categorization

specifies these scenarios as P90, P50, and P10 estimates or according to the terms proven, probable, and possible reserves, as shown in Figure F-4.

A P90 estimate generally indicates there is a 90% chance that the estimated quantity can indeed be recovered (ironically, the exact opposite naming convention also occurs, where a P10 estimate represents a 90% chance). These estimates are based on data regarding reservoir characteristics, but also on subjective criteria, such as expert judgment of reservoir engineers. In any case, these estimates represent a degree of uncertainty, but not in the rigorous sense as derived from a probability distribution.

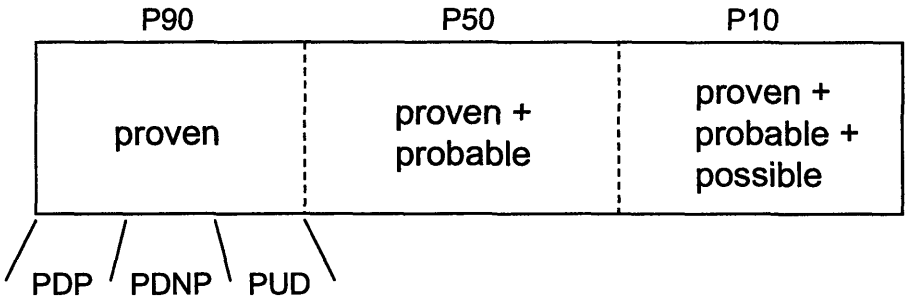


Figure F-4. Overview of reserves subcategories

Within the category of proven reserves, the state of the reserves regarding development and production is differentiated by a further categorization. The three categories in use are Proven Developed Producing (PDP), Proven Developed Non-Producing (PDNP), and Proven Undeveloped (PUD) reserves, as indicated in Figure F-4. The significant distinction between the developed and undeveloped reserves is that the former is expected to be exploited through existing wells with equipment in place, while the latter requires a capital investment before exploitation. Development of PUD reserves is typically done by 1) drilling of wells, 2) deepening of existing wells, or 3) installation of improved recovery systems.

F.3.3. The Uncertainty in Proven Reserves

As discussed, the proven reserves are generally considered equivalent to P90 estimates, indicating a 90% chance that at least the indicated quantity of petroleum can be recovered. Though this is an expression of uncertainty, the P90 value is usually given as a point estimate without information regarding the underlying distribution. Subsequent application of these P90 point estimates in deterministic calculations will lead to erroneous results and misrepresentation of the actual uncertainty in these results, as was illustrated in Section 2.3.2. Misinterpreting probabilistic information as deterministic relates to what is explained by the 'Flaw of Averages', stating that "plans based on *average* conditions are wrong on *average*" [96, 97].

Consistent treatment of uncertainty thus requires a probabilistic approach. The P90 value can be interpreted as a cumulative probability:

$$\text{P(reserves contain } x \text{ amount of oil)} = \text{P(oil amount} \geq x) = 0.9$$
$$\Rightarrow \text{ P(oil amount} \leq x) = 0.1$$

For the purpose of illustration, Figure F-5 relates the P90 value to the Gaussian distribution.

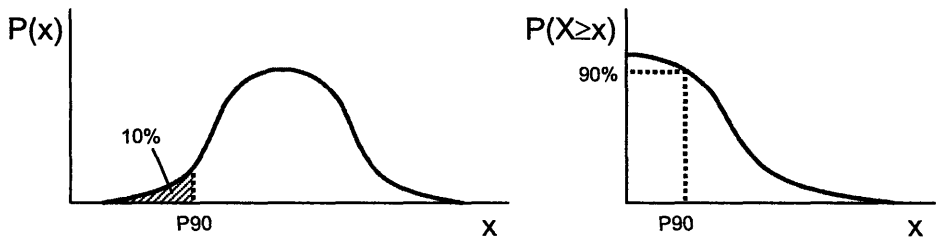


Figure F-5. The P90 estimate obtained from a normal probability and cumulative distribution

Again, knowledge of the underlying distribution characterizing the uncertainty in reserves quantities is important to properly propagate P90 values through any calculation where uncertainty regarding the outcomes is important.

F.3.4. Probable and Possible Reserves

The very definition of probable and possible reserves implies a relatively large uncertainty. Probable and possible reserves are areas of an oil field suspected to contain petroleum. Wells have not yet been drilled, so the presence of petroleum is not confirmed and thus underground conditions such as porosity and permeability are unknown. The only information available would be mostly seismic data and perhaps indirect clues because of nearby producing wells. In recent years, seismic analyses have become more advanced (3D and even 4D seismic is available, but at high cost), but these data can not yet guarantee identifying underground structures that contain petroleum. For example, even with strong indications for the presence of petroleum in a suitable structural trap, drilling can still lead to a dry hole, as hydrocarbons have migrated from the reservoir over time through an undetectable small crack.

F.3.5. SEC Definition of Proved Reserves

The criteria for proved reserves (equal to proven reserves) according to the SEC are part of the financial accounting regulations, which apply to publicly traded oil and gas companies. The goal is to provide consistent value assessments so that investors can compare financial performance among different companies. The SEC definition for proved reserves is:

> *"Proved oil and gas reserves are the estimated quantities of crude oil, natural gas, and natural gas liquids which geological and engineering data demonstrate with reasonable certainty to be recoverable in future years from known reservoirs under existing economic and operating conditions, i.e., prices and costs as of the date the estimate is made."*

Important to note is that the reserves estimate is dependent on existing economic and operating conditions. In other words, when market prices for oil and gas fluctuate or when technology improves, reserves estimates will change accordingly. However, for reserves reporting prices are established only at year-end and kept constant throughout a calendar year. Such requirements have recently been criticized as being outdated in the current dynamic oil and gas markets [98, 99]. This is a valid criticism as the SEC

regulation dates from 1978, while the petroleum industry and oil and gas trading have gone through significant developments since then.

Unfortunately, the SEC only demands a deterministic estimate, without any indication regarding uncertainty. With the mere definition that *estimated quantities are recoverable with reasonable certainty*, there are no guidelines on the methodology to determining reserves estimates. Each reporting petroleum company can implement a different estimation method, either more heavily relying on subjective valuation or on objective data analysis. Unfortunately, without the information on uncertainty, public information regarding reserves becomes less useful in assessing the risk regarding the company's petroleum assets.

F.4. Reserves Estimation and Uncertainty

Only recently the consideration of uncertainties in reserves estimation has become more prevalent. This section will, after a brief and general discussion on reserves estimation, illustrate the impact of reducing uncertainties according to two historical accounts. The discussion will then focus on probabilistic estimation, followed by a demonstration of the importance that the underlying distribution is known.

F.4.1. Uncertainties in Reserves Estimation

Estimation of petroleum reserves is needed internally to plan future exploration, development, and production activities, while externally it forms the basis for the valuation of the company's assets. As discussed, estimating reserves is a difficult undertaking, being both an art and a science. Hard data from seismic measurements, core analyses, and well logs have to be combined with information from expert judgment in the estimation procedures. Complicating the situation is that the data contain measurement errors and experts have different levels of experience (and thus different levels of accuracy/precision). Consequently, the estimated quantity of petroleum in the underground reservoir will always be uncertain. Typically, the uncertainty of a particular reserves estimate is largest at the start of an exploration project, and will decrease over

time when more data is collected (e.g. drilling, core sample analyses, and production performance).

F.4.2. Importance of Uncertainty in Reserves Estimates

Improved information leading to a reduction in the uncertainty regarding reserves has in the past shown to be of significant impact. This section will discuss the effect on financial markets for two such cases.

Reserve Value Disclosure and Bid-Ask Spreads

In 1978 the SEC mandated the disclosure of the discounted present value of reserves for oil and gas companies. Once the information regarding the reserves was publicly available from 10-K filings, the bid-ask spreads of the companies' common stock declined significantly [100]. While the previously informed trader with access to non-public information was able to earn above normal profits, the required disclosure reduced the information asymmetry. Market-makers experienced a reduction of losses to the informed traders (i.e. adverse selection costs), enabling them to reduce the bid-asks spreads. The lower spreads translate to value for investors as transaction costs will decrease. Additionally, the reduced bid-ask spreads can increase liquidity and trading volume, thereby reducing the illiquidity premium demanded by shareholders.

The Impact of Incorrect Reserves Estimation: Royal Dutch Shell

Royal Dutch Shell announced on January 9, 2004 to reclassify a part of their reserves, causing a decrease of 3.9 billion barrels of oil equivalent in their proven reserves. While the reclassification decreased the proven reserves by approximately 20% in amount of oil, the value of the proven reserves declined by 10% as calculated with the standardized discounting method of FAS69 [101, 102].

Since reported reserves are an important guideline for investors to forecast future earnings for companies in the oil and gas industry, the share price of Royal Dutch Shell traded down about 9% during the three days after the announcement. After this initial response by the financial markets, further investigation revealed improper practices with

regard to the reserves estimation. Management was eventually to blame, leading the ousting of the chairman and the chief executive of exploration and production.

Though most media attention went to blaming Shell's management for its malpractices, the difficulty of estimating reserves was also recognized and some called for an overhaul of the outdated and arcane estimation guidelines as imposed by the SEC [98].

F.4.3. Probabilistic Reserves Estimates

Deterministic estimates (a single 'hard number') are certainly easy to understand, but the inherent uncertainty in estimating petroleum reserves necessitates a probabilistic approach. Probabilistic estimation would entail full probability distributions, retaining all available information regarding the quantification of the reserves. Therefore, risk can realistically be assessed, so that the probabilistic approach generally leads to better informed decision-making.

The importance of a probabilistic approach was recognized by the 'Oil and Gas Reserves Committee' of the SPE in their recommendation to accommodate probabilistic assessment methods in the definitions and guidelines regarding oil and gas reserves [95]. In addition, some industry experts recently argued for the inclusion of probabilistic methods to complement the currently deterministic reserves estimation [99].

F.4.4. Knowledge of the Underlying Distribution

A deterministic estimate such as the expected value or a P90 value is only of limited use when the underlying distribution, from which this single number has been derived, is unknown. Characterizing uncertainty by a standard deviation σ is not sufficient, either, unless the Gaussian distribution is appropriate. For many systems, however, the Gaussian distribution is not applicable as values below 0 are physically impossible. For example, many of the uncertain parameters of a reservoir model (such as permeability, porosity, and saturation) are distributed according to lognormal distributions.

The following example illustrates the importance of knowing the actual shape of the underlying distribution, instead of merely the mean and variance obtained from a particular data set. With X~ N(μ,σ^2), the lognormal distribution can be derived by exp(X)~LogN(μ,σ^2) and can thus be specified by the μ and σ^2 of the normal distribution. However, note that μ and σ^2 of the normal distribution are not equal to the mean and variance of the lognormal distribution. The summary variables for the lognormal distribution, μ_{LogN} and σ_{LogN}^2, are defined as follows:

$$\mu_{LogN} = e^{\mu+\sigma^2/2}$$
$$\sigma_{LogN}^2 = e^{\sigma^2+2\mu}\left(e^{\sigma^2}-1\right)$$

With these relationships the difference in shape between the normal and lognormal distributions can be easily compared, as shown in Figure F-6 where $\mu_{LogN}=\mu$ and $\sigma_{LogN}^2=\sigma^2$.

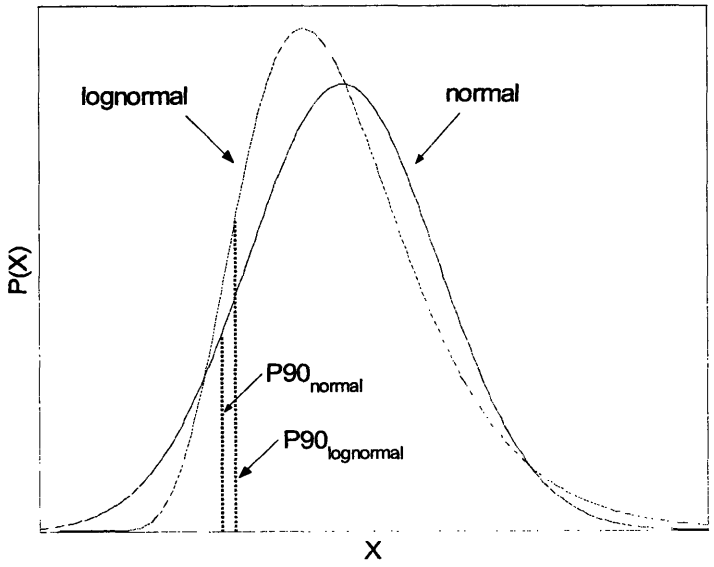


Figure F-6. Comparison of the normal and lognormal distribution and their P90 values

In addition to the normal and lognormal distributions, the respective P90 values are indicated. Though the mean and variance of both distributions are equal, clearly the P90

value derived from a normal distribution is incorrect when in fact the underlying distribution is lognormal. This is a strong argument in favor of the probabilistic approach to reserves estimation.

F.5. Uncertain Reserves and Financial Products

As mentioned, there has been increasing interest in structured financial products for risk management. The perception of risk is however dependent on the perspective of the stakeholder. Three perspectives will be discussed, before continuing the discussion with a focus on the objectives of a petroleum company. The financial product discussed in this section is the Volumetric Production Payment (VPP) contract. The VPP is already a relatively standard product and the opportunity to extend this product to more uncertain reserves will be illustrated. Extending the VPP requires a probabilistic approach to solving the optimal control problem regarding petroleum production from several uncertain reserves. The benefits of the probabilistic approach evaluating full probability distributions will be demonstrated according to an example.

F.5.1. Perspectives of Different Stakeholders

The valuation of reserves will be affected by the availability of information, particularly regarding the uncertainty of the reserves estimate. This section will discuss the perspective to valuing petroleum reserves by three stakeholders, namely a petroleum company, a financial institution, and an investor.

Petroleum Company

The ultimate goal of an exploration and production company is to maximize their NPV of petroleum production by deciding which projects to undertake. Obviously, the E&P activities require detailed information to generate probabilistic reserves estimates and production forecasts. Thorough characterization of the uncertainty is extremely important as the company's profitability is directly related to the projects they undertake.

Financial Institution

The valuation of reserves is relevant for a financial institution when having E&P companies as clients. A variety of services are provided by financial institutions, such as the arrangement of private loans, the pricing and placement of bonds, and the structuring of financial risk management products. For all of these services, reserves are often considered as collateral, so that an accurate valuation is important. However, financial institutions will be most interested in the aggregate petroleum reserves and production, so that the idiosyncratic risk at the level of individual reservoirs and wells is diversified away. Most importantly, petroleum production should provide cash flows to keep the company profitable, or at least solvent.

This being said, more accurate and precise estimates regarding the petroleum reserves and production would allow the financial institution to fine-tune their financial securities thereby creating better risk management tools.

Investor

The decision facing an investor is whether to invest in company A or B. Currently, the only information available regarding reserves are the quarterly and annually reported deterministic estimates without an indication regarding uncertainty. The effect of the absence of uncertainty information can be imagined by merely considering the expected or mean values of the optimal production profiles in the example below. Unfortunately, the investor will be unable to perform a systematic and coherent risk assessment regarding of his prospective investments.

F.5.2. Volumetric Production Payment

The Volumetric Production Payment (VPP) is a fairly established product to monetize proven reserves of a petroleum company, the perspective of which will be the focus in the following sections. The VPP involves an upfront payment to an oil company, which subsequently uses these funds to exploit its proven reserves and/or undertakes exploration and production projects. The oil company repays the buyer of the VPP with produced

petroleum according to a pre-arranged schedule of periodic oil payments. The receiver of the oil payments will have hedged the oil price over the duration of the payment schedule.

The VPP contract allows the petroleum company to take on a loan using its reserves as collateral. Compared to settling debt with cash obtained from petroleum sales, the company will not be affected by fluctuations in the price of oil as the commodity will be delivered as repayment. The second main advantage is that a VPP contract is by accounting standards not categorized as debt and therefore does not appear on the balance sheet. Thus, the company is provided with liquidity to engage in the capital intensive investments required for production, while the reported financial state of the company (represented by e.g. the debt/equity ratio) is not affected. Nevertheless, rating agencies have started to include VPP contracts when evaluating the credit rating of oil companies.

The VPP contract is mostly structured upon the information regarding Proved Developed Producing (PDP) reserves. The main reason for successful structuring of the VPP contract is that PDP reserves characteristics are a well-studied area within reservoir engineering, so that production forecasts and profiles contain little uncertainty. Figure F-7 schematically shows a forecasted production profile for PDP reserves (according to the production rate curve of Figure F-1), as well as the required production level for delivery of the periodic payments according to the VPP contract. In order to further minimize default risk, the VPP is structured with both a safety cushion regarding the expected production level, and the expected lifetime of the PDP reserves, as indicated by the arrows.

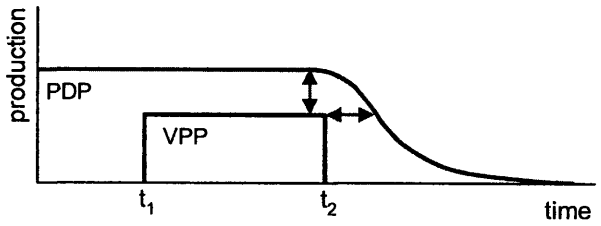


Figure F-7. Production profile of the PDP reserves and illustration of the VPP contract

F.5.3. Opportunity to Extend the VPP

As discussed above, the VPP is primarily structured on PDP reserves. The current challenge is to extend the concept of the VPP to include the next level in the proven reserves: Proven Developed Non-Producing (PDNP) and Proven Undeveloped (PUD) reserves. As of now, such a product is not on the market, thus providing an interesting opportunity.

Figure F-8 schematically illustrates the extension of the VPP contract to include each of the three components of the proven reserves: PDP, PDNP, and PUD reserves. The level of production payments between time t_1 and t_2 as specified by the VPP contract are raised to include production from the PDNP and PUD reserves, respectively. One advantage of extending the VPP contract thus entails the opportunity for providing larger loans, which could be of interest to certain petroleum companies.

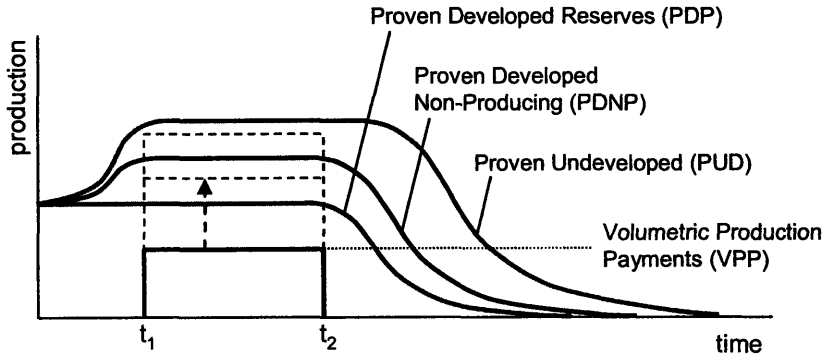


Figure F-8. Opportunity of including PDNP and PUD reserves into a VPP structure

However, since production from PDNP and PUD reserves is significantly more uncertain than production from PDP reserves, especially since the presence of petroleum has not even been proven in the case of PUD reserves, the extension of the VPP requires additional consideration. As the figure above does not portray any uncertainty, the production profiles in relation to their probability distributions are illustrated in Figure F-9 for the time interval between time t_1 and t_2.

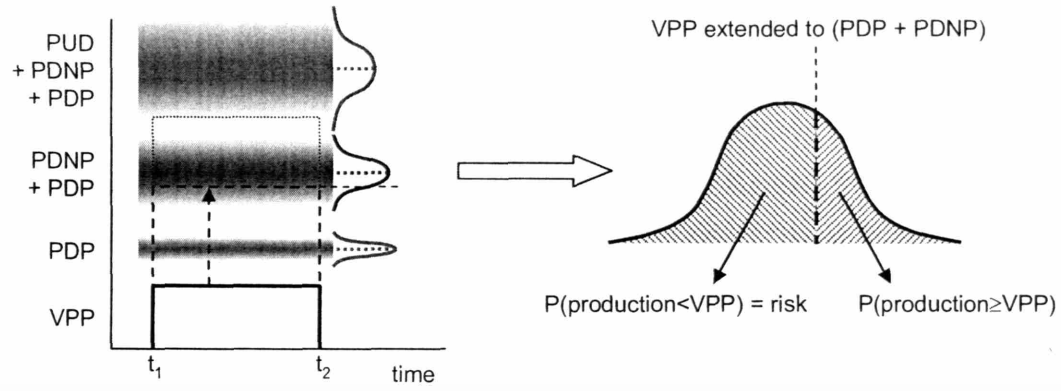


Figure F-9. Different degrees of uncertainty within proven reserves and the VPP risk definition

With the degree of uncertainty increasing upon including PDNP and PUD reserves in the VPP contract, the contract underwriter will be exposed to a higher degree of risk. As shown above, risk is in this case defined as the probability of default, occurring when the quantity of produced petroleum is insufficient to fulfill the periodic payment obligation. To minimize the probability of default, the VPP contract would be structured to leave a considerable margin of safety between the payment level and the expected production profile. There are currently no rigorous approaches to determine such margin of safety. Instead, petroleum payment levels for a VPP contract structured on PDP reserves are determined coarsely to stay well below the expected production profile.

The opportunity lies in properly characterizing the production profile uncertainty, generating full probability distributions, attaining an accurate risk assessment, and simply establishing the margin of safety according to an acceptable probability of default. The benefit of such a consistent probabilistic approach allows for structuring more attractive securities, thereby creating a competitive advantage for the VPP contract underwriter.

F.5.4. Access to E&P Information

To engage in structuring an extended VPP contract as discussed above, detailed information on the reserves, as well as E&P projects of the petroleum company would have to be available. Extension to include PDNP reserves should be fairly straightforward as production can be initiated easily. However, the further extension to

include PUD reserves requires considerable additional information to characterize the uncertainties regarding E&P activities.

Traditionally, the VPP contract underwriter obtains the necessary information regarding expected production levels of the PDP reserves from petroleum consultants who have (limited) access to the company's data. The extension of the VPP contract obviously demands a more intensive collaboration between the company as client and financial institution as underwriter.

F.5.5. Simulation-Based Optimal Design of a Production Schedule

Suppose a petroleum company has PDNP and PUD reserves and wishes to enter into a VPP contract for financing exploitation. The production profiles for these reserves are forecasted according to the curve in Figure F-1. Following a constant production plateau, the diminishing production is assumed to follow an exponential decay model. The exponential decay model is specified as

$$q(t) = q_i \exp(-d_i t) \tag{F-1}$$

where $q(t)$ is the production rate in bbl/year in year t calculated with the initial flow rate q_i in bbl/year and a constant decay rate d_i in year^{-1}. From the production rate $q(t)$, the cumulative annual production $Q(t)$ in bbl can be calculated by integration over time. The model parameters q_i and d_i specifying the production profile are known with uncertainty characterized by $q_i \sim N(\mu_q, \sigma_q^2)$ and $d_i \sim N(\mu_d, \sigma_d^2)$. These uncertainties in the model parameters propagate through equation (F-1) and cause uncertainty in the production profile.

Understandably, the company's main objective is to maximize the expected Present Value (PV) of their future production, to be calculated as

$$E[PV] = \frac{1}{J}\sum_{j=1}^{J}\left(\sum_{t}\frac{PQ(t)}{(1+r)^t}\right) \tag{F-2}$$

where the price P is assumed fixed over time, r is the discount rate, $Q(t)$ is the uncertain cumulative production of the reserves in year t, and J is the number of samples generated for parameters (q_i, d_i).

The challenge is to determine the order of exploiting the wells with the objective to maximize $E[PV]$, under the constraint that the periodic future payments according to the VPP contract can be fulfilled. Since the petroleum production is uncertain, this optimal control problem with constraints needs to be formulated probabilistically as follows

$$\begin{aligned} &\max_{t_1, t_2, \ldots, t_n} E[PV] \\ &s.t.\ P(Q(t) \geq VPP) \geq 0.95 \qquad \forall t \in [t_{start}, t_{end}] \end{aligned} \tag{F-3}$$

where t_i is the initiation year for well i, and VPP is the level of the annual production payment between t_{start} and t_{end} of the contract. An intricacy is that the existence of petroleum still has to be validated for PUD reserves. Thus exploration might lead to the realization that petroleum is not present after all. This uncertainty is introduced in the formulation as a binomial probability as

$$\begin{aligned} p(oil) &= p \\ p(dry) &= 1 - p \end{aligned} \tag{F-4}$$

where p is the probability of success in actually finding and producing petroleum at the forecasted rate $q(t)$. If the exploration results in a dry hole, the implemented production rate is $q(t) = 0$ for the particular well.

The diagram below in Figure F-10 schematically illustrates the optimal control problem with constraints under uncertainty for the implementation involving reserves A and B, categorized as PDNP and PUD reserves, respectively.

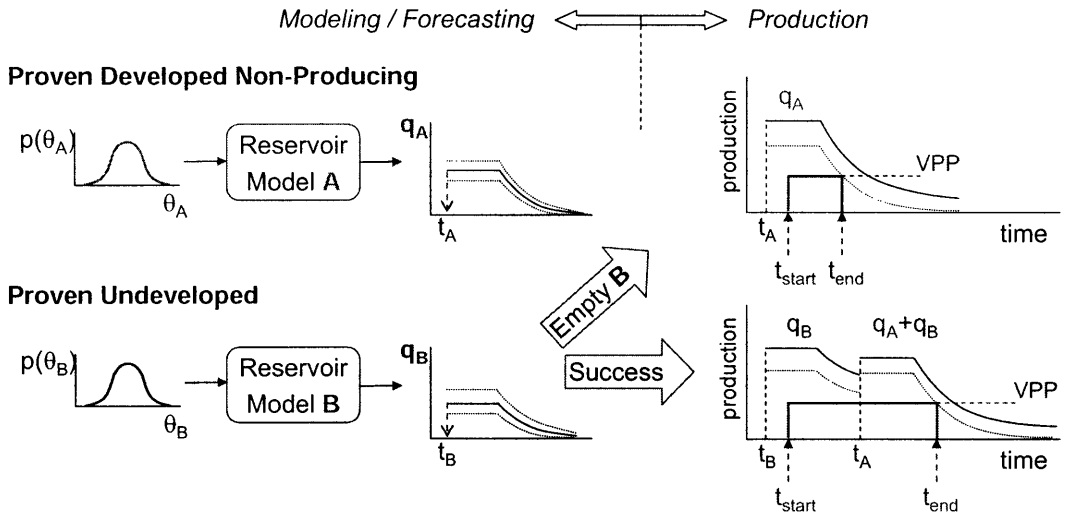


Figure F-10. Uncertain profiles for reserves A and B and their production under the constraint of the VPP

The two possible outcomes point out a significant difference with regards to structuring the VPP contract. For the same payment level, the duration of the contract is shorter if *B* turns out to be empty. The second important observation is that the payment level is lower under the one-sided probabilistic constraint of equation (F-3), compared to the payment level based on the constraint of expected value of petroleum production (as represented by the solid profiles).

The problem formulated above involving reserves *A* and *B* can be solved using the approach of simulation based optimal design as described in Muller [103, 104]. The algorithm is a variation of the MCMC simulation applied to solve the Bayesian parameter estimation problems as discussed in Chapter 4 of this thesis with the main difference that a utility function is evaluated at each simulation step instead of the posterior probability. The algorithm for simulation based optimal design regarding $\eta = \{t_1, t_2, \ldots, t_n\}$ for a number of n reserves is specified as follows:

1. Initialize counter $i=1$, specify an initial design η^0 for which $u^0 \neq 0$, and set $\eta^i = \eta^0$
2. With design η^i evaluate u^i

3. Generate the proposed design η^* from a probing distribution PD($\eta^* | \eta^i$) centered on η^i and evaluate u^*
4. Calculate the acceptance probability:

$$\alpha = \min\left\{1, \frac{u^*}{u^i}\right\} \tag{F-5}$$

5. Generate $v \sim$ Uniform(0,1) and accept or reject the proposed design:

$$\begin{aligned} &\textit{if } \alpha \geq v, \textit{ then accept}: \quad \eta^{i+1} = \eta^* \\ &\textit{if } \alpha < v, \textit{ then reject}: \quad \eta^{i+1} = \eta^i \end{aligned}$$

6. Increase counter i=i+1 and repeat step 2 to 5

This algorithm will step through the design space $\mathcal{H}$ while sampling in the direction towards a higher utility, calculated by the utility function u defined as

$$u = \begin{cases} E[PV] & \textit{if } P(Q(\eta,t) \geq VPP) \geq 0.95 \quad \forall t \in [t_{start}, t_{end}] \\ 0 & \textit{if } P(Q(\eta,t) < VPP) \geq 0.95 \quad \forall t \in [t_{start}, t_{end}] \end{cases} \tag{F-6}$$

The optimal design is identified as the design resulting in the largest value of $E[PV]$ achieved during the simulation.

F.5.6. Example Problem Results

Before showing the results obtained from implementation of the algorithm above, the details of the optimal control problem will first be specified. The characteristics of reserves A and B are given in Table F-1. The VPP contract requires an annual payment of 6,000 bbl/year for $[t_{start}, t_{end}] = [2, 14]$.

Table F-1. Specification of reserves A and B

	A	*B*
q_i (bbl/day)	110	80
σ_{q_i}	$0.1q_i$	$0.1q_i$
d_i (month^{-1})	0.011	0.018
σ_{d_i}	$0.1d_i$	$0.1d_i$
plateau $q(t)$	$0.25q_i$	$0.25q_i$
p(oil)	100%	40%

The resulting annual production $Q(t)$ for the optimal design (t_A = 0.12, t_B = 0.23) in Figure F-11 demonstrates that the VPP payments can each year be made with at least 95% probability. It is important to note that this one-sided 95% probability interval, as calculated by integrating over the full probability distribution according to equation (3-2), is different from the one-sided 95% confidence interval, based on the mean and standard deviation of the results.

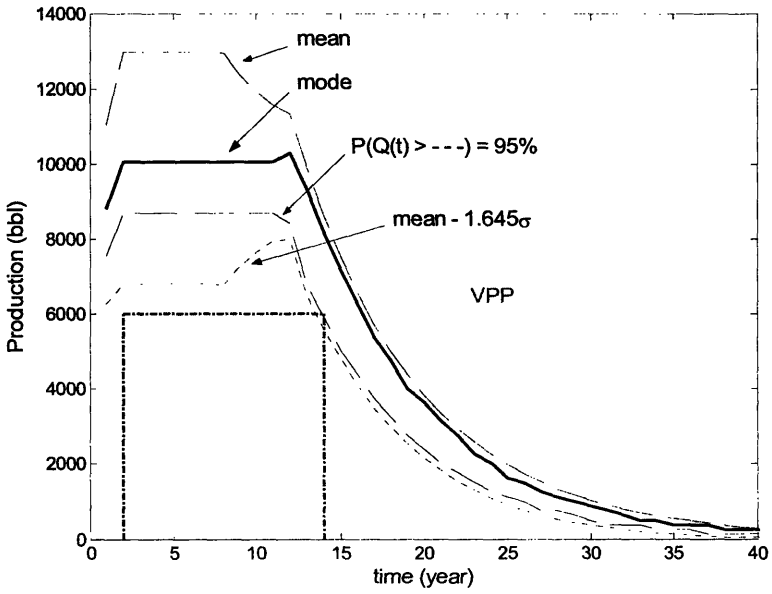


Figure F-11. Optimal production schedule determined by simulation based optimal design

The fact that the mean and mode of the annual production are different indicates the uncertainty can be described by an asymmetrical distribution. This is confirmed below by evaluating the full probability distributions.

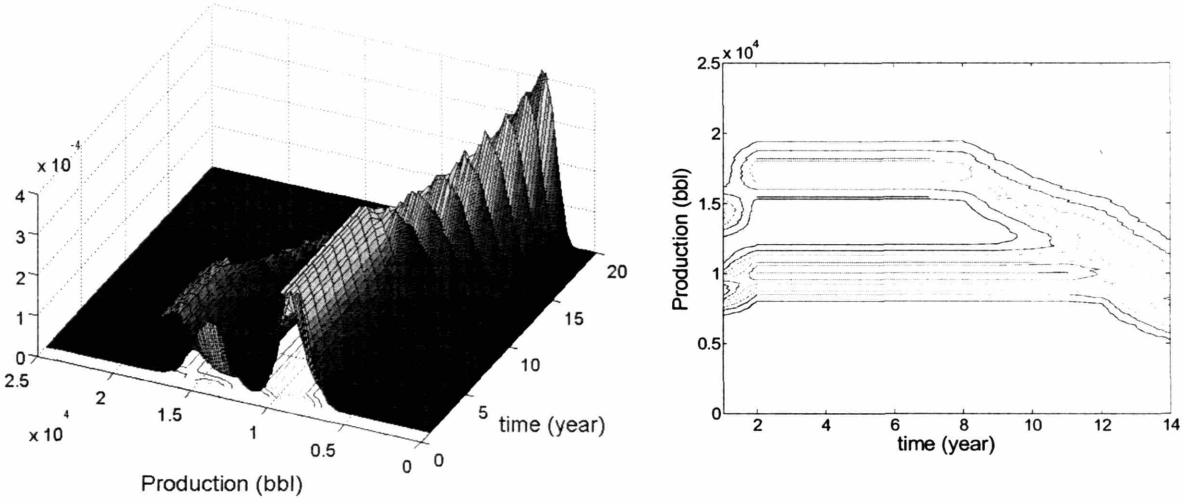


Figure F-12. Probability distribution and contour plot of production levels as a function of time

The 3-dimensional plot, as well as the supporting contour plot, show a bimodal distribution for the annual production $Q(t)$. This bimodality is introduced because of the binomial probability p(oil) regarding the presence of petroleum in B. A mere deterministic optimization would never reveal such information and would thus be inappropriate to be used for structuring the VPP contract. Only by probabilistically solving the optimal control problem will the information become available to quantify the default risk. Obviously, an additional margin of safety can be employed to further reduce the default risk regarding the VPP contract.

The example problem above showed the proof of concept regarding the simulation based optimal design to solve the optimal control problem under uncertainty. Without difficulty, additional reserves can be included in the estimation of the optimal design. Results for the annual production according to the optimal design for three PDNP (q_i = 110, 500, 750 bbl/month, and d_i = 0.011, 0.02, 0.05 year^{-1}) and three PUD (q_i = 80, 500,

750 bbl/month, and d_i = 0.018, 0.02, 0.05 year^{-1}, and p(oil) = 40%) reserves are shown below.

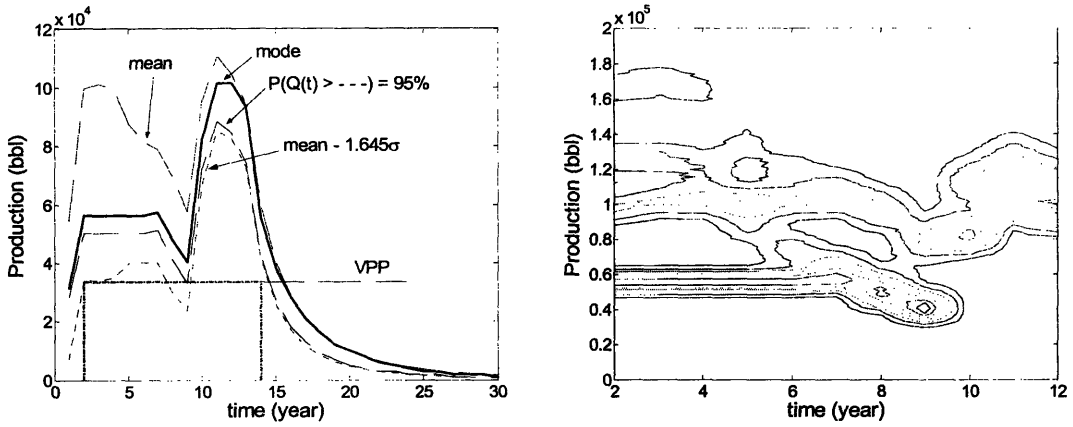


Figure F-13. Profile and contour plot for the annual production evaluating 3 PDNP and 3 PUD reserves

The optimal design for multiple reserves again demonstrates that information on uncertainty is benefial. With a probability of 95% the petroleum repayments will be produced, while at the same time a higher VPP level could be established compared to the one-sided 95% confidence interval. The importance of the probabilistic approach is also evident from the complexity of the obtained posterior distribution of the petroleum production as a function of time, as illustrated by the contour plot. A simple deterministic approach would clearly not have been adequate.

F.5.7. Including More Complexity

The example discussed above was still relatively straightforward and the problem could indeed become more realistic when including additional constraints, such as:

- An annual budget available for E&P activities
- A limit on the number of simultaneously producing wells
- A minimum rate production level for individual reserves

Also, the case study was rather simplistic in that the only control variables considered were the initiation times for production. Additional control variables could include:

- The production rate, to be determined by the reservoir engineer
- The start-up time, depending on the selection which equipment to install

Finally, the assumption that the production profiles are described by exponential models should be relaxed. Though in many cases the exponential production curve appears empirically representative for actual production profiles, it is possible that the exponential model is oversimplifying a petroleum reservoir, or that model parameters for the exponential model cannot be determined due to a lack of performance data from the reservoir. The approach presented here should be able to handly more complex models, obviously at the cost of an increased computation time.

F.6. Conclusion

Reinforcing the key points from the thesis, the discussion on petroleum reserves again demonstrated the superiority of the probabilistic approach when dealing with complex systems under uncertainty. The methodology and algorithms developed for Bayesian parameter estimation applied to chemical engineering problems were, with minor adjustments, successfully applied to the optimal control problem computing uncertain petroleum production profiles. Besides serving as an interesting example illustrating the benefits of the probabilistic approach to solving the optimal control problem, the analysis of extending the VPP to include PDNP and PUD reserves presented encouraging results that are worth pursuing further.

Script 'productionschedule'

```
% This script solves the optimal control problem determining the design
% for bringing reserves into production.
% The probabilistic constraint includes a minimum target production
% prescribed by a VPP contract
% The reserves are assumed to follow a constant production followed by
% diminishing production described by an exponential decay model

% Patrick de Man - MIT (June 2006)

%=========== SPECIFY THE OPTIMIZATION CHARACTERISTICS ===========
% Specify the VPP contract
target.Q = 3.35e4;
target.t = [2  14];

% Define othe system variables
percentUncertainty = 0.1;
tEnd = 60;
productionRange = linspace(0, 20e4, 200);

% Specify the number of samples to generate from uncertain parameters
% the parameter J refers to the algorithms in Muller (1998)
J = 100;

% Optional constraint: annual budget
budget = 25000;


%=========== SPECIFY THE RESERVES CHARACTERISTICS ===========

% Define the parameters for the PDNP reserves
qiPDNP = 365*[110  500   750 ];
diPDNP = 12*[ 0.011  0.02  0.05 ];

% Define the parameters for the PUD reserves
pDry = 0.6;
qiPUD = 365*[80  500  750  ];
diPUD = 12*[ 0.018  0.02  0.05 ];
```

```
% Specify the discount factor and generate the accompanying matrix
r = 0.15;
for t = 1 : tEnd
    discountFactor(1,t) = 1 / (1 + r)^t;
end % for t

for j = 1 : J
    DF(j,:) = discountFactor;
end % for j

% Solve the probabilistic optimal control problem
[MCMCoutput] = probabilisticoptimizationPDNP_PUD(target, ...
    qiPDNP, diPDNP, qiPUD, diPUD, pDry, percentUncertainty, ...
                                    tEnd, J, DF, budget, productionRange);

% Determine the optimal design by finding the maximum utility
sortedDesign = flipud(sortrows([MCMCoutput.utility/1e6 ...
                                              MCMCoutput.design],[1 2 3]));
optimalDesign = sortedDesign(1, 2:1+length(MCMCoutput.design(1,:)));

% Plot the information in 3D probability distribution, etc.
[Qtotal] = totalproductionprofilePDNP_PUD(optimalDesign, ...
    qiPDNP, diPDNP, qiPUD, diPUD, pDry, percentUncertainty, ...
                                          tEnd, target, productionRange);
```

Function 'probabilisticoptimizationPDNP_PUD'

```
% This function performs the algorithm for simulation based optimal design
% as specified in Appendix F and Muller (1998)

% Patrick de Man - MIT (June 2006)

function [MCMCoutput] = probabilisticoptimizationPDNP_PUD...
    (target, qiPDNP, diPDNP, qiPUD, diPUD, pDry, percentUncertainty, ...
                                        tEnd, J, DF, budget, productionRange);

% Determine the number of reserves involved
nPDNP = length(qiPDNP);
nPUD = length(qiPUD);
nTotal = nPDNP + nPUD;

% Declare the MCMC variables
n = 100;
burnIn = 5*n;
mcmcRun = 20000 + burnIn;

% include the prior domain for the time to bring the wells on-stream
prior.min = zeros(nTotal,1);
prior.max = target.t(2)*ones(nTotal,1);

% Initializer tracking vectors of the chain's performance
runCounter = zeros(mcmcRun+1, 1);  % storing # of MCMC runs
utility = zeros(mcmcRun + 1, 1);   % storing  utility at each run
alpha = zeros(mcmcRun + 1,1);      % storing alpha

% Specify the initial covariance matrix
% 'covarTuning' is used to tune the acceptRate
covarTuning = 1;
covarPD = covarTuning * eye(nTotal);

% Start the MCMC simulation
for iRun = 1 : mcmcRun

    % Report MCMC progress and calculate an updated covarPD matrix
    if iRun==n || iRun==2*n || iRun==3*n || iRun == 4*n || iRun == 5*n
        iRun
        mean(alpha(iRun-n+1:iRun))
```

```
    covarPD = covarTuning * cov(design(iRun-n+1 : iRun, :));
elseif iRun == 6*n
    iRun
    mean(alpha(5*n+1:6*n))
end

if iRun == 1
    currentUtility = 0;
    while currentUtility == 0;

        % Assign the initial design randomly
        design(1,:) = prior.max' .* rand(1,nTotal) ;
        [Qtotal] = productionprofilegenerator(design(1,:),...
            qiPDNP, diPDNP, qiPUD, diPUD, pDry, percentUncertainty,...
                                                        tEnd, J);
        [currentUtility] = getutilityPV(design(1,:), ...
                    target, Qtotal, DF, tEnd, budget, productionRange);
        % Notify when suitable initial values are found
        if currentUtility ~= 0
            fprintf('A suitable initial value has been found');
        end
    end

elseif iRun > 1
    if accept(iRun) == 1
        currentUtility = proposedUtility;
    end
end % if elseif iRun

% Store currentUtility
utility(iRun) = currentUtility;

% Propose a new design and calculate the utility
[proposedDesign] = getdesign(design(iRun,:), covarPD, ...
                                              prior.min, prior.max);
[Qtotal] = productionprofilegenerator(proposedDesign, ...
        qiPDNP, diPDNP, qiPUD, diPUD, pDry, percentUncertainty, tEnd, J);
[proposedUtility] = getutilityPV(proposedDesign, target, ...
                            Qtotal, DF, tEnd, budget, productionRange);

% Determine alpha
alpha(iRun+1,1) = min(1, (proposedUtility / currentUtility) );
```

```
    % Determine the acceptance/rejection + add next design tracking vectors
[design(iRun+1,:), accept(iRun+1)] = acceptreject(alpha(iRun+1, 1), ...
                                                  design(iRun,:),  proposedDesign);

end % for iRun

% Cut off the points that were sampled during the burn-in period
cleanDesign = design(burnIn : mcmcRun, :);
cleanUtility = utility(burnIn : mcmcRun);
cleanAlpha = alpha(burnIn : mcmcRun);
alphaAVG=mean(cleanAlpha)

% Create structure holding the results
MCMCoutput.mean = mean(cleanDesign);
MCMCoutput.alphaAvg = mean(cleanAlpha);
MCMCoutput.design = cleanDesign;
MCMCoutput.utility = cleanUtility;
MCMCoutput.rawDesign = design;

return; % function probabilisticoptimization
```

Function 'totalproductionprofilePDNP_PUD'

```
% This function calculates the profile obtained from the summation of the
% individual profiles for each well.
% The time to bring each well on-stream can be adjusted as desired.

% Patrick de Man - MIT (June 2006)

function [Qtotal] = totalproductionprofilePDNP_PUD...
    (design, qiPDNP, diPDNP, qiPUD, diPUD, pDry, percentUncertainty, ...
                                            tEnd, target, productionRange);

% Specify the number of samples to be drawn
J = 1000;

% Sample to include the effect of the dry hole probability
nDrill = 50;

% Initialize the structure to hold the samples
Qtotal = zeros(nDrill*J, tEnd);

% Perform the sampling of the drilling
for iDrill = 1 : nDrill
    [Qtotal( (iDrill - 1)*J + 1 : iDrill*J, :)] =
productionprofilegeneratorinclplateau...
    (design, qiPDNP, diPDNP, qiPUD, diPUD, pDry, percentUncertainty, tEnd, J);
end

% Calculate summary variables for the distribution over time
meanTotalProd = mean(Qtotal)';
sigTotalProd = std(Qtotal)';
confidenceLevel = 0.95;
productionConfLevel = productionprobabilityanalysis(Qtotal, productionRange,
tEnd, confidenceLevel);

% Generate a 3D profile of Qtotal
J = zeros(length(productionRange), tEnd);
modeTotalProd = zeros(tEnd,1);

% Determine the 3D probability profile of production
for t = 1:tEnd;
    n = hist(Qtotal(:,t), productionRange);
```

```
    p = n / sum((productionRange(2) - productionRange(1))*n);
    J(:,t) = p';

    % Find the mode over time
    modeTotalProd(t) = productionRange( find( p - max(p) == 0 ) );

end % for t

% Plot the production profile
figure; hold on;
plot(modeTotalProd, 'k', 'LineWidth', 2);
plot(meanTotalProd,'--');
plot(meanTotalProd - 1.645*sigTotalProd, ':');
plot(productionConfLevel,'--k');

plot([target.t(1); target.t(2)], [target.Q; target.Q], '-.r', 'LineWidth', 2);
plot([target.t(1); target.t(1)], [0 ; target.Q], '-.r', 'LineWidth', 2);
plot([target.t(2); target.t(2)], [0 ; target.Q], '-.r', 'LineWidth', 2);

ylabel('Production (bbl)');
xlabel('time (year)');

% Plot the 3D profile
tMax = 12;
figure;
surfc(1:tMax, productionRange, J(:,1:tMax));
xlabel('time (year)'); ylabel('Production (bbl)');

% Plot the contour plot of the probability distribution
figure;
colormap('jet');
contour(2:tMax, productionRange, J(:,2:tMax),7);
xlabel('time (year)'); ylabel('Production (bbl)');

return;
```

Function 'getutilityPV'

```
function [utility] = getutilityPV(design, target, Qtotal, DF, tEnd, ...
                                                    budget, productionRange)

% This function calculates the utility, which has been set equal to 0 if
% the design is infeasible (i.e. does not satisfy the constraint), and to
% E[NPV] of the oil production if the design is feasible

% Patrick de Man - MIT (June 2006)

% Determine the start year of each reservoir
startYear = ceil(design);

% Calculate the mean production cost
Q = mean(Qtotal);
producingCost = Q./365;

% Determine the one-sided probability level
confidenceLevel = 0.95;
productionConfLevel = productionprobabilityanalysis(Qtotal, ...
                                    productionRange, tEnd, confidenceLevel);

% Determine whether the design satisfies the production constraint
feasible = 1;
if any( (productionConfLevel(target.t(1) : target.t(2)) - target.Q) < 0)
    feasible = 0;
elseif any((budget - producingCost) < 0)
    feasible = 0;
end % if

% Determine the utility
if feasible == 1
    % Determine the Present Value (PV)
    PV = Qtotal .* DF;
    utility = mean(sum(PV'));
else
    utility = 0;
end % if feasible

return; % function getutilityPV
```

Function 'productionprofilegenerator'

```
% This file generates a probabilistic production profile with oil
% production given on an annual basis.  The profile is generated from an
% exponential flow model with uncertain parameters.  The uncertainty in the
% parameters is expressed as a standard deviation determined as a percentage
% of the parameter, and assuming a normal distribution.

% Patrick de Man - MIT (June 2006)

function [Qtotal] = productionprofilegeneratorinclplateau...
    (design, qiPDNP, diPDNP, qiPUD, diPUD, pDry, percentUncertainty, tEnd, J)

% Generate the samples to implement for qi and di
for iMC = 1 : J
    [qi, di] = drillingoutcome(qiPDNP, diPDNP, qiPUD, diPUD, pDry);
    qiMC(iMC,:) = qi + qi*percentUncertainty .* randn(1,length(qi));
    diMC(iMC,:) = di + di*percentUncertainty .* randn(1,length(qi));
end

% Check whether none of the qiMC or diMC are <0
if all(qiMC>0)
    dummy = 1;
elseif all(diMC>0)
    dummy = 1;
else
    error('qiMC or diMC below 0')
end

% Preceding the exponential decay of production, there is a constant level
% Determine tPlat, the time at which the plateau ends
plateau = 0.25;                % as in: q_plateau = plateau * qi
tPlat = -log(plateau) ./ di;

% Calculate the model output as a function of time for the pair of qi and di
qt = zeros(J, tEnd, length(design));
for iWell = 1 : length(design)
    for t = floor(tPlat(iWell)) - 1 : tEnd
        qt(:,t,iWell) = qiMC(:,iWell).* ...
                        exp(-diMC(:,iWell)*(t-ceil(design(iWell)) ) );
    end % for t
end % for iWell
```

```
% Calculate the cumulative production for each year
% MC values in columns, each column represents a year
Qt = zeros(J, tEnd, length(design));
for iWell = 1 : length(design)
    Qt(:, ceil(design(iWell)), iWell) = ...
        plateau*qiMC(:,iWell)*(ceil(design(iWell)) - design(iWell));

    Qt(:, ceil(design(iWell))+1 : ceil(design(iWell)) + ...
        floor(tPlat(iWell)), iWell) = plateau * repmat(qiMC(:,iWell),...
                                                1, floor(tPlat(iWell))) ;

    t = ceil(design(iWell)) + floor(tPlat(iWell)) + 1;
    Qt(:, t, iWell) = ...
        plateau*qiMC(:,iWell)*(tPlat(iWell) - floor(tPlat(iWell))) + ...
        ((qt(:,t-1,iWell) - qt(:,t,iWell)) ./ ...
                        diMC(:,iWell))*(ceil(tPlat(iWell)) - tPlat(iWell));

    for t = (ceil(design(iWell)) + ceil(tPlat(iWell)) + 1) : tEnd
        Qt(:,t,iWell) = (qt(:,t-1,iWell) - qt(:,t,iWell)) ./ diMC(:,iWell);
    end % for t

end % for iWell

Qtotal = sum(Qt,3);

return; % productionprofilegeneratorplateau
```

Function 'drillingoutcome'

```
% This file determines the actual overall vectors for qi and di that are
% fed to the simulation

% Patrick de Man - MIT (June 2006)

function [qi, di] = drillingoutcome(qiPDNP, diPDNP, qiPUD, diPUD, pDry)

qi = zeros(1, length(qiPDNP) + length(qiPUD));
qi(1:length(qiPDNP)) = qiPDNP;
di = [diPDNP  diPUD];

% Determine whether the PUD contain oil and incorporate in qi and di if so
for i = 1 : length(qiPUD)
    if pDry < rand(1)
        qi(length(qiPDNP)+i) = qiPUD(i);
    end
end

return; % function drillingoutcome
```

Bibliography

1. Bard, Y., *Nonlinear Parameter Estimation*, New York: Academic Press (1974).
2. Box, G.E.P. and G.C. Tiao, *Bayesian Inference in Statistical Analysis*, New York: John Wiley & Sons (1992).
3. Berger, J.O., *Bayesian analysis: A look at today and thoughts of tomorrow.* Journal of the American Statistical Association. **95**(452): p. 1269-1276 (2000).
4. Bernstein, P.L., *Against the Gods*, New York: John Wiley & Sons (1996).
5. Guide, *Guide to Expression of Uncertainty in Measurement*, International Organization for Standardization (ISO) (1993).
6. d'Agostini, G., *Bayesian Reasoning in Data Analysis - A Critical Introduction*, Singapore: World Scientific Publishing (2003).
7. Tatang, M.A., *Direct Incorporation of Uncertainty in Chemical and Environmental Engineering Systems.* PhD thesis, Cambridge, Massachussetts Institute of Technology (1995).
8. Wang, C., *Parametric Uncertainty Analysis for Complex Engineering Systems.* PhD thesis, Cambridge, Massachusetts Institute of Technology (1999).
9. Bertsekas, D.P. and J.N. Tsitsiklis, *Introduction to Probability*, Belmont, MA: Athena Scientific (2002).
10. Sen, A. and M. Srivastava, *Regression Analysis - Theory, Methods, and Applications*, New York: Springer-Verlag (1990).
11. Taylor, J.R., *An Introduction to Error Analysis*, Mill Valley, California: University Science Books (1982).
12. Sivia, D.S., *Data Analysis - a Bayesian Tutorial*, Oxford: Oxford University Press (1996).
13. Bayes, T., *An Essay towards solving a Problem in the Doctrine of Chances.* Philosophical Transactions of the Royal Society of London. **53**: p. 370 (1763).
14. Swinburne, R., ed. *Bayes's Theorem.* Proceedings of the British Academy. Vol. 113, Oxford University Press: Oxford(2003).
15. Laplace, P.S., *Philosophical Essay on Probability*, New York: John Wiley (1917).
16. Loredo, T.J., *From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics*, in *Maximum Entropy and Bayesian Methods*, P.F. Fougère, Editor. 1990, Kluwer Academic Publishers: Dordrecht. p. 81-142.
17. Bolstad, W.H., *Introduction to Bayesian Statistics*, Hoboken: John Wiley & Sons (2004).
18. Howson, C. and P. Urbach, *Bayesian Reasoning in Science.* Nature. **350**: p. 371-374 (1991).
19. Bernardo, J.M. and A.F.M. Smith, *Bayesian Theory*, Chichester: John Wiley & Sons (1994).
20. Jaynes, E.T., *Probability Theory - The Logic of Science*, Cambridge: Cambridge University Press (2003).

21. Cox, R.T., *Probability, Frequency, and Reasonable Expectation.* American Journal of Physics. **14**: p. 1-13 (1946).
22. Cox, R.T., *The Algebra of Probable Inference*, Baltimore: The Johns Hopkins Press (1961).
23. Getting, *Getting the Goat*, The Economist: February 20th, 1999, p. 72
24. El-Gamal, M.A. and D.M. Grether, *Are People Bayesian? Uncovering Behavioral Strategies.* Journal of the American Statistical Association. **90**(432): p. 1137-1145 (1995).
25. Körding, K.P. and D.M. Wolpert, *Bayesian Integration in Sensorimotor Learning.* Nature. **427**: p. 244-247 (2004).
26. Gelman, A., et al., *Bayesian Data Analysis*, New York: Chapman & Hall (1995).
27. Bryson, A.E. and Y.-C. Ho, *Applied Optimal Control*: Hemisphere Publishing Corporation (1975).
28. Lorenc, A.C., *Analysis Methods for Numerical Weather Prediction.* Quarterly Journal of the Royal Meteorological Society. **112**: p. 1177-1194 (1986).
29. Gamerman, D., *Markov Chain Monte Carlo*, London: Chapman & Hall (1997).
30. Strang, G., *Introduction to Applied Mathematics*, Wellesley: Wellesley-Cambridge Press (1986).
31. Carlin, B.P. and T.A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis.* 2nd ed, Boca Raton: Chapman & Hall (2000).
32. Smith, A.F.M., *Bayesian Computational Methods.* Philosophical Transactions: Physical Sciences and Engineering. **337**(1647): p. 369-386 (1991).
33. Gamerman, D., *Markov Chain Monte Carlo, Stochastic Simulation for Bayesian Inference.* Texts in Statistical Science, London: Chapman & Hall (1997).
34. Neal, R.M., *Probabilistic Inference using Markov Chain Monte Carlo Methods*, Department of Computer Science, University of Toronto (1993).
35. *The BUGS Project*, http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml
36. Gilks, W.R., S. Richardson, and D.J. Spiegelhalter, eds. *Markov Chain Monte Carlo in Practice.* Chapman & Hall: London(1996).
37. Draper, D., *Bayesian Hierarchical Modeling* (forthcoming).
38. Chib, S. and E. Greenberg, *Understanding the Metropolis-Hastings Algorithm.* The American Statistician. **49**(4): p. 327-335 (1995).
39. *MATLAB Toolbox for Density Estimation*, (1999). http://science.ntu.ac.uk/msor/ccb/densest.html
40. Metropolis, N., et al., *Equation of State Calculations by Fast Computing Machine.* Journal of Chemical Physics. **21**: p. 1087-1091 (1953).
41. Hastings, W.K., *Monte Carlo Sampling Methods using Markov Chains and their Applications.* Biometrika. **57**: p. 97-109 (1970).
42. Draper, D., *Bayesian Hierarchical Modeling*, New York: Springer-Verlag (2001).
43. Chen, M.-H., Q.-M. Shao, and J.G. Ibrahim, *Monte Carlo Methods in Bayesian Computation*, New York: Springer-Verlag (2000).
44. Press, W.H., *Numerical Recipes in C*, Cambridge: Cambridge University Press (1995).
45. Pouillot, R., et al., *Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to Listeria monocytogenes.* International Journal of Food Microbiology. **81**(2): p. 87-104 (2003).

46. Sivaganesan, M., E.W. Rice, and B.J. Marinas, *A Bayesian method of estimating kinetic parameters for the inactivation of Cryptosporidium parvum oocysts with chlorine dioxide and ozone.* Water Research. **37**(18): p. 4533-4543 (2003).
47. Borsuk, M.E. and C.A. Stow, *Bayesian parameter estimation in a mixed-order model of BOD decay.* Water Research. **34**(6): p. 1830-1836 (2000).
48. Pillonetto, G., G. Sparacino, and C. Cobelli, *Numerical non-identifiability regions of the minimal model of glucose kinetics: superiority of Bayesian estimation.* Mathematical Biosciences. **184**(1): p. 53-67 (2003).
49. Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, Singapore: McGraw-Hill (1991).
50. Cowles, M.K. and B.P. Carlin, *Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review.* Journal of the American Statistical Association. **91**(434): p. 883-904 (1996).
51. Best, N.G., M.K. Cowles, and S.K. Vines, *CODA Manual version 0.30.* Medical Research Council (MRC) Biostatistics Unit: Cambridge, UK. (1995).
52. Best, N.G., M.K. Cowles, and S.K. Vines, *CODA version 0.40 (Addendum to Manual).* Medical Research Council (MRC) Biostatistics Unit: Cambridge, UK. (1997).
53. *Econometrics Toolbox*, (2004). http://www.spatial-econometrics.com/
54. Steward, D.S., *On the Approach to Techniques for the Analysis of the Structure of Large Systems of Equations.* SIAM Review. **4**(4): p. 321-342 (1962).
55. Duff, I.S., *On Algorithms for Obtaining a Maximum Transversal.* ACM Transactions on Mathematical Software. **7**(3): p. 315-330 (1981).
56. Duff, I.S., *Algorithm 575. Permutations for a Zero-Free Diagonal.* ACM Transactions on Mathematical Software. **7**(3): p. 387-390 (1981).
57. Kwakernaak, H. and R. Sivan, *Linear Optimal Control Systems*, New York: John Wiley & Sons (1972).
58. Atkinson, A.C. and A.N. Donev, *Optimum Experimental Designs*, New York: Oxford University Press (1992).
59. Chaloner, K. and I. Verdinelli, *Bayesian Experimental Design: A Review.* Statistical Science. **10**(3): p. 273-304 (1995).
60. Clemen, R.T., *Making Hard Decisions: an Introduction to Decision Analysis*, Belmont, California: Duxbury Press (1996).
61. Dunlea, E.J., *Atmospheric Reactions of Electronically Excited Atomic and Molecular Oxygen.* PhD thesis, Boulder, University of Colorado at Boulder (2002).
62. Blitz, M.A., et al., *Laser Induced Fluoresence Studies of the Reactions of $O(^1D_2)$ with N_2, O_2, N_2O, CH_4, H_2, CO_2, Ar, Kr and n-C_4H_{10}.* Physical Chemistry Chemical Physics. **6**: p. 2162-2171 (2004).
63. Dunlea, E.J. and A.R. Ravishankara, *Kinetic Studies of the Reactions of $O(^1D)$ with Several Small Atmospheric Molecules.* Physical Chemistry Chemical Physics. **6**(9): p. 2152 - 2161 (2004).
64. Strekwoski, R.S., J.M. Nicovich, and P.H. Wine, *Temperature-Dependent Kinetics Study of the Reactions of $O(^1D_2)$ with N_2 and O_2.* Physical Chemistry Chemical Physics. **6**: p. 2145-2151 (2004).

65. Sander, S.P., et al., *Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies*. Jet Propulsion Laboratory, California Institute of Technology: Pasadena, California. (2003).
66. Seinfeld, J.H. and S.N. Pandis, *Atmospheric Chemistry and Physics: from Air Pollution to Climate Change*, New York: John Wiley (1998).
67. Shi, J. and J.R. Barker, *Kinetic Studies of the Deactivation of $O_2(^1O_g^+)$ and $O(^1D)$.* International Journal of Chemical Kinetics. **22**: p. 1283-1301 (1990).
68. Talukdar, R.K., et al., *Quantum Yields of O(1D) in the Photolysis of Ozone Between 289 and 329 nm as a Function of Temperature.* Geophysical Research Letters. **25**(2): p. 143 - 146 (1998).
69. Taylor, J.W., et al., *Direct Measurement of the Fast, Reversible Addition of Oxygen to Cyclohexadienyl Radicals in Nonpolar Solvents.* Journal of Physical Chemistry A. (2004).
70. Singer, A.B., *Global Dynamic Optimization*. PhD thesis, Cambridge, Massachusetts Institute of Technology (2004).
71. Effio, A., et al., *Studies on the Spiro[2.5]octadienyl Radical and the 2-Phenylethyl Rearrangement.* Journal of the American Chemical Society. **102**(19): p. 6063-6068 (1980).
72. Arends, I.W.C.E. and P. Mulder, *Rate Constants for Termination and TEMPO Trapping of Some Resonance Stabilized Hydroaromatic Radicals in the Liquid Phase.* Journal of Physical Chemistry. **99**: p. 8182-8189 (1995).
73. Pun, B.K.-L., *Treatment of Uncertanties in Atmospheric Chemical Systems: A Combined Modeling and Experimental Approach*. PhD thesis, Cambridge, Massachusetts Institute of Technology (1998).
74. Pilz, J., *Bayesian Estimation and Experimental Design in Linear Regression Models*, Chichester, England: John Wiley & Sons (1991).
75. Lindley, D.V., *Bayesian Statistics, A Review*. Regional Conference Series in Applied Mathematics, Philadelphia: Society for Industrial and Applied Mathematics (1971).
76. Janata, J., et al., *Chemical sensors.* Analytical Chemistry. **70**(12): p. 179R-208R (1998).
77. Wilson, D.M., et al., *Chemical sensors for portable, handheld field instruments.* IEEE Sensors Journal. **1**(4): p. 256-274 (2001).
78. Lee, D.D. and D.S. Lee, *Environmental gas sensors.* IEEE Sensors Journal. **1**(3): p. 214-224 (2001).
79. Meixner, H. and U. Lampe, *Metal oxide sensors.* Sensors and Actuators B-Chemical. **33**: p. 198-202 (1996).
80. Janata, J., *Apparatus and method for measuring the concentration of components in fluids*. US Patent 4,411,741 (1983).
81. Hagleitner, C., et al., *Smart single-chip gas sensor microsystem.* Nature. **414**: p. 293-296 (2001).
82. Tuller, H.L. and R. Mlcak, *Advanced sensor technology based on oxide thin film - MEMS integration.* Journal of Electroceramics. **4**(2/3): p. 415-425 (2000).
83. Penza, M. and L. Vasanelli, *SAW NOx gas sensor using WO3 thin-film sensitive coating.* Sensors and Actuators B: Chemical. **41**(1-3): p. 31-36 (1997).

84. Vogt, M.C., E.L. Shoemaker, and A.V. Fraioli, *Electrocatalytic Cermet Gas Detector/Sensor*. United States. US Patent 5,429,727 (1995).
85. Vogt, M., E. Shoemaker, and T. Turner, *A trainable cermet gas microsensor technology using cyclic voltammetry and neural networks.* Sensors and Actuators B-Chemical. **36**(1-3): p. 370-376 (1996).
86. Shoemaker, E.L. and M.C. Vogt, *Electrocatalytic Cermet Sensor*. United States. US Patent 5,772,863 (1998).
87. Kissinger, P.T. and W.R. Heineman, eds. *Laboratory Techniques in Electroanalytical Chemistry*. 2nd edition ed., Marcel Dekker: New York(1996).
88. Bard, A.J. and L.R. Faulkner, *Electrochemical Methods: fundamentals and applications*. 2nd edition ed: John Wiley & Sons (2001).
89. Laherrere, J., *Distribution of Field Sizes in a Petroleum System: Parabolic, Fractal, Lognormal or Stretched Exponential?* Marine and Petroleum Geology. **17**: p. 539-546 (2000).
90. Garb, F.A., *Oil and Gas Reserves Classification, Estimation, and Evaluation.* Journal of Petroleum Technology. p. 373-390 (1985).
91. *Production Forecasting with Uncertainty Quantification - PUNQ*, http://www.nitg.tno.nl/punq/
92. Lerche, I. and J.A. MacKay, *Economic Risk in Hydrocarbon Exploration*, San Diego: Academic Press (1999).
93. Simpson, G.S., et al., *The Application of Probabilistic and Qualitative Methods to Asset Management Decision Making.* SPE paper 59455, presented at the 2000 SPE Asia Pacific Conference, Yokohama, Japan, April 25-26. (2000).
94. Jonkman, R.M., et al., *Best Practices and Methods in Hydrocarbon Resource Estimation, Production and Emissions Forecasting, Uncertainty Evaluation and Decision Making.* SPE paper 65144, SPE Reservoir Evaluation & Engineering. **5**(2): p. 146-153 (2000).
95. Etherington, J., T. Pollen, and L. Zuccolo, *Comparison of Selected Reserves and Resource Classifications and Associated Definitions*, Society of Petroleum Engineers - Oil and Gas Reserves Committee (2005).
96. Savage, S., *The Flaw of Averages*, in *San Jose Mercury News*. (2000).
97. *The Flaw of Averages*, http://www.stanford.edu/~savage/flaw/
98. Cameron, D., *SEC asked to reconsider its reserves readings*, in *Financial Times*: London. p. 31 (2005).
99. Yergin, D. and D. Hobbs, *In Search of Reasonable Certainty - Oil and Gas Reserves Disclosure*, Cambridge Energy Research Associates (2005).
100. Boone, J.P., *Oil and Gas Reserve Value Disclosures and Bid-Ask Spreads.* Journal of Accounting and Public Policy. **17**: p. 55-84 (1998).
101. *Proved reserve recategorisation following internal review: No material effect on financial statements*, Royal Dutch Shell - News & Media Releases (2004).
102. *Statement of Financial Accounting Standards No. 69: Disclosures about Oil and Gas Producing Activities*, Financial Accounting Standards Board (1982).
103. Muller, P., *Simulation Based Optimal Design.* Bayesian Statistics. **6**: p. 459-474 (1998).
104. Ortega, J.M. and G.J. McRae, *Process Design and Optimization under Uncertainty: A Bayesian Decision Making Approach.* (to be published).