UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL AND HUMAN SCIENCES

Mathematics

ROBUST AND OPTIMAL EXPERIMENTAL DESIGNS FOR NON-LINEAR MODELS IN CHEMICAL KINETICS

By

Kieran James Martin

July 2012

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES

Mathematics

Doctor of Philosophy

ROBUST AND OPTIMAL EXPERIMENTAL DESIGNS FOR NON-LINEAR MODELS IN CHEMICAL KINETICS

By Kieran James Martin

This thesis considers the problem of selecting robust and optimal experimental designs for accurately estimating the unknown mean parameters of non-linear models in chemical kinetics. The design selection criteria used are local, Bayesian and maximin D-optimality. The thesis focuses on an example provided by GlaxoSmithKline which concerns a chemical reaction where the temperature at which runs of the reaction are conducted and the times at which observations can be made during the reaction are to be varied. Optimal designs for non-linear models are usually dependent on the unknown values of the model parameters. This problem may be overcome by finding designs whose performance is robust to a range of values for each model parameter.

Optimal designs are investigated for situations when observations are independent and when correlation exists between observations made on the same run of the process; different forms and strengths of correlation between observations are considered. Designs robust to the correlation and mean parameters are found and assessed via both theoretical measures and a large simulation study which compares the designs found to alternatives currently used in practice.

Designs for the situation when the error variables have non-constant variance are obtained by use of a model formed via a power transformation on the response and its expected value. Designs robust to the value of the transformation parameter as well as the correlation and mean parameters are found and assessed.

Analytic results are established for obtaining locally D-optimal designs when the model is assumed to have independent observations and the response and expected response have been transformed to remove heteroscedasticity. Where analytic results are not available, numerical methods are used to obtain optimal designs.

The differing costs of a run of a reaction and of making an observation on a run are incorporated into design selection. A criterion which includes the cost of the time taken to run a reaction in an experiment is formulated and used to find designs.

Contents

List of figures **ix**

List of tables **xi**

Declaration of authorship **xvii**

Acknowledgments **xix**

List of notation **xix**

1 Introduction **1**

1.1 Background 1

1.2 The chemistry model 2

1.3 Preliminaries 5

1.3.1 Statistical models 5

1.3.2 Designs 7

1.3.3 D-optimality and efficiency 8

1.3.4 Selection criteria for robust designs 11

1.4 Overview of the thesis 12

2 D-optimal designs for independent observations **15**

2.1 Introduction 15

2.2 Literature review 16

2.3 Analytical results for locally D-optimal designs 21

2.4 Exploration of locally optimal designs 26

2.5 Robust designs 30
2.5.1 Bayesian D-optimal designs 30
2.5.2 Maximin D-optimal designs 39
2.6 D-optimal designs when heteroscedasticity in the errors is present . . 44
2.6.1 Locally D-optimal designs when heteroscedasticity in the errors is present 45
2.6.2 Robust D-optimal designs when error heteroscedasticity is present 47
2.7 Conclusion 52

3 Locally and Bayesian D-optimal designs for non-linear models in the presence of correlation 55
3.1 Introduction 55
3.2 Literature review 56
3.3 Locally optimal designs in the presence of serial correlation 59
3.3.1 Investigating the D-optimality of designs $\xi_1 - \xi_6$ 62
3.4 Performance of locally D-optimal designs when the strength or form of serial correlation is mis-specified 64
3.4.1 Impact of the correlation structure 65
3.5 A comparison of constrained designs and locally optimal designs found assuming serial correlation 67
3.6 Locally optimal designs in the presence of serial correlation and run-to-run and measurement errors 68
3.6.1 Explaining the structure of optimal designs 72
3.7 Investigating the impact of varying r and q on an optimal design . . . 74
3.7.1 Assessing the impact of r and q on optimal designs when m, E_a and k_r are fixed 75
3.7.2 Assessing the impact of varying E_a, k_r and m 78
3.8 Bayesian D-optimal designs in the presence of serial correlation, and run-to-run and measurement errors 81
3.8.1 Prior distributions 81

3.8.2 Finding Bayesian D-optimal designs 83
3.8.3 Use of an equivalence theorem to establish Bayesian D- and near D-optimality 90
3.9 D-optimal designs when serial correlation and heteroscedasticity are believed to be present 90
3.9.1 Locally D-optimal designs 91
3.9.2 Bayesian D-optimal designs when correlation and heteroscedasticity are believed to be present 95
3.10 Conclusion 97

4 Investigation of design properties via a simulation study 99
4.1 Introduction 99
4.2 Literature review 101
4.3 Designs and scenarios used in the study 102
4.4 Methodology 106
4.5 The effect of analysing with and without considering correlation . . . 108
4.6 Impact of σ on relative mean square error for scenarios 1-21 112
4.7 Accuracy of the estimator of σ^2 114
4.8 Estimating bias using simulation 115
4.9 Performance of designs in the simulation 118
4.9.1 D-efficiency of designs under the edge scenarios 124
4.10 Simulating heteroscedasticity in the errors 125
4.10.1 Designs and scenarios used in the simulation 125
4.10.2 Methodology 127
4.10.3 Results 128
4.11 Conclusion 133

5 Cost considerations in design selection 135
5.1 Introduction 135
5.2 Literature review 136
5.3 Cost functions, criteria and parameter assumptions 138

5.4 Optimal designs for c_1 . . . 140
5.5 Locally D-optimal designs for c_2 . . . 141
5.6 Exact locally optimal designs . . . 147
5.7 Cost considerations for a Bayesian D-optimal design . . . 150
5.8 Conclusion . . . 151

6 Conclusions and Future Work **153**
6.1 Conclusions . . . 153
6.2 Recommendations for practitioners . . . 155
6.3 Future work . . . 156

A Proof of Lemmas in Chapter 2 **159**
A.1 Proof of Lemma 2.3.3 . . . 159
A.2 Proof of Lemma 2.3.4 . . . 161
A.3 Proof of Lemma 2.3.5 . . . 164
A.4 Proof of Lemma 2.3.6 . . . 165
A.5 Proof of Corollary 2.3.7 . . . 168

Bibliography **169**

List of figures

1.1 The model (1.5) plotted for temperatures T_1, T_2 and T_3 for $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ 5

2.1 Parameter combinations h/b and λ for which Theorem 2.3.2 has been shown to be true (shaded region) 25

2.2 $g(x)$ for $\lambda = 0.25$, $b = 1.01$ and $h = 1$ 25

2.3 $g(x)$ for $\lambda = 0.25$, $b = 1.01$ and $h = 0.96$ 26

2.4 The standardised variance against time and temperature 27

2.5 D-efficiencies of five designs for each of five scenarios 30

2.6 Value of the integrated standardised variance for d_6-d_{11} 37

2.7 Value of the integrated standardised variance for d_{12}-d_{16} 38

2.8 The D-efficiencies of, respectively, d_1, d_3, d_9 and d_{18} over Θ_2 43

2.9 The roots of $g(x)$ as $1/\lambda$ increases for $a_0 = 0.14$ and $b_0 = 1.336$ 45

2.10 The relative D-efficiency of designs for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, and $\lambda = 0.14$, 0.25, 0.5 and 1 47

2.11 The probability density function of π_{λ_1}, π_{λ_2} and π_{λ_3} 48

2.12 The Bayesian D-efficiencies attained by each design under each prior distribution 49

2.13 D-efficiency of designs over Θ_λ for d_{20} and d_{22} for $\lambda = 0.5$ 51

2.14 D-efficiency of designs over Θ_λ for d_{20} and d_{22} for $\lambda = 0.25$ 51

2.15 D-efficiency of designs over Θ_λ for d_{20} and d_{22} for $\lambda = 0.14$ 52

3.1 The strength of correlation between two observations 60

3.2 Average r values for designs demonstrating a switchover against q for $\tau = 1$, $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ 77
3.3 D-efficiency of designs for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\tau = 1, 10$, $\rho = 1$, $r = q = 1$ and $\lambda = 0.14, 0.25, 0.5, 1$ 94
3.4 D-efficiency of designs for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\tau = 1, 10$, $\rho = 1$, $r = 0.125$, $q = 8$ and λ=0.14, 0.25, 0.5, 1 94
3.5 D-efficiency of designs for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\tau = 1, 10$, $\rho = 1$, $r = 8$, $q = 0.125$ and λ=0.14, 0.25, 0.5, 1 95
3.6 Bayesian D-efficiencies of designs for the prior distributions $\pi_{\alpha b4}$ and π_{λ_1}, π_{λ_2} and π_{λ_3} when $m = 6$ 96

5.1 The optimal value for m (vertical axis) for a given α for each set of parameter assumptions: (a)-Sets 1,2 (b)-Sets 3,4 (c)-Sets 5,6 140
5.2 The optimal value of m for a Bayesian D-optimal design for given α and prior distributions given in Section 5.7 151

List of tables

1.1 Notation for observations, temperature and times 6

2.1 Locally D-optimal designs for values of k_r and E_a 28

2.2 Designs and scenarios for the D-efficiency study 29

2.3 Bayesian D-optimal designs d_i, $(i = 1, \dots 16)$ 34

2.4 Bayesian D-efficiencies of the designs d_1-d_{16} 36

2.5 Maximin D-optimal designs obtained for parameter spaces Θ_1 and Θ_2, $T_l = 70$, $T_u = 100$ 40

2.6 Minimum D-efficiency $\phi_M(d)$ over each parameter range, Θ_1 and Θ_2 for designs d_1-d_{18} 42

2.7 Bayesian D-efficiencies of the maximin D-optimal designs 44

2.8 Locally D-optimal designs for varying values of E_a, k_r and λ 46

2.9 Bayesian D-optimal and maximin D-optimal designs when $m = 1$ for model (1.8); $T_l = 70$, $T_u = 100$ 50

3.1 Approximate locally D-optimal designs for correlation structure (3.1) 61

3.2 Approximate locally D-optimal designs under correlation structure (3.1), where $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ 62

3.3 Approximate locally D-optimal designs in the presence of an exponential correlation structure (3.1) for $m = 3$ where $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ 63

3.4 D-efficiencies of locally D-optimal designs for a grid of values for τ and ρ where $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ 64

3.5 Approximate locally D-optimal designs in the presence of a quadratic or linear correlation structure where $E_a = 4.91\times10^4$ and $k_r = 4.59\times10^{-4}$ 66

3.6 D-efficiencies of approximate designs given in Table 3.2 where $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ and the correlation structure is either linear or quadratic, and ρ_l and ρ_q were varied over three different values 66

3.7 Approximate locally D-optimal design for $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ when observations are constrained to be at least two minutes apart 67

3.8 D-efficiencies of the constrained approximate design for differing values of τ and ρ when $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ 67

3.9 Approximate locally D-optimal designs for $q = 0$ and differing values of r and τ, with $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ 70

3.10 Approximate locally D-optimal designs for differing values of r, τ and q, with $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ 71

3.11 D-efficiencies of the labelled designs in Table 3.9 and 3.10 with $\tau = \rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ 72

3.12 Locally D-optimal designs exhibiting 'switchovers' for $\tau = 1$, $q = 0$, $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ for r spaced evenly across $0.1 \leq r \leq 1$ and $q = 0$ 75

3.13 Occurrence of switchover for $m = 4$, $\tau = 1$, $\rho = 1$ and $q = 0$ 79

3.14 Occurrence of Switchover for $m = 8$, $\tau = 1$, $\rho = 1$ and $q = 0$ 80

3.15 Bayesian D-optimal designs 84

3.16 Mean Bayesian D-efficiency obtained when the scenario and designs for the prior distributions on E_a and k_r are varied 86

3.17 Mean, median and minimum averaged Bayesian D-efficiencies of designs found under different prior distributions 87

3.18 Bayesian D-efficiencies of Bayesian D-optimal designs for $\xi_{\alpha a1}$-$\xi_{\alpha c4}$ . . 88

3.19 Bayesian D-efficiencies of Bayesian D-optimal designs for $\xi_{\beta a1}$-$\xi_{\beta c4}$ . . 89

3.20 Maximum integrated standardised variance observed over a grid search for each Bayesian D-optimal design 90
3.21 Locally D-optimal designs for differing values of τ and λ when $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\rho = 1$ and $r = q = 1$ 92
3.22 Locally D-optimal designs for differing values of τ and λ when $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\rho = 1$, $r = 0.125$ and $q = 8$ 92
3.23 Locally D-optimal designs for differing values of τ and λ when $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\rho = 1$, $r = 8$ and $q = 0.125$ 93
3.24 Bayesian D-optimal designs for prior distributions $\pi_{\alpha b4}$ and π_{λ_1}, π_{λ_2}, π_{λ_3} 96

4.1 Designs used in the simulation study 105
4.2 Scenarios s_{14}-s_{24} 105
4.3 The ratio of maximum error/total error for E_a 108
4.4 Value of average relative DSE, defined in Equation (4.2) for $\sigma = 2 \times 10^{-6}$ and designs $d_1 - d_{15}$ 111
4.5 Comparison of $ADSE$ relative to the minimum $ADSE$ obtained per scenario when, in the data analysis, correlation is assumed not present and when correlation is present and known for $\sigma = 2 \times 10^{-6}$ 111
4.6 Values of (4.3) when $u = 2 \times 10^{-6}$ and $v = 2 \times 10^{-5}$ 113
4.7 Values for (4.3) when $u = 2 \times 10^{-5}$ and $v = 2 \times 10^{-4}$ 113
4.8 DSE relative to the minimum DSE for each scenario, for the variance estimates with $\sigma = 2 \times 10^{-6}$ and the data analysed assuming correlation is known 115
4.9 Relative bias for the estimator of E_a (all values multiplied by 10^8) . 116
4.10 Relative bias for the estimator of k_r (all values multiplied by 10^8) . . 117
4.11 Relative bias for the estimator of σ^2 (all values multiplied by 10^3) . . 117
4.12 Relative bias for the estimator of σ^2 when correlation is not included during analysis 118
4.13 $ADSE$ relative to minimum $ADSE$ over scenarios $s_1 - s_{13}$, with the mean, median and minimum obtained by each design also given . . . 119

4.14 $ADSE$ relative to minimum $ADSE$ over scenarios $s_{14} - s_{21}$, with the mean, median and minimum obtained by each design also given . . . 120
4.15 $ADSE$ relative to minimum $ADSE$ over scenarios $s_{22} - s_{24}$ 120
4.16 Minimum mean and median $ADSE$ relative to minimum $ADSE$ for all designs across all scenarios 121
4.17 $ADSE$ of designs relative to the minimum $ADSE$ obtained across all scenarios and designs (all values multiplied by 10^2) 122
4.18 $ADSE$ of designs relative to the minimum $ADSE$ obtained across all scenarios and designs, excluding s_{18}-s_{21} 123
4.19 Comparison of D-efficiency over scenarios $s_{14} - s_{21}$ 124
4.20 Designs used during simulation 125
4.21 Scenarios used in the simulation study, to investigate the impact of error heteroscedasticity on accuracy of parameter estimation 127
4.22 DSE relative to minimum DSE for scenarios $s_{25} - s_{27}$, when the data were analysed ignoring heteroscedasticity 129
4.23 DSE relative to minimum DSE for scenarios $s_{25} - s_{27}$ when the data were analysed considering heteroscedasticity 129
4.24 DSE obtained when the data were analysed ignoring heteroscedasticity relative to the DSE obtained when the data were analysed considering heteroscedasticity for scenarios $s_{25} - s_{27}$ 130
4.25 DSE relative to minimum DSE for scenarios $s_{28} - s_{31}$ 131
4.26 DSE relative to minimum DSE for scenarios $s_{32} - s_{35}$ 132
4.27 DSE relative to minimum DSE for scenarios $s_{36} - s_{39}$ 132

5.1 Values for τ, r and q used to find designs 139
5.2 Designs for cost function c_2 and Set 1 of parameter assumptions . . . 142
5.3 Designs for cost function c_2 and Set 2 of parameter assumptions . . . 143
5.4 Designs for cost function c_2 and Set 3 of parameter assumptions . . . 144
5.5 Designs for cost function c_2 and Set 4 of parameter assumptions . . . 145
5.6 Designs for cost function c_2 and Set 5 of parameter assumptions . . . 146
5.7 Designs for cost function c_2 and Set 6 of parameter assumptions . . . 146

5.8 Optimal values of m for the approximate designs for α=1, 2, 4, 6 and Sets $1, \ldots, 6$ 147
5.9 Optimal values of N and m for α=1, 2, 4, 6, when $c_T = 6$ and Sets $1, \ldots, 6$ 148
5.10 Under/over efficiency of rounded approximate designs for all six sets of parameter assumptions 148
5.11 Relative efficiency of 'under spending' exact designs with randomly appended observation times for all six sets 149
5.12 Relative efficiency of 'under spending' exact designs with appended observation times for all six sets 150

Declaration of authorship

I, Kieran James Martin, declare that the thesis entitled "Robust and optimal designs for non-linear models in chemical kinetics" and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission

`Signed:`

`Date:`

Acknowledgments

I would like to thank my supervisors, Dr Stefanie Biedermann, Professor Susan Lewis and Dr David Woods for their support and patience throughout the last three and a half years working on this thesis. I would also like to thank EPSRC and GlaxoSmithKline for funding the project through a CASE award. In particular, I am grateful to Gillian Amphlett, Neil Hodnett, Mike Denham and Mohammad Yahyah for their advice and input.

Finally, a special thanks to my wife, Alice, who has put up with me keeping weird hours, and occasionally getting up at two in the morning to scribble some calculations on a piece of paper which, in the cold light of day, would usually end up being utter nonsense.

List of notation

A, B, C, D, E	Names of chemical compounds
$[A], [B], [C], [D], [E]$	Concentration of compounds, $A, \ldots, E$, in the chemistry model
a, b, c, h, l	Scalar constants in the model (2.1)
$\boldsymbol{y}$	Vector of observed values of $[A]$
η	Expected value of y
ε	Random error in y
σ	Standard error of ε
σ_s^2	Variance due to serial correlation between observations
σ_m^2	Variance due to measurement error
σ_r^2	Variance due to run-to-run error
t	Time after the start of the process when an observation is taken; units are in minutes in each design specification and in seconds in each model
T	Temperature at which a run of the process is conducted; units are in °C in each design specification and in °Kelvin in each model
a_0	Starting concentration of A
b_0	Starting concentration of B
k	Reaction coefficient (Joules/second)
T_r	Reference temperature (°Kelvin)
R	The universal gas constant, 8.34
k_r	Reaction coefficient at temperature T_r

E_a	Activation energy
w	Weight of a support point in an approximate design
W	Diagonal matrix holding the weights
n	Number of support points
N	Number of process runs
m	Number of observations per run
F	Sensitivity matrix
M	Information matrix
p	Number of parameters in the model
p_m	Number of parameters in the expected response
p_c	Number of parameters in the correlation model
$\mathcal{X}$	Space of all possible values at which observations are allowed to be taken
$\boldsymbol{\theta}_m$	Vector of parameters in the expected response
$\boldsymbol{\theta}_c$	Vector of correlation parameters
ϕ	The objective function to be maximised for D-optimality
ϕ_B	The objective function to be maximised for Bayesian D-optimality
ϕ_M	The objective function to be maximised for maximin D-optimality
π	A prior distribution for a parameter
Θ	Range of potential parameter values for maximin D-optimal designs
Q	Matrix of correlations between observations in the same run
r	σ_r^2/σ_s^2
q	σ_m^2/σ_s^2
τ, ρ	Parameters of the correlation structure in Equation (3.1)
ρ_l	Parameter of the correlation structure defined in Equation (3.2)
ρ_q	Parameter of the correlation structure in Equation (3.3)
d	Label for a design for an experiment in Chapter 2 and Chapter 4

ξ	Label for a design for an experiment in Chapter 3
λ	Transformation parameter in Equation (1.8)
s	Scenario used in the simulation study
S	Number of simulations conducted
DSE	Median relative squared error
$ADSE$	DSE averaged over $\boldsymbol{\theta}_m$
$RDSE$	Relative DSE
c_i	A function determining the cost of experimentation ($i = 1, 2$)
c_T	Total cost of an experiment
$\mathbf{1}_m$	Column m-vector with each entry equal to one
$\otimes$	Kronecker product
$\forall$	For all
$WLOG$	Without Loss of Generality
$\lvert . \rvert$	Determinant of a matrix

Chapter 1

Introduction

1.1 Background

There has been an increasing uptake in the use of design of experiment methods in chemistry and pharmaceutical research in academia and industry. In particular, response surface methodology for linear models is now used routinely for optimising the yield from chemical processes in production, for example, by GlaxoSmithKline (GSK), see Owen et al. (2001). There has been much less work until recently on selecting and implementing experimental designs for chemical development where the mathematical/chemistry models are non-linear in the parameters. Designs currently applied in practice are often ad hoc and do not take advantage of current theory and methods.

The overall aim of this thesis is concerned with finding efficient and effective designs for experiments in chemical kinetics to enable reaction processes to be better understood, while allocating experimental resources most cost-effectively. The focus throughout the thesis is on finding designs which will provide accurate estimates of the parameters of the mean response, regardless of the true values of parameters that occur in the mean response or in the error structure.

The work is motivated by, and focuses on, an example of an experiment provided by GSK of a particular chemical reaction. The chemists had available a well-established, deterministic, mathematics/chemistry model for the concentration of

a chemical during a reaction. This model is non-linear in its parameter values and is described in Section 1.2.

Observations in an experiment are gathered by conducting several trials of the process with a temperature specified for each trial. Within each trial, observations of the concentrations of the chemicals of interest are taken at pre-specified time points.

This chapter introduces the GSK example in Section 1.2 and, in Section 1.3, statistical models and aspects of design of experiments used in the thesis. The contents and aims of the thesis are given in Section 1.4.

1.2 The chemistry model

The example concerns a chemical reaction in which concentrations, $[A]$, $[B]$, $[C]$, $[D]$ and $[E]$, with units moles/litre, of five chemical compounds, A, B, C, D and E, are observed at a number of time points during each trial of a chemical reaction. The reaction is believed to be characterised by:

$$A + B \xrightarrow{k_1} C + D,$$
$$B + D \xrightarrow{k_2} E,$$

where k_1 and k_2 are the reaction coefficients for the two stages of reaction. Coefficient k_2 is much larger than k_1, and the second reaction can thus be treated as instantaneous. Under this assumption, the reaction can be written, using a single reaction coefficient k, as:

$$A + 2B \xrightarrow{k} C + E. \tag{1.1}$$

This reaction description now consists of only four chemical compounds as D is used up as rapidly as it is produced. Equation (1.1) can be described by the following

differential equations:

$$-\frac{d[A]}{dt} = -\frac{1}{2}\frac{d[B]}{dt} = \frac{d[C]}{dt} = \frac{d[E]}{dt} = k[A][B],$$

where t is the time (in seconds) elapsed since the start of the reaction.

Throughout this thesis, each trial of the reaction process will be called a run. For each run, initial concentrations of the chemicals A and B are specified. These are chosen by the chemists on scientific grounds and are regarded as fixed for the problem of designing the experiment. A value for the temperature, T (in °Kelvin), at which each run is conducted, must be chosen as part of the design, as well as the times at which observations are made during each run.

The concentration $[A]$ is of primary interest to the chemists and is treated as the response. The closed form solution for the value of $[A]$ after t seconds can be shown to be:

$$\eta(t) = \frac{a_0(b_0 - 2a_0)}{b_0 e^{k(b_0-2a_0)t} - 2a_0}, \tag{1.2}$$

where the values a_0 and b_0 are the initial concentrations of the chemicals A and B, and are fixed at 0.140 and 1.336 moles/litre respectively.

The unknown reaction coefficient k usually depends on the temperature T chosen for the run. This dependence is generally believed to be described by an empirical relationship called the Arrhenius equation (see for example Laidler, 1984) for which there are several forms in the literature. We use a form chosen by chemists at GSK as appropriate for the process:

$$k(T) = z \exp(-E_a/RT), \tag{1.3}$$

where $R = 8.314$ is the universal gas constant and E_a is the activation energy for

the reaction. The unknown parameters are z and E_a.

Chemists prefer to eliminate z from Equation (1.2) by using a reference temperature T_r (chosen by the chemists in our example to be 350°Kelvin) in order to obtain a model with an easier chemistry interpretation. From Equation (1.3):

$$k(T_r) = z \exp(-E_a/RT_r). \tag{1.4}$$

By taking logarithms across each of Equations (1.3) and (1.4) we obtain

$$\log\left(k(T)\right) - \log\left(k(T_r)\right) = -\frac{E_a}{R}\left(\frac{1}{T} - \frac{1}{T_r}\right).$$

Hence

$$k(T) = k_r \exp\left[-\frac{E_a}{R}\left(\frac{1}{T} - \frac{1}{T_r}\right)\right],$$

where $k_r = k(T_r)$. Substitution of this expression for k into Equation (1.2) gives the form of the chemistry model used from now on for the concentration:

$$\eta(t, T, \boldsymbol{\theta}_m) = \frac{a_0(b_0 - 2a_0)}{b_0 \exp\left\{k_r(b_0 - 2a_0)t \exp[-\frac{E_a}{R}(\frac{1}{T} - \frac{1}{T_r})]\right\} - 2a_0}, \tag{1.5}$$

where the $p_m \times 1$ vector $\boldsymbol{\theta}_m = (E_a, k_r)'$, T has units of °Kelvin, t has units of seconds and the parameters E_a and k_r are to be estimated.

Throughout this thesis, other than where noted, the set of all possible values of combinations of T and t that might be used in an experiment, is defined by:

$$\mathcal{X} = \{(T, t); 70 \leq T \leq 100; 1 \leq t \leq 200\}, \tag{1.6}$$

where, for clarity in the specification of a design, the units of T and t are °C and

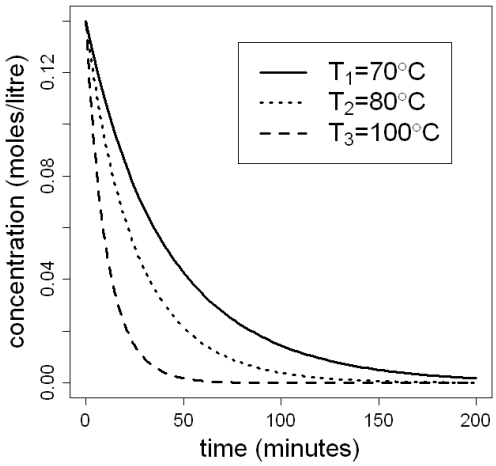


Figure 1.1: The model (1.5) plotted for temperatures T_1, T_2 and T_3 for $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$

minutes, respectively. Note that when a model is defined, the International System of Units (SI) of °Kelvin and seconds are retained to preserve the chemistry model.

The value of $\eta(t, T, \boldsymbol{\theta}_m)$ is plotted for three different temperatures for $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ in Figure 1.1; these mean parameter values were suggested from analysing data from a previous experiment.

1.3 Preliminaries

1.3.1 Statistical models

Throughout the thesis, observations made on different runs are assumed to be independent; there are N runs of the process, each with $m \geq 1$ observations. One form of a statistical model which will be considered is now described. Let the observations be held in an $Nm \times 1$ vector, $\boldsymbol{y}$, with elements ordered lexicographically, so that $\boldsymbol{y} = (y_{11}, \ldots, y_{1m}, y_{21}, \ldots, y_{Nm})'$. Let

$$\boldsymbol{y} = \boldsymbol{\eta}(\boldsymbol{t}, \boldsymbol{T}, \boldsymbol{\theta}_m) + \boldsymbol{\varepsilon}, \tag{1.7}$$

where $\boldsymbol{\eta}$, $\boldsymbol{T}$ and $\boldsymbol{t}$ are, respectively, the $Nm \times 1$ vector of expected responses with elements obtained from Equation (1.5), the $Nm \times 1$ vector holding the temperature at which each observation is made, and the $Nm \times 1$ vector holding the observation times, where $\boldsymbol{t} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N)'$, $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{im})'$ (where $i = 1, \ldots, N$) and $\boldsymbol{T} = (T_1, \ldots, T_N)' \otimes \mathbf{1}_m$, where $\otimes$ denotes the Kronecker product. A schematic of the data is given in Table 1.1.

We assume that the vector of random errors $\boldsymbol{\varepsilon} = (\varepsilon_{ij}) \sim MVN(\mathbf{0}, Q(\boldsymbol{\theta}_c)\sigma^2)$, where $Q(\boldsymbol{\theta}_c)$ is an $Nm \times Nm$ block diagonal matrix. The matrix $Q(\boldsymbol{\theta}_c)$ has $m \times m$ matrices $Q_i(\boldsymbol{\theta}_c)$ $(i = 1, \ldots, N)$ on the leading diagonal, where $Q_i(\boldsymbol{\theta}_c)$ is the correlation matrix for observations on the ith run; $\boldsymbol{\theta}_c$ is the $p_c \times 1$ vector of correlation parameters and σ^2 is the unknown variance, assumed constant.

In Chapter 2, designs for experiments are investigated in which a single observation is made on each run of the process so that $m = 1$, the observations are independent and $Q(\boldsymbol{\theta}_c) = I_N$ (the $N \times N$ identity matrix) in model (1.7). In later chapters, designs for experiments with $m > 1$ are investigated and correlations between observations taken on the same run of the process are then described through $Q(\boldsymbol{\theta}_c)$.

Table 1.1: Notation for observations, temperature and times

	Observation	Temperature	Times
Run 1	y_{11}	T_1	t_{11}
	$\vdots$	$\vdots$	$\vdots$
	y_{1m}	T_1	t_{1m}
Run 2	y_{21}	T_2	t_{21}
	$\vdots$	$\vdots$	$\vdots$
	y_{2m}	T_2	t_{2m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Run N	y_{N1}	T_N	t_{N1}
	$\vdots$	$\vdots$	$\vdots$
	y_{Nm}	T_N	t_{Nm}

When non-constant variance and asymmetry of the error distribution are likely

to occur, Atkinson (2003) proposed that in the design and analysis of an experiment both the response and the expected response be transformed by a power λ, where $0 < \lambda \leq 1$. Atkinson proposed transforming both $\boldsymbol{y}$ and $\boldsymbol{\eta}$ to preserve the link between the chemistry model and the response, while controlling for heteroscedasticity.

We therefore consider also a generalisation of the statistical model (1.7),

$$\boldsymbol{y}^{\lambda} = [\boldsymbol{\eta}(\boldsymbol{t}, \boldsymbol{T}, \boldsymbol{\theta}_m)]^{\lambda} + \boldsymbol{\varepsilon}, \tag{1.8}$$

where $\boldsymbol{y}^{\lambda} = (y_{11}^{\lambda}, \ldots, y_{Nm}^{\lambda})$ and similarly for $[\boldsymbol{\eta}(\boldsymbol{t}, \boldsymbol{T}, \boldsymbol{\theta}_m)]^{\lambda}$. The same assumptions were made on the error structure as for model (1.7). For this model, optimal designs found depend on the value of $\boldsymbol{\theta} = (\boldsymbol{\theta}_m; \boldsymbol{\theta}_c; \lambda)$, a $p \times 1$ vector, where $p_m + p_c + 1 = p$.

Note that finding D-optimal designs (introduced in the next subsection) for model (1.8) is equivalent (for least squares estimation, provided λ is assumed known) to finding D-optimal designs for a statistical model with the same expected response as model (1.7) and errors having a multivariate normal distribution with mean $\mathbf{0}$ and variance matrix given by:

$$\Delta Q(\boldsymbol{\theta}_c) \Delta \sigma^2,$$

where Δ is an $Nm \times Nm$ matrix whose diagonal entries are $[\boldsymbol{\eta}(\boldsymbol{t}, \boldsymbol{T}, \boldsymbol{\theta}_m)]^{1-\lambda}$ and all other entries are zero.

1.3.2 Designs

This subsection introduces the form of the designs, both approximate and exact, and terminology which will be used throughout this thesis.

A design point is a specification of a run in the experiment and gives the temperature and the times of the observations made on that run. Design points have the form $(T_i, \boldsymbol{t}_i)$, $i = 1, \ldots, N$, and are not necessarily distinct, where $\boldsymbol{t}_i$ is defined above. The support points are a set of $1 \leq n \leq N$ distinct design points and without loss of generality are labeled $(T_i, \boldsymbol{t}_i)$, $i = 1, \ldots, n$.

We will find approximate designs for models (1.7) and (1.8), defined as follows.

An approximate design is made up of n support points, and can be written for our problem as:

$$\xi = \left\{ \begin{array}{cccc} (T_1, \boldsymbol{t}_1) & (T_2, \boldsymbol{t}_2) & \dots & (T_n, \boldsymbol{t}_n) \\ w_1 & w_2 & \dots & w_n \end{array} \right\},$$

where the weights w_i give the proportion of experimental effort allocated to each support point and $w_1 + \ldots + w_n = 1$, $0 < w_i \leq 1$, $i = 1, \ldots, n$. For example, if a support point has weight $1/3$, and $N = 6$ runs are to be made, two runs should be conducted at that support point. For a similar formulation for block designs see, for example, Cheng (1995) for linear models and Woods and van de Ven (2011) for generalised linear models.

In practice all designs are exact, i.e. the number of times the support point $(T_i, \boldsymbol{t}_i)$ is used is an integer. An exact design can be regarded as having each support point $(T_i, \boldsymbol{t}_i)$ with a weight of $w_i = r_i/N$ where r_i is an integer, $r_1 + \ldots + r_n = N$ and N is the total number of runs that will be made. Pukelsheim and Rieder (1992) provided methods to obtain efficient exact designs from rounding approximate designs.

1.3.3 D-optimality and efficiency

Most methods of finding optimal designs optimise a function of the information matrix which is dependent on the design. When the expected response is non-linear in the parameters, it can be linearised to obtain the information matrix. For model (1.7), a Taylor expansion around $\boldsymbol{\theta}_m^0$ yields, for an observation y_{ij}:

$$\begin{aligned} y_{ij} =& \eta(t_{ij}, T_i, \boldsymbol{\theta}_m) + \varepsilon_{ij}, \\ \approx& \eta(t_{ij}, T_i, \boldsymbol{\theta}_m^0) + \sum_{u=1}^{p_m} \left[\frac{\partial \eta(t_{ij}, T_i, \boldsymbol{\theta}_m)}{\partial \theta_{mu}} \right]_{\boldsymbol{\theta}_m = \boldsymbol{\theta}_m^0} (\theta_{mu} - \theta_{mu}^0) + \varepsilon_{ij}, \end{aligned} \tag{1.9}$$

where η is defined in Equation (1.5) and θ_{mu} is the uth element of $\boldsymbol{\theta}_m$. If we set

$$\begin{aligned}
\eta_{ij}^0 =& \eta(t_{ij}, T_i, \boldsymbol{\theta}_m^0), \\
y_{ij}^0 =& y_{ij} - \eta_{ij}^0, \\
U_{iju} =& \left[\frac{\partial \eta(\boldsymbol{t}_{ij}, T_i, \boldsymbol{\theta}_m)}{\partial \theta_{mu}}\right]_{\boldsymbol{\theta}_m = \boldsymbol{\theta}_m^0}, \\
\omega_u =& (\theta_{mu} - \theta_{mu}^0),
\end{aligned}$$

then we can write Equation (1.9) as:

$$y_{ij}^0 \approx \sum_{u=1}^{p_m} \omega_u U_{iju} + \varepsilon_{ij}. \tag{1.10}$$

This is a linear regression model with unknown parameters ω_u, $u = 1, \ldots, p_m$.

In matrix notation Equation (1.10) is

$$\boldsymbol{y}^0 \approx U\boldsymbol{\omega} + \boldsymbol{\varepsilon}. \tag{1.11}$$

If the errors are assumed to be independent and normally distributed then the least squares estimate of $\boldsymbol{\omega}$ is:

$$\hat{\boldsymbol{\omega}} = \left(U'U\right)^{-1} U'\boldsymbol{y}^0,$$

where $U'U$ is called the information matrix. When correlation is present and the form of $Q(\boldsymbol{\theta}_c)$ is known, pre-multiplication of the terms in Equation (1.11) by L^{-1}, where $Q(\boldsymbol{\theta}_c) = L'L$, makes the resulting transformed errors independent. This gives an information matrix for estimation of $\boldsymbol{\omega}$ of $U'Q^{-1}(\boldsymbol{\theta}_c)U$.

For an approximate design, the additivity of information matrices can be used (see, for example, Silvey 1980, page 14) to obtain the information matrix, $M(\xi, \boldsymbol{\theta})$, for estimating $\boldsymbol{\theta}_m$. For model (1.8), for a given design ξ, we get:

$$M(\xi, \boldsymbol{\theta}) = \sum_{i=1}^{n} w_i F_i' Q_i^{-1}(\boldsymbol{\theta}_c) F_i, \tag{1.12}$$

where F_i is the $m \times p_m$ sensitivity matrix for the ith support point with (u, v)th entry

$$f_{u,v}(t_{iu}, T_i, \boldsymbol{\theta}_m) = \lambda \eta(t_{iu}, T_i, \boldsymbol{\theta}_m)^{\lambda-1} \frac{\partial \eta(t_{iu}, T_i, \boldsymbol{\theta}_m)}{\partial \theta_{mv}}. \tag{1.13}$$

$u = 1, \ldots, m$ and $v = 1, \ldots, p_m$.

As practitioners are usually interested in increasing process understanding, accurate estimation of the mean parameters is a high priority. We therefore use D-optimality to find designs. A design ξ is called D-optimal if it maximises the objective function:

$$\phi(\xi) = |M(\xi, \boldsymbol{\theta})|, \tag{1.14}$$

where $|.|$ denotes determinant. For non-linear models, this criterion is asymptotically equivalent to minimising the volume of the joint confidence ellipsoid for the mean parameters. The model is non-linear in the parameters, so any optimal design may depend on the values of $\boldsymbol{\theta}_m$, $\boldsymbol{\theta}_c$ and λ. When values are assumed for the parameters, a D-optimal design is called locally D-optimal.

Throughout the thesis, a design ξ will be assessed via its D-efficiency, which is defined as

$$D_{\text{eff}} = \left\{ \frac{|M(\xi, \boldsymbol{\theta})|}{|M(\xi^*, \boldsymbol{\theta})|} \right\}^{\frac{1}{p_m}},$$

where p_m is the number of parameters in the mean response, and ξ^* is a D-optimal design.

If the relative performances of two designs ξ_1 and ξ_2 are of interest, their relative D-efficiency can be calculated, and is given by:

$$D_{\text{Reff}} = \left\{ \frac{|M(\xi_1, \boldsymbol{\theta})|}{|M(\xi_2, \boldsymbol{\theta})|} \right\}^{\frac{1}{p_m}}.$$

1.3.4 Selection criteria for robust designs

Two criteria for choosing designs robust to parameter values are considered in this thesis: Bayesian D-optimality and maximin D-optimality. From a scientist's knowledge or experimentation carried out in preparation for an experiment, we can usually obtain some prior information on the parameter values. Hence a prior probability distribution $\pi(\boldsymbol{\theta})$ over the parameters can be specified to take account of prior knowledge in design selection.

This prior distribution can be incorporated into the optimal design criterion to obtain a "pseudo-Bayesian" criterion (see Chaloner and Verdinelli, 1995) which seeks to maximise the objective function:

$$\phi_B(\xi) = \int_{\Theta} \log(|M(\xi, \boldsymbol{\theta})|) d\pi(\boldsymbol{\theta}), \tag{1.15}$$

where Θ is the space of all possible parameter values. This function averages the log determinant of the information matrix with respect to the prior distribution assigned to $\boldsymbol{\theta}$. Designs that maximise the objective function $\phi_B(\xi)$ are called Bayesian D-optimal.

Designs can be assessed by their Bayesian D-efficiency (as used by Song and Wong 1998), given by:

$$\exp\left(\frac{\phi_B(\xi) - \phi_B(\xi^*)}{p_m}\right), \tag{1.16}$$

where ξ^* is a Bayesian D-optimal design.

An alternative approach for finding designs which are robust to parameter values is to find designs which achieve the highest minimum D-efficiency over specified ranges of parameter values. This approach requires less prior knowledge than for Bayesian D-optimal designs, the only requirement being a range of potential values for each of the parameters. The designs found tend to have a lower average D-efficiency than Bayesian D-optimal designs but offer greater protection against a worst case scenario. The objective function to be maximised is:

$$\phi_M(\xi) = \min_{\boldsymbol{\theta}\in\Theta} \left\{ \frac{|M(\xi,\boldsymbol{\theta})|}{|M(\xi^*,\boldsymbol{\theta})|} \right\}^{\frac{1}{p_m}}, \tag{1.17}$$

where ξ^* is a locally D-optimal design for $\boldsymbol{\theta}$. A design which maximises $\phi_M(\xi)$ is called maximin D-optimal.

1.4 Overview of the thesis

In this thesis we aim to achieve several goals which are motivated by the example in particular, and are relevant to non-linear models with responses which vary in time (dynamic models) more generally. These are:

1. To find and assess locally D-optimal designs for model (1.8), using both analytic and numerical search techniques, when observations are made on independent runs of the process, having:

   (a) one observation per run, and $0 < \lambda \leq 1$

   (b) more than one observation per run, where observations on the same run are assumed to be correlated, for various different forms and strengths of correlation and $0 < \lambda \leq 1$

2. To investigate the impact of misspecification of the parameters $\boldsymbol{\theta}_m$, $\boldsymbol{\theta}_c$ and λ

3. To find and evaluate designs whose efficiencies are robust or insensitive to the values of $\boldsymbol{\theta}_m$, $\boldsymbol{\theta}_c$ and λ

4. To investigate how well D-optimal designs, motivated by asymptotic theory, perform for finite experiments

5. To investigate methods of finding the best allocation of experimental resource between the number of runs and number of observations taken per run when information on relative costs is available

Chapter 2 addresses goals 1a, 2 and 3, through finding analytic results for model (1.8), with $Q(\boldsymbol{\theta}_c) = I_N$, to assist in obtaining locally, Bayesian and maximin D-optimal designs. Assessments and comparisons of the designs are made.

Chapter 3 addresses goals 1b, 2 and 3. Designs for different types and strengths of correlation are explored, through finding locally D-optimal and Bayesian D-optimal designs. The interaction between the impact on optimal design of correlation between observations made on the same run and heteroscedasticity of the errors is also investigated by finding and comparing locally D-optimal and Bayesian D-optimal designs for different values of λ.

Chapter 4 addresses the fourth goal. The efficiency of the designs found in Chapter 3 and ad hoc designs used by practitioners in the past are compared and assessed via a large simulation study across multiple scenarios, which vary $\boldsymbol{\theta}_m$, $\boldsymbol{\theta}_c$ and λ.

Chapter 5 addresses the final goal. The optimal allocation of experimental effort between the number of runs and the number of observations per run is determined for several locally D-optimal designs and a Bayesian D-optimal design. A criterion for cost which includes the cost of the run time of the reaction is formulated and applied.

In Chapter 6, the findings of the research are summarised, and recommendations for designing studies and areas for future research are given.

Chapter 2

D-optimal designs for independent observations

2.1 Introduction

In this chapter, we consider the special case of design of experiments for the non-linear dynamic model (1.8) for which $p_c = 0$ and the observations are independent, so that $\boldsymbol{\theta} = (\boldsymbol{\theta}_m; \lambda)$. In practice, experimenters usually want to take several observations per run on the grounds of cost and thus the assumption of independence may not be valid. Assuming that observations are independent allows analytical results to be obtained which will prove useful in later chapters, when finding optimal designs for models where the assumption of independence does not hold. For example, Algorithm 3.1 in Section 3.3 uses the analytical results found in this chapter to help find optimal designs when the statistical model includes serial correlations between observations.

Initially designs are found under the assumption that $\lambda = 1$, but this assumption is relaxed in Section 2.6.

The literature for design of experiments for non-linear models which include the assumption of independent errors is reviewed in Section 2.2. We then investigate, in Section 2.3, how results from optimal design theory (Atkinson et al., 2007, for example) can be used to obtain analytical forms for locally D-optimal designs

for the special case of independent normally distributed errors. The application of these results to find locally D-optimal designs for the motivating example is presented in Section 2.4, and the D-efficiencies of the designs are compared. Designs which are robust to parameter misspecification are found and assessed in Section 2.5 through two approaches: Bayesian D-optimal designs and maximin D-optimal designs. In Section 2.6, local and robust D-optimal designs are found for when heteroscedasticity in the errors is reduced via transformation of the response and the mean response. In the final section, the results of the chapter are summarised and discussed. Throughout the chapter, designs are compared using D-efficiency and a critical assessment of their properties is made.

2.2 Literature review

This review is focused mainly on the design of experiments for models similar to those considered in this thesis, where the mechanistic model is non-linear in the parameters, although not necessarily dynamic. Additive independent errors are assumed and the goal is to find designs which accurately estimate the mean parameters.

Designs for the Arrhenius equation: Rodriguez-Aragón and López-Fidalgo (2005) found locally D-optimal designs for the Arrhenius equation, with rate coefficient as the response. In practice, the value of the rate coefficient is usually found by experiment. The authors obtained analytic solutions for the optimal temperatures at which to make observations of the reaction coefficient. They demonstrated that their designs provided more accurate parameter estimates than ad hoc alternatives, but did not consider the performance of locally D-optimal designs when the mean parameters are misspecified. The model for the expected response was expanded by Rodríguez-Díaz and Santos-Martín (2009), who found locally D-optimal designs for the so called modified Arrhenius equation:

$$k(T) = \frac{z}{T^{\iota}} e^{-E_a/T},$$

where z and E_a are as defined in Chapter 1, and $\iota \in \mathbb{R}$; these three parameters require estimation. The modified form has an additional term, T^{ι}, in the Arrhenius equation as given in Equation (1.3). The authors found analytic results for locally D-optimal designs, and used these to compare optimal designs for the modified Arrhenius equation as ι was varied to the optimal design for the unmodified Arrhenius equation. The authors demonstrated that, for many values of $\iota \neq 0$, a locally D-optimal design for the unmodified Arrhenius equation was outperformed by designs found using ad hoc methods, and thus the correct specification of ι is important. The authors also found the maximin D-optimal design with respect to E_a, but assumed ι was known while designing. For our example, the chemists were clear that a modified Arrhenius equation was not appropriate.

Mannaswamy et al. (2010) found locally D-optimal designs for a chemistry model of a viscosity reaction called the "cross equation" which includes the modified Arrhenius equation as a term in the model, but is not dynamic. The model had six parameters and three controllable variables: temperature, shear rate and concentration of the chemical. They investigated the impact of parameter misspecification on design selection by finding locally D-optimal designs for a full factorial of parameter values and analysing the D-efficiencies obtained to determine which parameters had the most influence on the choice of optimal design. They found that only two of the six mean parameters had a large impact on D-efficiency. It will be seen in this chapter that, for the example under study, E_a has less influence on D-efficiency than the reaction coefficient k_r over the parameter ranges considered.

Designs for the Michaelis-Menten model: Although not a dynamic model, the Michaelis-Menten model has been thoroughly investigated in the literature and therefore is reviewed here. It is a model for reaction rate in enzyme kinetics, of the form:

$$\frac{V_{max}[S]}{K_m + [S]},$$

where V_{max} and K_m are parameters requiring estimation, and $[S]$ is the concentration of a particular substrate. Observations are assumed to be of the reaction rate, where

the concentration of the substrate is the controllable variable. A locally D-optimal design depends only on the value of K_m, as the model is linear in the parameter V_{max}.

Dette and Biedermann (2003) found analytically maximin D-optimal designs for the Michaelis-Menten model with two support points. They demonstrated that the designs are highly efficient, with a D-efficiency of over 80% throughout the large range assumed for K_m. This is not always the case, however, and for our example the maximin D-optimal designs found have a much lower average D-efficiency (see Section 2.5).

Matthews and Allcock (2004) found Bayesian D-optimal designs using a uniform and also a log uniform prior distribution for K_m in the Michaelis-Menten model. They discovered that, for each prior distribution, the Bayesian D-optimal design changed from having three support points to having only two as the domain of the prior distribution was decreased. They also allowed for heteroscedastic errors by modeling the variance as a function of the expected response raised to the power of $2(1-\lambda)$, with $0 \leq \lambda \leq 0.5$, where λ was assumed known. They found that λ did not have a large influence on the Bayesian D-optimal design obtained, except for $\lambda = 0$, for which the optimal design no longer depended on the mean parameters and consisted of one support point at each end of the permissible interval for $[S]$.

Bogacka et al. (2011) found locally D-optimal designs for four different models describing enzyme kinetics, all of which are extensions of the Michaelis-Menten model. Some exploration of the effects of parameter misspecification was conducted, by considering the D-efficiency of locally optimal designs over plausible ranges for the mean parameters. The authors found that locally D-optimal designs which assumed high parameter values had a performance which was robust to parameter misspecification and recommended that, given a range for each parameter value, practitioners should design an experiment using the largest allowable parameter values. Similar arguments are used in this thesis in Chapter 3 when investigating the impact of serial correlation, where evidence is presented that assuming stronger serial correlation between observations from the same run leads to designs whose

performance is robust to the serial correlation being weaker than anticipated.

Designs for dynamic models in chemical engineering: Experimental design has been taken up by practitioners working with dynamic models. There are many papers in the chemistry engineering literature which build on the statistical literature and apply the methods in various fields. Franceschini and Macchietto (2008a) reviewed the state of experimental design, focusing on non-linear dynamic models. They discussed novel design criteria which incorporate aspects not considered in traditional optimal design, such as finding designs which lead to reduced correlation between parameter estimates. Other work by practitioners include Chu and Hahn (2008), who found designs aiming to incorporate parameter selection and D-optimality when not all parameters can be estimated, and Barz et al. (2010) who found modified A-optimal designs, i.e. designs that minimise the trace of the inverse of the information matrix. The authors remarked that A-optimal designs are better at reducing correlations between parameter estimates than D-optimal designs but no evidence was given. For a two parameter model, a D-optimal design will minimise this correlation (Franceschini and Macchietto, 2008b), so we do not investigate A-optimality in this thesis.

Designs for dynamic biological models: There has been a great amount of interest in applying design theory in the context of dynamic biological models, where samples of observations usually come from patients. Some examples include Dette et al. (2005) who explored optimal designs for the Monod model, a non-linear dynamic model which is similar in form to the Michaelis-Menten model, but where time of observation is controlled. They found maximin D-optimal designs, and demonstrated that for all parameters the design performed more efficiently than a uniformly spaced design. They also compared their design to a locally D-optimal design found by Dette et al. (2003) and demonstrated that the maximin D-optimal design provided better parameter estimates than the locally optimal design for most values of the mean parameters.

Matthews and James (2005) explored locally and Bayesian D-optimal designs for a non-linear dynamic model describing the rate of blood flow in the brain which can

be measured by taking blood samples from one of two sets of veins and arteries. They discussed the limitations of standard optimal designs for their problem, specifically, that rounding approximate designs to obtain an exact design can be impractical, and that observations taken on the same subject (run) cannot be taken too closely together in time. They addressed these problems by finding restricted designs in which a specified number of observations spaced evenly in time were taken, where the experimenter controlled which vein blood was drawn from. These designs were less efficient for parameter estimation, but more efficient for estimating a particular linear combination of the mean parameters that was considered of interest. Concerns about designs having observations too close together in time are addressed briefly in Section 3.5, and efficient rounding of designs is discussed in Chapter 5.

Designs for non-linear models with heteroscedastic errors: There has been some investigation into the problem of how to design an experiment when heteroscedasticity in the errors is believed to be present. For the situation when the variance is proportional to the mean raised to the power $2(1-\lambda)$, Atkinson (2003) suggested use of a transformation of both the response and the expected response. He described the Horwitz rule, which provides a method of choosing the value of λ, but requires the use of previous data. He found locally D-optimal designs for given values of λ for three example models, an exponential decay over time, a dynamic multi-response model and a dynamic model where the form of the expected response was not analytically available. Atkinson (2005a) investigated the efficiencies of designs under transformations for the first two models considered by Atkinson (2003) and demonstrated that designs which assumed incorrect values for λ had a very low D-efficiency. Atkinson (2004) proposed prior distributions over λ and found Bayesian D-optimal designs for the first two examples in Atkinson (2003), assuming that $\boldsymbol{\theta}_m$ was known. For the same examples Atkinson (2005b) found Bayesian D-optimal designs for when prior distributions were proposed for $\boldsymbol{\theta}_m$ and λ was assumed known, but did not find Bayesian D-optimal designs for prior distributions over both $\boldsymbol{\theta}_m$ and λ. In this thesis we find designs whose performance is robust to both $\boldsymbol{\theta}_m$ and λ.

When the form of the variance is known, Bogacka and Wright (2005) pointed out that weighted least squares can be used for parameter estimation and experimental design, rather than using transformation. They found several designs to expand results from Atkinson and Bogacka (1997) for a particular example of a dynamic non-linear model with two mean parameters. They also found locally D-optimal designs for a non-linear model where the variance is assumed to be proportional to the expected response raised to the power of $2(1-\lambda)$. This assumption provides optimal designs (for least squares regression, assuming λ is known) that are the same as those found for transformation of both the expected response and response by λ.

A general comment is that designs for non-linear models are usually model specific. Methods which succeed for one model may not be appropriate for another. The particular model form considered in this thesis has not been studied before and results obtained may be different from those for other models.

The main approaches to obtaining parameter robust designs, maximin and Bayesian D-optimality are usually very effective, although both involve a certain amount of approximation. In particular, maximin D-optimal designs are often quite difficult to find if analytical results for locally D-optimal designs are not available.

2.3 Analytical results for locally D-optimal designs

In this section, analytic results for locally D-optimal designs are obtained under certain restrictions on the model constants, and the times and temperatures of observation. For simplicity, we reparameterise the expected response in the model (1.5) using the following: $a = h(b-h)/2$, $b = b_0$, $h = 2a_0$, $c = (b-h)$, $l = 1/R$, $\theta_2 = E_a$ and $\theta_1 = k_r \exp(E_a/(RT_r))$.

Then, under the assumption of normal, independent and additive errors, the model for an observation (T, t) is:

$$y^\lambda = [\eta(t, T, \boldsymbol{\theta}_m)]^\lambda + \varepsilon, \text{ where} \tag{2.1}$$

$$\eta(t, T, \boldsymbol{\theta}_m) = \frac{a}{b\exp(ct\theta_1 \exp(-l\theta_2/T)) - h}, \tag{2.2}$$

with $a, b, c, l, h, \lambda \in \mathbb{R}$ being known constants, where $a > 0$, $b > 0$, $c > 0$, $l > 0$, $\lambda > 0$, $b > h$ with the restrictions that for $\lambda < 0.5$, $\lambda\exp(1) > h/b$ or $\lambda < \frac{b-h}{b}$. The unknown parameters are $\theta_1 > 0$ and $\theta_2 > 0$, and $t > 0$ and $T > 0$ are variables whose values are set in experimentation. The random variable $\varepsilon \sim N(0, \sigma^2)$ where σ^2 is an unknown constant.

It can be shown that, given that the optimal values for the observation times are within the permitted interval, the locally D-optimal design has two support points. Any D-optimal design with as many support points as there are mean parameters to be estimated has equal weights for each support point (see, for example, pages 42-43, Silvey 1980). As both support points of a D-optimal two point design have the same weights, we use $d(T_1, t_1; T_2, t_2)$ as a shorthand for the design, i.e.

$$\xi = \left\{ \begin{array}{cc} (T_1, t_1) & (T_2, t_2) \\ 1/2 & 1/2 \end{array} \right\} = d(T_1, t_1; T_2, t_2).$$

Our strategy for obtaining D-optimal designs for model (2.1) is as follows: the D-optimal two support point design is found, and it is then shown, via the general equivalence theorem, that this is the D-optimal design.

We use the following equivalence theorem, to obtain sufficient conditions for a design to be D-optimal:

Theorem 2.3.1 *(Kiefer and Wolfowitz, 1960): Given a model with p_m mean parameters and a space, $\mathcal{X}$ of possible combinations of T and t, as in Equation* (1.6)*, a design ξ with information matrix $M(\xi, \boldsymbol{\theta})$ is locally D-optimal if*

$$v(T, t, \boldsymbol{\theta}, \xi) = f(T, t, \boldsymbol{\theta})' M(\xi, \boldsymbol{\theta})^{-1} f(T, t, \boldsymbol{\theta}) \leq p_m \quad \forall\ (T, t) \in \mathcal{X}, \tag{2.3}$$

where $f(T, t, \boldsymbol{\theta}_m)$ is the vector of partial derivatives of the expected response with respect to the elements of $\boldsymbol{\theta}_m$ at (T, t), and equality in Equation (2.3) *occurs at the*

support points of ξ.

The following result characterises locally D-optimal designs for model (2.1) by providing the times and temperatures at which observations are made for each of the runs of the process, when the temperature range and the range for the time elapsed are as specified.

Theorem 2.3.2 *Let* $T_1,\ T_2 \in [T_{min}, T_{max}]$ *and* $t_1,\ t_2 \in [t_{min}, t_{max}]$ *and assume* $x, b, h, \lambda \in \mathbb{R}$ *and* $x \geq 0,\ b, \lambda > 0,\ b > h$, *with the restrictions that for* $\lambda < 0.5$, $\lambda \exp(1) > b/h$ *or* $\lambda < \frac{b-h}{b}$. *Define*

$$t_{T_i} = \frac{x_0}{c\theta_1 \exp(-l\theta_2/T_i)} \qquad (i = 1, 2),$$

where x_0 *is the root of the function*

$$g(x) = b(\lambda x - 1)\exp(x) + h(x + 1), \tag{2.4}$$

with $x \geq 0$. *Then a D-optimal design for model* (2.1) *is the two support point design:* $d(T_{max}, t_{T_{max}}; T_{min}, t_{T_{min}})$, *provided that, for all* $T \in [T_{min}, T_{max}]$, *the time* t_T *remains in the interval* $[t_{min}, t_{max}]$.

For clarity, we prove this theorem by splitting it into four lemmas, the proofs of which are given in Appendix A. We assume throughout that the conditions assumed in the Theorem are true.

The properties of the function $g(x)$ are first established.

Lemma 2.3.3 *The function g(x) defined in Equation* (2.4) *has exactly one root,* x_0. *In addition,* $g(x) < 0\ \forall x < x_0$ *and* $g(x) > 0\ \forall x > x_0$.

The optimal observation time t_i, $i = 1, 2$, when both T_1 and T_2 are fixed can now be established. This time point is determined by the value of T_i.

Lemma 2.3.4 *For a two point design* $d(T_1, t_1; T_2, t_2)$, *with fixed positive values for* $T_1,\ T_2$ *and* $t_i\ (i = 1, 2)$, *the value of* $t_j,\ j \neq i\ (j = 1, 2)$, *which maximises the*

determinant of the information matrix $M(\xi, \boldsymbol{\theta})$ *is*

$$t_{T_j} = \frac{x_0}{c\theta_1 \exp(-l\theta_2/T_j)},$$

where x_0 *is the root of the function*

$$g(x) = b(\lambda x - 1)\exp(x) + h(x + 1).$$

We next establish the D-optimal two point design.

Lemma 2.3.5 *If the time* t_{T_i} *is in the interval* $[t_{min}, t_{max}]$ *for all* T_i $(i = 1, 2)$*, then the* D*-optimal two point design is* $d(T_{max}, t_{T_{max}}; T_{min}, t_{T_{min}})$.

The two support point design is now shown to be D-optimal which completes the proof of Theorem 2.3.2.

Lemma 2.3.6 *The* D*-optimal design for model* (2.1) *is the two support point design* $d(T_{max}, t_{T_{max}}; T_{min}, t_{T_{min}})$.

The following corollary provides the D-optimal design for the situation where the values of T_1 and T_2 are predetermined, and are chosen such that $t_{T_j} < t_{min}$ or $t_{T_j} > t_{max}$ $(j = 1, 2)$. As might be expected, the optimal time at which to make an observation is on the lower or upper boundary of the specified time range

Corollary 2.3.7 *For a two point design* $d(T_1, t_1; T_2, t_2)$*, with fixed positive values for* T_1*,* T_2 *and* t_i $(i = 1, 2)$*, if the value of* t_j $(j = 1, 2)$ $j \neq i$ *is constrained to be within* $[t_{min}, t_{max}]$*, then the value of* $t_j = t_{opt}$ *that maximises the determinant of the information matrix* M *is*

$$t_{opt} = \begin{cases} t_{min} & \text{if } t_{T_j} < t_{min} \\ t_{max} & \text{if } t_{T_j} > t_{max}. \end{cases}$$

The parameter combinations h/b and λ for which Theorem 2.3.2 holds are displayed in Figure 2.1.

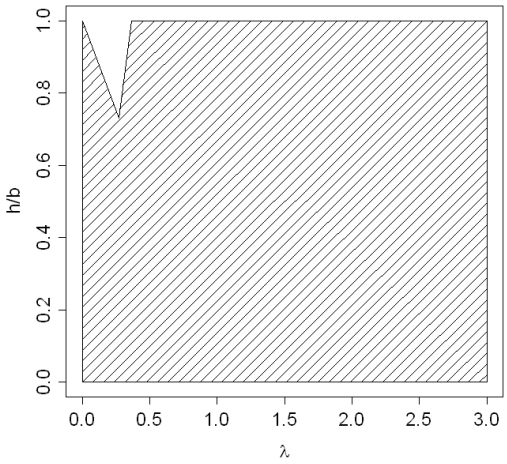


Figure 2.1: Parameter combinations h/b and λ for which Theorem 2.3.2 has been shown to be true (shaded region)

Outside this permitted region, the function $g(x)$ will occasionally have three roots, for instance Figure 2.2 displays $g(x)$ for $\lambda = 0.25$, $b = 1.01$ and $h = 1$.

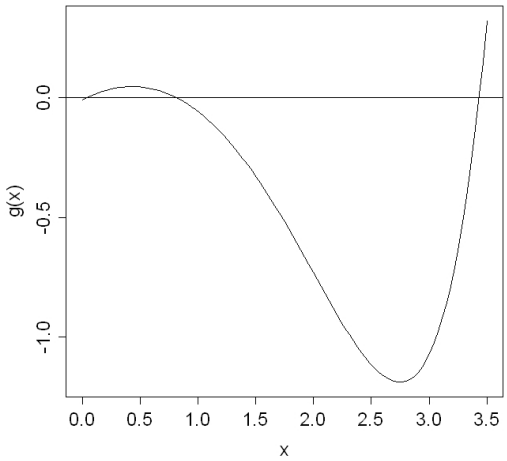


Figure 2.2: $g(x)$ for $\lambda = 0.25$, $b = 1.01$ and $h = 1$

D-optimal designs for several parameter combinations which are not in the shaded region were found numerically. It was found that for $\theta_2 = 4 \times 10^4$, $\theta_1 = 2128327$, $\lambda = 0.25$, $b = 1.01$ and $h = 1$ the optimal times of observation were at the first root in Figure 2.2, although multiple numerical design searches with different starting values would often give the third root as the optimal time to

make observations, implying that when searching for designs in this region numerical searches could easily become trapped at local optima. The first root is not always the optimal point at which to make observations. When $h = 0.98$ in the above example the optimal times of observation are given by the third root.

While the Theorem provides sufficient conditions for b, h and λ so that the design obtained is optimal, there are values of b, h and λ outside the specified region where the theorem's results still hold. For instance, Figure 2.3 demonstrates that for $\lambda = 0.25$, $b = 1.01$ and $h = 0.96$, $g(x)$ still only has one root despite this parameter combination being outside the shaded region given in Figure 2.1.

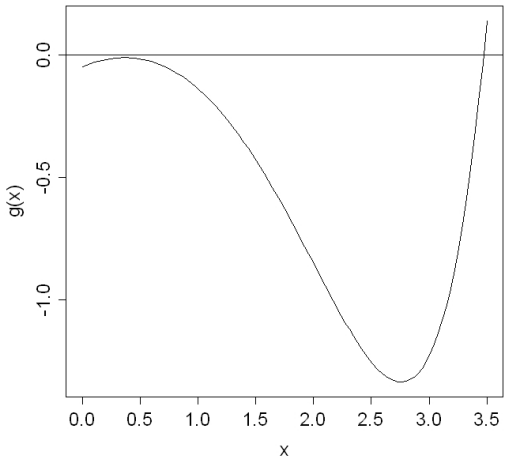


Figure 2.3: $g(x)$ for $\lambda = 0.25$, $b = 1.01$ and $h = 0.96$

The theorem and corollary enable locally optimal designs for this model to be quickly obtained. While there are some conditions on the parameter values for which these results apply, it is easy to ascertain when these conditions hold.

2.4 Exploration of locally optimal designs

The results of the previous section can be applied to obtain locally D-optimal designs for model (1.8) with independent errors. In what follows, we will first consider the special case where $\lambda = 1$ (for which the theorem holds for all values of h and b), which will subsequently be relaxed in Section 2.6.

We assume that $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$, values obtained by previous experimentation. Then as $b = 1.336$ and $h = 2 \times 0.14$, the approximate value of the root, x_0, is 0.8329874. From Theorem 2.3.2 the optimal times (in minutes) at which to make observations for temperature $T_{min} = 70°C$ and $T_{max} = 100°C$ is obtained from:

$$t_T = \frac{x_0}{(b_0 - 2a_0)60k_r \exp\left(-\frac{E_a}{R}\left[\frac{1}{T} - \frac{1}{T_r}\right]\right)}. \tag{2.5}$$

This gives the following approximate design:

$$\xi = \left\{ \begin{array}{cc} (70, 40.4) & (100, 10.1) \\ 0.5 & 0.5 \end{array} \right\}.$$

Figure 2.4 shows the standardised variance $v(T, t, \boldsymbol{\theta}, \xi)$ (defined in Equation (2.3)) for this design over $\mathcal{X}$, confirming that the design is optimal, by Theorem 2.3.1.

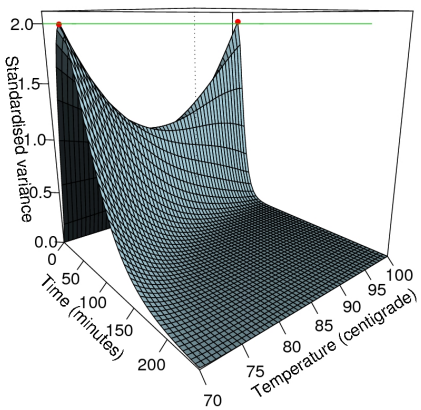


Figure 2.4: The standardised variance against time and temperature

While there may be some previous experimentation available to provide estimates of the parameter values for this model, if the estimates are poor then the locally D-

optimal design may take observations a considerable distance from the optimal times for the "true", unknown, parameter values. Collected in Table 2.1 are the optimal designs for some other possible values of E_a and k_r, where t_1 is the observation time on a run at temperature $T_1 = 70$ and t_2 is the observation time on a run at temperature $T_2 = 100$.

Table 2.1: Locally D-optimal designs for values of k_r and E_a

k_r	E_a	t_1	t_2
1×10^{-4}	2×10^4	151.27	86.04
	4×10^4	174.04	56.32
	4.91×10^4	185.51	46.44
	6×10^4	200.25	36.86
	7×10^4	214.80	29.82
4×10^{-4}	2×10^4	37.82	21.51
	4×10^4	43.51	14.08
	4.91×10^4	46.38	11.61
	6×10^4	50.06	9.22
	7×10^4	53.70	7.46
4.59×10^{-4}	2×10^4	32.96	18.75
	4×10^4	37.92	12.27
	4.91×10^4	40.42	10.12
	6×10^4	43.63	8.03
	7×10^4	46.80	6.50
5×10^{-4}	2×10^4	30.25	17.21
	4×10^4	34.81	11.26
	4.91×10^4	37.10	9.29
	6×10^4	40.05	7.37
	7×10^4	42.96	5.96
10×10^{-4}	2×10^4	15.13	8.60
	4×10^4	17.40	5.63
	4.91×10^4	18.55	4.64
	6×10^4	20.03	3.69
	7×10^4	21.48	2.98

Two clear trends can be seen from this study, which are apparent from examination of Equation (2.5). The optimal observation time becomes lower as k_r increases. This is intuitive, as higher values for k_r indicates that the reaction is faster than for a smaller value of k_r. As E_a increases, the optimal observation time increases or decreases depending on whether $T = 70$ or 100 respectively. In addition

the magnitude of the effect E_a has on the optimal observation times decreases as k_r increases. For $k_r = 10 \times 10^{-4}$, the difference between t_1 for $E_a = 2 \times 10^4$ and $E_a = 7 \times 10^4$ is only 6.35, while for $k_r = 1 \times 10^{-4}$ the difference is 63.53.

For an extremely slow reaction ($k_r = 1 \times 10^{-4}$ and $E_a > 6 \times 10^4$) the optimal value for t_1 is greater than 200. For the purposes of design comparison this is permitted, but in practice the optimal design would need to be found numerically, as the results of the previous section have only been proven to hold when the optimal time points are within the given region for time.

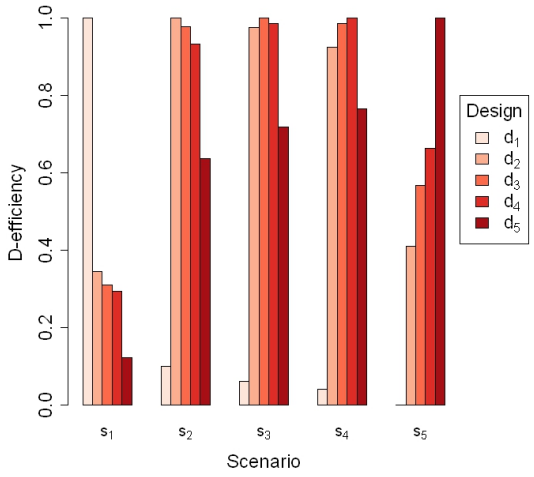
The robustness of the locally optimal designs to discrepancies between the true and assumed parameter values was assessed via D-efficiency (Section 1.3) for five combinations of k_r and E_a, shown in Table 2.2.

Table 2.2: Designs and scenarios for the D-efficiency study

Design	Scenario	k_r	E_a	t_1	t_2
d_1	s_1	1×10^{-4}	7×10^4	214.80	29.82
d_2	s_2	4×10^{-4}	6×10^4	50.06	9.22
d_3	s_3	4.59×10^{-4}	4.91×10^4	40.42	10.12
d_4	s_4	5×10^{-4}	4×10^4	34.81	11.26
d_5	s_5	10×10^{-4}	2×10^4	15.13	8.60

The results of the D-efficiency study are shown in Figure 2.5. While the D-efficiency of d_3 remains reasonably high for scenarios where the true and assumed parameters value are similar, under the more extreme scenarios, s_1 and s_5, the D-efficiency of d_3 drops to below 0.6. In particular, in s_1, where the reaction is happening much more slowly than anticipated when the experiment was planned, the efficiency of d_3 is only just over 0.3. There is no locally optimal design which performs well across all five scenarios. The design d_2, which assumed a slightly slower reaction than anticipated from previous experimentation, has the largest minimum D-efficiency of 0.34 (to 2 d.p.).

Figure 2.5: D-efficiencies of five designs for each of five scenarios



2.5 Robust designs

In the previous section, evidence was provided that mis-specifying the true value of the mean parameters can lead to low D-efficiency for the obtained design. Two criteria which might render designs robust to the impact of parameter misspecification, Bayesian and maximin D-optimality, are now explored. Unlike for locally D-optimal designs, the number of support points for Bayesian and maximin D-optimal designs is not guaranteed to be two, but is unbounded, see Braess and Dette (2007).

2.5.1 Bayesian D-optimal designs

In this subsection Bayesian D-optimal designs are found for a number of different prior distributions, chosen to reflect likely prior knowledge possessed before experimentation. The designs found are confirmed to be optimal via use of an equivalence theorem, then assessed via their Bayesian D-efficiency.

Bayesian D-optimal designs, maximise the objective function (1.15) (in Section 1.3). The integral in the objective function (1.15) usually cannot be solved analytically, and hence must be estimated numerically. Orthogonal array based Latin hypercubes (Fang et al., 2006) of 20^2 points across the parameters E_a and k_r

were found. Orthogonal array based Latin hypercubes are grids of points designed to fill a possibly multidimensional space and ensure no area is left uncovered, by stratifying by each dimension simultaneously. Their space filling properties have been demonstrated by such authors as McKay et al. (1979) and Tang (1993). These hypercubes were used to provide Monte Carlo estimates of the integral, and are known to be reasonably accurate for two dimensions. Analytical results are not available for these Bayesian D-optimal designs and there is no guarantee that the optimal design will have two support points. To search across the set of all possible optimal designs `optim` in `R`, was utilised, using algorithm `L-BFGS-B`, as given by Byrd et al. (1995), which uses gradients to search across a constrained region, and, given a design ξ returns a design, $\xi^{\dagger}$, which achieves a greater or equal value for the given objective function. The design search was then performed using the following algorithm:

Algorithm 2.1 Algorithm to find Bayesian D-optimal designs

1: Set the number of support points $n = n_0$, where $n_0 \geq 2$
2: Generate 10 random starting designs, $\{d_1, \ldots, d_{10}\}$ each with n_0 support points having equal weighting
3: For d_i, $(i = 1, \ldots, 10)$, apply `L-BFGS-B` to obtain $d_i^{\dagger}$. For each $d_i^{\dagger}$ if the weight of any support points is less than δ, $0 < \delta < 1$, then the weight is set to 0.
4: If the number of nonzero support points for at least one $d_i^{\dagger}$ $(i = 1, \ldots, 10)$ is equal to n_0, set $n_0 = n_0 + 2$ and go to step 2
5: Otherwise, let d be the design in $\left\{d_1^{\dagger}, \ldots, d_{10}^{\dagger}\right\}$ that has the highest value obtained for the objective function (1.15)
6: Let n_0 be the number of support points with non-zero weight for d, and apply `L-BFGS-B` to d to obtain $d^{\dagger}$, which is accepted as the optimal design

The value of δ was chosen to be sufficiently small that said support point would have little impact on the design: for our search, this was 0.01. Searching over a wide range of values for each parameter can cause computational problems. For certain combinations of parameter values and designs, the value of the determinant of the information matrix can be extremely small, indicating that this design is providing very little information about that combination of parameters. Rounding errors in the numerical evaluation can then cause the calculated determinant to be negative which is not possible for the true value of the determinant (of a non

negative definite matrix). In those situations a value is assigned to the determinant for that combination of parameter values of 1% of the minimum determinant value (for that design) obtained for all combinations of parameter values that have non negative calculated determinant.

Prior distributions for the unknown parameter values E_a and k_r were chosen to reflect the following scenarios:

1. Previous experiments provided information on the parameter values

2. No previous experiments are available, and scientific knowledge is used to choose the prior distributions

The parameters E_a and k_r are assumed to be independent throughout this thesis. Hence, from Equation (1.4), there is an implied distribution for z, which is dependent on both E_a and k_r. We choose to specify distributions for E_a and k_r as these parameters have a straightforward chemistry interpretation (the activation energy and reaction rate at reference temperature 77°C, respectively) which facilitates elicitation of prior distributions.

Clearly, dependent prior distributions for E_a and k_r could be chosen if this better reflects chemists' prior knowledge. However, even in this situation, assuming independent distributions for E_a and k_r may lead to a wider range of parameter combinations having substantive prior density. Hence, a more robust design may result.

We now define and justify the prior distributions for E_a and k_r used for each scenario.

Scenario 1: We assumed that some information on the mean and variance for each parameter was available. A normal prior distribution was chosen for E_a and a gamma prior distribution for k_r. A gamma prior distribution for k_r was chosen to ensure that k_r does not become negative. Neither E_a or k_r are allowed to be negative, but the mean and variance for the normal prior distribution for E_a ensures that the probability of $E_a < 0$ is negligibly small.

Distributions of $E_a \sim \mathrm{N}(4.91\times10^4,(1000K)^2)$ and $k_r \sim \mathrm{G}(m_k^2/v_k, v_k/m_k)$, with a mean of $m_k = 4.59\times10^{-4}$ and variance of $v_k = (1\times10^{-4}K)^2$ were chosen. The prior distributions are labeled $\pi_1(K)$ and $\pi_2(K)$ respectively. These reflected plausible mean values for E_a and k_r. For π_1, $K = 1$, 4 and 10. For π_2, $K = 1$, 4 and 8. The values of K were chosen to reflect the possible level of prior confidence on the accuracy of the estimates of k_r and E_a obtained from prior experimentation.

Scenario two: Two uniform prior distributions were chosen:
$E_a \sim \mathrm{U}[4\times10^4,6\times10^4]$ and $k_r \sim \mathrm{U}[4\times10^{-4},6\times10^{-4}]$, and $E_a \sim \mathrm{U}[2\times10^4,7\times10^4]$ and $k_r \sim \mathrm{U}[1\times10^{-4},10\times10^{-4}]$ which reflected typical prior knowledge possessed by practitioners. These are labelled π_3, π_4, π_5 and π_6 respectively.

Findings: The Bayesian D-optimal designs obtained for each prior distribution are given in Table 2.3. To save space, we label $T_l = 70$ and $T_u = 100$.

For scenario one, the impact on design of altering the variance on the prior distribution of k_r is much more pronounced than altering the variance on the prior distribution of E_a. The optimal designs when the variance of k_r is $(1 \times 10^{-4})^2$ $(d_6 - d_9)$ have two support points, and are very similar to the locally optimal design (d_3), obtained when E_a and k_r were assumed to be the means of π_1 and π_2, 4.91×10^4 and 4.59×10^{-4} respectively.

Once the variance of the prior distribution π_2 of k_r is $(4 \times 10^{-4})^2$ or greater, the optimal design has six or seven support points. All designs with six or seven support points, while having very different weightings on each support point, have several things in common. All of them have observations at the upper value for time, 200, and the other support points occur mainly at the start of the process: most of the optimal time of observations are before 25, with observations at 47.1 and 173.7 in d_9, d_{10} and d_{11}. As might be expected, the support points are distributed fairly evenly between the two temperatures; while in the seven support point designs there are four support points at T_u and only three at T_l, the weights of the three add up to ≈ 0.5.

For scenario two, the impact of the choice of a uniform prior distribution for E_a and k_r when designing is simpler. Both d_{15} and d_{16} have only two support points,

Table 2.3: Bayesian D-optimal designs d_i, $(i = 1, \ldots 16)$ found via Algorithm 2.1 with $T_l = 70$, $T_u = 100$ and $\pi_j(K)$ defined in text

Scenario	Prior distribution for E_a	Prior distribution for k_r	Design
1	$\pi_1(1)$	$\pi_2(1)$	$d_6 = \left\{ \begin{array}{cc} (T_l, 40.6) & (T_u, 10.2) \\ 0.500 & 0.500 \end{array} \right\}$
1	$\pi_1(4)$	$\pi_2(1)$	$d_7 = \left\{ \begin{array}{cc} (T_l, 40.6) & (T_u, 10.1) \\ 0.500 & 0.500 \end{array} \right\}$
1	$\pi_1(10)$	$\pi_2(1)$	$d_8 = \left\{ \begin{array}{cc} (T_l, 40.5) & (T_u, 10.0) \\ 0.500 & 0.500 \end{array} \right\}$
1	$\pi_1(1)$	$\pi_2(4)$	$d_9 = \left\{ \begin{array}{cccccc} (T_l, 31.9) & (T_l, 47.1) & (T_l, 200.0) & (T_u, 8.9) & (T_u, 25.0) & (T_u, 173.7) \\ 0.228 & 0.211 & 0.0612 & 0.387 & 0.0987 & 0.0146 \end{array} \right\}$
1	$\pi_1(4)$	$\pi_2(4)$	$d_{10} = \left\{ \begin{array}{cccccc} (T_l, 32.0) & (T_l, 47.1) & (T_l, 200.0) & (T_u, 8.8) & (T_u, 25.0) & (T_u, 173.7) \\ 0.229 & 0.210 & 0.0612 & 0.382 & 0.104 & 0.0137 \end{array} \right\}$
1	$\pi_1(10)$	$\pi_2(4)$	$d_{11} = \left\{ \begin{array}{ccccccc} (T_l, 31.9) & (T_l, 47.1) & (T_l, 200.0) & (T_u, 8.05) & (T_u, 18.8) & (T_u, 25.0) & (T_u, 173.7) \\ 0.230 & 0.127 & 0.0617 & 0.333 & 0.0828 & 0.0686 & 0.0154 \end{array} \right\}$
1	$\pi_1(1)$	$\pi_2(8)$	$d_{12} = \left\{ \begin{array}{ccccccc} (T_l, 8.8) & (T_l, 23.4) & (T_l, 200.0) & (T_u, 2.2) & (T_u, 5.0) & (T_u, 20.9) & (T_u, 200.0) \\ 0.0362 & 0.161 & 0.303 & 0.0308 & 0.132 & 0.141 & 0.196 \end{array} \right\}$
1	$\pi_1(4)$	$\pi_2(8)$	$d_{13} = \left\{ \begin{array}{ccccccc} (T_l, 10.5) & (T_l, 24.7) & (T_l, 200.0) & (T_u, 2.4) & (T_u, 5.5) & (T_u, 21.9) & (T_u, 200.0) \\ 0.0526 & 0.145 & 0.303 & 0.0409 & 0.129 & 0.134 & 0.196 \end{array} \right\}$
1	$\pi_1(10)$	$\pi_2(8)$	$d_{14} = \left\{ \begin{array}{ccccccc} (T_l, 9.5) & (T_l, 24.4) & (T_l, 200.0) & (T_u, 3.0) & (T_u, 8.1) & (T_u, 23.3) & (T_u, 200.0) \\ 0.0439 & 0.154 & 0.302 & 0.0900 & 0.0993 & 0.115 & 0.196 \end{array} \right\}$
2	π_3	π_4	$d_{15} = \left\{ \begin{array}{cc} (T_l, 37.2) & (T_u, 9.1) \\ 0.500 & 0.500 \end{array} \right\}$
2	π_5	π_6	$d_{16} = \left\{ \begin{array}{cc} (T_l, 32.8) & (T_u, 9.0) \\ 0.500 & 0.500 \end{array} \right\}$

and as the support on the prior distributions for the mean parameters becomes wider, the optimal designs have observations earlier in time.

Checking Bayesian D-optimality: The designs in Table 2.3 were obtained by numerical optimisation, not by analytic results, which may lead to a local optimum being found in the search instead of the global optimum. We can check if a design is Bayesian D-optimal by applying an equivalence theorem (see Firth and Hinde 1997).

Theorem 2.5.1 *A Bayesian D-optimal design ξ^* satisfies the following inequality, with $p_m = 2$,*

$$\int f(T,t,\boldsymbol{\theta})' M(\xi^*,\boldsymbol{\theta})^{-1} f(T,t,\boldsymbol{\theta}) d\pi(\boldsymbol{\theta}) \leq p_m \quad \forall\ (T,t) \in \mathcal{X},$$

with equality to p_m occurring at each support point.

We cannot solve this inequality analytically, but we can check if the inequality is violated over a wide grid of values for T and t. The value of the integrated standardised variance can then be plotted, as in Figures 2.6 and 2.7

These results are only approximate, but show that the designs found are either optimal or near optimal. For designs $d_9 - d_{14}$, some of the support points did not attain the value of $p_m = 2$, although they were very close.

The performance of the Bayesian D-optimal designs under each assumed prior distribution can be compared via their Bayesian D-efficiency, as in Equation (1.16) (given in Section 1.3).

Table 2.4 displays the Bayesian D-efficiencies for the designs $d_1 - d_{16}$. As might be anticipated from Table 2.3, to two decimal places there is no difference in efficiency between the Bayesian D-optimal designs with the same prior distribution for k_r but different prior distributions for E_a. As with the results from the D-efficiency study for locally optimal designs, d_1 performs poorly throughout all scenarios. The best performing designs on average appear to be d_9-d_{11}, whose Bayesian D-efficiency remains reasonably high throughout, only dropping to 0.53 when the variance for

Table 2.4: Bayesian D-efficiencies of the designs d_1-d_{16}

	Prior distribution										
Designs	$\pi_1(1),\pi_2(1)$	$\pi_1(4),\pi_2(1)$	$\pi_1(10),\pi_2(1)$	$\pi_1(1),\pi_2(4)$	$\pi_1(4),\pi_2(4)$	$\pi_1(10),\pi_2(4)$	$\pi_1(1),\pi_2(8)$	$\pi_1(4),\pi_2(8)$	$\pi_1(10),\pi_2(8)$	π_3,π_4	π_5,π_6
d_1	0.06	0.06	0.06	0.06	0.06	0.06	0.02	0.02	0.02	0.04	0.02
d_2	0.98	0.98	0.98	0.95	0.95	0.94	0.28	0.29	0.28	0.96	0.91
d_3	1.00	1.00	1.00	0.96	0.96	0.96	0.29	0.29	0.29	0.99	0.97
d_4	0.99	0.99	0.98	0.95	0.95	0.94	0.28	0.28	0.28	0.98	0.97
d_5	0.72	0.72	0.72	0.67	0.67	0.67	0.19	0.19	0.19	0.76	0.82
d_6	1.00	1.00	1.00	0.96	0.96	0.96	0.29	0.29	0.29	0.99	0.97
d_7	1.00	1.00	1.00	0.96	0.96	0.96	0.29	0.29	0.29	0.99	0.97
d_8	1.00	1.00	1.00	0.96	0.96	0.96	0.29	0.29	0.28	0.99	0.97
d_9	0.86	0.86	0.86	1.00	1.00	1.00	0.53	0.53	0.53	0.85	0.89
d_{10}	0.85	0.85	0.86	1.00	1.00	1.00	0.53	0.53	0.53	0.84	0.89
d_{11}	0.84	0.84	0.84	1.00	1.00	1.00	0.54	0.55	0.54	0.83	0.89
d_{12}	0.34	0.34	0.35	0.69	0.69	0.70	1.00	1.00	1.00	0.33	0.48
d_{13}	0.34	0.34	0.35	0.69	0.69	0.70	1.00	1.00	1.00	0.33	0.48
d_{14}	0.35	0.35	0.35	0.69	0.69	0.70	1.00	1.00	1.00	0.34	0.48
d_{15}	0.99	0.99	0.99	0.95	0.95	0.95	0.28	0.28	0.28	1.00	0.99
d_{16}	0.98	0.98	0.98	0.93	0.93	0.93	0.27	0.27	0.27	0.99	1.00

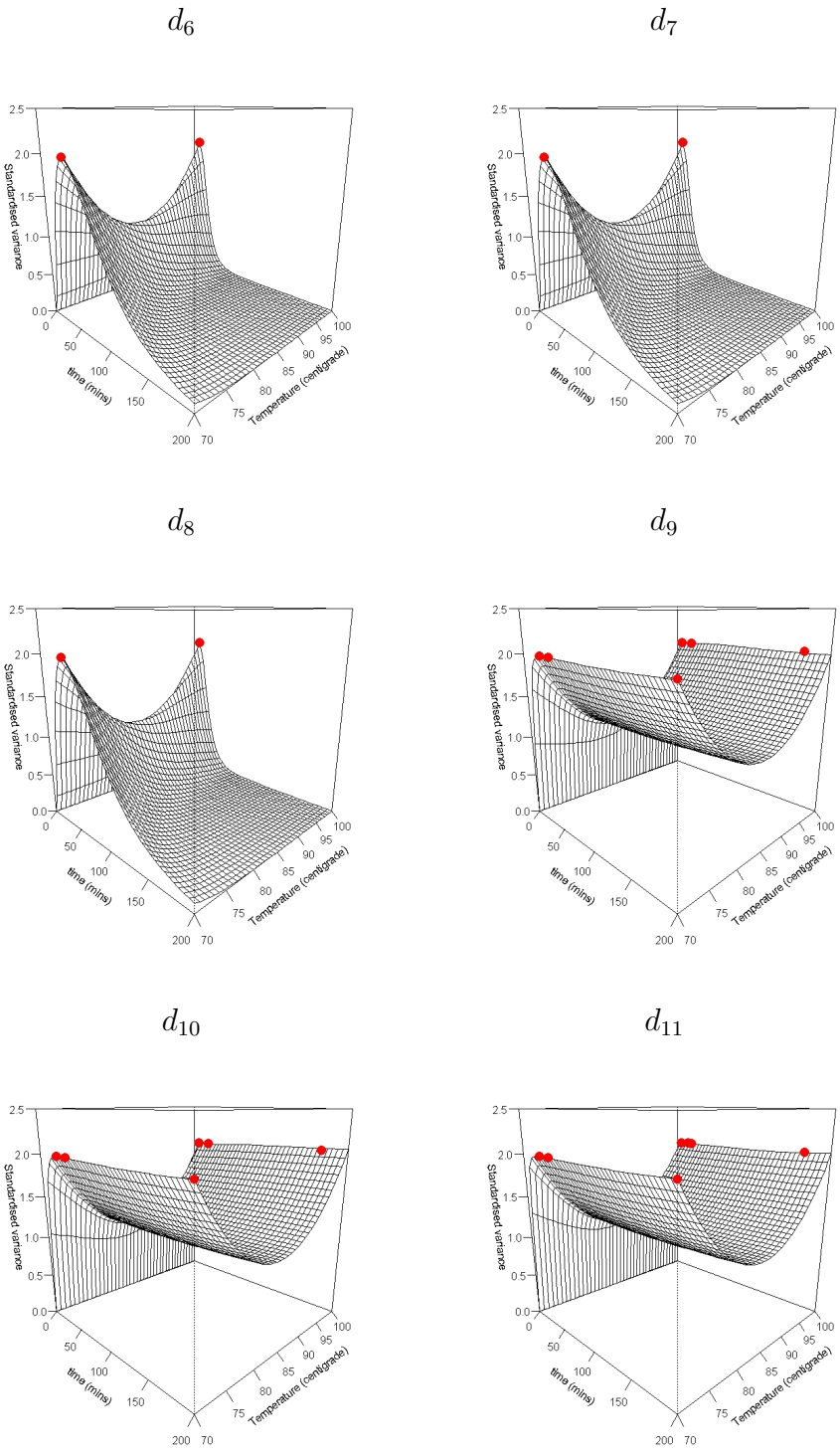
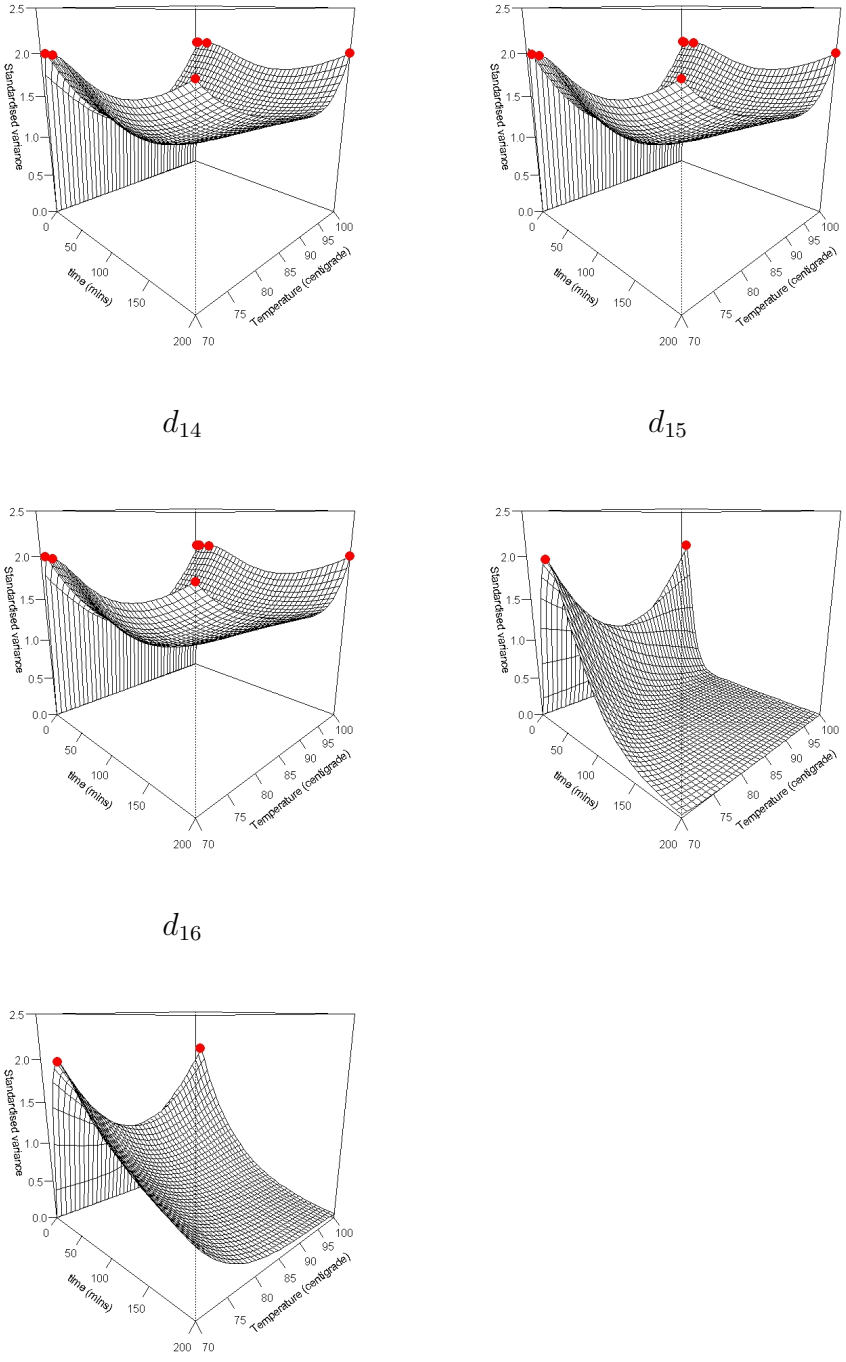


Figure 2.6: Value of the integrated standardised variance for d_6-d_{11}, support points given by dots on each figure

the prior distribution of k_r is four times that anticipated while designing (under the prior distribution $\pi_2(8)$ for k_r). For that prior distribution, with the exception of

d_{12} d_{13}



d_{14} d_{15}

d_{16}

Figure 2.7: Value of the integrated standardised variance for d_{12}-d_{16}, support points given by dots on each figure

designs $d_{12} - d_{14}$, all designs have a low Bayesian D-efficiency.

Bayesian D-optimal designs are usually found when prior information is

available on the parameter values and this information is used to choose the prior distributions. If only limited information is available, and practitioners wish to find a design whose performance is robust to extreme values of k_r and E_a, then $\pi_1(4)$ and $\pi_2(4)$ might be selected as the respective prior distributions for E_a and k_r because, in this study, d_9 achieved a high average Bayesian D-efficiency.

2.5.2 Maximin D-optimal designs

Maximin D-optimal designs, defined in Section 1.3, are used less than Bayesian D-optimal designs in practice as they can be difficult to obtain when locally D-optimal designs have to be found numerically.

Finding maximin D-optimal designs can be burdensome, as it requires finding the optimal design for every single combination of parameter values. In practice, without analytic results the objective function for maximin D-optimality can only be approximated, as with Bayesian D-optimality. A grid of 20^2 parameter values generated via an orthogonal array based Latin hypercube was searched across, as in the previous section.

As before, the number of support points in the optimal design is not guaranteed to be two. Algorithm 2.1 was used to obtain optimal designs. Two possible ranges for the parameters values, Θ_1 and Θ_2 were considered, corresponding to the domains of the prior distributions $\pi_3 - \pi_6$ used in the previous section.

- Θ_1: $E_a \in [4\times10^4, 6\times10^4]$ and $k_r \in [4\times10^{-4}, 6\times10^{-4}]$
- Θ_2: $E_a \in [2\times10^4, 7\times10^4]$ and $k_r \in [1\times10^{-4}, 10\times10^{-4}]$

Maximin D-optimal designs, labeled d_{17} and d_{18} were found for Θ_1 and Θ_2 respectively. For d_{17} the optimal design was easily obtained, but for d_{18} the search proved more difficult, so a hybrid search algorithm was utilised, as given by Algorithm 2.2. The genetic algorithm generates 10 random designs for a given number of support points and then, through 100 generations "evolves" to obtain an optimal design $\hat{\xi}$ via methods described in Hamada et al. (2001).

Table 2.5: Maximin D-optimal designs obtained for parameter spaces Θ_1 and Θ_2, $T_l = 70$, $T_u = 100$

Parameter space	Design
Θ_1	$d_{17} = \begin{Bmatrix} (T_l, 38.29) & (T_u, 9.06) \\ 0.5 & 0.5 \end{Bmatrix}$
Θ_2	$d_{18} = \begin{Bmatrix} (T_l, 137.3) & (T_u, 2.8) & (T_u, 4.6) & (T_u, 6.6) & (T_u, 37.8) & (T_u, 78.8) \\ 0.284 & 0.008 & 0.147 & 0.256 & 0.201 & 0.104 \end{Bmatrix}$

Algorithm 2.2 Algorithm to find maximin D-optimal designs

1: Set the number of support points, $n = n_0$, where $n_0 > 2$
2: Conduct the genetic algorithm for n_0, obtaining a design $\hat{d}$
3: Apply `L-BFGS-B` in `optim` to $\hat{d}$, obtaining $\hat{d}^{\dagger}$
4: If the weight of a support point in $\hat{d}^{\dagger}$ is less than δ, $0 < \delta < 1$, then the weight is set to 0
5: If the number of nonzero support points in $\hat{d}^{\dagger}$ is equal to n_0, set $n_0 = n_0 + 2$ and go to step 2
6: Otherwise, apply the genetic algorithm where one of the starting designs is $\hat{d}^{\dagger}$, obtaining $\hat{d}_2$,
7: Apply `L-BFGS-B` to $\hat{d}_2$, obtaining $\hat{d}_2{}^{\dagger}$
8: If the difference in value for (1.17) between $\hat{d}_2{}^{\dagger}$ and $\hat{d}_2$ is greater than $\delta^c > 0$, set $\hat{d}^{\dagger} = \hat{d}_2{}^{\dagger}$ and go to step 6
9: Otherwise, accept $\hat{d}_2{}^{\dagger}$ as the optimal design

These methods gave the maximin D-optimal designs given in Table 2.5. The design d_{17} is very similar to d_3. Examining d_{18} indicates that as the range the parameter values could lie within becomes larger more support points are required, as with the Bayesian D-optimal designs found in the previous section. Particularly of note is that, unlike with the Bayesian D-optimal designs, the best maximin D-optimal design gives more weighting to observations taken at T_u than T_l, with over 70% of observations being taken at the higher temperature T_u. While observations in d_{18} are spread out in time, none are at the permitted end points for time, unlike the Bayesian D-optimal designs of the previous sub-section. To compare these maximin D-optimal designs to those found so far the value for $\phi_M(d)$ for all designs was calculated.

The results are given in Table 2.6. As anticipated, the locally optimal design d_3 attains a minimum very close to that of the maximin D-optimal designs for Θ_1. Design d_1 performs poorly under this criterion for both Θ_1 and Θ_2. Other than d_{18}, the design which attains the highest minimum over the wider parameter space Θ_2 is d_{14}, which was the Bayesian D-optimal design with the most uncertainty specified over its parameter space.

The D-efficiencies over the parameter space for any given design can be plotted. Designs d_1, d_3, d_9 and d_{18} are plotted over Θ_2, they are a design with a very low

Table 2.6: Minimum D-efficiency $\phi_M(d)$ over each parameter range, Θ_1 and Θ_2 for designs d_1-d_{18}

Category	Design	Θ_1	Θ_2
Locally optimal designs	d_1	0.00982	2.63×10^{-5}
	d_2	0.861	0.215
	d_3	0.877	0.202
	d_4	0.837	0.199
	d_5	0.638	0.0806
Bayesian D-optimal designs	d_6	0.875	0.195
	d_7	0.875	0.194
	d_8	0.882	0.192
	d_9	0.764	0.260
	d_{10}	0.763	0.261
	d_{11}	0.750	0.263
	d_{12}	0.312	0.282
	d_{13}	0.312	0.284
	d_{14}	0.317	0.284
	d_{15}	0.924	0.167
	d_{16}	0.902	0.151
Maximin D-optimal designs	d_{17}	0.927	0.178
	d_{18}	0.462	0.454

minimum efficiency, the locally D-optimal design based on previous experimentation, a Bayesian D-optimal design, and the maximin D-optimal design respectively. These are plotted in Figure 2.8.

Misspecification of k_r has more impact on D-efficiency than misspecification of E_a. The design d_1 performs well for low values of k_r but its efficiency drops rapidly as k_r increases. While d_3 and d_1 attain the maximum value of one somewhere in the parameter region, neither d_9 or d_{18} do so, instead both are more flat over the parameter region. This means that both represent a compromise. Design d_9 maintains a good average performance across the parameter space, while it does not attain an efficiency much higher than 0.8, it gets a reasonably high efficiency elsewhere, only losing most of its efficiency when both E_a and k_r are small. The maximin D-optimal design has an almost flat efficiency plane with respect to the mean parameters. It will never perform too poorly, but will never perform very well either.

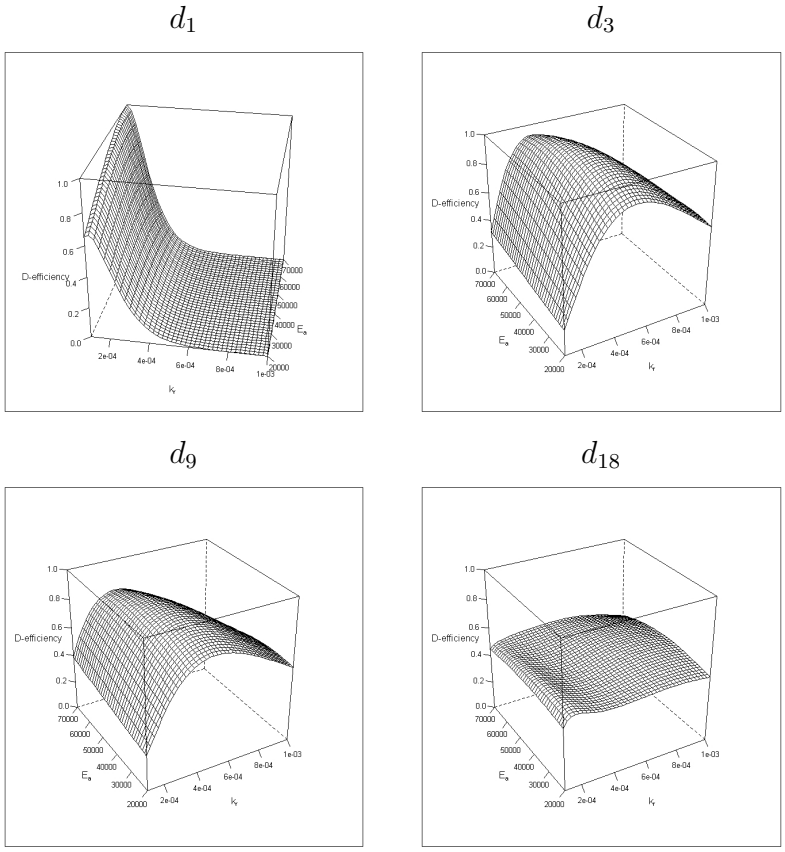


Figure 2.8: The D-efficiencies of, respectively, d_1, d_3, d_9 and d_{18} over Θ_2

Finally the Bayesian D-efficiency the maximin D-optimal designs achieve under the scenarios considered in the previous subsection can be calculated. The results are given in Table 2.7.

Table 2.7: Bayesian D-efficiencies of the maximin D-optimal designs

	Assumed prior distribution					
Designs	$\pi_1(1),\pi_2(1)$	$\pi_1(4),\pi_2(1)$	$\pi_1(10),\pi_2(1)$	$\pi_1(1),\pi_2(4)$	$\pi_1(4),\pi_2(4)$	$\pi_1(10),\pi_2(4)$
d_{17}	0.99	0.99	0.99	0.95	0.95	0.95
d_{18}	0.51	0.51	0.52	0.89	0.89	0.90
	Assumed prior distribution					
Designs	$\pi_1(1),\pi_2(8)$	$\pi_1(4),\pi_2(8)$	$\pi_1(10),\pi_2(8)$	π_3,π_4	π_5,π_6	
d_{17}	0.28	0.28	0.28	1.00	0.99	
d_{18}	0.70	0.70	0.70	0.51	0.67	

As d_{17} is very similar to d_3, it has a similar performance in terms of its Bayesian D-efficiency. The maximin D-optimal design d_{18} has a high Bayesian D-efficiency when the prior distribution for k_r has a large variance, but drops to only 0.5 efficiency when the prior distribution on k_r has a small variance.

2.6 D-optimal designs when heteroscedasticity in the errors is present

In this section there is assumed to be some heteroscedasticity present in the errors, and the data have been transformed by λ as in model (1.8). The transformation parameter λ is assumed to be between 0 and 1, and as $b_0/2a_0 < 1/2$, the results in Section 2.3 can be applied to obtain locally D-optimal designs when observations are assumed to come from independent runs of the process. For this section, the space of all possible values of combinations of T and t that might be used in experiments, is defined by:

$$\mathcal{X} = \{(T,t); (70 \leq T \leq 100, 1 \leq t \leq 1400)\},$$

which is expanded, as, as will be seen, optimal designs have observations much later in time for small values of λ. This implies that practitioners will need to be willing to take observations later in time if heteroscedasticity is suspected. In particular, for certain combinations of parameter values the D-optimal design will have observations at $t = 1400$, and to allow maximin designs to be obtained this wider region is used.

2.6.1 Locally D-optimal designs when heteroscedasticity in the errors is present

Table 2.8 gives the locally D-optimal designs for three different E_a and k_r values, and $\lambda = 1, 0.5, 0.25$ and 0.14. The λ values were chosen as they represent, respectively, the untransformed case, the square root, the fourth root, and 0.14 the mean value found in an empirical study conducted by Lischer (1999).

The impact of λ is to increase the optimal times of observation as λ decreases. The rate of increase is a function of $(1/\lambda)$, as can be seen in Figure 2.9, which plots the roots of $g(x)$ against $(1/\lambda)$.

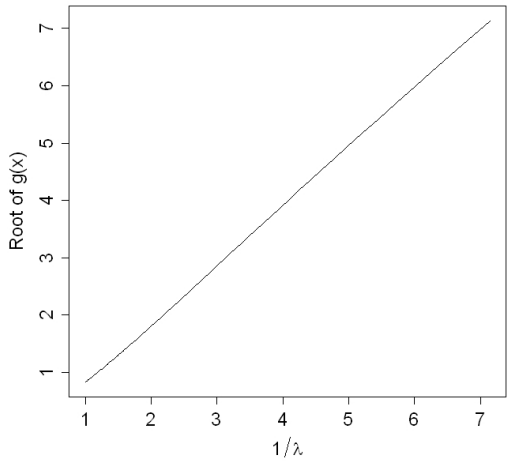


Figure 2.9: The roots of $g(x)$ as $1/\lambda$ increases for $a_0 = 0.14$ and $b_0 = 1.336$

Empirical investigation indicated that the root of $g(x)$, x_0, follows $x_0 = 1/\lambda^{o(a_0/b_0)}$, where o is an unknown function. For the particular values of a_0 and

Table 2.8: Locally D-optimal designs for varying values of E_a, k_r and λ

$k_r \times 10^4$	$E_a \times 10^{-3}$	λ	t_1 ($T_1 = 70$)	t_2 ($T_2 = 100$)
2.00	30.0	0.14	694.7	298.1
2.00	30.0	0.25	381.6	163.7
2.00	30.0	0.50	176.0	75.5
2.00	30.0	1.00	81.1	34.8
2.00	49.1	0.14	794.3	198.9
2.00	49.1	0.25	436.3	109.2
2.00	49.1	0.50	201.2	50.4
2.00	49.1	1.00	92.8	23.2
2.00	60.0	0.14	857.4	157.9
2.00	60.0	0.25	471.0	86.7
2.00	60.0	0.50	217.2	40.0
2.00	60.0	1.00	100.1	18.4
4.59	30.0	0.14	302.7	129.9
4.59	30.0	0.25	166.3	71.3
4.59	30.0	0.50	76.7	32.9
4.59	30.0	1.00	35.3	15.2
4.59	49.1	0.14	346.1	86.7
4.59	49.1	0.25	190.1	47.6
4.59	49.1	0.50	87.7	21.9
4.59	49.1	1.00	40.4	10.1
4.59	60.0	0.14	373.6	68.8
4.59	60.0	0.25	205.2	37.8
4.59	60.0	0.50	94.6	17.4
4.59	60.0	1.00	43.6	8.0
7.00	30.0	0.14	198.5	85.2
7.00	30.0	0.25	109.0	46.8
7.00	30.0	0.50	50.3	21.6
7.00	30.0	1.00	23.2	9.9
7.00	49.1	0.14	226.9	56.8
7.00	49.1	0.25	124.7	31.2
7.00	49.1	0.50	57.5	14.4
7.00	49.1	1.00	26.5	6.6
7.00	60.0	0.14	245.0	45.1
7.00	60.0	0.25	134.6	24.8
7.00	60.0	0.50	62.1	11.4
7.00	60.0	1.00	28.6	5.3

b_0 used in our motivating example, the value of $o(a_0/b_o)$ is close to 1, which is why the plot looks close to linear.

The D-efficiencies of the designs obtained assuming that $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, for the four values of λ as each λ value is assumed true in turn is given in Figure 2.10.

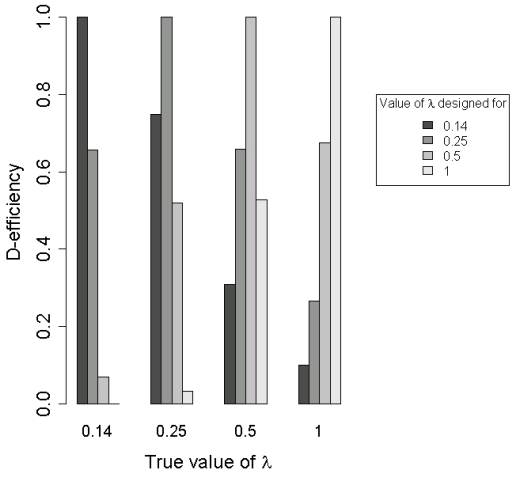


Figure 2.10: The relative D-efficiency of designs for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, and $\lambda = 0.14$, 0.25, 0.5 and 1

This plot demonstrates that incorrectly specifying the value of λ greatly impacts on the efficiency of the design obtained. The design with the highest minimum efficiency assumed that $\lambda = 0.25$, but still has a fairly low minimum efficiency of 0.27 when the true value of $\lambda = 1$.

2.6.2 Robust D-optimal designs when error heteroscedasticity is present

Designs whose performance is robust to the mean parameters when $\lambda = 1$ were found in Section 2.5. We now carry out a similar investigation with the additional goal that a design's performance should be robust to the value of λ.

In Section 2.5, the prior distributions on E_a and k_r which were found to lead to designs whose performance was robust to misspecification of the prior distribution

were $E_a \sim N(4.91 \times 10^4, 4000^2)$ and $k_r \sim \mathrm{G}(m_k^2/v_k, v_k/m_k)$, with a mean of $m_k = 4.59 \times 10^{-4}$ and variance of $v_k = (4 \times 10^{-4})^2$.

These prior distributions were used to find Bayesian D-optimal designs for three prior distributions for λ, π_{λ_1}, π_{λ_2} and π_{λ_3}, plotted in Figure 2.11. They are, respectively, $\lambda \sim \mathrm{Beta}(4,1)$, $\lambda \sim U(0,1)$ and $\lambda \sim \mathrm{Beta}(2, 12.8571)$. The prior distributions were chosen to reflect likely prior beliefs. The prior distribution π_{λ_1} was chosen to reflect the belief that the data are unlikely to be heteroscedastic, but the experimenter wishes to select a design that is robust to heteroscedasticity if it is present, π_{λ_2} is an ignorance prior distribution, and π_{λ_3} is centered around the mean of 0.14, found empirically.

A larger orthogonal array based Latin hypercube of 20^3 points was used to provide Monte Carlo estimates of the integral, as the integral to be evaluated is now across three dimensions.

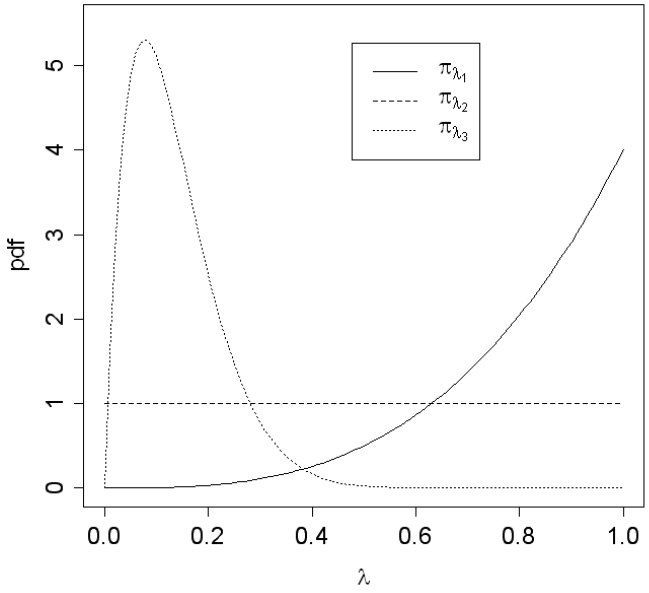


Figure 2.11: The probability density function of π_{λ_1}, π_{λ_2} and π_{λ_3}

As with Bayesian designs, prior information on λ can be specified to find a maximin D-optimal design over the space Θ_λ defined as $E_a \in [2{\times}10^4, 7{\times}10^4]$, $k_r \in [1{\times}10^{-4}, 10{\times}10^{-4}]$ and $\lambda \in [0.14, 1]$. As before, the full space was approximated by

a Latin hypercube over a grid of 20^3 points.

The designs were found using Algorithm 2.2, and are collected in Table 2.9. As might be anticipated, the Bayesian D-optimal designs have optimal observations later in time as the mean value for λ becomes closer to 0. The designs for the uniform prior distribution and the maximin D-optimal design have observations spread further out in time. All Bayesian designs assign roughly equal weight to support points at $T = 70$ and $T = 100$, while the maximin D-optimal design assigns more weight (≈ 0.65) to support points for $T = 100$. The maximin design found in Subsection 2.5.2 when λ was assumed equal to one similarly assigned more weight to support points at $T = 100$.

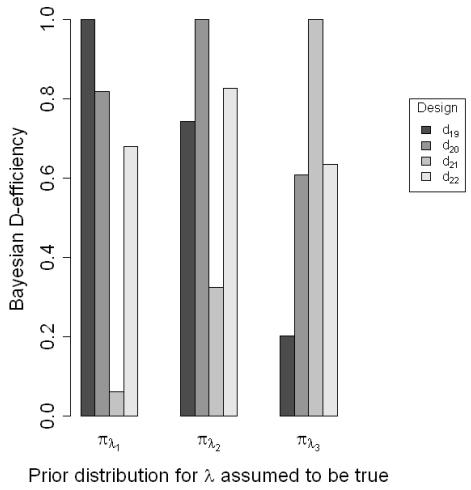


Figure 2.12: The Bayesian D-efficiencies attained by each design under each prior distribution

The relative Bayesian D-efficiencies attained by each design under each prior distribution are displayed in Figure 2.12. The choice of prior distribution for λ has a large impact on the efficiency obtained, as choosing an incorrect prior distribution provides a design with very low efficiency. The maximin D-optimal design d_{22} and the Bayesian D-optimal design d_{20} assuming $\lambda \sim$U(0,1) maintain a reasonably high Bayesian D-efficiency across all three prior distributions, as might be expected.

The minimum local D-efficiencies attained by $d_{19} - d_{22}$ were 0.0312, 0.0886, 1.89×10^{-6} and 0.284 respectively, so the maximin D-optimal design is much more

Table 2.9: Bayesian D-optimal and maximin D-optimal designs when $m = 1$ for model (1.8); $T_l = 70$, $T_u = 100$

Design (prior distribution/parameter space for λ given in brackets)

$$d_{19}(\pi_{\lambda_1}) = \left\{ \begin{array}{ccccccc} (T_l, 39.8) & (T_l, 41.3) & (T_l, 120.0) & (T_l, 1008.7) & (T_u, 7.3) & (T_u, 12.8) & (T_u, 51.3) \\ 0.0302 & 0.331 & 0.130 & 0.0123 & 0.0695 & 0.365 & 0.0620 \end{array} \right\}$$

$$d_{20}(\pi_{\lambda_2}) = \left\{ \begin{array}{ccccccccccc} (T_l, 35.6) & (T_l, 39.5) & (T_l, 77.5) & (T_l, 244.2) & (T_l, 1192.6) & (T_u, 8.1) & (T_u, 14.9) & (T_u, 19.9) & (T_u, 87.8) & (T_u, 893.5) \\ 0.0373 & 0.0508 & 0.212 & 0.121 & 0.00714 & 0.0437 & 0.129 & 0.144 & 0.151 & 0.0315 \end{array} \right\}$$

$$d_{21}(\pi_{\lambda_3}) = \left\{ \begin{array}{cccccccc} (T_l, 134.0) & (T_l, 318.2) & (T_l, 1196.7) & (T_u, 47.4) & (T_u, 145.6) & (T_u, 251.5) & (T_u, 1192.1) & (T_u, 1198.0) \\ 0.0774 & 0.218 & 0.205 & 0.195 & 0.160 & 0.103 & 0.0247 & 0.0165 \end{array} \right\}$$

$$d_{22}(\Theta_\lambda) = \left\{ \begin{array}{ccccccccccc} (T_l, 25.4) & (T_l, 59.3) & (T_l, 135.7) & (T_u, 3.5) & (T_u, 8.9) & (T_u, 35.9) & (T_u, 126.2) & (T_u, 274.4) & (T_u, 294.2) & (T_u, 615.5) \\ 0.127 & 0.108 & 0.107 & 0.106 & 0.196 & 0.00636 & 0.0209 & 0.0678 & 0.0651 & 0.196 \end{array} \right\}$$

robust to the worst case scenario than the Bayesian D-optimal designs found. In particular, designing for a prior distribution for λ centred around a mean of 0.14, as in d_{21}, can lead to a very low D-efficiency in some scenarios.

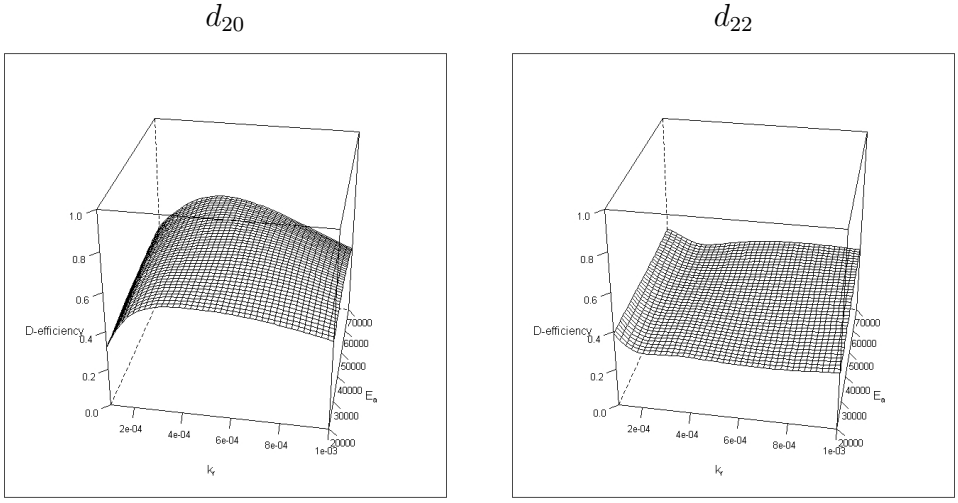


Figure 2.13: D-efficiency of designs over Θ_λ for d_{20} and d_{22} for $\lambda = 0.5$

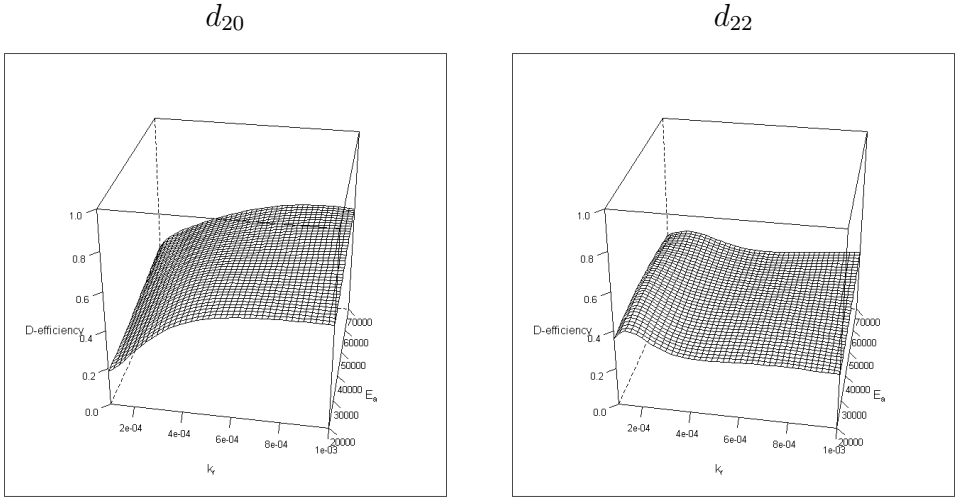


Figure 2.14: D-efficiency of designs over Θ_λ for d_{20} and d_{22} for $\lambda = 0.25$

Figures 2.13, 2.14 and 2.15 compare the D-efficiency of the optimal design for the uniform prior distribution over λ and the maximin D-optimal design across E_a and k_r for $\lambda = 0.5$, 0.25 and 0.14 respectively. The maximin D-optimal design has a mostly flat efficiency over the considered space, while the Bayesian D-optimal design's efficiency varies as the parameters change. Similarly to the case where

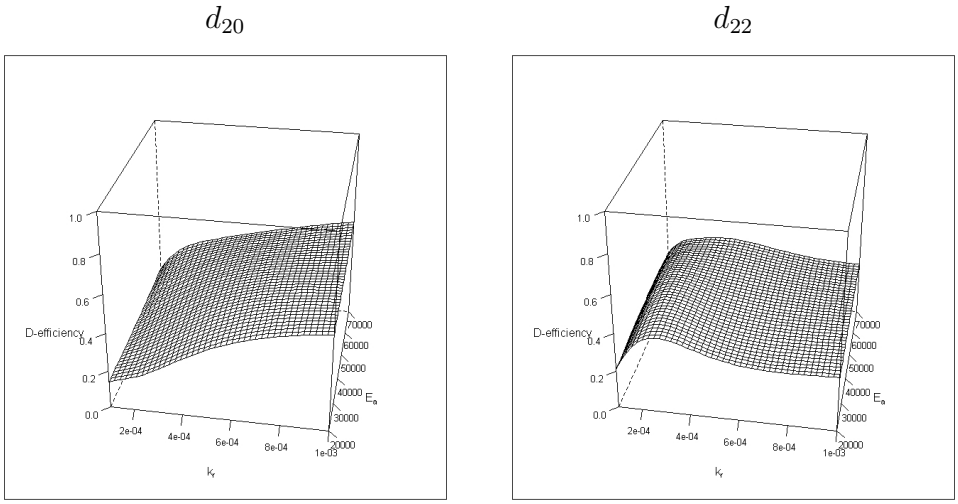


Figure 2.15: D-efficiency of designs over Θ_λ for d_{20} and d_{22} for $\lambda = 0.14$

$\lambda = 1$, the value of k_r has more impact on D-efficiency than E_a.

The prior distribution chosen when designing an experiment depends on the objectives of the practitioner and their concerns as to the possibility of mis-specifying prior information. It seems reasonable to suppose that the practitioner may have less information on λ than the mean parameters, so finding a maximin D-optimal design over λ or a Bayesian D-optimal design assuming a uniform distribution over λ might be reasonable choices to make while designing.

2.7 Conclusion

In this chapter various designs have been found for an example from chemistry for accurately estimating the mean parameters when independent errors are assumed. This assumption is removed in the remainder of the thesis.

Analytic results for finding locally D-optimal designs were obtained when only one observation is taken on each run of the process. D-optimal designs for the example then have two support points, with each run of the process being conducted at an end point of the allowed interval for temperature, and the observation times found by Equation (2.5). This finding was used to find various locally optimal designs, and it was demonstrated that if the prior knowledge of the parameter values

was misspecified, the design obtained would have a very low D-efficiency.

Robust designs were explored by finding and evaluating both Bayesian and maximin D-optimal designs for a range of prior distributions for the mean parameter values. These designs were compared to investigate which design's performance would remain the most robust to mis-specification of the mean parameters. It was demonstrated that the Bayesian D-optimal designs had a higher average D-efficiency than the maximin D-optimal designs. The efficiency of the maximin D-optimal design remained roughly constant across all parameter values, so might be a better choice when there is no prior information as to the mean of the parameter values.

Design of experiments was also investigated when heteroscedasticity in the statistical model was considered by varying λ in model (1.8). Locally, Bayesian and maximin D-optimal designs were found for a variety of values of λ. It was discovered that the impact of λ on all three types of designs was for the optimal observation times to increase as λ decreases.

The designs found in this chapter will often not be practically applicable, as experimenters will usually wish to make multiple observations in a given run of a process as this is more cost efficient. The next chapter will address the problem of finding optimal designs when multiple observations are made during each run. Additionally, it was assumed that the form of the expected response is known and has good predictive power across the region studied. If an alternative model for the expected response has better predictive power, there is no guarantee that the designs found in this chapter will provide good estimates of the mean parameters for the alternative model. For the specific case of independent observations when the expected response in this chapter is a good fit for the data, this chapter provides a fairly comprehensive guide as to which designs should be used in practice to obtain accurate estimates of the mean parameters.

Chapter 3

Locally and Bayesian D-optimal designs for non-linear models in the presence of correlation

3.1 Introduction

The designs found in Chapter 2 do not often represent a practical solution for scientists in the context of dynamic models, as each observation is made on a separate, independent run of the process. Taking multiple observations during a run of the process is much more cost efficient and, in this chapter, we investigate designs for this type of experiment. For these experiments, we can no longer assume that all observations are independent, and must take this dependence into account in the model assumed for the observations and the design of the experiment. Hence, potentially present are serial correlation between observations made on the same run, and run-to-run error induced by unknown sources of variation between the runs of the process.

It is assumed, initially, that the errors are homoscedastic and hence that $\lambda = 1$ in model (1.8) so that the impact of correlation on the selection of an optimal experimental design can be assessed. In Section 3.9, we relax this assumption by letting $0 < \lambda \leq 1$ in model (1.8), i.e. we allow for a transformation to remove error

heteroscedasticity.

Section 3.2 reviews the literature on D-optimal designs for non-linear models when correlation between observations is believed to be present. In Section 3.3, we find locally D-optimal designs when serial correlation between observations on the same process run is believed to be present and investigate the impact of correlation on the selection and performance of locally D-optimal designs for different values of the mean and correlation parameters. In Section 3.4, we investigate the robustness of the performance of the designs found in Section 3.3 to the assumed correlation parameter values. We also study the impact of incorrectly specifying the form of the model for serial correlation between observations made on the same process run. In Section 3.5, we find locally D-optimal designs when observations are constrained to be taken at least two minutes apart, and compare these designs with the locally optimal designs obtained when serial correlation is included in the statistical model.

We expand the statistical model, in Section 3.6, to include measurement and run-to-run error, and investigate the robustness of locally D-optimal designs to the values of the parameters in the mean and correlation functions. We extend this study, in Section 3.7, to find values of the correlation parameters for which the D-optimal design switches from one form to another.

In Section 3.8, we find Bayesian D-optimal designs, by specifying prior distributions for the mean and correlation parameters. In Section 3.9, local and robust D-optimal designs are found for situations when both error heteroscedasticity and correlation between observations may be present, and in the final section we draw together findings from the chapter.

3.2 Literature review

Designs for non-linear models when serial correlation between observations is present has received some attention in the literature. Hughes-Oliver (1998) explored the effect of such correlation on locally D-optimal two point designs found for two non-linear dynamic models where the experimenter controls the time at which

observations are made. She focused on an exponential correlation structure, as in Equation (3.1), in Section 3.3, with $\rho = 1$, where the correlation parameters are assumed to be known.

Ucinski and Atkinson (2004) provided an exchange algorithm for finding locally D-optimal designs for non-linear models with a known correlation structure present between the observations. The algorithm assumes that the optimal design is exact and that all observations are made on a single run, which is not the case for experiments considered in this thesis. They applied this algorithm to find designs for a non-linear dynamic model with the same exponential correlation structure as in Hughes-Oliver (1998). They demonstrated that the stronger the serial correlation, the further apart the optimal observation times.

Patan and Bogacka (2007) studied the design of experiments for a non-linear multi-response dynamic model, where responses are correlated and there is also the same exponential correlation structure between observations taken over time. As before, the impact of the correlation was to push optimal observations times further apart as the correlation became stronger. Stehlík et al. (2008) found designs for a non-linear dynamic model, where the form of correlation was either exponential or linear, as explored in Subsection 3.4.1. They found locally D-optimal designs for two situations: when the correlation parameters are nuisance parameters, and when their estimation is a goal of the experiment. The second case is not considered in this thesis as, usually, in practice correlation parameters are treated as nuisance parameters. The authors demonstrated that in the second case, correct specification of the form of the correlation is very important. We show in this chapter that when the correlation parameters are regarded as nuisance parameters, correctly specifying the form of the correlation matters less.

When serial correlation has been considered in the literature it has usually been assumed to be known, with not much concern thus far paid towards robustness to misspecification of the correlation parameters.

In pharmacokinetic models, the impact of drugs on different patients is modelled by a non-linear mixed model where the parameters in the mean function are random

effects, inducing a correlation function on the response. Mentré et al. (1997) considered finding locally D-optimal designs for non-linear random effects models, and used a linearisation to obtain the information matrix. This approach is the one taken in most of the literature in this area and was extended by Retout and Mentré (2003), through deriving a better approximation of the information matrix which is more computationally intensive. These methods have been applied to find locally D-optimal designs, for example, by Dufful et al. (2001), Retout et al. (2007), Ogungbenro and Aarons (2008) and Geudj et al. (2011).

Gueorguieva et al. (2006) extended this approach to a non-linear mixed model with multiple responses. Atkinson (2008b) gave an equivalence theorem for D-optimal designs for non-linear models with mixed effects. Dette et al. (2010) presented a method for obtaining optimal designs when random effects and serial correlation between observations is present. More recently Wang et al. (2012) found locally D-optimal designs for non-linear models with random effects where the closed form solution of the differential equations describing the expected response is not available. These approaches all assume that the distributions of the random effects are known when the experiment is being designed.

The adoption of a mixed effects model is reasonably intuitive when the different runs correspond to different patients, but less so for the type of experiment considered in this thesis. Instead, runs of the process are assumed to have an additive blocking effect, inducing run-to-run error, as outlined in Section 3.6. We consider this to be a more appropriate model when each run is a separate chemical reaction rather than made on a different individual. A similar approach was taken by Woods and van de Ven (2011) who found Bayesian D-optimal blocked designs for discrete responses. They found designs that are robust to the mean parameters, and established that these designs are robust to misspecification of the correlation parameters.

3.3 Locally optimal designs in the presence of serial correlation

In this section the statistical model is assumed to include only serial correlation between observations on the same run, so that the impact of serial correlation on design selection can be studied. An exponential correlation structure is assumed, and locally D-optimal designs are found for a range of mean and correlation parameter values. The optimality of the designs is assessed via an equivalence theorem.

It is assumed m observations are made on each run. A value of $m = 6$ was chosen for this study as this is similar to those values used by practitioners in the types of experiments under consideration. The value of m that would be an optimal choice under cost restriction is investigated in Chapter 5.

The correlation structure was assumed to be of the following form (see also Pinheiro and Bates, 2004, pages 230-232):

$$\text{cor}(y_{il}, y_{hu}) = \begin{cases} \exp(-\tau|t_{il} - t_{hu}|^{\rho}) & \text{if } i = h, \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

where y_{il} is the lth observation in the ith run, t_{il} is the time at which y_{il} is observed and $\tau > 0$ and $\rho > 0$ are parameters which determine the strength of the correlation. Note that, for convenience, time is considered to be in units of minutes in the correlation structure throughout this thesis. Structure (3.1) reflects the belief that observations further apart will be less correlated. For other models and applications, different correlation structures may be appropriate.

Model (1.8) given in Section 1.3 is assumed, where the correlation matrix $Q(\boldsymbol{\theta}_c) = diag\{Q_i\}$, $i = 1, \ldots, N$, with Q_i, the $m \times m$ correlation matrix for observations on the ith run, $\boldsymbol{\theta}_c = (\tau, \rho)'$ and $\lambda = 1$.

The values of correlation obtained from (3.1) can be calculated for different values of τ and ρ to visualise how the strength of correlation changes as observations are taken further apart in time. Figure 3.1 shows how the strength of correlation changes

as the distance between observations increases. For higher values of τ, ρ has little effect, but for small values of τ the difference between $\rho = 1$ and $\rho = 2$ is much larger.

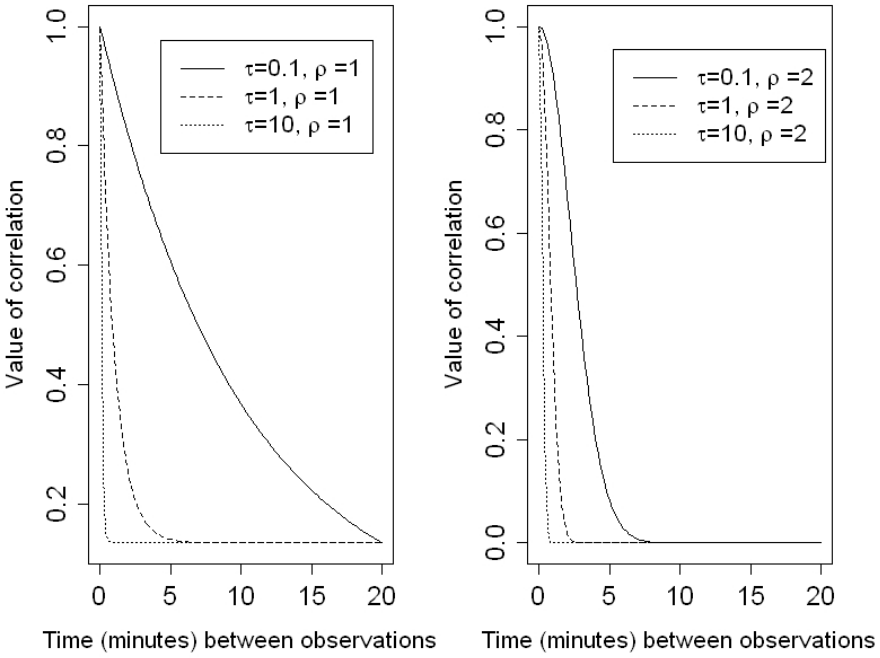


Figure 3.1: The strength of correlation between two observations

Analytic results are difficult to obtain for this problem. Hence optimal designs have to be found by numeric search, as for the Bayesian and maximin D-optimal designs in Section 2.5. It was anticipated that the introduction of correlation would alter the distance between observation times in an optimal design, and that the times would still be clustered around the optimal observation times in Chapter 2. To investigate this hypothesis, designs for several combinations of E_a, k_r, τ and ρ values were explored.

We found locally D-optimal designs using search with a constrained optimisation algorithm, `L-BFGS-B` (as in Section 2.5). The procedure is summarised in Algorithm 3.1 below. It uses the knowledge acquired in Chapter 2, but allows a D-optimal design to be different from the designs found analytically for independent errors in Section 2.4. The value of a chosen was three, chosen to give a reasonable distance

in time between observations in the starting design.

Algorithm 3.1 Algorithm to find locally D-optimal designs

1: Generate 50 random designs, $\{\xi_1, \ldots, \xi_{50}\}$, each with four equally weighted support points and randomly chosen values for temperature and time
2: Generate a design ξ_{51} with two equally weighted support points: temperatures of 70 and 100°C and equidistant observation times from the interval $[t_T - a, t_T + a]$, with a a chosen constant and t_T given by (2.5) in Section 2.3
3: For each ξ_i, $(i = 1, \ldots, 51)$, use `L-BFGS-B` to obtain $\xi_i^\dagger$, then set $\xi_i = \xi_i^\dagger$
4: Select the design in $\{\xi_1, \ldots, \xi_{51}\}$, $\hat{\xi}$ which achieves the highest value of objective function (1.14), $|M(\xi, \boldsymbol{\theta})|$
5: For $\hat{\xi}$ use `L-BFGS-B` to obtain $\hat{\xi}^\dagger$
6: If $|M(\hat{\xi}^\dagger, \boldsymbol{\theta})| - |M(\hat{\xi}, \boldsymbol{\theta})| > \delta$, for some small $\delta > 0$, set $\hat{\xi} = \hat{\xi}^\dagger$ and go to step 5
7: Otherwise, select $\hat{\xi}$

The designs found are given in Table 3.1, including the vectors of observation times $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$. For comparison, this table also gives the times, t_1^*, and t_2^* which are the optimal times of observation when the two runs of each design are constrained to have only one observation.

Table 3.1: Approximate locally D-optimal designs for correlation structure (3.1); t_1^*, and t_2^* are the optimal observation times for each run when only one observation per run is allowed

E_a	k_r	τ	ρ	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)						$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)						t_1^*	t_2^*
2×10^4	10×10^{-4}	1	2	10.5	12.8	15.0	17.2	19.5	21.9	5.0	7.0	9.0	11.0	13.0	15.2	15.1	8.6
4×10^4	5×10^{-4}	1	2	28.9	31.6	34.1	36.7	39.3	41.9	7.2	9.3	11.4	13.5	15.6	17.9	34.8	11.3
6×10^4	4×10^{-4}	1	2	43.7	46.5	49.2	51.9	54.6	57.4	5.5	7.6	9.5	11.5	13.6	15.8	50.1	9.2
2×10^4	10×10^{-4}	10	1	13.5	14.2	14.9	15.6	16.3	17.0	7.2	7.9	8.4	9.0	9.6	10.3	15.1	8.6
4×10^4	5×10^{-4}	10	1	32.7	33.6	34.5	35.3	36.1	37.0	9.8	10.4	11.1	11.7	12.3	13.0	34.8	11.3
6×10^4	4×10^{-4}	10	1	47.8	48.8	49.7	50.6	51.5	52.4	7.8	8.4	9.0	9.6	10.2	10.9	50.1	9.2

As anticipated, the observation times in the locally D-optimal design for $m = 6$ observations per run cluster around the optimal times for when correlation is not present (as given in Equation (2.5)). The stronger the correlation, the further apart are the observation times.

The impact on the selection of a D-optimal design of the choice of values of the two parameters controlling the strength of correlation was then explored. Designs were found for 6 different combinations of τ and ρ in a small factorial experiment, with $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ (values seen previously experimentally)

using Algorithm 3.1. They are presented in Table 3.2, and labelled for later use as $\xi_1, \ldots, \xi_6$.

Table 3.2: Approximate locally D-optimal designs under correlation structure (3.1), where $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$

Design	τ	ρ	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)						$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)					
ξ_1	0.1	1	19.8	34.6	49.3	65.1	83.2	106.1	1.0	3.5	6.8	11.4	17.9	27.9
ξ_2	1	1	30.2	35.2	39.8	44.4	49.1	54.4	5.6	8.3	10.8	13.4	16.3	19.6
ξ_3	10	1	38.3	39.2	40.0	40.9	41.7	42.7	8.7	9.3	9.90	10.5	11.2	11.8
ξ_4	10	2	38.1	39.0	40.0	41.0	41.9	42.9	8.3	9.1	9.90	10.7	11.5	12.4
ξ_5	1	2	34.3	37.0	39.6	42.2	44.9	47.6	6.2	8.3	10.4	12.4	14.5	16.7
ξ_6	0.1	2	26.9	33.8	40.6	47.3	54.2	61.6	4.6	9.2	13.9	18.8	23.9	29.7

It can be observed that the effect of strong correlation (low values of τ) is to make observation times in the D-optimal design further apart. As can be anticipated from Figure 3.1, the value of ρ has a large effect on the optimal times only when the correlation is very strong ($\tau = 0.1$). Then the optimal times are much more spread out for lower values of ρ, especially for the support point at 70°C.

3.3.1 Investigating the D-optimality of designs $\xi_1 - \xi_6$

The designs $\xi_1 - \xi_6$ were found numerically and therefore may not be locally D-optimal designs. This subsection introduces an equivalence theorem which can be used to confirm the D-optimality of designs when observations from the same run are assumed to be correlated. An equivalence theorem is then applied to investigate the optimality of $\xi_1 - \xi_6$.

An equivalence theorem to check the optimality of designs for non-linear models with correlated observations (see, for example, page 446, Atkinson et al., 2007) is as follows:

Theorem 3.3.1 *The following conditions are equivalent for a design ξ^* with $\Upsilon(T, \boldsymbol{t}) = F(T, \boldsymbol{t})' Q(\boldsymbol{\theta}_c)^{-1} F(T, \boldsymbol{t})$, where $\boldsymbol{t} = (t_1, \ldots, t_m)'$ and F is an $n \times p$ matrix holding the sensitivity equations:*

1. *The design ξ^* maximises $|M(\xi, \boldsymbol{\theta})|$,*

2. *The design* ξ^* *minimises* $max_{(T,\boldsymbol{t})\in\mathcal{X}} tr[M^{-1}(\xi^*, \boldsymbol{\theta})\Upsilon(T, \boldsymbol{t})]$

3. $max_{(T,\boldsymbol{t})\in\mathcal{X}} tr[M^{-1}(\xi^*, \boldsymbol{\theta})\Upsilon(T, \boldsymbol{t})] = p_m$.

We can obtain evidence for designs being D-optimal or near D-optimal by applying a search method to confirm condition 3. For $m = 6$, a thorough search over a seven-dimensional space (e.g. a grid of 20^7 points) would need to be conducted. Such a search is not practical and any search would not be exhaustive, and could give only a weak indication of optimality. We conducted a small study, however, by searching over a grid of 6^7 points. The points were chosen via Latin hypercubes based on orthogonal arrays, as in Section 2.5. For all six designs, the support points achieved the maximum value of $p_m = 2$, while the grid search found only maximum values of 1.86, 1.74, 1.75, 1.75, 1.71 and 1.73 (to 2 d.p.), for
$\xi_1, \ldots, \xi_6$ respectively. This is not strong evidence of optimality, as a large amount of the space was unexplored.

Table 3.3: Approximate locally D-optimal designs in the presence of an exponential correlation structure (3.1) for $m = 3$ where $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$

τ	ρ	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)			$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)		
0.1	1	26.96	48.62	73.70	7.14	13.29	23.23
1	1	35.2	41.0	46.9	7.5	11.0	14.7
10	1	39.4	40.4	41.4	9.4	10.2	10.9
10	2	37.7	40.6	43.4	8.2	10.5	12.9
1	2	39.4	40.4	41.5	9.3	10.2	11.1
0.1	2	33.77	41.46	49.24	6.67	12.35	18.36

To provide evidence that the designs found were locally D-optimal, equivalence theorem 3.3.1 was applied to a smaller example with $m = 3$. The designs found are shown in Table 3.3. The four dimensional space (T, t_1, t_2, t_3) was searched using a grid of 40^4 points for each design in Table 3.3, and maximum values of 1.97, 1.99, 1.97, 1.97, 1.97 and 1.98 (to 2.d.p.) were found. This study provides good evidence that the designs for $m = 3$ are D-optimal or near D-optimal. As the designs for $m = 6$ include similar observation times, this example provides some support for the designs in Section 3.3 being D-optimal or near D-optimal.

3.4 Performance of locally D-optimal designs when the strength or form of serial correlation is mis-specified

In this section the impact of mis-specifying the strength or form of correlation when finding locally optimal designs is investigated. The D-efficiency of designs $\xi_1 - \xi_6$ is found when the correlation parameters τ and ρ or the form of the correlation are mis-specified.

The D-efficiency of the designs $\xi_1 - \xi_6$ was calculated for six combinations of values of τ and ρ under the assumption that $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$. The efficiencies, together with the strength of correlation calculated from (3.1), are given in Table 3.4.

Table 3.4: D-efficiencies of locally D-optimal designs for a grid of values for τ and ρ where $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$

True value		Value of correlation (3.1) for time separation (minutes)			D-efficiency					
τ	ρ	0.5	1	2	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6
0.1	1	0.951	0.905	0.819	1.00	0.76	0.53	0.55	0.66	0.83
1	1	0.607	0.368	0.135	0.75	1.00	0.54	0.60	0.95	0.91
10	1	6.74×10^{-3}	4.54×10^{-5}	2.06×10^{-9}	0.65	0.91	1.00	1.00	0.95	0.79
10	2	8.21×10^{-2}	4.54×10^{-5}	4.25×10^{-18}	0.65	0.92	0.99	1.00	0.96	0.79
1	2	0.779	0.368	1.83×10^{-2}	0.69	0.97	0.54	0.60	1.00	0.84
0.1	2	0.975	0.905	0.670	0.82	0.87	0.45	0.46	0.66	1.00

The results indicate that underestimating the strength of the correlation leads to large drops in efficiency, but if the strength of the correlation is overestimated then not as much efficiency is lost. For example, design ξ_3 is D-optimal when $\tau = 10$ and $\rho = 1$, but has a D-efficiency of only 0.44 when the true correlation parameter values are $\tau = 1$ and $\rho = 2$. Design ξ_2, which is D-optimal for $\tau = 1$ and $\rho = 1$, has a minimum D-efficiency of 0.76 for the values of ρ and τ explored. Over the values searched, it seems that $\tau = 0.1$ and $\rho = 2$ (ξ_6) or $\tau = 1$ and $\rho = 1$ (ξ_2) would be good values to choose for practitioners who wished to find designs whose performance is robust to correlation parameter misspecification.

3.4.1 Impact of the correlation structure

The true correlation structure may not be well described by the exponential form in Equation (3.1), leading to a loss in the efficiency of the design selected. Other potential correlation structures (see, for example, Pinheiro and Bates, 2004) include the linear and quadratic forms. In this section, we begin by defining these two forms and then conduct a small study to investigate the impact of correlation structure on the efficiency of the designs obtained.

The linear correlation function can be expressed as:

$$\text{cor}(y_{il}, y_{hu}) = \begin{cases} \left(\frac{\rho_l - |t_{il} - t_{hu}|}{\rho_l}\right) \text{I}(|t_{il} - t_{hu}| < \rho_l) & \text{if } i = h, \\ 0 & \text{otherwise,} \end{cases} \tag{3.2}$$

where ρ_l is a parameter influencing the strength of correlation; the correlation drops to 0 beyond a distance ρ_l, and $I(.)$ is an indicator function being one if the conditions within the brackets are satisfied, and zero otherwise.

The quadratic correlation function can be expressed as:

$$\text{cor}(y_{il}, y_{hu}) = \begin{cases} \frac{1}{1+\rho_q(t_{il} - t_{hu})^2} & \text{if } i = h, \\ 0 & \text{otherwise,} \end{cases} \tag{3.3}$$

where ρ_q is a parameter influencing the strength of correlation.

Designs were found where the parameters ρ_l and ρ_q were chosen such that the strength of correlation between observations, made at a unit distance apart, was the same for all three correlation structures (for $\tau = 0.1$, $\tau = 1$ and $\tau = 10$, with $\rho = 1$). Locally optimal designs were found using Algorithm 3.1, and are given in Table 3.5. The corresponding designs for exponential correlation (3.1) are given in Table 3.2.

The optimal observation times are similar for all three correlation structures when each correlation structure is calibrated to have the same correlation for observations one minute apart. Locally D-optimal designs found for the linear correlation structure tend to have observations closer and further apart in time than

Table 3.5: Approximate locally D-optimal designs in the presence of a quadratic or linear correlation structure where $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$

Correlation structure	Correlation at distance 1	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)						$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)					
Linear	0.905	20.1	30.6	51.7	41.1	62.2	72.7	1.0	4.7	6.4	25.7	15.2	43.5
Linear	0.368	38.0	39.9	41.7	43.3	44.9	46.6	7.0	8.7	10.3	11.9	13.5	15.1
Linear	4.54×10^{-5}	37.7	38.8	39.8	40.8	41.8	42.8	7.7	8.7	9.7	10.7	11.7	12.7
Quadratic	0.905	22.9	33.1	42.8	52.8	63.8	76.9	5.0	9.3	13.7	18.5	23.9	30.7
Quadratic	0.368	29.7	35.0	39.8	44.6	49.7	55.7	5.8	8.3	10.6	13.0	15.6	18.7
Quadratic	4.54×10^{-5}	39.2	39.7	40.2	40.7	41.2	41.7	9.5	9.8	10.0	10.3	10.5	10.8

designs obtained for the other correlation structures for weak and strong correlation respectively. This finding might be due to observations being uncorrelated when they are further apart than ρ_l under a linear correlation structure, resulting in a D-optimal design tending to take observations exactly ρ_l minutes apart.

The D-efficiencies attained by the designs found for an exponential structure (Table 3.2) when the true correlation structure is quadratic or linear were calculated. These results are presented in Table 3.6.

From this table, we see that if a fairly strong correlation is assumed, then a robust design is obtained. Design ξ_2, which is D-optimal for $\tau = \rho = 1$, has a high average D-efficiency no matter what the structure or strength of correlation. These results again indicate that mis-specifying the strength of serial correlation has more impact on the D-efficiency than mis-specifying the structure. In the remainder of the thesis, we will adopt the exponential correlation structure.

Table 3.6: D-efficiencies of approximate designs given in Table 3.2 where $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ and the correlation structure is either linear or quadratic, and ρ_l and ρ_q were varied over three different values

Correlation structure	Correlation at distance 1	Efficiency of design					
		ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6
Linear	0.905	0.90	0.70	0.37	0.38	0.57	0.82
Linear	0.368	0.67	0.94	0.54	0.61	0.98	0.81
Linear	4.54×10^{-5}	0.66	0.92	0.73	0.86	0.96	0.79
Quadratic	0.905	0.93	0.85	0.46	0.48	0.68	0.96
Quadratic	0.368	0.76	1.00	0.55	0.61	0.95	0.92
Quadratic	4.54×10^{-5}	0.65	0.91	1.00	1.00	0.95	0.78

3.5 A comparison of constrained designs and locally optimal designs found assuming serial correlation

A possible advantage of designs which assume a strong correlation is that the observations tend to be far apart in time. In an experimental setting the scientists may have difficulty taking observations too close together, so a design which assumes strong correlation may be easier to implement. For our motivating example it was indicated that designs should be constrained so that observations from the same run should be at least two minutes apart. Finding constrained designs can be difficult, so designs which meet these requirements without imposing a constraint are preferred. In this section we find a constrained design and compare its performance to designs found without constraints, under the assumption that serial correlation is present.

An approximate constrained design for $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ was found with an assumption of no correlation, and a constraint that observations from the same run must be two minutes apart. This gave the design in Table 3.7, which is similar to the optimal design found for $\tau = 1$ and $\rho = 2$, ξ_5.

Table 3.7: Approximate locally D-optimal design for $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ when observations are constrained to be at least two minutes apart

$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)						$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)					
35.6	37.6	39.6	41.6	43.6	45.6	5.0	7.0	9.0	11.0	13.0	15.0

Table 3.8: D-efficiencies of the constrained approximate design for differing values of τ and ρ when $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$

True values of τ and ρ	$\tau = 0.1$ $\rho = 1$	$\tau = 1$ $\rho = 1$	$\tau = 1$ $\rho = 2$	$\tau = 10$ $\rho = 1$	$\tau = 10$ $\rho = 2$	$\tau = 0.1$ $\rho = 2$
D-efficiency	0.63	0.89	0.98	0.95	0.96	0.59

The D-efficiencies of this constrained design in the presence of correlation are given in Table 3.8. The results indicate that the constrained design is reasonably robust to the values of correlation explored, and maintains a fairly high D-efficiency.

3.6 Locally optimal designs in the presence of serial correlation and run-to-run and measurement errors

In the previous sections the impact of serial correlation on the choice of an optimal design for an experiment was investigated independently of other types of correlation. When observations are taken on the same run of a process, we might anticipate a blocking effect to be present. This effect may be modelled by including run-to-run error in the statistical model. Once run-to-run error and serial correlation is present in the statistical model, we also need to make allowance for errors within each run. To do this we incorporate measurement error into the statistical model.

In this section, we begin by explaining how the model (1.8) is further developed to incorporate these types of error. We then find locally D-optimal designs for different magnitudes of types of error in the statistical model, and assess the designs using D-efficiency.

To allow three types of error in the model, the variance σ^2 is split into three components:

$$\sigma^2 = \sigma_m^2 + \sigma_r^2 + \sigma_s^2,$$

where σ_m^2, σ_r^2 and σ_s^2 are the variance components giving the magnitude of measurement error, run-to-run error and serial correlation. We take as the vector of unknown correlation parameters $\boldsymbol{\theta}_c = (\tau, \rho, r, q)'$, where $r = \sigma_r^2/\sigma_s^2$ and $q = \sigma_m^2/\sigma_s^2$. We use ratios r and q as, as is shown below, only their ratios have an impact on the correlation between observations from the same run. The correlation between observations is of the form

$$\text{cor}(y_{il}, y_{hu}) = \begin{cases} \frac{\sigma_r^2 + \sigma_s^2 K(t_{il}, t_{hu}) + \sigma_m^2 \delta_{lu}}{\sigma_r^2 + \sigma_s^2 + \sigma_m^2} & \text{if } i = h, \\ 0 & \text{otherwise,} \end{cases} \tag{3.4}$$

where

$$K(y_{il}, y_{hu}) = \begin{cases} \exp(-\tau|t_{il} - t_{hu}|^{\rho}) & \text{if } i = h, \\ 0 & \text{otherwise,} \end{cases}$$

and δ_{lu} is the Kronecker delta function, which is 1 if $l = u$ and 0 otherwise. Equation (3.4) can be rewritten as

$$\text{cor}(y_{il}, y_{hu}) = \begin{cases} \frac{r+K(t_{il},t_{hu})+q\delta_{lu}}{r+1+q} & \text{if } i = h, \\ 0 & \text{otherwise} \end{cases}$$

The impact on a locally D-optimal design of varying r and q was investigated by finding locally D-optimal designs for differing values of r and q. Designs were found for $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$, with $\rho = 1$ and for each of $\tau = 1$ and $\tau = 10$. The value of ρ was held constant as the results in previous sections had demonstrated that its value did not usually have a large impact on the designs obtained.

A large search scheme was used for the optimal design; as the results will show, for some values of r and q the optimal designs are very similar. Initial investigation determined that the optimal design has two support points with $T_1 = 70$ and $T_2 = 100$. Then, to ensure that for each value of r and q explored, a distinct optimal design was obtained, Algorithm 3.2 was implemented. Algorithm 3.2 uses a genetic algorithm, as in Algorithm 2.2, but where one of the starting designs was pre-selected to be a given design ξ, obtaining $\hat{\xi}$.

Designs for $q = 0$: Table 3.9 shows the designs obtained when $q = 0$, i.e. no measurement error is assumed. Some of the designs are labelled for future reference.

The optimal designs obtained indicate the effect of increasing the variance component for run-to-run error with respect to the variance component for serial correlation. As r increases, optimal designs take observations at the end points of the allowed time intervals. There is not much difference between the designs for $r = 1$ and $r = 8$. There is little evidence of 'interaction' in how varying τ and r affects the design selected: the effect of altering τ is to take observations further

Algorithm 3.2 Algorithm to find locally optimal designs when run-to-run error, measurement error and serial correlation are believed to be present

1: Define $\boldsymbol{r}$ and $\boldsymbol{q}$ as vectors of length n_p holding the values of r and q of interest. Let τ, ρ, E_a and k_r be fixed.

2: For all $1 \leq i \leq n_p$

a) Generate random designs $\{\xi_1, \ldots, \xi_9\}$, with two equally weighted support points at 70 and 100 °C, with observation times chosen randomly.

b) Generate a design ξ_{10} with two support points at 70 and 100°C with equal weighting, and the times of observation centered on the optimal value given by (2.5) in Section 2.3

c) Label $S_i = \{\xi_1, \ldots, \xi_{10}\}$

3: For each $i = 1 \ldots n_p$

a) Let r and q be the ith element of $\boldsymbol{r}$ and $\boldsymbol{q}$ respectively

b) Apply `L-BFGS-B` to each design in S_i, to give $S_i^\dagger$

c) Apply a genetic algorithm to each design in S_i obtaining $\hat{S}_i$

4: For each $i = 1 \ldots n_p$

a) Let r and q be the ith element of $\boldsymbol{r}$ and $\boldsymbol{q}$ respectively

b) Find the design $\tilde{\xi}_i$ which achieves the highest value of (1.14) among all designs in $S_1 \cup \ldots \cup S_{n_p} \cup S_1^\dagger \cup \ldots \cup S_{n_p}^\dagger \cup \hat{S}_1 \cup \ldots \cup \hat{S}_{n_p}$

5: If for all $1 \leq i, j \leq n_p$ $\tilde{\xi}_i \neq \tilde{\xi}_j$, $i \neq j$, stop

6: Otherwise set $S_i = \left\{\tilde{\xi}_1, \ldots, \tilde{\xi}_{n_p}\right\}$ and go to step 3

Table 3.9: Approximate locally D-optimal designs for $q = 0$ and differing values of r and τ, with $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$

Design	τ	r	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)						$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)					
	10	1/8	38.4	39.3	40.1	40.9	41.7	42.5	8.8	9.4	9.9	10.5	11.1	11.7
		1	1.0	39.2	40.0	40.9	41.7	200.0	9.5	10.2	10.9	196.7	198.4	200.0
		8	1.0	39.5	40.4	41.4	199.3	200.0	9.5	10.2	10.8	196.6	198.3	200.0
ξ_{10}	1	1/8	31.1	35.6	39.7	43.8	48.1	52.9	6.0	8.3	10.5	12.8	15.2	18.1
ξ_{20}		1	1.0	34.2	38.9	43.3	48.1	200.0	7.7	10.8	14.1	175.2	187.3	200.0
ξ_{30}		8	1.0	34.5	38.9	43.2	47.8	200.0	7.8	10.7	13.9	175.1	187.4	200.0

apart no matter the value of r.

Designs for $q \neq 0$: The value of q was then allowed to be nonzero. Locally D-optimal designs were found for each combination of values of $r = 1/8, 1, 8$ and $q = 1/8, 1, 8$, with $\tau = 1$ and $\tau = 10$, $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$. The designs found are presented in Table 3.10. These results demonstrate that the impact of altering the value of q on the optimal design is less significant than the impact of r. For large values of q, the impact of r on the choice of optimal design is reduced, with the optimal design for $r = 1$ no longer having observations at the extremes of the range for observation times. Increasing q does impact slightly on the effect of the serial correlation on the choice of optimal design, as observation times are closer together for higher values of q. These results are intuitive, as a high value of q corresponds to a reduction in correlation between two observations from the same run, see Equation (3.4).

Table 3.10: Approximate locally D-optimal designs for differing values of r, τ and q, with $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$

Design	τ	q	r	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)						$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)					
	10	1/8	1/8	38.4	39.3	40.1	40.9	41.7	42.5	8.8	9.4	9.9	10.5	11.1	11.7
		1/8	1	1.0	39.2	40.0	40.9	41.7	200.0	9.5	10.2	10.8	196.8	198.4	200.0
		1/8	8	1.0	39.5	40.4	41.3	199.3	200.0	9.5	10.2	10.8	196.6	198.3	200.0
		1	1/8	38.5	39.3	40.1	40.8	41.6	42.4	8.8	9.4	9.9	10.5	11.0	11.6
		1	1	38.9	39.7	40.4	41.2	42.0	200.0	9.3	9.9	10.4	11.0	198.4	200.0
		1	8	1.0	39.6	40.4	41.3	199.3	200.0	9.5	10.1	10.8	196.7	198.3	200.0
		8	1/8	38.8	39.5	40.1	40.8	41.4	42.1	9.1	9.5	10.0	10.4	10.8	11.3
		8	1	38.9	39.5	40.1	40.8	41.4	42.0	9.1	9.6	10.0	10.4	10.8	11.2
		8	8	1.0	39.4	40.1	40.8	41.4	200.0	9.6	10.1	10.6	197.0	198.4	200.0
ξ_{11}	1	1/8	1/8	31.2	35.6	39.7	43.8	48.0	52.8	6.0	8.3	10.5	12.7	15.0	17.9
ξ_{21}		1/8	1	1.0	34.3	38.9	43.3	48.0	200.0	7.7	10.7	13.9	175.0	187.8	200.0
ξ_{31}		1/8	8	1.0	34.6	38.9	43.1	47.6	200.0	7.8	10.7	13.8	174.8	187.4	200.0
ξ_{12}		1	1/8	31.7	35.9	39.7	43.4	47.4	51.8	6.3	8.4	10.3	12.2	14.3	16.9
ξ_{22}		1	1	1.0	34.7	39.0	43.0	47.4	200.0	7.5	9.7	11.8	14.3	187.0	200.0
ξ_{32}		1	8	1.0	35.1	39.0	42.8	46.8	200.0	8.0	10.5	13.2	175.7	187.8	200.0
ξ_{13}		8	1/8	33.6	36.8	39.6	42.5	45.4	48.8	7.3	8.7	10.0	11.3	12.6	14.4
ξ_{23}		8	1	34.2	37.1	39.7	42.2	44.9	48.0	7.6	8.8	9.9	11.0	12.2	13.7
ξ_{33}		8	8	1.0	36.3	39.3	42.1	45.2	200.0	8.6	10.3	12.0	176.4	187.8	200.0

The importance of correctly specifying values for q and r before finding a design was assessed by comparing the D-efficiencies of the optimal designs obtained when the values are misspecified. To reduce the number of comparisons, it was assumed that $\tau = \rho = 1$ and the designs investigated were $\xi_{10} - \xi_{33}$, in Tables 3.9 and 3.10.

Their efficiencies were compared assuming that each value of r and q used in finding the designs was true. The results of this study are given in Table 3.11.

Table 3.11: D-efficiencies of the labelled designs in Table 3.9 and 3.10 with $\tau = \rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$

q	0	0	0	1/8	1/8	1/8	1	1	1	8	8	8
r	1/8	1	8	1/8	1	8	1/8	1	8	1/8	1	8
ξ_{10}	1.00	0.52	0.10	1.00	0.56	0.11	1.00	0.77	0.18	0.97	0.97	0.56
ξ_{20}	0.79	1.00	1.00	0.77	1.00	1.00	0.70	0.98	1.00	0.61	0.73	0.99
ξ_{30}	0.79	1.00	1.00	0.77	1.00	1.00	0.70	0.98	1.00	0.61	0.73	0.99
ξ_{11}	1.00	0.52	0.10	1.00	0.57	0.11	1.00	0.77	0.18	0.98	0.97	0.56
ξ_{21}	0.79	1.00	1.00	0.77	1.00	1.00	0.70	0.98	1.00	0.61	0.73	0.99
ξ_{31}	0.79	1.00	1.00	0.77	1.00	1.00	0.70	0.98	1.00	0.61	0.73	0.99
ξ_{12}	1.00	0.53	0.10	1.00	0.57	0.11	1.00	0.78	0.18	0.99	0.98	0.56
ξ_{22}	0.84	0.98	0.94	0.83	0.99	0.94	0.77	1.00	0.95	0.70	0.81	0.99
ξ_{32}	0.79	0.99	1.00	0.77	1.00	1.00	0.70	0.98	1.00	0.62	0.74	0.99
ξ_{13}	0.95	0.53	0.10	0.96	0.57	0.11	0.98	0.78	0.18	1.00	1.00	0.58
ξ_{23}	0.92	0.53	0.10	0.93	0.57	0.11	0.96	0.78	0.18	1.00	1.00	0.58
ξ_{33}	0.74	0.96	0.96	0.73	0.96	0.97	0.68	0.96	0.99	0.62	0.74	1.00

These results indicate that, as was anticipated from observing the optimal designs, the impact of incorrectly specifying r is much greater on design efficiency than incorrect specification of q. The efficiency of ξ_{10}, which was D-optimal for $r = 1/8$, is 0.1 when $r = 8$. The design which achieves the highest mean efficiency (0.90) and the highest minimum efficiency (0.70) is ξ_{22} which is D-optimal for $q = r = 1$.

3.6.1 Explaining the structure of optimal designs

It is not immediately apparent why, for the designs in Table 3.10, the optimal design has observations at $t = 1$ and 200 when r is large, i.e. at the extremes of the range for observation times. Such observations, if taken by themselves, would provide very little information on the mean parameter values, E_a and k_r. This subsection aims to give some insight into why the presence of run-to-run error makes such "low information" points useful when included in the optimal design.

To illustrate this phenomenon, we explore a simple example as an analogy to our more complex problem. Suppose there are N runs with m observations per run and a simple linear model to describe a response y_{ij} observed on the ith run of a process for a value $x_{ij} \in [0,1]$ of an independent variable, where $i = 1\ldots, N$ and $j = 1, \ldots, m$. The statistical model assumed is:

$$\boldsymbol{y} = \beta \boldsymbol{x} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$=$(y_{11}, \ldots, y_{1m}, \ldots, y_{Nm})'$, $\boldsymbol{x}$=$(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)'$, $\boldsymbol{x}_i = (x_{i1}, \ldots x_{im})'$,
$\boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, (\sigma_r^2 + \sigma_m^2)Q)$ and β is an unknown parameter requiring estimation, with Q a block diagonal matrix with matrices Q_i on the leading diagonal which hold the correlation between any pair of observations on run i $(i = 1, \ldots, N)$. Then, for two observations y_{ij} and y_{il} $(j \neq l)$, the (j,l)th element of Q_i is

$$\kappa = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_m^2}.$$

The model has been chosen to have no intercept, in line with the non-linear model considered in this thesis. An approximate design for this example is of the form:

$$\xi = \left\{ \begin{array}{cccc} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_n \\ w_1 & w_2 & \ldots & w_n \end{array} \right\},$$

where w_i is the weight of support point $\boldsymbol{x}_i$ and $w_1 + \ldots + w_n = 1$. If observations are taken from independent runs, then a D-optimal design for estimating β is a single support point at $x = 1$. Suppose a practitioner knew that $\kappa = 0.8$, and had the choice of taking two independent observations on different runs at 1 (i.e. $N = 2$, $m = 1$, with $x_{11} = x_{21} = 1$), or two observations from the same run (i.e. $N = 1$, $m = 2$, with $x_{11} \neq x_{12}$). For the first choice, the value of the determinant of the

information matrix is 2. For the second choice, the determinant can be shown to be

$$\frac{x_{11}^2 + x_{12}^2 - 2\kappa x_{11} x_{12}}{(1-\kappa^2)^2}.$$

The design which maximises this expression (for $\kappa > 0.5$) has $x_{11} = 1$ and $x_{12} = 0$, which make the determinant have value 7.72. The value of κ for which the two designs give the same value for the determinant is $\kappa^2 = 1 - 1/\sqrt{2} \approx 0.293$.

This demonstrates two things for the simple example: for sufficiently high values of the variance component for run-to-run error, the optimal design might have a single support point with two observations from the same run (a fairly intuitive strategy), and that some of those points might provide "low information". The observation at $x_{12} = 0$ provides no information on the value of β when it is the only observation made, as the predicted value for y_{12} is 0. However, the high value of the variance component for run-to-run error relative to σ_m^2 means that these "low information" points are valuable; any observation made at 0 will be purely of the statistical error, so the run-to-run error can be cancelled by taking $y_{11} - y_{12}$. When $\kappa = 0.8$, cancelling run-to-run error will greatly improve the accuracy of estimation of β.

For our more complex problem, this argument may give some insights into why optimal designs with high values for r have observations at the end points of the allowed range for time: the information content of an observation made at those points is nearly 0 for estimating $\boldsymbol{\theta}_m$ but, being observations of the statistical error, they allow most of the run-to-run error to be cancelled out when $\boldsymbol{\theta}_m$ is estimated.

3.7 Investigating the impact of varying r and q on an optimal design

In this section the values of r and q for which a locally D-optimal design has observations at the two extreme values for time are explored as, from Table 3.10, we observe that there seems to be a discrete "jump" between such points being present or

absent in such a design. A grid of values for r and q is explored, and the relationship between their values and the optimal times of observation is investigated. The values of the number of observations, m, per run and the mean parameters are then varied to investigate whether the relationship between r, q and the form of optimal designs obtained is affected by varying m and the values assumed for the mean parameters.

3.7.1 Assessing the impact of r and q on optimal designs when m, E_a and k_r are fixed

Locally optimal designs were found under the assumption that $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$, with τ having two values 1 or 10 (meaning strong or weak correlation respectively), and r varied uniformly over 100 values between 0.1 and 1, with $q = 0$. Designs found for $\tau = 1$ were very similar to those for $\tau = 10$, so only the former are presented here. It was found that there were several "switchovers", where the optimal design radically changed. We say a switchover has occurred between two optimal designs, ξ_a and ξ_b when there exists at least one (i, j) such that $|t^a_{ij} - t^b_{ij}| > 10$ minutes, where $\{t^a_{ij}\}$ are the observation times for ξ_a, and similarly for ξ_b.

Table 3.12: Locally D-optimal designs exhibiting 'switchovers' for $\tau = 1$, $q = 0$, $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ for r spaced evenly across $0.1 \leq r \leq 1$ and $q = 0$. The D-efficiency of each design is calculated for the adjacent r value (e.g. the D-efficiency of the first design is calculated for $r = 0.17$)

r	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)						$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)						D-efficiency
0.16	31.2	35.6	39.7	43.7	47.9	52.6	6.1	8.4	10.5	12.6	15.0	17.8	0.998
0.17	31.4	35.7	39.7	43.6	47.8	52.4	6.5	9.0	11.3	13.8	16.8	200.0	0.998
0.24	31.5	35.8	39.7	43.6	47.6	52.1	6.5	9.0	11.3	13.8	16.6	200.0	0.999
0.25	1.0	32.7	37.2	41.4	45.6	50.3	7.0	9.7	12.4	15.5	186.6	200.0	0.995
0.46	1.0	32.9	37.2	41.3	45.5	50.1	7.1	9.7	12.4	15.4	186.5	200.0	0.999
0.47	1.0	34.0	38.8	43.5	48.4	200.0	7.1	9.7	12.4	15.4	186.5	200.0	0.999
0.69	1.0	34.1	38.9	43.4	48.3	200.0	7.1	9.7	12.3	15.2	186.4	200.0	1.000
0.70	1.0	34.1	38.8	43.4	48.3	200.0	7.7	10.8	14.1	174.7	187.8	200.0	1.000

The optimal designs and r values at which these switchovers occurred are given in Table 3.12 for $\tau = 1$. Note that the pairs of designs in these switchovers have very high and equal or very close D-efficiencies. We see that when $r > 0.47$ it seems

sensible to choose ξ_{10} in the set of starting designs in Algorithm 3.2 to include three endpoints (1 and 200 for $\boldsymbol{t}_1$; 200 for $\boldsymbol{t}_2$) in the observation times.

The switchovers were also investigated when q is nonzero. A grid of combinations of values of q and r of size 100×100 where q and r were evenly spaced between 0 and 6 and $\tau = 1$, $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$ was considered. For each combination of r and q an optimal design was found, and then the set of designs obtained was searched for the values of r and q at which the optimal design changed to include the extreme time values. A fifth switchover was discovered for low values of q.

To investigate how the occurrence of a switchover varies with the assumed values of r and q, the average of the two r values which the switchover occurred between was plotted against q, see Figure 3.2. The relationship is roughly linear and hence linear regression was used to estimate the equation of each line. These are, for each switchover,

$$\begin{aligned}
\text{Switchover 1:} &\quad r = 0.1605 + 0.1956q \\
\text{Switchover 2:} &\quad r = 0.2353 + 0.2670q \\
\text{Switchover 3:} &\quad r = 0.4587 + 0.4796q \\
\text{Switchover 4:} &\quad r = 0.6290 + 0.9278q \\
\text{Switchover 5:} &\quad r = 1.795 + 1.721q.
\end{aligned}$$

From the $q = 0$ case, the values for the intercepts of the first four equations should be 0.165, 0.245, 0.465 and 0.695 respectively, which implies that the approximations are reasonably accurate.

These results demonstrate that the switchovers appear to occur in a linear and predictable fashion as r and q are varied. This should mean that searches for the locally D-optimal design can be conducted more rapidly, by beginning searches for optimal designs with starting designs that include the appropriate observation times for the given values of r and q.

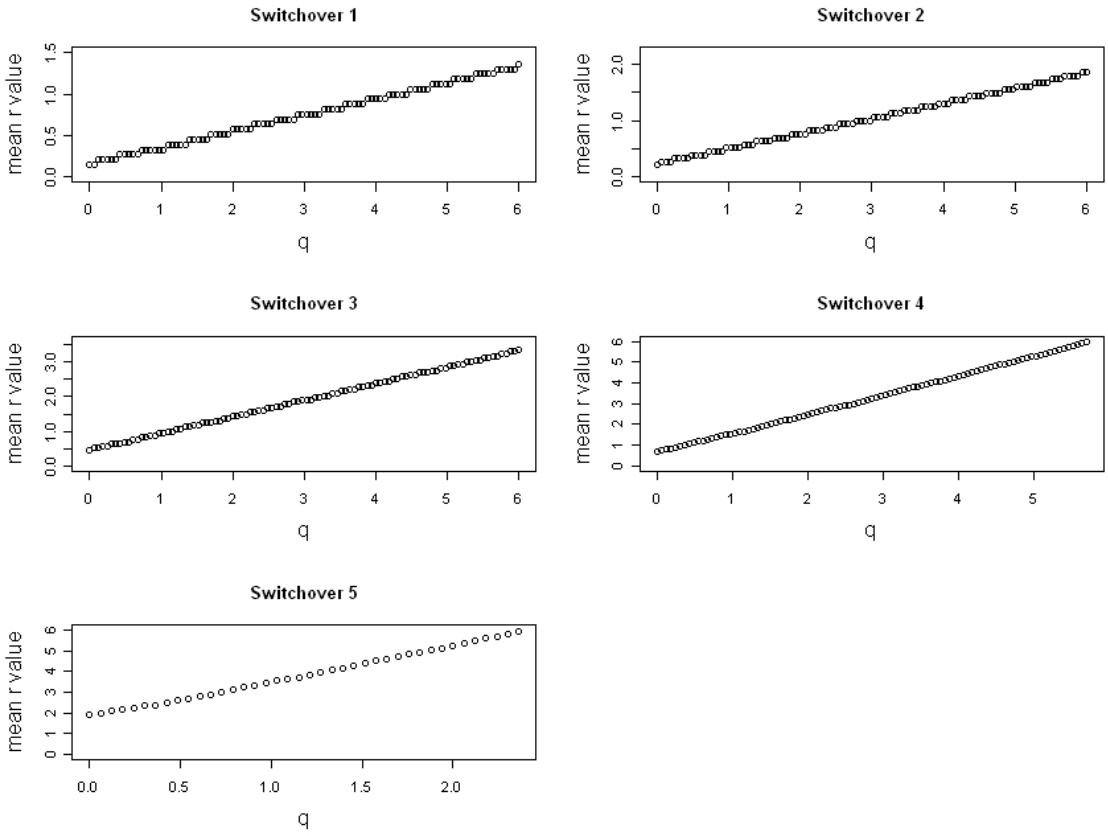


Figure 3.2: Average r values for designs demonstrating a switchover against q for $\tau = 1$, $\rho = 1$, $E_a = 4.91 \times 10^4$ and $k_r = 4.59 \times 10^{-4}$

3.7.2 Assessing the impact of varying E_a, k_r and m

The results so far show switchovers in the features of optimal designs as r increases, with the 'centre' of the observation times being around the optimal times of observation given by Equation (2.5). We also investigated whether this finding remains true when the number, m, of observations and the value of the mean parameters, E_a and k_r, are altered.

The values considered were:

- (E_a, k_r)=$(4\times10^4, 6\times10^{-4})$, $(6\times10^4, 4\times10^{-4})$ or $(4.91 \times 10^4, 4.59 \times 10^{-4})$
- $m = 4$ or 8
- $\tau = \rho = 1$ and $q = 0$
- Twenty values of r equally spaced between 0.1 and 1

For comparison, the optimal temperatures and times at which to take observations when $m = 1$ are found from Equation (2.5) to be (70, 29.0; 100, 9.4) and (70, 50.1; 100, 9.2) for (E_a,k_r)=$(4\times10^4, 6\times10^{-4})$ and $(6\times10^4, 4\times10^{-4})$ respectively.

The study in Subsection 3.7 was repeated and the resulting designs where switchovers occur are shown in Tables 3.13 and 3.14 for $m = 4$ and 8, respectively. For $m = 4$ the results are as anticipated, with the centre of the observations times centered around the single optimal time for the case of no correlation.

Whether the optimal design for $m = 4$ includes $t = 1$ appears to depend on the speed of the reaction. When the reaction is occurring rapidly, as for $E_a = 4 \times 10^4$ and $k_r = 6 \times 10^{-4}$, the optimal design does not have an observation at $t = 1$, but otherwise the optimal design does include observations at $t = 1$ for sufficiently high values of r. The switchovers occur for higher values of r than for the switchovers found for $m = 6$ given in Table 3.12.

For $m = 8$, Table 3.14 shows that when there are more observations times in each support point, the optimal design includes the maximum value $t = 200$ in at least one run for smaller values of r than for the m=6 designs given in Table 3.12.

Table 3.13: Occurrence of switchover for $m = 4$, $\tau = 1$, $\rho = 1$ and $q = 0$

r	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)				$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)			
	$E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$							
0.29	34.1	38.8	43.4	48.4	7.1	9.7	12.3	15.2
0.34	34.1	38.9	43.4	48.2	7.7	10.8	14.1	200.0
0.43	34.3	38.9	43.3	48.0	7.7	10.8	14.0	200.0
0.48	35.7	40.9	46.1	200.0	7.7	10.7	13.9	200.0
0.57	35.8	40.9	46.1	200.0	7.8	10.7	13.9	200.0
0.62	1.0	35.8	40.9	46.1	8.6	12.4	186.7	200.0
	E_a=4×10^4, k_r=6×10^{-4}							
0.24	23.40	27.7	31.9	36.4	6.5	9.0	11.6	14.5
0.29	23.50	27.8	31.8	36.3	7.1	10.1	13.3	200.0
0.34	23.6	27.8	31.8	36.1	7.1	10.0	13.2	200.0
0.38	24.8	29.5	34.3	200.0	7.1	10.0	13.2	200.0
0.76	25.0	29.5	34.1	200.0	7.2	10.0	12.9	200.0
0.81	25.0	29.5	34.1	200.0	8.0	11.6	185.7	200.0
	$E_a = 6 \times 10^4$, $k_r = 4 \times 10^{-4}$							
0.24	43.0	48.2	53.3	58.6	6.3	8.9	11.4	14.3
0.29	43.1	48.3	53.2	58.5	6.9	9.9	13.1	200.0
0.34	43.2	48.3	53.2	58.4	6.9	9.9	13.0	200.0
0.38	1.0	45.1	50.5	56.0	7.0	9.8	12.7	200.0
0.76	1.0	45.1	50.5	56.0	7.0	9.8	12.7	200.0
0.81	1.0	45.1	50.5	55.9	7.8	11.4	185.4	200.0

For all parameter values explored except $k_r = 4.59 \times 10^{-4}$ and $E_a = 4.91 \times 10^4$, the optimal design has $t = 200$ for the lowest value of r investigated.

This short study demonstrates that, while the influence of r on the occurrence of switchovers is similar for different values of m, E_a and k_r, it is not identical. To apply the equations found in the previous subsection may be misleading about where searches for an optimal design should be conducted, when m, E_a or k_r have different values to those assumed during the study in the previous subsection.

Table 3.14: Occurrence of Switchover for $m = 8$, $\tau = 1$, $\rho = 1$ and $q = 0$

r	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)								$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)							
	$E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$															
0.10	28.9	33.1	36.9	40.5	44.2	48.0	52.0	56.6	5.3	7.4	9.3	11.2	13.1	15.1	17.4	20.1
0.15	29.2	33.3	36.9	40.5	44.0	47.6	51.5	55.9	5.7	7.8	9.8	11.8	13.9	16.2	18.9	200.0
0.15	29.2	33.3	36.9	40.5	44.0	47.6	51.5	55.9	5.7	7.8	9.8	11.8	13.9	16.2	18.9	200.0
0.19	1.0	30.2	34.4	38.2	41.9	45.7	49.7	54.2	6.4	9.0	11.4	14.0	17.0	175.4	187.9	200.0
0.24	1.0	30.4	34.5	38.2	41.9	45.6	49.5	53.9	6.4	9.0	11.4	13.9	16.9	175.3	187.8	200.0
0.29	1.0	31.4	35.7	39.7	43.6	47.8	52.4	200.0	6.5	9.0	11.4	13.9	16.8	174.9	187.5	200.0
0.48	1.0	31.6	35.8	39.7	43.5	47.5	51.9	200.0	6.5	9.0	11.3	13.7	16.6	174.8	187.5	200.0
0.53	1.0	32.6	37.2	41.4	45.7	50.4	195.0	200.0	7.0	9.7	12.5	15.6	163.9	175.6	187.5	200.0
	$E_a = 4 \times 10^4$, $k_r = 6 \times 10^{-4}$															
0.10	19.3	23.0	26.3	29.5	32.7	36.0	39.5	43.6	5.1	7.2	9.2	11.2	13.3	15.6	18.4	200.0
0.15	20.2	24.0	27.4	30.7	34.1	37.7	41.8	200.0	5.2	7.3	9.2	11.1	13.1	15.3	17.9	200.0
0.15	20.2	24.0	27.4	30.7	34.1	37.7	41.8	200.0	5.2	7.3	9.2	11.1	13.1	15.3	17.9	200.0
0.19	20.4	24.1	27.4	30.6	33.9	37.4	41.5	200.0	5.5	7.7	9.8	11.9	14.2	17.0	185.6	200.0
0.19	20.4	24.1	27.4	30.6	33.9	37.4	41.5	200.0	5.5	7.7	9.8	11.9	14.2	17.0	185.6	200.0
0.24	21.2	25.1	28.6	32.2	35.8	40.0	194.0	200.0	5.9	8.3	10.6	13.1	16.0	173.2	186.4	200.0
0.34	21.4	25.1	28.6	32.0	35.7	39.7	193.9	200.0	5.9	8.3	10.6	13.0	15.9	173.1	186.4	200.0
0.38	22.2	26.3	30.1	34.0	38.3	189.0	194.8	200.0	6.0	8.3	10.6	13.0	15.8	173.0	186.4	200.0
0.53	22.3	26.3	30.1	33.9	38.2	188.9	194.7	200.0	6.0	8.3	10.5	12.9	15.7	173.0	186.4	200.0
0.57	22.3	26.3	30.1	33.9	38.2	188.8	194.7	200.0	6.4	9.1	11.7	14.7	161.8	174.2	186.6	200.0
	$E_a = 6 \times 10^4$, $k_r = 4 \times 10^{-4}$															
0.10	37.2	41.8	45.9	49.9	53.9	58.0	62.4	67.3	5.0	7.1	9.1	11.0	13.1	15.4	18.1	200.0
0.15	37.6	42.0	46.0	49.9	53.7	57.7	61.9	66.6	5.1	7.1	9.0	10.9	12.9	15.0	17.7	200.0
0.19	1.0	38.8	43.3	47.5	51.5	55.6	59.9	64.8	5.7	8.2	10.5	13.0	15.9	173.4	186.80	200.0
0.43	1.0	39.5	43.7	47.6	51.3	55.1	59.1	63.6	5.8	8.2	10.4	12.7	15.5	173.0	186.6	200.0
0.48	1.0	40.4	44.9	49.1	53.2	57.5	62.3	200.0	6.3	8.9	11.5	14.5	161.9	174.4	187.1	200.0

3.8 Bayesian D-optimal designs in the presence of serial correlation, and run-to-run and measurement errors

The investigations so far in this chapter have only considered locally optimal designs, where full knowledge of the parameter values is assumed before the experiment is designed. This assumption is not realistic and, as in Chapter 2, we will now investigate designs that are robust to the values of the parameters and ratios of parameters that are involved in the correlation and mean functions of model (1.8): The criterion used to find the designs is Bayesian D-optimality.

In this section, we investigate the most appropriate distributions to use as prior distributions when finding Bayesian D-optimal designs. We begin by defining the prior distributions for the parameters used in the study and then describe how the designs are obtained and investigate their optimality via an equivalence theorem.

3.8.1 Prior distributions

We investigate Bayesian D-optimal designs through a variety of scenarios formed from different combinations of choices of prior distributions for the parameters. As in Chapter 2, we make use of prior knowledge, either scientific or from previous experiments, of the mean parameters. We define prior distributions first, for the mean function parameters; second, for the parameters controlling serial correlation and, third, for ratios of the variance components, r and q. We assume independence of all the prior distributions.

(i): Two joint prior distributions for E_a and k_r were chosen:

π_α: $E_a \sim \mathrm{N}(4.91\times10^4,(4\times10^3)^2)$, $k_r \sim \mathrm{G}(m_k^2/v_k, v_k/m_k)$, with mean $m_k = 4.59\times10^{-4}$ and variance $v_k = (4\times10^{-4})^2$

π_β: $E_a \sim \mathrm{U}(2.5\times10^4, 6.5\times10^4)$, $k_r \sim \mathrm{U}(1\times10^{-4}, 1\times10^{-3})$

The choice for π_α was made because it led to designs whose Bayesian D-efficiency was

robust to mis-specification of the prior distribution (see Section 2.5); π_β was chosen to reflect likely ranges in which E_a and k_r might fall, as suggested by practitioners.

Practitioners are not expected to have prior knowledge of the magnitude of correlation between observations within the same run, as this is typically not estimated from previous experiments. Hence designs are required which are robust to the values of the correlation parameters.

(ii): For serial correlation, three prior distributions were considered: a point prior distribution, using values of τ and ρ which give strong correlation; a gamma distribution for τ with probability concentrated on lower values of τ (with mean 2 and variance 4) to give stronger correlation and a uniform distribution for each of τ and ρ.

We define:

π_a: $\tau = \rho = 1$

π_b: $\tau \sim$G(1,2), $\rho \sim$U(0,2)

π_c: $\tau \sim$U(0,10), $\rho \sim$U(0,2)

(iii): It was also important to find designs whose performance is robust to the values of the variance components for run-to-run and measurement error. As both r and q are positive, sensible prior distributions which might be used are fixed values (point prior distributions), inverse gamma (IG) distributions, or log uniform distributions. A scenario where r was much larger than q on average was suggested by practitioners and included in the study (π_3).

The following four prior distributions were used:

π_1: $r = q = 1$

π_2: r and $q \sim$IG(10,2)

π_3: $r \sim$IG(100,2), $q \sim$IG(10,2)

π_4: log(r) and log(q) $\sim$U(-2.5,2.5)

A prior distribution for E_a, k_r, τ, ρ, r and q can be formed as a combination of a choice from of each of (i), (ii) and (iii), i.e. from the $2 \times 3 \times 4 = 24$ different possible prior distributions. We use $\pi_{\alpha a 1}$ to denote the prior distribution obtained by selecting π_α, π_a and π_1, and so on.

3.8.2 Finding Bayesian D-optimal designs

To find a Bayesian D-optimal design requires a numerical estimate of the integral in Equation (1.15). When prior distributions for the correlation parameters are included, this requires a six dimensional estimate. As the Latin hypercube method used in the previous chapter is less efficient in terms of number of function evaluations required at higher dimensions, we use the method given by Gotwalt et al. (2009). This method uses transformation of the function to be integrated and then evaluation by quadrature. As originally described, the method assumes that the prior distribution is a normal distribution, but the method can be applied to different prior distributions using further transformation, provided the parameters have independent prior distributions. The method evaluates the integral in the objective function (1.15) via methods described by Monahan and Genz (1997), Cassity (1965) and Press et al. (1992).

Gregson (2009) found that the method produced much more accurate estimates than Monte Carlo sampling for similar numbers of function evaluations and, depending on the problem being studied, performed similarly to orthogonal array Latin hypercube sampling of integrals. For the work in this thesis, my method in `R` built on the work of Gregson (2009).

For each of the 24 prior distributions formed from (i)-(iii), a Bayesian D-optimal design was found by computer search. These designs are given in Table 3.15. These designs are generally very similar. The largest difference is between the observation times for the two prior distributions for E_a and k_r, with the optimal observation times for π_β being smaller on average than for designs found for π_α. In particular, when the prior distribution for E_a and k_r is π_α the fifth observation time in the optimal design on $\boldsymbol{t}_1$ is usually around 190, while for π_β the respective observation

Table 3.15: Bayesian D-optimal designs

Prior distribution	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)						$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)					
$\pi_{\alpha a1}$	1.0	25.6	32.4	39.5	192.7	200.0	4.8	8.1	12.1	22.8	66.8	200.0
$\pi_{\alpha a2}$	13.3	28.9	37.6	48.0	193.4	200.0	4.6	8.0	12.1	18.1	54.5	200.0
$\pi_{\alpha a3}$	1.0	25.9	32.6	39.6	189.9	200.0	4.8	8.3	12.6	18.5	58.4	200.0
$\pi_{\alpha a4}$	1.0	25.7	32.5	39.5	189.8	200.0	4.6	8.0	12.1	18.3	57.2	200.0
$\pi_{\alpha b1}$	1.0	22.6	33.7	47.7	184.5	200.0	4.8	8.1	12.9	23.9	90.3	200.0
$\pi_{\alpha b2}$	1.0	21.5	31.4	44.1	101.9	200.0	4.7	8.0	12.5	20.9	68.3	200.0
$\pi_{\alpha b3}$	1.0	22.8	32.9	45.8	193.1	200.0	5.4	8.2	12.7	37.5	183.5	200.0
$\pi_{\alpha b4}$	1.0	21.2	31.2	44.1	94.9	200.0	4.7	8.0	12.4	21.1	77.3	200.0
$\pi_{\alpha c1}$	1.0	24.7	33.2	43.5	192.8	200.0	5.3	7.8	11.5	23.2	89.1	200.0
$\pi_{\alpha c2}$	1.0	25.0	32.9	42.1	183.0	200.0	5.2	7.7	11.3	18.9	65.1	200.0
$\pi_{\alpha c3}$	1.0	23.7	32.5	44.3	196.5	200.0	5.7	8.1	11.9	38.0	192.7	200.0
$\pi_{\alpha c4}$	1.0	24.9	33.2	42.6	189.2	200.0	5.2	7.7	11.1	20.0	72.3	200.0
$\pi_{\beta a1}$	1.0	25.8	30.2	34.5	39.2	200.0	5.9	8.6	11.6	15.5	192.8	200.0
$\pi_{\beta a2}$	1.0	25.6	30.2	34.6	39.5	200.0	5.6	8.2	11.0	14.4	55.9	200.0
$\pi_{\beta a3}$	1.0	26.7	30.2	33.6	37.4	200.0	6.5	8.5	10.7	34.4	194.2	200.0
$\pi_{\beta a4}$	1.0	25.6	30.1	34.5	39.3	200.0	5.9	8.6	11.6	15.4	192.4	200.0
$\pi_{\beta b1}$	1.0	24.6	29.7	34.8	41.3	200.0	5.9	8.1	10.6	14.2	68.9	200.0
$\pi_{\beta b2}$	1.0	24.3	29.	35.1	42.2	200.0	5.8	8.0	10.4	13.8	50.2	200.0
$\pi_{\beta b3}$	1.0	26.6	30.1	33.6	37.9	200.0	6.8	8.4	10.3	37.9	190.1	200.0
$\pi_{\beta b4}$	1.0	24.5	29.7	34.8	41.5	200.0	5.9	8.0	10.4	13.7	57.2	200.0
$\pi_{\beta c1}$	1.0	27.1	30.4	33.5	37.4	200.0	6.6	8.1	9.6	11.6	63.1	200.0
$\pi_{\beta c2}$	1.0	27.1	30.4	33.5	37.7	200.0	6.6	8.1	9.6	11.8	52.1	200.0
$\pi_{\beta c3}$	1.0	28.3	30.6	32.9	35.5	200.0	7.2	8.4	9.6	37.6	195.1	200.0
$\pi_{\beta c4}$	1.0	27.0	30.3	33.5	37.5	200.0	6.6	8.1	9.6	11.6	56.6	200.0

time is around 40.

The Bayesian D-efficiency of each design was calculated from Equation (1.16) (Section 1.3) assuming each of the 24 prior distributions was 'correct' in turn. These distributions are here called 'scenarios' and labelled as s_1-s_{24} using the same lexicographical ordering for the $\pi_{\alpha a1}, \ldots, \pi_{\beta c4}$ as in Table 3.15. The efficiencies of the designs are given in Tables 3.18 and 3.19.

All the optimal designs achieve fairly high Bayesian D-efficiencies across all scenarios, which follows from the similarities of the designs in Table 3.15. The lowest Bayesian D-efficiency attained was 0.93 in Scenario 1 where the true prior distribution is $\pi_{\alpha a1}$, for design $\xi_{\beta c3}$; under this scenario all types of prior distributions from (i), (ii) and (iii) are mis-specified in the design search.

The study can be viewed as a full factorial experiment, where the factors are the types of prior distributions, (i), (ii) and (iii), assumed for each design and scenario, with 2, 3 and 4 levels respectively, and the deterministic response is the Bayesian D-efficiency. In order to assess the relative importance of the effect of each factor, and their joint effects, we assess the variability in the 'responses' in an analogous method to an analysis of variance. Both the Bayesian D-efficiency and the logarithm of the efficiency were considered as the response, but gave similar results for the amount of variation determined by each factor. This determined that 50% of the variation between Bayesian D-efficiencies can be explained by the interaction between the prior distribution assumed on E_a and k_r when designing and the prior distribution on E_a and k_r assumed when calculating Bayesian D-efficiency. This implies that the most important prior distributions to choose when designing if high Bayesian D-efficiency is required are the prior distributions for E_a and k_r. The term which explained the second most variation was the interaction between the prior distributions assumed on τ and ρ and the prior distribution assumed for r and q, which account for 18% of the variation. All other terms explain a negligible amount of the variance.

Table 3.16 gives the mean Bayesian D-efficiencies obtained from the study when the prior distributions for E_a and k_r were varied during design and calculation of

Bayesian D-efficiency. This demonstrates that, as should be anticipated, the highest value obtained is when the scenario and design match. The lowest mean is obtained when normal and gamma prior distributions are assumed for E_a and k_r and the 'true' prior distributions are actually uniform distributions.

Table 3.16: Mean Bayesian D-efficiency obtained when the scenario and designs for the prior distributions on E_a and k_r are varied

		Prior distribution for E_a and k_r designed for	
		π_α	π_β
Prior distribution	π_α	0.998	0.981
assumed to be 'true'	π_β	0.978	0.995

The impact on design efficiency of the choice of prior distribution for a particular set of parameters E_a, k_r or τ, ρ or r, q was also investigated by calculating the average of the Bayesian D-efficiencies of all the designs found using a particular choice of prior distribution and a given scenario. For example, there are six designs found for the prior distribution π_4 for r and q (see (iii) in Subsection 3.8.1). These are labelled π_{**4}. For the first scenario (where the prior distribution was $\pi_{\alpha a1}$), the performance of a design found assuming this prior distribution can be assessed by averaging the Bayesian D-efficiencies of all six designs found for π_{**4} ($\pi_{\alpha a4}$, $\pi_{\alpha b4}$, $\pi_{\alpha c4}$, $\pi_{\beta a4}$, $\pi_{\beta b4}$ and $\pi_{\beta c4}$). Table 3.17 gives these value averaged over all 24 scenarios for each prior distribution.

These results demonstrate that all prior distributions give designs with a similar average performance, but some are preferred over others. For r and q there seems to be a slight advantage to using the log uniform prior distributions (π_4) in finding a design. Table 3.17 shows that such designs have a slightly higher mean average Bayesian D-efficiency than designs obtained for the three other prior distributions for r and q in the study, although the minimum value obtained for designs found for π_2 is slightly higher. The prior distributions π_a and π_b for τ and ρ provide designs which are superior to designs found for π_c (uniform distributions over τ and ρ). Optimal designs found for π_α and π_β have very similar average performances.

Table 3.17: Mean, median and minimum averaged Bayesian D-efficiencies of designs found under different prior distributions

Parameters	Prior	Mean	Median	Minimum
r,q	π_1	0.988	0.991	0.941
	π_2	0.988	0.991	0.945
	π_3	0.986	0.988	0.926
	π_4	0.989	0.991	0.943
τ,ρ	π_a	0.988	0.990	0.948
	π_b	0.988	0.993	0.943
	π_c	0.986	0.989	0.926
E_a,k_r	π_α	0.988	0.990	0.943
	π_β	0.988	0.990	0.926

The prior distributions chosen for E_a and k_r would typically be those which best reflect the prior belief of the practitioner. For τ, ρ, r and q there is usually less prior information available than for the mean parameters, so the prior distributions which give designs robust to prior misspecification of the values of these parameters should be chosen. This study suggests that the prior distributions which provide the most robust designs are π_b (which had a higher median than π_a, although a lower minimum) for τ and ρ, and π_4 for r and q. In practice, all prior distributions will give very similar performances, however.

Table 3.18: Bayesian D-efficiencies of Bayesian D-optimal designs for $\xi_{\alpha a1}$-$\xi_{\alpha c4}$

	Design											
Scenario	$\xi_{\alpha a1}$	$\xi_{\alpha a2}$	$\xi_{\alpha a3}$	$\xi_{\alpha a4}$	$\xi_{\alpha b1}$	$\xi_{\alpha b2}$	$\xi_{\alpha b3}$	$\xi_{\alpha b4}$	$\xi_{\alpha c1}$	$\xi_{\alpha c2}$	$\xi_{\alpha c3}$	$\xi_{\alpha c4}$
s_1	1.00	0.99	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	0.99	1.00
s_2	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00
s_3	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00
s_4	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
s_5	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s_6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
s_7	1.00	0.98	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00
s_8	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
s_9	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s_{10}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s_{11}	1.00	0.98	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00
s_{12}	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s_{13}	0.96	0.95	0.96	0.96	0.94	0.95	0.95	0.95	0.96	0.96	0.96	0.96
s_{14}	0.97	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.98
s_{15}	0.98	0.96	0.98	0.98	0.98	0.97	0.99	0.97	0.98	0.98	0.99	0.98
s_{16}	0.98	0.97	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.98
s_{17}	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.98	0.99
s_{18}	0.99	0.98	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.98	0.99
s_{19}	0.99	0.97	0.99	0.99	0.98	0.98	0.99	0.98	0.99	0.99	0.99	0.99
s_{20}	0.99	0.98	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.98	0.99
s_{21}	0.98	0.97	0.98	0.98	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98
s_{22}	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.98	0.98	0.98	0.98
s_{23}	0.99	0.97	0.99	0.99	0.98	0.98	0.99	0.98	0.99	0.99	0.99	0.99
s_{24}	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98

Table 3.19: Bayesian D-efficiencies of Bayesian D-optimal designs for $\xi_{\beta a1}$-$\xi_{\beta c4}$

	Design											
Scenario	$\xi_{\beta a1}$	$\xi_{\beta a2}$	$\xi_{\beta a3}$	$\xi_{\beta a4}$	$\xi_{\beta b1}$	$\xi_{\beta b2}$	$\xi_{\beta b3}$	$\xi_{\beta b4}$	$\xi_{\beta c1}$	$\xi_{\beta c2}$	$\xi_{\beta c3}$	$\xi_{\beta c4}$
s_1	0.96	0.97	0.96	0.96	0.97	0.97	0.95	0.97	0.94	0.94	0.93	0.94
s_2	0.98	0.99	0.98	0.98	0.99	0.99	0.97	0.99	0.97	0.97	0.96	0.97
s_3	0.98	0.99	0.98	0.98	0.99	0.99	0.97	0.99	0.97	0.97	0.96	0.97
s_4	0.98	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.97	0.97	0.96	0.97
s_5	0.97	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.97
s_6	0.98	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.98	0.98	0.97	0.98
s_7	0.98	0.98	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.98
s_8	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.98	0.98	0.98	0.98
s_9	0.98	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.98
s_{10}	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
s_{11}	0.98	0.98	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.99	0.98
s_{12}	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
s_{13}	1.00	1.00	0.98	1.00	1.00	0.99	0.98	1.00	0.98	0.98	0.95	0.98
s_{14}	1.00	1.00	0.99	1.00	1.00	1.00	0.98	1.00	0.99	0.99	0.96	0.99
s_{15}	1.00	0.99	1.00	1.00	0.99	0.99	1.00	0.99	1.00	0.99	1.00	0.99
s_{16}	1.00	1.00	0.99	1.00	1.00	1.00	0.98	1.00	0.99	0.99	0.97	0.99
s_{17}	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.98	1.00
s_{18}	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.98	1.00
s_{19}	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
s_{20}	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00
s_{21}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
s_{22}	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00
s_{23}	1.00	0.99	1.00	1.00	0.99	0.99	1.00	0.99	1.00	1.00	1.00	1.00
s_{24}	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00

3.8.3 Use of an equivalence theorem to establish Bayesian D- and near D-optimality

We investigate if the designs, $\xi_{\alpha a1}, \ldots, \xi_{\beta c4}$, found are Bayesian D-optimal via Theorem 2.5.1. As before, when assessing designs over so many dimensions, a full search is impractical, so a Latin hypercube over a grid of 4^7 points was used. The maximum values of the integrated standardised variance obtained over the grid search for each design are given in Table 3.20, and should be at most 2, as explained in Theorem 2.5.1. The support points of each design achieved the value of 2 when the integrated standardised variance was evaluated.

Table 3.20: Maximum integrated standardised variance observed over a grid search for each Bayesian D-optimal design

Design	$\xi_{\alpha a1}$	$\xi_{\alpha a2}$	$\xi_{\alpha a3}$	$\xi_{\alpha a4}$	$\xi_{\alpha b1}$	$\xi_{\alpha b2}$	$\xi_{\alpha b3}$	$\xi_{\alpha b4}$	$\xi_{\alpha c1}$	$\xi_{\alpha c2}$	$\xi_{\alpha c3}$	$\xi_{\alpha c4}$
Value	1.98	2.00	2.00	1.97	1.95	1.97	1.95	1.96	1.95	2.00	1.95	1.98
Design	$\xi_{\beta a1}$	$\xi_{\beta a2}$	$\xi_{\beta a3}$	$\xi_{\beta a4}$	$\xi_{\beta b1}$	$\xi_{\beta b2}$	$\xi_{\beta b3}$	$\xi_{\beta b4}$	$\xi_{\beta c1}$	$\xi_{\beta c2}$	$\xi_{\beta c3}$	$\xi_{\beta c4}$
Value	1.86	1.89	1.89	1.86	1.85	1.88	1.89	1.87	1.82	1.88	1.88	1.87

These results provide reasonably strong evidence that the designs are Bayesian D-optimal or near D-optimal.

3.9 D-optimal designs when serial correlation and heteroscedasticity are believed to be present

In this section, the impact of heteroscedasticity on the locally and Bayesian D-optimal designs found previously is explored, when correlation between observations from the same run is present. The model we adopt is (1.8), with correlation between observations given by Equation (3.4) and $0 < \lambda \leq 1$. Locally D-optimal designs are obtained for a range of values of λ and the D-efficiencies of the designs obtained are compared. Then Bayesian D-optimal designs for three different prior distributions for λ are found.

As in Section 2.6, to accommodate the impact of λ on the selection of observation times, the range of observation times is extended, but only to 300 minutes here, so that:

$$\mathcal{X} = \{(T, t); 70 \leq T \leq 100, 1 \leq t \leq 300\}.$$

Note that in practice, the model for the expected response applies only to a limited range of observation times. Observations taken a long time after the start of the reaction will not provide any information, as the reaction will be complete.

3.9.1 Locally D-optimal designs

As before, designs were found for $m = 6$, using model (1.8). Locally D-optimal designs were found for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\tau = 1$ and 10, $\rho = 1$, $\lambda = 0.14$, 0.25, 0.5 and 1 and $r = q = 1$, $r = 0.125$, $q = 8$ and $r = 8$, $q = 0.125$.

The values of r and q were chosen to reflect the scenarios when the variance component for run-to-run error has, respectively:

- Similar magnitude to other variance components
- Negligible magnitude compared to other variance components
- Greater magnitude than other variance components

The values of λ were chosen to be similar to those investigated in Section 2.6. The values of $\tau = 1$ and 10 represent strong and weak serial correlation respectively.

Table 3.21 gives the results for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\rho = 1$ and $r = q = 1$. For the support point at $T = 70$, the third and fourth optimal observation times increase as λ decreases for each of $\tau = 1$ and 10. The presence of end points depend on the value of λ in a less linear manner than observed in Section 2.6. For $\lambda = 1$ and both τ=1 and 10, the optimal design requires observations at 300 when $T = 70$. For $\lambda = 0.25$, 0.5 the optimal design does not, instead having observations at $t = 1$. While the optimal design for $\lambda = 0.14$ does have observations at 300, a

Table 3.21: Locally D-optimal designs for differing values of τ and λ when $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\rho = 1$ and $r = q = 1$

τ	λ	$\boldsymbol{t}_1$, $T_1 = 70$						$\boldsymbol{t}_2$, $T_2 = 100$					
1	0.14	1.0	4.2	283.7	289.6	295.0	300.0	1.0	2.9	79.1	84.5	89.6	95.1
10	0.14	1.0	297.0	297.8	298.6	299.3	300.0	1.0	84.9	85.8	86.7	87.5	88.4
1	0.25	1.0	3.6	180.2	187.1	193.6	200.5	41.4	45.9	50.2	54.8	294.1	300.0
10	0.25	1.0	1.5	188.7	189.7	190.8	191.9	46.3	47.2	48.0	48.8	299.2	300.0
1	0.50	1.0	2.9	80.4	85.9	91.2	96.9	17.6	20.9	24.1	27.7	291.0	300.0
10	0.50	1.0	1.4	86.2	87.2	88.1	89.0	20.9	21.6	22.3	23.1	255.9	279.3
1	1.00	34.7	38.9	43.0	47.4	293.9	300.0	7.5	9.7	11.8	14.3	207.8	289.6
10	1.00	39.2	40.0	40.8	41.7	299.2	300.0	9.3	9.9	10.4	11.0	285.0	294.1

possible explanation is that this is due to the restricted observation times rather than the impact of run-to-run error.

The impact of reducing maximum observation time: To investigate how much information would be lost if the upper bound on observation times was restricted to 200 minutes, the locally optimal designs for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\rho = 1$ and $r = q = 1$, $\lambda = 0.14$ were found for $\tau = 1$ and 10. These designs have D-efficiencies of 0.88 and 0.94 respectively, when compared to the optimal design found when the observation times are restricted to ≤ 300. These efficiencies indicate that, while the information loss from not allowing observations to be taken after 200 minutes is not massive, there is still an appreciable loss in accuracy of parameter estimation.

Table 3.22: Locally D-optimal designs for differing values of τ and λ when $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\rho = 1$, $r = 0.125$ and $q = 8$

τ	λ	$\boldsymbol{t}_1$, $T_1 = 70$						$\boldsymbol{t}_2$, $T_2 = 100$					
1	0.14	279.1	284.1	288.5	292.6	296.4	300.0	77.0	81.3	85.2	89.1	93.0	97.4
10	0.14	296.6	297.4	298.1	298.8	299.4	300.0	84.7	85.5	86.3	87.1	87.8	88.7
1	0.25	177.0	182.7	187.9	193.1	198.3	204.1	40.2	43.6	46.6	49.6	52.7	56.3
10	0.25	187.8	188.7	189.7	190.6	191.5	192.5	45.9	46.6	47.3	48.0	48.6	49.4
1	0.50	77.7	82.2	86.2	90.2	94.3	98.9	17.0	19.4	21.5	23.6	25.8	28.4
10	0.50	85.7	86.5	87.3	88.1	88.9	89.7	20.6	21.2	21.7	22.3	22.8	23.4
1	1.00	33.6	36.8	39.7	42.5	45.4	48.8	7.3	8.7	10.0	11.3	12.6	14.4
10	1.00	38.8	39.5	40.1	40.8	41.4	42.1	9.1	9.5	10.0	10.4	10.8	11.3

Table 3.22 gives the locally D-optimal designs when $E_a = 4.91 \times 10^4$, $k_r =$

4.59×10^{-4}, $\rho = 1$, $r = 0.125$ and $q = 8$. When the variance component for run-to-run error, σ_r^2 is negligible compared to the other variance components, σ_s^2 and σ_m^2, the optimal design is similar to that obtained when only serial correlation is present. The designs in this case require observations later in time as λ decreases, as seen in Section 2.6.

Table 3.23: Locally D-optimal designs for differing values of τ and λ when $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\rho = 1$, $r = 8$ and $q = 0.125$

τ	λ	$\boldsymbol{t}_1$, $T_1 = 70$						$\boldsymbol{t}_2$, $T_2 = 100$					
1	0.14	1.0	4.2	8.0	287.3	294.0	300.0	1.0	3.0	5.6	80.9	87.0	93.2
10	0.14	1.0	1.6	2.2	298.3	299.2	300.0	1.0	1.4	1.9	79.9	86.1	98.8
1	0.25	1.0	3.6	6.9	182.6	190.3	198.0	42.6	47.9	53.3	288.0	294.3	300.0
10	0.25	1.0	1.6	2.1	179.5	181.9	197.0	46.6	49.7	50.4	290.3	298.7	300.0
1	0.50	1.0	3.0	5.5	81.9	88.1	94.4	18.3	22.5	26.7	253.4	276.9	296.1
10	0.50	1.0	1.4	1.9	86.5	87.6	88.7	21.2	22.0	22.8	229.7	251.5	295.4
1	1.00	35.8	40.9	46.0	287.6	294.1	300.0	7.8	10.7	13.8	266.2	281.2	296.3
10	1.00	39.5	40.4	41.3	298.3	299.2	300.0	9.5	10.2	10.8	280.4	282.2	300.0

Table 3.23 gives the D-optimal designs obtained when $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\rho = 1$, $r = 8$ and $q = 0.125$. These results are similar to those in Table 3.21 and indicate that increasing r does not have a large influence on the effect of λ on the optimal design.

Both the variance component for run-to-run error and the magnitude of λ influence whether or not observations are made at the end points of the interval for time. As found in Section 2.6, the optimal observation times increase as λ decreases.

The D-efficiencies of each design compared to the optimal design for differing assumed values for λ, r and q are plotted in Figures 3.3, 3.4 and 3.5. The results for each set of values for r and q are very similar. Incorrectly specifying the value of λ has a large impact on the D-efficiency, more than the impact of incorrectly specifying τ as 1 or 10. The assumption that $\lambda = 0.5$ leads to a design which is the most robust to mis-specification of λ, but the design still has very poor efficiency of between 0.1 and 0.3 (depending on the value of r and q) when $\lambda = 0.14$.

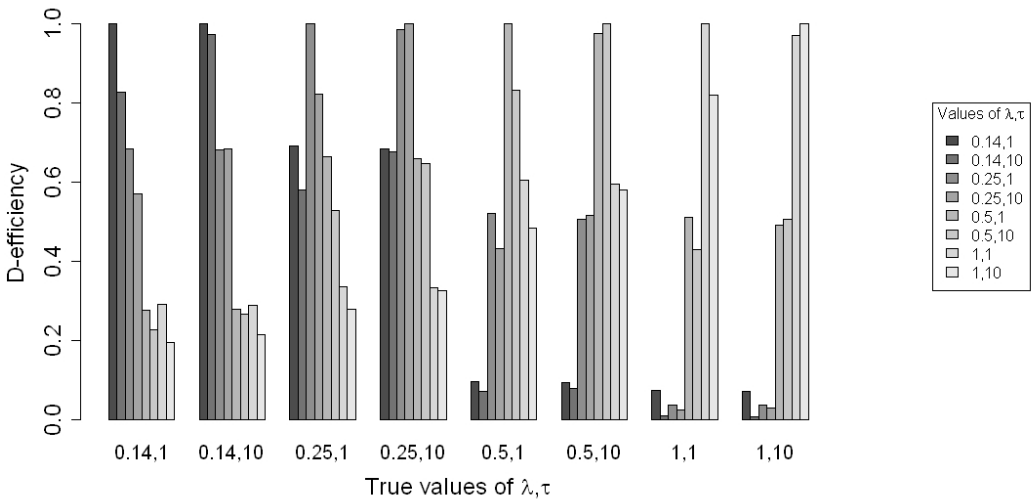


Figure 3.3: D-efficiency of designs for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\tau = 1, 10$, $\rho = 1$, $r = q = 1$ and $\lambda = 0.14$, 0.25, 0.5, 1

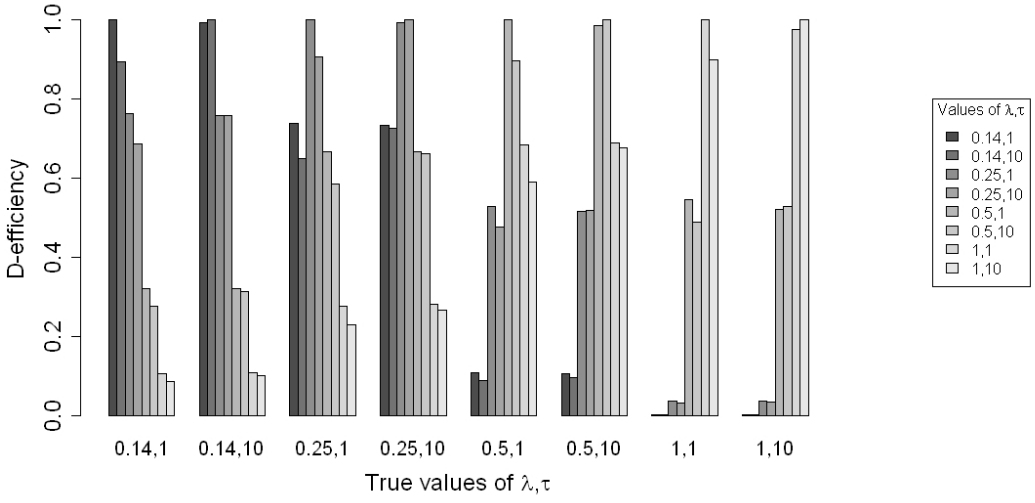


Figure 3.4: D-efficiency of designs for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\tau = 1, 10$, $\rho = 1$, $r = 0.125$, $q = 8$ and λ=0.14, 0.25, 0.5, 1

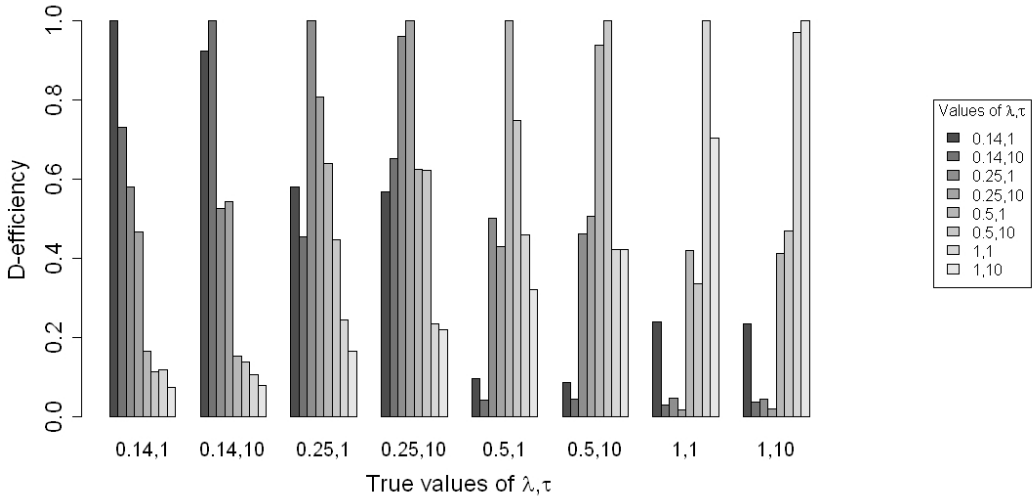


Figure 3.5: D-efficiency of designs for $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$, $\tau = 1, 10$, $\rho = 1$, $r = 8$, $q = 0.125$ and λ=0.14, 0.25, 0.5, 1

3.9.2 Bayesian D-optimal designs when correlation and heteroscedasticity are believed to be present

In this subsection we find designs robust to the mean, correlation and transformation parameters by finding Bayesian D-optimal designs. Bayesian D-optimal designs are found for three different priors on λ to explore the impact of different prior assumptions on the transformation parameter, and the Bayesian D-efficiencies of each design obtained are compared.

We adopt the following prior distributions which were defined in Subsection 3.8.1 (i)-(iii): π_α for E_a, k_r, π_b for τ, ρ and π_4 for $\log(r)$, $\log(q)$.

We find designs for each choice of prior distribution for λ as given in Section 2.6.2:

π_{λ_1}: $\lambda \sim \text{Beta}(4, 1)$, π_{λ_2}: $\lambda \sim U(0, 1)$ and π_{λ_3}: $\lambda \sim \text{Beta}(2, 12.8571)$ with corresponding mean values 0.8, 0.5 and 0.14.

The prior distribution $\pi_{\alpha b4}$ was chosen as it was shown to provide designs with a high average Bayesian D-efficiency in Section 3.8. The respective prior distributions for λ were chosen to reflect a range of scenarios as described in Subsection 2.6.2: a prior distribution with higher probability for values of λ near to 1, a uniform

distribution and a prior distribution with mean 0.14, a value found empirically by Lischer (1999).

The designs are given in Table 3.24.

Table 3.24: Bayesian D-optimal designs for prior distributions $\pi_{\alpha b4}$ and π_{λ_1}, π_{λ_2}, π_{λ_3}

Prior distribution for λ (mean)	$\boldsymbol{t}_1$, $T_1 = 70$						$\boldsymbol{t}_2$, $T_2 = 100$					
$\pi_{\lambda_1}(0.8)$	17.5	35.2	56.1	83.6	284.8	300.0	5.8	10.0	16.0	28.3	90.5	300.0
$\pi_{\lambda_2}(0.5)$	1.0	31.0	55.8	104.9	288.1	300.0	7.4	15.1	27.4	69.2	267.7	300.0
$\pi_{\lambda_3}(0.14)$	1.0	42.5	276.8	287.8	295.1	300.0	1.0	31.1	63.1	110.7	286.9	300.0

The results indicate that the nature of a prior distribution for λ has a large impact on the optimal observation times. Generally, a smaller value for the mean of the prior distribution for λ leads to an optimal design which has larger observation times. An exception is the first observation time for $T = 70$, which has value 1.

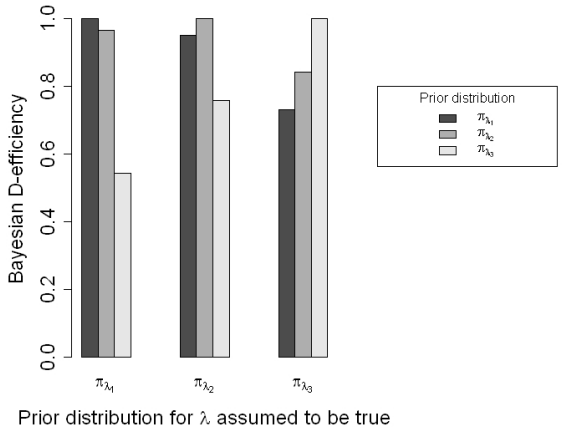


Figure 3.6: Bayesian D-efficiencies of designs for the prior distributions $\pi_{\alpha b4}$ and π_{λ_1}, π_{λ_2} and π_{λ_3} when $m = 6$

The Bayesian D-efficiencies of each design in Table 3.24 was calculated assuming that each prior distribution was true and the results are shown in Figure 3.6. We see that π_{λ_2} is the best prior distribution to adopt for a design to be robust to the three choices of prior distributions for λ. The prior distribution π_{λ_3}, which has a mean of $\lambda = 0.14$, produces an optimal design that is not very robust when the true value of λ is larger than 0.14.

3.10 Conclusion

In this chapter both local and Bayesian D-optimal designs have been investigated for the non-linear dynamic model (1.8), when correlation between observations on the same run may be present. The main finding is that, for the examples studied, Bayesian designs are preferred to locally D-optimal designs.

Other findings are:

- Stronger serial correlation between observations leads to optimal designs with observations spread further apart in time. Hence correctly specifying the correlation strength is more important than specifying the form of the correlation function when considering the D-efficiency of the designs obtained
- When the variance component for run-to-run error σ_r^2, becomes large relative to σ_s^2 and σ_m^2, optimal designs have observation times at the start and end of the allowed time interval
- The size of the transformation parameter, λ, influences whether or not an optimal design has observation times at extreme values, and, as λ increases, the optimal observation times decrease.

Bayesian D-optimal designs that are robust to the values of parameters in the mean and correlation functions in the model were also investigated.

Recommendations from the studies in this chapter are to use Bayesian D-optimal designs found using the following prior distributions:

- The prior distributions on the mean parameters should be informed, as far as possible, by prior experimentation or scientific knowledge
- For the exponential correlation function (3.1), a gamma prior distribution should be chosen for τ which has probability concentrated on lower τ values, and a uniform prior distribution between 0 and 2 chosen for ρ
- Log uniform prior distributions should be used for q and r, the ratios of the variance components for measurement error and run-to-run error with respect to the serial correlation

- When heteroscedasticity is believed likely to be present a uniform prior distribution should be adopted for λ

In this chapter, assumptions were made on the form of the correlation function including that, for the example studied, the serial correlation may be described by a continuous function of time, and that run-to-run error is best incorporated into the statistical model as an additive random effect. Alternatives might be to model serial correlation via an auto-regressive correlation function, and to include run-to-run error in the statistical model via random effects on the mean parameters, as mentioned in Section 3.2.

For the statistical models explored, the work in this chapter provides designs enabling accurate mean parameter estimation regardless of assumptions made on the mean and correlation parameters and the transformation parameter. These results will not necessarily apply to models which have different expected responses or correlation functions.

Chapter 4

Investigation of design properties via a simulation study

4.1 Introduction

In this chapter the findings from two large simulation studies are presented. In the main study, data from experiments using 15 different designs are simulated and the data analysed for each of 24 scenarios. It is initially assumed for the simulation that heteroscedasticity is not present in the errors, and $\lambda = 1$ in model (1.8). This assumption is then relaxed in Section 4.10, where a second, smaller study is conducted, where data for experiments using 7 different designs are simulated for each of 15 scenarios, where $0 < \lambda \leq 1$.

Simulation studies can be used to assess design performance in ways that efficiency measures cannot. In particular, when the model is non-linear, results about the optimality of D-optimal designs for maximising parameter accuracy are only asymptotic, and will not necessarily hold for the small experiments employed in practice.

The following questions are of interest.

- How accurate are parameter estimates obtained using locally and Bayesian D-optimal designs compared to those obtained using ad hoc designs similar to those used in industry? The comparisons may be made using:

  - likely scenarios defined through prior information
  - scenarios which examine an optimal design's performance for conditions that differ from those assumed at the design stage

- What is the impact of analysing the data assuming correlation is or is not present?
- What is the impact of different magnitudes of the total variance, σ^2?
- How accurate are the estimators of σ^2?
- How biased are the parameter estimators ?
- What is the impact of analysing the data assuming error heteroscedasticity is or is not present?

Many of these questions could not be answered without simulation, as real experiments would not provide information on the bias (the true parameter values are unknown) and analytic results are difficult or impossible to obtain for non-linear models.

Section 4.2 reviews papers in which simulation studies have been used to assess designs for similar models to those in this thesis. Section 4.3 introduces and motivates the designs being considered and the simulation scenarios selected. Section 4.4 discusses the methodology used to generate and analyse the data in the simulation study. Section 4.5 considers the impact of analysing the data assuming correlation is or is not present. In Section 4.6, the impact of varying the values of σ^2 on the relative accuracy of the parameter estimates is investigated. The accuracy of estimation of σ^2 is investigated in Section 4.7 and the size of the bias of parameter estimates is explored in Section 4.8. In Section 4.9, the performance of the designs for parameter estimation for each of the scenarios is compared, and the results are related to the D-efficiency of the designs, where possible. In Section 4.10, a smaller simulation study is conducted to assess the impact of heteroscedasticity on estimation. In the final section, conclusions are drawn on the findings from the chapter.

4.2 Literature review

Simulation studies are not typically considered in the literature when assessing the performance of designs under a non-linear model. There are some examples, which include Atkinson et al. (1993), who considered a non-linear dynamic model, a compartmental model with three parameters, where an experiment consists of a run where the times of observation are chosen. They assessed the asymptotic efficiency results for the Bayesian c-optimal designs found, where c-optimality is a criterion which finds designs which minimise the variance of a specified combination of the mean parameters. They simulated observations for three designs: an ad hoc design, and two Bayesian c-optimal designs found for uniform prior distributions of differing widths. For the study, the true parameter values of the two parameters affecting optimal design were drawn from a grid of likely values. They found that the Bayesian c-optimal designs provided the most accurate parameter estimates, with the more conservative Bayesian c-optimal design performing better when the true parameter values were on the edges of the anticipated interval for the parameter values.

Asprey and Macchietto (2002) compared a locally D-optimal design, a Bayesian D-optimal design and a maximin D-optimal design for a similar model to the Michaelis-Menten model (as described in Section 2.2), but with four parameters, by running a small simulation study with only one scenario, in which the true parameter values were fixed. They concluded that the maximin D-optimal design provided the most accurate parameter estimates. Only simulating for one set of parameter values limits the usefulness of this study, because different parameter values would give different results. Dette and Biedermann (2003) assessed maximin D-optimal designs for the Michaelis-Menten model by comparing them to locally D-optimal designs and ad hoc designs. They varied the parameter k_m between two values, where one was outside the region used for finding the maximin D-optimal designs. They found that the maximin D-optimal design gave the smallest MSE for both values of the parameters. Exploring a wider range of parameter values might have given differing results, however.

Retout et al. (2007), investigated designs for a non-linear mixed model from pharmacokinetics, with four parameters, where designs which maximised an asymptotic measure of power were selected. They assessed whether the asymptotic estimate of power would be accurate in practice via a simulation study, in which error was simulated for two different sets of assumptions on the parameter values, chosen to reflect likely values in practice. They demonstrated that the theoretical results were similar to those found by simulation.

Also in the pharmacokinetic literature, Wang et al. (2012) found locally and Bayesian D-optimal designs for a non-linear mixed model with three parameters. They simulated to assess how accurate asymptotic results would be at estimating the variance for each parameter when the parameters designed for were assumed to be correct, and also simulated for three sets of parameter values which were different to those designed for. They demonstrated that the asymptotic results were reasonably accurate, but also that the Bayesian and locally D-optimal designs gave quite biased estimates when the parameters were misspecified, with the locally D-optimal designs being the most sensitive to parameter misspecification. Such bias was not observed in our simulation study, other than when correlation was ignored during analysis.

Simulation is usually used to confirm theoretical results and to develop results that are difficult to investigate using efficiency measures. In the examples found in the literature, the studies were typically small, and only considered a few scenarios. The goal of the work in this chapter is to conduct a much more thorough investigation of design performance.

4.3 Designs and scenarios used in the study

In this section, the designs (labelled $d_1, \ldots, d_{15}$) and scenarios ($s_1, \ldots, s_{24}$) used in the first simulation study are motivated and defined. It was assumed throughout that the practitioners have the resource to conduct four process runs, each with $m = 6$ observations.

The locally optimal designs chosen for the study are defined and labelled below. They were the designs whose performance was found to be the most robust to misspecification of the correlation parameters, τ, ρ, r and q. Design d_1 is most robust when run-to-run error and measurement error were not present. Designs d_2 and d_3 were included to try to investigate the result in Section 3.4.1 that correlation structure has less impact on the choice of an optimal design than correlation strength. Designs d_4 and d_5 were the most robust locally optimal designs when q is zero or non-zero, respectively.

The Bayesian D-optimal designs chosen for the study were those found in Section 3.8 to be most robust when the prior distribution for the parameters was misspecified. Two ad hoc designs, $d_{14} - d_{15}$, were included to assess how well the optimal designs would perform compared to designs used in practice by industry (given in Table 4.1).

The model assumptions made that led to each of the designs in the study are given below. When a corresponding approximate design has already been presented and labelled in Chapter 3, the label is given in brackets after the design description. Note that the designs labelled here are distinct from designs found in Chapter 2, despite there being some labels in common.

a) Locally optimal designs with $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$:

   d_1: Exponential correlation structure (Equation (3.1)), $\tau = 1$ and $\rho = 1$, $r = 0$, $q = 0$ (ξ_1)

   d_2: Linear correlation structure (as in (3.2)) with equal correlation between observations at distance 1 to the exponential correlation structure used when finding d_1

   d_3: Quadratic correlation structure (as in (3.3)) with equal correlation between observations at distance 1 to the exponential correlation structure used when finding d_1, for the given values of ρ and τ

   d_4: $\tau = 1$ and $\rho = 1$, $r = 1$, $q = 0$ (ξ_{20})

d_5: $\tau = 1$ and $\rho = 1$, $r = 1$, $q = 1$ (ξ_{22})

b) Bayesian D-optimal designs (assuming an exponential correlation structure)

d_6: $E_a \sim$N(4.91×10^4, 4000^2), $k_r \sim$ G(m_k^2/v_k,v_k/m_k), mean $m_k = 4.59 \times 10^{-4}$, variance $v_k = (4 \times 10^{-4})^2$, $\tau = \rho = r = q = 1$ ($\xi_{\alpha a1}$)

d_7: $E_a \sim$N(4.91×10^4, 4000^2), $k_r \sim$ G(m_k^2/v_k,v_k/m_k), mean $m_k = 4.59 \times 10^{-4}$, variance $v_k = (4 \times 10^{-4})^2$, $\tau = \rho = 1$, log(r) and log(q) ∼U(-2.5,2.5) ($\xi_{\alpha a4}$)

d_8: $E_a \sim$N(4.91×10^4, 4000^2), $k_r \sim$ G(m_k^2/v_k,v_k/m_k), mean $m_k = 4.59 \times 10^{-4}$, variance $v_k = (4 \times 10^{-4})^2$, $\tau \sim$ G(1,2), $\rho \sim$ U(0,2), $r = q = 1$ ($\xi_{\alpha b1}$)

d_9: $E_a \sim$N(4.91×10^4, 4000^2), $k_r \sim$ G(m_k^2/v_k,v_k/m_k), mean $m_k = 4.59 \times 10^{-4}$, variance $v_k = (4 \times 10^{-4})^2$, $\tau \sim$ G(1,2), $\rho \sim$U(0,2), log(r) and log(q) ∼U(-2.5,2.5) ($\xi_{\alpha b4}$)

d_{10}: $E_a \sim$ U(2.5×10^4,6.5×10^4), $k_r \sim$ U(1×10^{-4},1×10^{-3}), $\tau = \rho = r = q = 1$ ($\xi_{\beta a1}$)

d_{11}: $E_a \sim$ U(2.5×10^4,6.5×10^4), $k_r \sim$ U(1×10^{-4},1×10^{-3}), $\tau = \rho = 1$, log(r) and log(q) $\sim$ U(-2.5,2.5) ($\xi_{\beta a4}$)

d_{12}: $E_a \sim$ U(2.5×10^4,6.5×10^4), $k_r \sim$ U(1×10^{-4},1×10^{-3}), $\tau \sim$G(1,2), $\rho \sim$ U(0,2), $r = q = 1$ ($\xi_{\beta b1}$)

d_{13}: $E_a \sim$ U(2.5×10^4,6.5×10^4), $k_r \sim$ U(1×10^{-4},1×10^{-3}), $\tau \sim$G(1,2), $\rho \sim$ U(0,2), log(r) and log(q) $\sim$ U(-2.5,2.5) ($\xi_{\beta b4}$)

c) Ad hoc designs

d_{14}: Uniformly spaced design

d_{15}: Design used by chemists in the past

These designs are presented in Table 4.1.

Table 4.1: Designs used in the simulation study

Design	T_1	T_2	$\boldsymbol{t}_1 = \boldsymbol{t}_2$						T_3	T_4	$\boldsymbol{t}_3 = \boldsymbol{t}_4$					
d_1	70	70	30.2	35.2	39.8	44.4	49.1	54.4	100	100	5.6	8.3	10.8	13.4	16.3	19.6
d_2	70	70	38.0	39.9	41.7	43.3	44.9	46.6	100	100	7.0	8.7	10.3	11.9	13.5	15.1
d_3	70	70	29.7	35.0	39.8	44.6	49.7	55.7	100	100	5.8	8.3	10.6	13.0	15.6	18.7
d_4	70	70	1.0	34.2	38.9	43.3	48.1	200.0	100	100	7.7	10.8	14.1	175.2	187.3	200.0
d_5	70	70	1.0	34.7	39.0	43.0	47.4	200.0	100	100	7.5	9.7	11.8	14.3	187.0	200.0
d_6	70	70	1.0	25.6	32.4	39.5	192.7	200.0	100	100	4.8	8.1	12.1	22.8	66.8	200.0
d_7	70	70	1.0	25.7	32.5	39.5	189.8	200.0	100	100	4.6	8.0	12.1	18.3	57.2	200.0
d_8	70	70	1.0	22.6	33.7	47.7	184.5	200.0	100	100	4.8	8.1	12.9	23.9	90.3	200.0
d_9	70	70	1.0	21.2	31.2	44.1	94.9	200.0	100	100	4.7	8.0	12.4	21.1	77.3	200.0
d_{10}	70	70	1.0	25.8	30.2	34.5	39.2	200.0	100	100	5.9	8.6	11.6	15.5	192.8	200.0
d_{11}	70	70	1.0	25.6	30.2	34.6	39.5	200.0	100	100	5.6	8.2	11.0	14.4	55.9	200.0
d_{12}	70	70	1.0	24.6	29.7	34.8	41.3	200.0	100	100	5.9	8.1	10.6	14.2	68.9	200.0
d_{13}	70	70	1.0	24.3	29.6	35.1	42.2	200.0	100	100	5.8	8.0	10.4	13.8	50.2	200.0
d_{14}	70	80	1.0	40.8	80.6	120.4	160.2	200.0	90	100	1.0	40.8	80.6	120.4	160.2	200.0
d_{15}	70	80	1.0	4.0	8.0	16.0	100.0	200.0	90	100	1.0	4.0	8.0	16.0	100.0	200.0

The 24 data generation scenarios arise as follows. Scenarios s_1-s_{13} are the assumptions in (a) and (b) above, under which designs d_1-d_{13} were obtained. Scenarios s_{14}-s_{21} have the true mean parameter values chosen in the extremes of the prior distributions assumed in finding designs $d_6 - d_{13}$. In addition, scenarios $s_{18} - s_{21}$ have run-to-run error as the dominating source of error. Scenarios s_{22}-s_{24} correspond to three different scenarios for $\boldsymbol{\theta}_m$ and have uncorrelated errors. The scenarios s_{14}-s_{24} are given in Table 4.2.

Table 4.2: Scenarios s_{14}-s_{24}

Scenario	$E_a \times 10^{-4}$	$k_r \times 10^4$	τ	ρ	r	q
s_{14}	2	1	0.1	1	1	1
s_{15}	7	1	0.1	1	1	1
s_{16}	2	10	0.1	1	1	1
s_{17}	7	10	0.1	1	1	1
s_{18}	2	1	0.1	1	100	1
s_{19}	7	1	0.1	1	100	1
s_{20}	2	10	0.1	1	100	1
s_{21}	7	10	0.1	1	100	1
	Scenarios with no correlation					
s_{22}	$E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$					
s_{23}	$E_a \sim$N(4.91×10^4,4000^2), $k_r \sim$ G(m_k^2/v_k,v_k/m_k)					
	mean $m_k = 4.59 \times 10^{-4}$ variance $v_k = (4 \times 10^{-4})^2$					
s_{24}	$E_a \sim$U(2.5×10^4,6.5×10^4), $k_r \sim$U(1×10^{-4},1×10^{-3})					

4.4 Methodology

This section describe the methodology used to conduct the simulation study.

When analysing non-linear models, analytic solutions for the least square estimates of the parameters are not available. Instead numerical approximation must be used. One method is Gauss-Newton estimation (page 40, Bates and Watts, 1988), which uses a first order Taylor approximation of the expected response and then iterates guesses of the parameters to minimise the residual sum of squares until the change is sufficiently small that the algorithm is considered to have converged. This algorithm is available in `R` as `nls`.

Gauss-Newton estimation requires that the errors are independent, which is not the case for s_1-s_{21}, where correlation is present via a specified structure. If the form of the correlation matrix $Q(\boldsymbol{\theta}_c)$ is assumed to be known, then Gauss-Newton approximation can be applied by transforming the simulated data by pre-multiplication of both the response and expected response (Equation (1.7), Section 1.3) by the transpose of $Q(\boldsymbol{\theta}_c)^{-1/2}$.

Non-linear estimation requires a starting value, which was chosen to be the simulated parameter values. In a real experiment the "true" parameter values are not known, but in the analysis of the data a wide variety of starting parameter values can be tried to find the best fitting model.

In the study, observations were simulated for an experiment under each scenario s_i (as given in (a), (b), and Table 4.2) when design d_k (Table 4.1), $k = 1, \ldots, 15$, is used. The following steps were used to generate each observation using model (1.7) (Section 1.3):

1. The parameter values were chosen from the given distribution for s_i

2. The expected response was calculated using (1.5)

3. The value of the random error was obtained by a random draw from a MVN(0,$\sigma^2 Q(\boldsymbol{\theta}_c)$) distribution, where $Q(\boldsymbol{\theta}_c)$ is defined in Section 1.3

4. The simulated observation was obtained as the sum of the values from steps

2 and 3

5. For a specified value of σ, $S = 100,000$ simulations of the experiment were made for each design and scenario

The simulations were repeated three times for $\sigma = 2{\times}10^{-4}$, $2{\times}10^{-5}$ and $2{\times}10^{-6}$, magnitudes similar to those observed experimentally.

The observations were analysed using `nls`. When correlation was present in the model used for data generation, the data were analysed in two ways: assuming correlation was known, and assuming correlation was not present. Each approach gave an estimate, $\hat{\boldsymbol{\theta}}_m$, of the mean parameters, E_a and k_r. For each run, j, of the simulation the squared difference between the "true" parameter $\boldsymbol{\theta}_m$ and the estimate $\hat{\boldsymbol{\theta}}_m$ was calculated component-wise. This value was taken relative to the square of the true parameter value, as below:

$$\left(\frac{(\theta_{m_i}^{(j)} - \hat{\theta}_{m_i}^{(j)})^2}{\left(\theta_{m_i}^{(j)}\right)^2} \right) \quad i = 1, 2,$$

where $\theta_{m_i}^{(j)}$ is the true value of the ith parameter in the jth simulation, and $\hat{\theta}_{m_i}^{(j)}$ is the estimate of the ith parameter obtained from the jth simulation.

In practice, it was discovered that parameter estimation for $s_6 - s_9$ where parameter values $\boldsymbol{\theta}_m$ were drawn from a normal/gamma distribution could be very inaccurate for some simulated parameter values. This resulted in extremely large squared error, which dominated all other squared errors obtained for that design and scenario. This is illustrated, for s_7 and $\sigma = 2 \times 10^{-4}$, by comparing the ratio of the largest squared error obtained for E_a for each design, given by:

$$\max_{j \in (1,\ldots,S)} \left(\frac{(E_a^{(j)} - \hat{E}_a^{(j)})^2}{(E_a^{(j)})^2} \right),$$

with the total squared error observed for E_a for each design, i.e.

$$\sum_{j=1}^{S} \left(\frac{(E_a^{(j)} - \hat{E}_a^{(j)})^2}{(E_a^{(j)})^2} \right),$$

where the data were analysed assuming the correlation was known. The results for each design are given in Table 4.3.

Table 4.3: The ratio of maximum error/total error for E_a

d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
0.11	0.22	0.36	0.33	0.24	0.31	0.16	0.98	0.40	0.31	0.28	0.47	0.47	0.33	0.89

The results indicate that the mean of the relative squared errors does not provide useful information, as the mean will be dominated by this maximum error, which can be as large as 0.98 for d_8. This effect may be due to certain values of E_a and k_r causing reactions to go far too slowly or quickly for any information to be gathered. In practice, such reactions are very unlikely to occur.

This problem was overcome by using the median error instead of the mean. To make the results comparable, we used the median throughout. For scenarios where these large errors do not occur, the mean and median should be similar.

We define the median relative squared error (DSE) by:

$$DSE(\hat{\theta}_{m_i}) = \text{median}_{j \in S} \left(\frac{(\theta_{m_i}^{(j)} - \hat{\theta}_{m_i}^{(j)})^2}{\left(\theta_{m_i}^{(j)}\right)^2} \right). \tag{4.1}$$

4.5 The effect of analysing with and without considering correlation

In this section, we study the effect of incorporating correlation into the data analysis, using designs $d_1 - d_{15}$ and scenarios $s_1 - s_{21}$. Most practitioners would not include correlation in the statistical model for analysis, which could lead to a loss in the accuracy of parameter estimates. In the study the two mean parameters were each estimated in two ways: when correlation is present (and known) in the model, leading to an estimate $\hat{\theta}_{m_i}$ and when there is no correlation in the model, giving an estimator $\tilde{\theta}_{m_i}$. Results were obtained for all three values of σ^2 studied.

The median relative squared error, DSE (Equation (4.1)), was calculated for each of $\hat{\theta}_{m_i}$ and $\tilde{\theta}_{m_i}$ for $i = 1, 2$. The ratios corresponding to each of the two

parameters were found to have similar values. Hence the average of these two DSE ratios was used to summarise the results, i.e.:

$$\frac{1}{2}\left(\frac{DSE(\hat{E}_a)}{DSE(\tilde{E}_a)}+\frac{DSE(\hat{k}_r)}{DSE(\tilde{k}_r)}\right). \tag{4.2}$$

Here a value less than one indicates that the DSE averaged over both parameters was lower when a model incorporating correlation was used in the analysis. The results were similar for all values of σ^2 in the study. Therefore, only results for $\sigma = 2 \times 10^{-6}$ are presented in Table 4.4.

For scenarios s_1 and s_3, most designs have value of (4.2) close to one. For d_{15}, the uniform design, the relative value is exactly one as the observation times are spaced far enough apart for correlation to have no impact on the analysis. For scenario s_2 the value of (4.2) is 1 for every design, indicating that observation times are so far part that correlation between observations from the same run is 0 (as the correlation structure assumed is linear in s_2).

These results demonstrate that optimal designs can be found which effectively eliminate the impact of serial correlation on the analysis. The relative accuracies, as measured by Equation (4.2), of parameter estimates for $s_1 - s_3$ are very close to 1.

When run-to-run error is present, analysing data for a model with or without correlation has a large impact on the relative accuracy of the parameter estimates obtained. In particular, if the data are analysed assuming correlation is not present, then designs $d_1 - d_3$, which were found assuming that run-to-run error was not present, tend to perform better than the theoretically optimal designs in scenarios $s_4 - s_{13}$.

This is demonstrated in Table 4.5, which presents the DSE averaged over E_a and k_r ($ADSE$) relative to the minimum $ADSE$ observed for each scenario, for scenarios $s_4 - s_5$. The values obtained for analysis with and without correlation in the model, for $\sigma = 2 \times 10^{-6}$, are presented. When correlation is not included in the model during analysis, the designs which provide the most accurate parameter

estimates, even when run-to-run error and measurement error are present, are d_2 and d_3. When the data are analysed assuming correlation is known, the relative accuracy of the parameter estimates obtained by designs d_2 and d_3 drops to just over 0.5 in scenario s_4. These results might be expected as designs $d_6 - d_{13}$ have observations taken at the end points of the allowed time interval. These observations provide little information unless run-to-run error is included in the model in the analysis, see Subsection 3.6.1.

These results suggest that if the practitioner does not intend to account for correlation in the analysis of an experiment, then they might as well not include run-to-run error in the model used to select a design. It is recommended that the practitioner does include correlation during analysis, as they can achieve much better accuracy in mean parameter estimation by doing so. As run-to-run error is not eliminated by design, it should be possible to estimate it. The tools for estimating correlation are not well established, but the methods presented in Pinheiro and Bates (2004), Chapter 7, could be applied.

The remaining results presented in this chapter are those obtained by assuming that the magnitude and form of correlation between observations on a process run is known.

Table 4.4: Value of average relative DSE, defined in Equation (4.2) for $\sigma = 2 \times 10^{-6}$ and designs $d_1 - d_{15}$

	Design														
Scenario	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
s_1	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s_2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s_3	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
s_4	0.99	1.00	0.99	0.49	0.49	0.47	0.47	0.46	0.48	0.50	0.50	0.50	0.50	0.45	0.46
s_5	1.00	1.00	1.00	0.68	0.69	0.66	0.67	0.66	0.69	0.69	0.70	0.70	0.69	0.65	0.67
s_6	0.96	0.99	0.95	0.69	0.70	0.68	0.68	0.68	0.69	0.70	0.71	0.71	0.70	0.68	0.65
s_7	0.91	0.97	0.91	0.58	0.59	0.56	0.56	0.56	0.57	0.58	0.59	0.58	0.59	0.55	0.54
s_8	0.95	0.99	0.94	0.68	0.69	0.67	0.67	0.66	0.67	0.69	0.69	0.69	0.69	0.67	0.63
s_9	0.91	0.97	0.90	0.57	0.58	0.55	0.55	0.54	0.55	0.57	0.58	0.57	0.57	0.55	0.52
s_{10}	0.96	0.99	0.95	0.69	0.70	0.67	0.67	0.66	0.69	0.70	0.71	0.69	0.71	0.68	0.65
s_{11}	0.92	0.97	0.92	0.57	0.58	0.55	0.55	0.55	0.56	0.58	0.58	0.58	0.58	0.55	0.54
s_{12}	0.95	0.98	0.95	0.68	0.68	0.66	0.65	0.65	0.67	0.68	0.68	0.68	0.68	0.65	0.63
s_{13}	0.91	0.96	0.92	0.56	0.56	0.54	0.54	0.54	0.55	0.56	0.57	0.56	0.56	0.53	0.52
s_{14}	0.87	0.91	0.86	0.73	0.71	0.72	0.70	0.70	0.69	0.70	0.67	0.67	0.67	0.75	0.67
s_{15}	0.93	0.97	0.92	0.68	0.69	0.71	0.70	0.70	0.68	0.65	0.64	0.66	0.65	0.71	0.66
s_{16}	0.88	0.96	0.88	0.74	0.76	0.71	0.71	0.68	0.68	0.75	0.74	0.74	0.73	0.66	0.61
s_{17}	0.75	0.77	0.75	0.67	0.67	0.67	0.65	0.64	0.65	0.66	0.66	0.66	0.67	0.68	0.62
s_{18}	0.12	0.33	0.11	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
s_{19}	0.14	0.46	0.13	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
s_{20}	0.08	0.28	0.08	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01
s_{21}	0.10	0.27	0.09	0.02	0.02	0.02	0.02	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.01

Table 4.5: Comparison of $ADSE$ relative to the minimum $ADSE$ obtained per scenario when, in the data analysis, correlation is assumed not present and when correlation is present and known for $\sigma = 2 \times 10^{-6}$

	Design														
Scenario	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
	Analysing assuming correlation is not present														
s_4	0.99	1.00	1.00	0.91	0.92	0.76	0.77	0.74	0.75	0.88	0.87	0.88	0.86	0.13	0.26
s_5	0.98	1.00	0.99	0.86	0.88	0.71	0.73	0.69	0.71	0.85	0.84	0.84	0.83	0.12	0.25
	Analysing assuming correlation is present and known														
s_4	0.53	0.53	0.53	1.00	0.98	0.86	0.87	0.86	0.83	0.95	0.93	0.94	0.91	0.15	0.30
s_5	0.77	0.78	0.78	0.98	1.00	0.84	0.85	0.83	0.81	0.96	0.94	0.94	0.93	0.14	0.29

4.6 Impact of σ on relative mean square error for scenarios 1-21

In this section, the degree to which DSE, defined in Equation (4.1), is affected by the value of the standard error, σ, is explored. As the volume of the asymptotic confidence ellipsoid for the mean parameters is only proportional to the standard error, it might be supposed that the DSE increases in proportion to the value of σ when the bias in the estimates is negligible. This effect will occur asymptotically, but may not be true for small experiments.

We examine ratios of DSE values defined as follows:

$$RDSE(\hat{\theta}_{m_i}(\sigma_u), \hat{\theta}_{m_i}(\sigma_v)) = \frac{DSE(\hat{\theta}_{m_i}(\sigma_u))}{DSE(\hat{\theta}_{m_i}(\sigma_v))},$$

where $(\sigma_u, \sigma_v) \in \{(2 \times 10^{-6}, 2 \times 10^{-5}), (2 \times 10^{-5}, 2 \times 10^{-4})\}$.

If the change in DSE is only proportional to the change in σ, then the ratio should be roughly constant for all designs and scenarios. To investigate if this is the case, the ratio of $RDSE(\hat{\theta}_{m_i}(\sigma_u), \hat{\theta}_{m_i}(\sigma_v))$ to $RDSE^*(\hat{\theta}_{m_i}(\sigma_u), \hat{\theta}_{m_i}(\sigma_v))$ is calculated, where $RDSE^*$ is the highest $RDSE$ observed for that scenario. This ratio was calculated for each of E_a and k_r and the values obtained were found to be very similar. Hence their average was used, given by

$$\frac{1}{2}\left(\frac{RDSE(\hat{E}_a(\sigma_u), \hat{E}_a(\sigma_v))}{RDSE^*(\hat{E}_a(\sigma_u), \hat{E}_a(\sigma_v))} + \frac{RDSE(\hat{k_r}(\sigma_u), \hat{k_r}(\sigma_v))}{RDSE^*(\hat{k_r}(\sigma_u), \hat{k_r}(\sigma_v))}\right) \tag{4.3}$$

for $(\sigma_u, \sigma_v) \in \{(2 \times 10^{-6}, 2 \times 10^{-5}), (2 \times 10^{-5}, 2 \times 10^{-4})\}$.

Tables 4.6 and 4.7 indicate that the value of (4.3) is not very sensitive to the value of σ and is close to one throughout. This supports the above supposition that the DSE for each parameter increases in proportion to σ. From now on, all simulations presented will be for $\sigma = 2 \times 10^{-6}$.

Table 4.6: Values of (4.3) when $u = 2 \times 10^{-6}$ and $v = 2 \times 10^{-5}$

	Design														
Scenario	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
s_1	1.00	0.99	0.99	0.98	0.99	0.97	0.98	0.99	0.99	0.97	0.98	0.99	0.98	0.98	0.99
s_2	0.98	0.97	0.96	0.98	0.97	0.98	0.98	0.99	0.96	0.99	0.96	0.96	0.98	0.98	0.97
s_3	0.99	0.98	0.99	0.97	0.98	0.98	0.97	0.99	0.98	0.97	0.98	0.98	1.00	0.99	0.97
s_4	0.98	0.99	0.98	0.96	0.98	1.00	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.98	0.97
s_5	1.00	0.98	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.97	0.98	0.98
s_6	0.98	0.98	0.98	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.98	0.99	0.99	0.98	0.98
s_7	0.97	0.99	0.97	0.97	0.97	0.98	0.99	0.98	0.99	0.98	0.99	0.97	0.99	0.98	0.99
s_8	0.99	0.98	0.97	0.98	1.00	0.98	1.00	0.98	0.98	0.98	0.99	0.98	0.97	0.98	0.98
s_9	0.99	0.99	0.97	0.99	1.00	0.97	0.99	0.99	0.97	0.99	1.00	0.98	0.99	0.99	0.99
s_{10}	0.97	0.99	0.97	0.98	0.98	0.98	0.97	0.98	0.99	0.98	0.98	0.97	0.97	0.98	0.96
s_{11}	0.98	0.96	0.97	0.97	0.98	0.98	0.98	0.99	0.96	0.98	0.98	0.98	0.99	0.98	0.98
s_{12}	0.98	0.98	0.99	0.99	0.97	0.99	0.99	0.98	0.98	0.97	0.99	0.98	0.98	0.96	0.96
s_{13}	0.99	0.98	0.98	0.99	0.98	0.98	0.99	0.99	0.99	0.98	0.97	0.98	0.99	0.99	0.99
s_{14}	0.97	0.99	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.99	0.98	0.99	0.98	0.97	0.98
s_{15}	0.98	0.98	0.96	0.98	0.98	0.98	0.98	0.98	0.98	0.99	1.00	0.98	0.99	0.99	0.99
s_{16}	0.99	0.98	1.00	0.98	0.99	0.98	1.00	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.97
s_{17}	0.97	0.99	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.99
s_{18}	0.97	1.00	0.98	0.98	0.98	0.99	0.98	0.98	0.99	0.98	0.98	0.98	0.99	0.97	0.98
s_{19}	0.98	1.00	0.96	0.96	0.97	0.97	0.98	0.96	0.97	0.98	0.98	0.98	0.97	0.98	0.97
s_{20}	1.00	0.98	0.98	0.98	0.99	0.97	0.99	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.98
s_{21}	0.98	0.99	0.99	0.99	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.98	0.99	1.00	0.99

Table 4.7: Values for (4.3) when $u = 2 \times 10^{-5}$ and $v = 2 \times 10^{-4}$

	Design														
Scenario	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
s_1	0.97	0.98	0.99	0.99	0.98	0.97	0.99	0.97	0.98	0.99	0.99	0.98	0.99	0.98	0.97
s_2	0.98	0.98	0.99	0.98	0.99	0.97	0.98	0.98	0.99	0.99	0.99	0.99	0.98	0.97	0.99
s_3	0.97	0.97	0.97	0.98	0.99	0.98	0.99	0.96	0.97	0.98	0.98	0.99	0.96	0.96	0.98
s_4	0.98	0.98	0.97	0.99	0.97	0.97	0.99	0.98	0.97	0.97	0.97	0.97	0.97	0.96	0.99
s_5	0.97	0.97	0.96	0.98	0.98	0.97	0.97	0.97	0.97	0.98	0.97	0.99	1.00	0.98	0.99
s_6	0.98	0.99	0.98	0.98	0.96	0.99	0.99	0.99	0.98	0.99	1.00	0.98	0.98	0.98	0.99
s_7	0.98	0.96	0.99	0.97	0.98	0.98	0.97	0.97	0.98	0.97	0.99	0.98	0.97	0.98	0.96
s_8	1.00	0.99	0.99	0.97	0.99	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99
s_9	0.98	0.98	0.99	0.99	0.98	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.98	0.97	0.98
s_{10}	1.00	0.98	0.99	0.99	0.97	0.97	0.99	0.98	0.97	0.98	0.98	0.99	1.00	0.98	1.00
s_{11}	0.98	0.99	0.99	1.00	0.98	0.97	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.98	0.98
s_{12}	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.99	0.99	1.00	0.97	0.99	1.00	0.99	1.00
s_{13}	0.98	0.98	0.99	0.98	0.99	0.99	0.98	0.98	0.99	0.99	0.99	0.98	0.99	0.98	0.97
s_{14}	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.96	0.99	0.98	0.99	0.98	0.99
s_{15}	0.98	0.99	0.98	0.99	0.98	0.98	0.98	0.99	0.99	0.98	0.97	0.99	0.97	0.96	0.97
s_{16}	0.98	0.99	0.99	0.99	0.99	1.00	0.98	0.99	0.99	0.99	0.99	0.98	1.00	0.98	0.99
s_{17}	0.99	0.98	0.98	0.97	0.99	0.98	0.98	0.98	0.99	0.97	0.98	0.98	0.99	0.98	0.97
s_{18}	0.98	0.97	0.99	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	1.00	0.99
s_{19}	0.99	0.97	0.99	0.98	0.99	0.98	0.98	0.99	0.98	0.98	0.98	1.00	0.98	0.99	0.98
s_{20}	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.99	0.97	0.99	0.98	0.99	0.98	0.97	0.99
s_{21}	0.97	0.98	0.99	0.99	0.98	1.00	0.98	0.99	0.99	0.98	0.98	1.00	0.99	0.99	1.00

4.7 Accuracy of the estimator of σ^2

The accuracy and sensitivity of the estimates of σ^2 to the choice of experimental design are explored in this section. The estimates of σ^2, $\hat{\sigma}^2$, are obtained, given a data set $\boldsymbol{y}$ with length Nm and parameter estimates $\hat{\boldsymbol{\theta}}_m$ of length p_m, by:

$$\hat{\sigma}^2 = \frac{(\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{t}, \boldsymbol{T}, \hat{\boldsymbol{\theta}}_m))'(\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{t}, \boldsymbol{T}, \hat{\boldsymbol{\theta}}_m))}{Nm - p_m}$$

as given by page 6, Bates and Watts (1988). Note that this assumes that the data has already been transformed by pre-multiplication by the transpose of $Q(\boldsymbol{\theta}_c)^{-1/2}$.

Table 4.8 gives the DSE for the estimate $\hat{\sigma}^2$, relative to the minimum DSE obtained for each scenario, when $\sigma = 2 \times 10^{-6}$ and the statistical model (1.8) fitted assuming the correlation is known.

The designs perform very similarly in estimating the variance, with no design being preferred over another. In particular the D-optimal designs achieve similar results to the ad hoc designs, d_{14} and d_{15}. This is perhaps not surprising as the optimal designs found were intended for parameter estimation, rather than estimation of the variance or correlation parameters.

The error associated with estimating the variance is much larger than that for estimating the mean parameters, which is what might be expected. The value is typically about 0.04 for all values of σ, with the squared error for parameter estimation being of the order of 10^{-6} for $\sigma = 2 \times 10^{-4}$, and 10^{-10} for $\sigma = 2 \times 10^{-6}$.

Table 4.8: DSE relative to the minimum DSE for each scenario, for the variance estimates with $\sigma = 2 \times 10^{-6}$ and the data analysed assuming correlation is known

	Design														
Scenario	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
s_1	0.96	0.98	0.99	0.99	0.98	1.00	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.99	0.98
s_2	0.99	0.99	0.99	1.00	1.00	1.00	0.98	0.98	0.99	0.98	0.98	0.99	0.99	0.99	0.98
s_3	0.99	0.97	1.00	0.97	0.99	1.00	0.99	1.00	0.98	0.98	1.00	1.00	0.99	0.99	0.98
s_4	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.99	0.98	0.99	1.00
s_5	0.99	0.99	0.99	0.99	0.98	0.99	1.00	0.99	1.00	1.00	0.99	0.98	0.99	0.99	0.99
s_6	0.98	0.99	1.00	0.98	0.98	1.00	0.99	0.99	0.99	0.98	0.99	0.99	0.97	0.98	1.00
s_7	0.99	0.99	0.98	0.98	0.99	0.99	0.99	0.98	1.00	0.99	1.00	0.98	0.99	0.98	1.00
s_8	0.98	1.00	0.99	0.98	0.98	0.99	0.99	0.99	0.98	1.00	0.99	0.98	0.99	0.99	1.00
s_9	0.98	1.00	0.99	0.99	0.99	0.98	1.00	0.98	1.00	1.00	0.98	0.99	1.00	1.00	0.99
s_{10}	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	1.00	0.98	0.98	0.98	0.98	0.98
s_{11}	0.99	0.98	0.98	0.99	0.99	1.00	0.97	0.97	0.98	0.96	0.99	0.98	0.99	0.98	0.98
s_{12}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s_{13}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s_{14}	0.98	1.00	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.98
s_{15}	1.00	1.00	0.98	1.00	0.99	0.99	0.99	0.98	1.00	0.99	1.00	0.99	0.99	0.99	1.00
s_{16}	0.98	0.99	0.99	0.99	0.99	0.99	0.99	1.00	0.99	0.98	0.98	0.99	1.00	0.99	0.99
s_{17}	0.97	0.99	0.99	0.99	1.00	0.99	1.00	0.98	0.98	0.99	0.99	0.99	0.98	0.99	0.99
s_{18}	0.99	0.99	0.99	1.00	1.00	0.99	0.98	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99
s_{19}	0.99	0.98	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.99	1.00	0.98	1.00	0.99	0.99
s_{20}	0.98	1.00	0.99	0.98	0.99	1.00	0.98	0.99	0.98	1.00	0.98	0.98	0.98	1.00	0.98
s_{21}	0.99	0.99	0.99	0.98	0.98	0.97	0.99	0.98	0.98	0.98	0.99	0.99	0.97	0.99	1.00

4.8 Estimating bias using simulation

In this section, the bias in the parameter estimators of the mean parameters is assessed, and the results are presented. We consider only scenarios in which the true values of the parameters are fixed, i.e. $\theta_{m_i}^{(j)} = \theta_{m_i}$, $j = 1, \ldots, S$ so that an estimate of the bias can be obtained. The relative bias was calculated, and is:

$$\text{Relative Bias}(\hat{\theta}_{m_i}) = \frac{1}{S} \sum_{j=1}^{S} \frac{\theta_{m_i} - \hat{\theta}_{m_i}^{(j)}}{\theta_{m_i}}.$$

Results are presented for $\sigma = 2 \times 10^{-6}$ and a model fitted in the analysis in which the correlation is known, given in Tables 4.9, 4.10 and 4.11.

There are no systematic patterns of bias here, and the values are very small. This indicates that any bias, if it exists, is negligibly small, at least over the scenarios explored here. This is not the case when data are analysed without correlation, as the estimates of σ^2 become negatively biased, as shown in Table 4.12.

The estimates for E_a and k_r do not become substantially more biased if a model is fitted in which correlation is ignored. There is a strong negative bias on variance estimation if we do not incorporate correlation in the analysis, especially when correlation is strongly affecting our results. In some cases the bias is 0.4 of the true value of σ^2. This provides another reason to account for correlation when analysing experiments for this model.

Table 4.9: Relative bias for the estimator of E_a (all values multiplied by 10^8)

	Scenario												
Design	s_1	s_2	s_3	s_4	s_5	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}	s_{21}
d_1	-4.38	-0.91	2.95	1.33	-9.54	2.39	8.89	-5.19	-4.48	3.83	-6.69	-1.06	-5.15
d_2	3.88	-2.18	-2.46	-2.80	-3.43	-3.77	9.75	-0.79	3.68	-19.20	-8.75	-2.42	-9.56
d_3	-6.07	-5.24	0.82	5.18	1.56	-0.14	-18.40	0.04	9.20	-1.54	0.84	2.33	0.40
d_4	-3.92	10.80	2.40	-0.67	2.51	1.08	-5.93	-1.74	5.91	2.24	-1.17	0.49	1.33
d_5	-5.64	3.36	-5.72	-8.96	0.07	-16.70	-6.33	5.55	10.70	0.10	0.13	0.61	-2.60
d_6	5.53	7.95	2.45	0.14	-5.61	16.30	1.87	-5.84	5.59	-2.29	0.29	0.52	1.57
d_7	6.52	1.00	-2.67	6.85	-12.70	15.10	-7.01	8.39	-6.75	0.31	-0.07	-0.46	2.34
d_8	-1.16	-6.16	5.69	-1.63	-1.03	1.58	-6.04	18.90	4.93	-1.47	-1.18	0.43	-0.32
d_9	-7.16	-1.06	-5.32	-4.83	-0.56	-6.05	-9.41	17.40	8.37	-1.21	-0.73	-3.00	0.25
d_{10}	1.55	-5.88	8.29	-7.86	8.18	-0.19	-9.86	6.03	2.19	0.21	1.24	-1.09	0.47
d_{11}	6.21	11.70	-9.33	-4.07	-3.34	6.76	11.90	-5.04	7.00	-1.54	0.77	-2.56	1.15
d_{12}	-2.17	6.40	-3.06	0.05	-3.12	6.54	-10.10	-14.20	0.07	-0.57	-2.01	-0.14	0.78
d_{13}	3.97	1.28	-6.84	-3.89	9.47	-7.77	5.39	13.10	13.20	3.03	-0.69	1.79	1.76
d_{14}	6.95	4.00	34.00	20.90	-4.02	1.20	8.61	4.60	-16.70	-1.66	-3.54	3.44	-2.65
d_{15}	3.28	0.50	2.88	2.96	20.10	-10.70	15.00	1.96	-5.44	-2.41	0.85	-0.56	-0.04

Table 4.10: Relative bias for the estimator of k_r (all values multiplied by 10^8)

	Scenario												
Design	s_1	s_2	s_3	s_4	s_5	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}	s_{21}
d_1	0.79	6.04	-2.99	-6.36	-6.74	2.31	-1.01	-16.20	2.75	-1.24	2.14	1.56	3.14
d_2	-5.00	3.54	-1.15	-0.47	-1.14	3.57	-0.37	5.64	21.60	4.12	2.23	-6.74	26.20
d_3	4.19	0.51	3.31	-5.09	-5.94	0.50	2.34	1.59	-6.79	1.85	-0.44	-11.40	-7.38
d_4	2.98	-5.70	2.60	-0.28	-5.38	1.27	1.42	-10.20	1.21	-0.02	0.22	-4.65	1.19
d_5	8.15	-2.43	7.28	4.77	3.38	-0.48	1.61	-1.64	-7.86	0.11	0.09	-0.39	8.15
d_6	-4.66	-2.01	2.94	0.40	11.90	-1.33	0.22	-4.81	3.21	-0.37	-0.58	0.70	2.04
d_7	4.13	6.12	-4.25	-10.90	4.90	-2.70	-0.28	-10.80	-10.80	0.15	0.01	1.76	-0.63
d_8	-2.09	3.52	-2.37	-2.49	-0.30	1.44	2.01	-12.20	-11.70	-0.37	0.10	-4.21	-3.68
d_9	0.87	-1.05	1.70	-2.65	-1.01	-0.90	1.05	-14.90	-20.60	-0.32	-0.08	-1.03	-0.64
d_{10}	2.97	3.25	-0.48	3.11	-3.82	0.04	0.45	-17.60	3.61	0.32	-0.16	1.74	0.13
d_{11}	-1.87	-2.39	3.64	-1.27	4.21	1.10	-0.01	-8.82	-13.50	-0.15	0.01	1.28	1.40
d_{12}	-1.25	-6.83	-3.69	1.34	-0.65	-0.88	0.77	13.40	-8.48	0.75	-0.24	0.69	-0.80
d_{13}	1.59	-1.41	-0.23	1.49	-2.96	0.91	0.51	-19.10	-8.62	0.01	0.34	-3.39	-1.63
d_{14}	3.62	4.93	-4.00	-13.50	7.69	0.03	0.34	87.70	13.80	-0.11	0.48	7.83	-2.42
d_{15}	-3.87	-1.99	-2.89	2.25	-8.09	0.39	-2.18	-17.10	5.84	0.07	-0.21	1.78	0.86

Table 4.11: Relative bias for the estimator of σ^2 (all values multiplied by 10^3)

	Scenario												
Design	s_1	s_2	s_3	s_4	s_5	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}	s_{21}
d_1	1.01	-0.17	0.74	-0.60	1.30	0.15	1.76	1.16	-1.05	0.64	-0.38	-2.25	0.71
d_2	-1.29	-0.67	-0.49	0.12	0.37	0.12	0.12	-0.61	1.02	0.13	1.22	0.35	1.16
d_3	0.33	-1.71	-0.26	-1.67	-2.25	-1.41	1.32	0.31	-1.44	0.01	-0.88	-1.29	-0.09
d_4	-0.68	-1.28	0.67	1.35	0.73	0.51	-1.28	0.48	0.86	1.06	0.86	-1.24	-0.50
d_5	-1.28	1.87	0.02	0.87	-1.88	-0.28	0.33	0.28	-0.53	-1.50	-0.28	-0.00	0.54
d_6	-0.05	0.45	0.44	1.09	-1.10	0.30	-0.50	-0.93	0.46	-1.09	-0.24	0.11	-1.16
d_7	0.53	2.60	-1.88	1.16	0.10	-0.01	0.93	-0.25	-0.37	1.18	1.19	-1.21	0.12
d_8	0.37	-1.44	-0.03	0.22	-0.06	0.48	0.12	-0.43	0.74	1.62	1.45	-0.78	-0.56
d_9	1.29	-0.43	0.09	1.54	-0.37	1.24	-0.39	0.42	-0.77	0.31	-2.74	-0.95	0.33
d_{10}	0.64	0.04	0.69	-1.18	0.99	0.16	-0.13	1.95	-0.03	1.06	0.22	0.99	2.15
d_{11}	-0.08	-1.06	-0.85	0.27	1.29	-0.70	-1.27	-0.14	-1.14	-0.33	0.11	0.82	-0.97
d_{12}	-0.25	0.06	-1.55	-0.06	0.09	1.19	0.23	0.18	0.01	1.39	-0.88	0.35	0.16
d_{13}	-0.30	-1.52	0.27	1.27	-3.12	0.19	0.47	0.86	0.12	1.56	-2.06	0.67	1.51
d_{14}	1.50	0.61	-0.67	0.66	-1.30	0.45	0.80	0.36	-1.36	-0.46	0.94	-0.01	-0.96
d_{15}	0.43	-0.38	-0.50	-0.38	-0.44	-1.51	1.51	0.29	0.41	-0.55	0.13	0.88	0.10

Table 4.12: Relative bias for the estimator of σ^2 when correlation is not included during analysis (all values multiplied by 10^2)

	Scenario												
Design	s_1	s_2	s_3	s_4	s_5	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}	s_{21}
d_1	-0.49	-0.02	-0.92	-22.98	-15.24	-20.76	-21.09	-21.26	-17.16	-41.84	-43.19	-42.89	-34.11
d_2	-3.55	-0.07	-3.63	-24.47	-16.24	-25.03	-25.45	-25.56	-22.53	-43.51	-44.29	-44.49	-39.23
d_3	-0.67	-0.17	-1.10	-23.19	-15.53	-21.04	-21.16	-21.34	-17.48	-42.05	-43.21	-42.48	-35.02
d_4	-0.39	-0.13	-0.53	-12.19	-8.22	-14.25	-11.88	-12.41	-10.79	-35.68	-26.31	-24.21	-21.05
d_5	-1.02	0.19	-1.13	-14.98	-10.10	-14.54	-14.29	-15.18	-13.46	-35.50	-29.73	-28.60	-25.11
d_6	-0.15	0.04	-0.30	-13.24	-8.88	-12.00	-12.56	-11.48	-8.51	-31.46	-30.74	-23.89	-17.80
d_7	-0.08	0.26	-0.54	-13.66	-9.03	-11.92	-12.57	-11.83	-8.89	-32.04	-30.39	-24.25	-17.66
d_8	-0.07	-0.14	-0.26	-13.04	-8.78	-11.61	-11.93	-10.28	-7.66	-31.65	-30.16	-22.60	-16.83
d_9	0.01	-0.04	-0.27	-14.56	-9.77	-11.18	-12.01	-11.04	-8.73	-31.46	-30.97	-23.99	-19.08
d_{10}	-0.35	0.00	-0.64	-14.80	-9.78	-13.98	-12.72	-14.38	-12.33	-34.02	-27.35	-28.46	-23.79
d_{11}	-0.48	-0.11	-0.85	-15.39	-10.02	-12.60	-13.33	-14.86	-12.61	-31.81	-31.38	-29.39	-24.44
d_{12}	-0.59	0.01	-0.96	-14.98	-10.08	-12.26	-13.48	-14.51	-12.69	-31.31	-31.19	-28.36	-24.53
d_{13}	-0.62	-0.15	-0.80	-15.46	-10.63	-12.37	-13.24	-14.59	-12.85	-31.87	-31.19	-29.04	-24.94
d_{14}	0.15	0.06	-0.07	-9.35	-6.28	-12.16	-10.25	-3.43	-2.45	-35.45	-29.92	-9.92	-6.96
d_{15}	-0.07	-0.04	-0.33	-12.64	-8.50	-7.66	-8.47	-11.02	-10.75	-21.21	-22.19	-22.14	-21.46

4.9 Performance of designs in the simulation

The accuracy of parameter estimates (as measured by Equation (4.1)) that were obtained for each design and each scenario is investigated in this section, to assess which designs lead to the 'best' parameter estimates over a wide range of potential scenarios.

A model was fitted assuming the form and magnitude of the correlation function is known, for $\sigma = 2 \times 10^{-6}$. The value of DSE, defined in Equation (4.1), was calculated for each of E_a and k_r. As the two values were similar, their average ($ADSE$) was used as a measure of accuracy of parameter estimation. The ratio of the value of $ADSE$ relative to the minimum value of $ADSE$ obtained for each scenario was calculated. Table 4.13 presents results for scenarios $s_1 - s_{13}$, which simulate assuming that the true parameters come from the prior distributions assumed when finding $d_1 - d_{13}$. The minimum, median and mean values across all scenarios are also presented.

The Bayesian D-optimal designs performed well throughout. In particular, designs $d_{10} - d_{13}$ found for uniform prior distributions for E_a and k_r have high values for a wide range of scenarios. Over the scenarios $s_1 - s_{13}$ both the ad hoc

Table 4.13: *ADSE* relative to minimum *ADSE* over scenarios $s_1 - s_{13}$, with the mean, median and minimum obtained by each design also given

	Scenario															
Design	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_7	s_9	s_{10}	s_{11}	s_{12}	s_{13}	Min	Mean	Median
d_1	1.00	0.94	1.00	0.53	0.77	0.75	0.67	0.71	0.64	0.79	0.71	0.77	0.68	0.53	0.77	0.75
d_2	0.82	1.00	0.84	0.53	0.78	0.70	0.60	0.66	0.57	0.74	0.66	0.73	0.64	0.53	0.71	0.70
d_3	1.00	0.95	1.00	0.53	0.78	0.75	0.67	0.73	0.65	0.79	0.71	0.77	0.68	0.53	0.77	0.75
d_4	0.64	0.59	0.65	1.00	0.98	0.94	0.94	0.93	0.91	0.94	0.93	0.93	0.92	0.59	0.87	0.93
d_5	0.69	0.67	0.69	0.98	1.00	0.96	0.95	0.93	0.93	0.96	0.96	0.96	0.94	0.67	0.89	0.95
d_6	0.56	0.50	0.56	0.86	0.84	0.99	0.99	0.99	0.99	0.95	0.95	0.96	0.95	0.50	0.86	0.95
d_7	0.57	0.51	0.58	0.87	0.85	1.00	1.00	0.98	0.99	0.96	0.95	0.95	0.95	0.51	0.86	0.95
d_8	0.55	0.49	0.56	0.86	0.83	0.99	0.99	1.00	0.98	0.94	0.92	0.95	0.93	0.49	0.85	0.93
d_9	0.59	0.54	0.61	0.83	0.81	1.00	0.99	0.99	1.00	0.94	0.94	0.95	0.95	0.54	0.86	0.94
d_{10}	0.69	0.63	0.69	0.95	0.96	0.96	0.96	0.95	0.94	1.00	1.00	0.99	0.99	0.63	0.90	0.96
d_{11}	0.69	0.65	0.69	0.93	0.94	0.98	0.98	0.96	0.96	1.00	1.00	0.99	1.00	0.65	0.90	0.96
d_{12}	0.68	0.64	0.69	0.94	0.94	0.98	0.99	0.97	0.97	1.00	1.00	1.00	0.99	0.64	0.91	0.97
d_{13}	0.69	0.64	0.68	0.91	0.93	0.98	0.97	0.97	0.97	0.99	0.99	0.99	0.98	0.64	0.90	0.97
d_{14}	0.08	0.07	0.08	0.15	0.14	0.32	0.32	0.33	0.32	0.20	0.19	0.21	0.20	0.07	0.20	0.20
d_{15}	0.20	0.18	0.20	0.30	0.29	0.49	0.49	0.50	0.49	0.42	0.41	0.43	0.42	0.18	0.37	0.42

designs $d_{14} - d_{15}$ perform less well, achieving their best (slightly over 0.3 or under 0.5 respectively) relative accuracy of parameter estimation when the prior distributions for E_a and k_r are normal and gamma respectively ($s_6 - s_9$). The design with the highest minimum of 0.67 is d_5. The highest mean value achieved by a design was 0.91, by d_{12} and the highest median value was 0.97, achieved by d_{12} and d_{13}.

Table 4.14 gives the performance of the designs under the "edge scenarios", s_{14}-s_{21}. These are scenarios where the assumed value for the mean parameters is much larger or smaller than anticipated during design.

Under the edge scenarios the ad hoc designs have a higher mean and median accuracy than for $s_1 - s_{13}$, but are still outperformed by the Bayesian D-optimal designs. Both d_7 and d_8 perform extremely well across these scenarios, with d_8 having the highest minimum, mean and median of 0.96, 0.99 and 1. The designs $d_{10} - d_{13}$ which were found for assumed uniform prior distributions on E_a and k_r do not perform as well here, which is to be expected; scenarios $s_{14} - s_{21}$ assume the true parameter values are either on the edge of, or outside the intervals assumed for the uniform distributions for E_a and k_r when designing.

Table 4.15 gives the performance of the designs under the scenarios when correlation is not present, s_{22}-s_{24}. As might be expected the design which assumes the least correlation, d_2, performs the best across all three scenarios.

Table 4.14: *ADSE* relative to minimum *ADSE* over scenarios $s_{14} - s_{21}$, with the mean, median and minimum obtained by each design also given

	Scenario										
Design	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}	s_{21}	Min	Mean	Median
d_1	0.31	0.33	0.62	0.72	0.10	0.06	0.13	0.12	0.06	0.30	0.22
d_2	0.22	0.27	0.46	0.53	0.03	0.02	0.04	0.04	0.02	0.20	0.13
d_3	0.30	0.35	0.62	0.70	0.11	0.07	0.12	0.13	0.07	0.30	0.22
d_4	0.74	0.71	0.65	0.58	0.56	0.76	0.58	0.54	0.54	0.64	0.61
d_5	0.64	0.72	0.62	0.59	0.48	0.74	0.55	0.55	0.48	0.61	0.61
d_6	0.92	0.97	0.90	0.97	0.93	0.96	0.90	0.97	0.90	0.94	0.94
d_7	0.89	0.99	0.90	0.98	0.91	0.96	0.90	0.99	0.89	0.94	0.94
d_8	1.00	1.00	0.96	1.00	1.00	1.00	0.97	1.00	0.96	0.99	1.00
d_9	0.98	0.91	1.00	1.00	0.99	0.90	1.00	0.97	0.90	0.97	0.98
d_{10}	0.60	0.70	0.90	0.86	0.52	0.76	0.81	0.79	0.52	0.74	0.78
d_{11}	0.77	0.75	0.91	0.90	0.77	0.75	0.83	0.83	0.75	0.82	0.80
d_{12}	0.80	0.72	0.96	0.89	0.81	0.71	0.87	0.81	0.71	0.82	0.81
d_{13}	0.76	0.77	0.94	0.90	0.74	0.77	0.86	0.82	0.74	0.82	0.80
d_{14}	0.84	0.79	0.09	0.27	0.77	0.72	0.10	0.29	0.09	0.48	0.50
d_{15}	0.58	0.59	0.65	0.62	0.73	0.66	0.69	0.61	0.58	0.64	0.63

Table 4.15: *ADSE* relative to minimum *ADSE* over scenarios $s_{22} - s_{24}$

	Scenario		
Design	s_{22}	s_{23}	s_{24}
d_1	0.93	0.99	0.98
d_2	1.00	1.00	1.00
d_3	0.94	1.00	0.98
d_4	0.60	0.70	0.62
d_5	0.67	0.77	0.70
d_6	0.50	0.73	0.62
d_7	0.51	0.75	0.63
d_8	0.49	0.71	0.59
d_9	0.54	0.74	0.64
d_{10}	0.63	0.76	0.72
d_{11}	0.63	0.79	0.73
d_{12}	0.64	0.77	0.73
d_{13}	0.64	0.79	0.74
d_{14}	0.07	0.21	0.10
d_{15}	0.18	0.32	0.25

Finally the minimum, mean and median across all scenarios $s_1 - s_{21}$ of the $ADSE$ relative to minimum $ADSE$ for each scenario are presented in Table 4.16.

The highest minimum relative accuracy of parameter estimation, 0.64, is achieved by d_{12} and d_{13}, with the highest mean and median, 0.87 and 0.95, achieved by d_8 and d_9. This implies that when finding designs which aim to provide parameter estimates with high relative accuracy across a wide range of potential scenarios, the best assumptions to make according to this study are to use normal prior distributions over E_a, a gamma distribution for k_r and τ, a uniform prior on ρ and a log uniform prior distribution for r and q.

Table 4.16: Minimum mean and median $ADSE$ relative to minimum $ADSE$ for all designs across all scenarios

	Min	Mean	Median
d_1	0.06	0.64	0.71
d_2	0.02	0.58	0.65
d_3	0.07	0.64	0.70
d_4	0.54	0.76	0.72
d_5	0.48	0.78	0.73
d_6	0.50	0.85	0.94
d_7	0.51	0.86	0.93
d_8	0.49	0.86	0.95
d_9	0.54	0.87	0.95
d_{10}	0.52	0.82	0.84
d_{11}	0.63	0.85	0.87
d_{12}	0.64	0.85	0.88
d_{13}	0.64	0.85	0.88
d_{14}	0.07	0.29	0.20
d_{15}	0.18	0.45	0.46

For practitioners, it may be of interest under which scenarios designs give the most or the least accurate mean parameter estimates. To investigate this problem we calculated the $ADSE$ for each design relative to the minimum $ADSE$ across all scenarios and designs. The results are presented in Table 4.17.

It can be observed from Table 4.17 that the most accurate parameter estimates are obtained under scenarios where the variance component for run-to-run error is

Table 4.17: *ADSE* of designs relative to the minimum *ADSE* obtained across all scenarios and designs (all values multiplied by 10^2)

	Design														
Scenario	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
s_1	8.33	6.84	8.32	5.34	5.77	4.65	4.76	4.55	4.95	5.76	5.72	5.70	5.73	0.63	1.63
s_2	8.79	9.32	8.88	5.51	6.24	4.70	4.78	4.53	5.04	5.85	6.03	5.98	5.96	0.63	1.64
s_3	8.07	6.80	8.07	5.25	5.61	4.54	4.65	4.53	4.92	5.56	5.53	5.55	5.52	0.63	1.62
s_4	2.54	2.52	2.54	4.77	4.69	4.12	4.17	4.09	3.95	4.52	4.42	4.48	4.34	0.73	1.44
s_5	3.28	3.35	3.32	4.21	4.28	3.60	3.62	3.53	3.48	4.11	4.03	4.03	3.99	0.60	1.25
s_6	2.06	1.91	2.07	2.59	2.65	2.73	2.75	2.72	2.74	2.63	2.69	2.70	2.69	0.89	1.35
s_7	2.41	2.14	2.41	3.37	3.40	3.54	3.57	3.55	3.53	3.44	3.48	3.54	3.47	1.14	1.75
s_8	1.96	1.82	1.99	2.56	2.55	2.72	2.68	2.74	2.72	2.60	2.62	2.65	2.66	0.91	1.36
s_9	2.29	2.05	2.31	3.24	3.31	3.54	3.53	3.51	3.57	3.37	3.42	3.48	3.46	1.16	1.76
s_{10}	2.34	2.19	2.34	2.76	2.84	2.81	2.82	2.77	2.77	2.94	2.93	2.95	2.92	0.59	1.23
s_{11}	2.72	2.54	2.73	3.58	3.69	3.68	3.67	3.55	3.64	3.84	3.85	3.84	3.83	0.75	1.59
s_{12}	2.24	2.14	2.23	2.70	2.78	2.79	2.77	2.75	2.76	2.89	2.87	2.91	2.88	0.61	1.24
s_{13}	2.60	2.43	2.61	3.50	3.60	3.64	3.63	3.57	3.61	3.77	3.82	3.80	3.75	0.77	1.60
s_{14}	0.18	0.13	0.17	0.42	0.36	0.53	0.51	0.57	0.56	0.34	0.44	0.46	0.44	0.48	0.33
s_{15}	1.21	1.00	1.27	2.59	2.64	3.56	3.61	3.65	3.34	2.57	2.76	2.65	2.82	2.89	2.16
s_{16}	0.36	0.27	0.36	0.38	0.36	0.52	0.53	0.56	0.58	0.52	0.53	0.56	0.54	0.05	0.38
s_{17}	2.56	1.88	2.50	2.05	2.10	3.45	3.51	3.57	3.56	3.08	3.22	3.19	3.20	0.97	2.20
s_{18}	1.03	0.35	1.05	5.57	4.83	9.29	9.11	10.00	9.86	5.17	7.75	8.13	7.44	7.74	7.30
s_{19}	4.59	1.33	5.07	58.60	57.10	74.20	74.80	77.50	69.70	58.80	57.90	54.70	59.60	55.70	51.20
s_{20}	1.93	0.56	1.87	8.72	8.28	13.60	13.50	14.60	15.10	12.30	12.60	13.10	13.00	1.54	10.40
s_{21}	12.30	3.79	13.40	53.90	54.60	97.50	98.60	100.00	97.10	79.20	82.70	81.20	82.10	28.70	61.20
s_{22}	8.72	9.35	8.81	5.61	6.25	4.70	4.80	4.61	5.04	5.93	5.93	5.97	5.97	0.63	1.64
s_{23}	5.12	5.15	5.15	3.60	3.98	3.77	3.84	3.68	3.83	3.93	4.09	3.96	4.05	1.07	1.67
s_{24}	5.90	6.02	5.92	3.73	4.21	3.73	3.80	3.58	3.87	4.32	4.39	4.40	4.43	0.63	1.53

relatively very high, as in s_{21}. A possible explanation follows from the argument made in Subsection 3.6.1, that the run-to-run error can be effectively 'cancelled' during analysis, leading to highly accurate parameter estimates. If, in fact, the data are analysed as if correlation was not present, the error on parameter estimation is higher for s_{21} than for any other scenario. If s_{18}-s_{21} are excluded then the accuracy obtained under other scenarios can be compared more clearly. This is given in Table 4.18.

The most accurate parameter estimation is now obtained by d_2 in s_2, the scenario with the lowest serial correlation in the error structure. The most inaccurate estimation is obtained for scenarios s_{14} and s_{16}, which both assumed a low value for $E_a = 2 \times 10^4$ and $k_r = 1 \times 10^{-4}$ or $k_r = 1 \times 10^{-3}$. Scrutiny of the *DSE* for both parameters indicates that this is due to less accurate estimation of E_a, so very low values of E_a may be difficult to estimate accurately. The relative accuracy of the parameter estimates obtained for $s_5 - s_{13}$, when the parameters were varied according to the prior distributions designed for when finding $d_5 - d_{13}$, is generally

quite low. This is probably due to the impact of run-to-run error, as the same effect can be seen in s_4 and s_5. While the result in Subsection 3.6.1 implies that the variance component for run-to-run error being large can lead to more accurate parameter estimates, the accuracy will probably depend on the magnitude of the variance component for run-to-run error: for lower values, less accurate parameter estimates might be obtained. This implies that practitioners need to be confident about whether run-to-run error is present or not before experimentation, as making the wrong decision while designing can lead to efficiency being needlessly lost. If the variance component for run-to-run error is negligible then more information is gained by using a design similar to d_1, but when r is higher, accuracy of parameter estimation is lost by doing so, as seen in Section 4.5.

Table 4.18: *ADSE* of designs relative to the minimum *ADSE* obtained across all scenarios and designs, excluding s_{18}-s_{21}

	Design														
Scenario	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
s_1	0.89	0.73	0.89	0.57	0.62	0.50	0.51	0.49	0.53	0.62	0.61	0.61	0.61	0.07	0.17
s_2	0.94	1.00	0.95	0.59	0.67	0.50	0.51	0.48	0.54	0.63	0.65	0.64	0.64	0.07	0.18
s_3	0.86	0.73	0.86	0.56	0.60	0.49	0.50	0.48	0.53	0.59	0.59	0.59	0.59	0.07	0.17
s_4	0.27	0.27	0.27	0.51	0.50	0.44	0.45	0.44	0.42	0.48	0.47	0.48	0.46	0.08	0.15
s_5	0.35	0.36	0.36	0.45	0.46	0.39	0.39	0.38	0.37	0.44	0.43	0.43	0.43	0.06	0.13
s_6	0.22	0.20	0.22	0.28	0.28	0.29	0.29	0.29	0.29	0.28	0.29	0.29	0.29	0.10	0.14
s_7	0.26	0.23	0.26	0.36	0.36	0.38	0.38	0.38	0.38	0.37	0.37	0.38	0.37	0.12	0.19
s_8	0.21	0.19	0.21	0.27	0.27	0.29	0.29	0.29	0.29	0.28	0.28	0.28	0.28	0.10	0.15
s_9	0.25	0.22	0.25	0.35	0.35	0.38	0.38	0.38	0.38	0.36	0.37	0.37	0.37	0.12	0.19
s_{10}	0.25	0.23	0.25	0.29	0.30	0.30	0.30	0.30	0.30	0.31	0.31	0.32	0.31	0.06	0.13
s_{11}	0.29	0.27	0.29	0.38	0.40	0.39	0.39	0.38	0.39	0.41	0.41	0.41	0.41	0.08	0.17
s_{12}	0.24	0.23	0.24	0.29	0.30	0.30	0.30	0.29	0.30	0.31	0.31	0.31	0.31	0.07	0.13
s_{13}	0.28	0.26	0.28	0.37	0.39	0.39	0.39	0.38	0.39	0.40	0.41	0.41	0.40	0.08	0.17
s_{14}	0.02	0.01	0.02	0.05	0.04	0.06	0.05	0.06	0.06	0.04	0.05	0.05	0.05	0.05	0.04
s_{15}	0.13	0.11	0.14	0.28	0.28	0.38	0.39	0.39	0.36	0.27	0.29	0.28	0.30	0.31	0.23
s_{16}	0.04	0.03	0.04	0.04	0.04	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.01	0.04
s_{17}	0.27	0.20	0.27	0.22	0.22	0.37	0.38	0.38	0.38	0.33	0.34	0.34	0.34	0.10	0.24
s_{22}	0.93	1.00	0.94	0.60	0.67	0.50	0.51	0.49	0.54	0.63	0.63	0.64	0.64	0.07	0.18
s_{23}	0.55	0.55	0.55	0.38	0.43	0.40	0.41	0.39	0.41	0.42	0.44	0.42	0.43	0.11	0.18
s_{24}	0.63	0.64	0.63	0.40	0.45	0.40	0.41	0.38	0.41	0.46	0.47	0.47	0.47	0.07	0.16

4.9.1 D-efficiency of designs under the edge scenarios

For the edge scenarios, s_{14}-s_{21}, the D-optimal design can be found and thus the D-efficiency of the designs in the simulation study can be calculated. The results are given in Table 4.19.

The results obtained are reasonably similar to those from the simulation study, but in s_{14} the ad hoc design which took observations spread uniformly across time, d_{14}, has the highest D-efficiency of 0.66. The Bayesian D-optimal designs achieve a fairly high D-efficiency across all scenarios, especially d_6-d_9, which have a D-efficiency higher than 0.5 in all scenarios. This implies that the low relative accuracy of parameter estimates observed in s_{14} and s_{16} is likely to be evident no matter which design is used, and is not something which can be compensated for with design.

Table 4.19: Comparison of D-efficiency over scenarios $s_{14} - s_{21}$

	Scenario							
Design	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}	s_{21}
d_1	0.19	0.30	0.52	0.52	0.06	0.07	0.09	0.17
d_2	0.14	0.25	0.43	0.38	0.02	0.03	0.03	0.07
d_3	0.19	0.30	0.53	0.51	0.06	0.07	0.09	0.17
d_4	0.44	0.46	0.58	0.41	0.33	0.47	0.51	0.38
d_5	0.39	0.47	0.57	0.42	0.30	0.46	0.47	0.39
d_6	0.55	0.65	0.73	0.71	0.52	0.58	0.66	0.71
d_7	0.54	0.66	0.73	0.72	0.51	0.59	0.66	0.73
d_8	0.59	0.65	0.76	0.72	0.56	0.60	0.71	0.74
d_9	0.58	0.61	0.79	0.73	0.55	0.54	0.73	0.72
d_{10}	0.37	0.47	0.74	0.61	0.31	0.47	0.61	0.57
d_{11}	0.46	0.52	0.75	0.64	0.43	0.47	0.63	0.60
d_{12}	0.48	0.49	0.77	0.63	0.45	0.43	0.66	0.59
d_{13}	0.45	0.54	0.76	0.64	0.42	0.48	0.64	0.59
d_{14}	0.66	0.56	0.09	0.21	0.55	0.49	0.10	0.23
d_{15}	0.45	0.41	0.66	0.54	0.53	0.42	0.65	0.54

4.10 Simulating heteroscedasticity in the errors

In this section the results are presented of a smaller study intended to assess the performance of optimal designs when error heteroscedasticity is present. We investigated how accurate parameter estimates obtained from the optimal designs in Section 3.9 are under the assumption of error heteroscedasticity.

4.10.1 Designs and scenarios used in the simulation

Seven designs were compared. These were five Bayesian D-optimal designs, each found for the same prior distributions on the mean and correlation parameters: $E_a \sim N(4.91 \times 10^4, 4000^2)$, $k_r \sim \text{G}(m_k^2/v_k, v_k/m_k)$, with a mean of $m_k = 4.59 \times 10^{-4}$ and variance of $v_k = (4 \times 10^{-4})^2$, $\tau \sim$G(1,2), $\rho \sim$U(0,2), log(r) and log(q) $\sim$U(-2.5,2.5). There were also two ad hoc designs. The designs were:

1. Design d_9, a Bayesian D-optimal design assuming $\lambda = 1$
2. Design d_{16}, an extended version of d_9, where the optimal design was permitted to take observations up to $t = 300$
3. Designs $d_{17} - d_{19}$, the three Bayesian D-optimal designs obtained in Section 3.9, with three different priors on λ: $\lambda \sim B(4,1)$, $\lambda \sim U(0,1)$ and $\lambda \sim B(2, 12.8571)$
4. Designs d_{20} and d_{21}, extended versions of the ad hoc designs d_{14} and d_{15}, which take observations up to $t = 300$

The designs are given in Table 4.20.

Table 4.20: Designs used during simulation

Design	T_1	T_2	$\boldsymbol{t}_1 = \boldsymbol{t}_2$						T_3	T_4	$\boldsymbol{t}_3 = \boldsymbol{t}_4$					
d_9	70	70	1.0	21.2	31.2	44.1	94.9	200.0	100	100	4.7	8.0	12.4	21.1	77.3	200.0
d_{16}	70	70	1.0	21.6	32.0	46.7	92.3	300.0	100	100	4.8	8.2	12.7	22.0	85.4	300.0
d_{17}	70	70	17.5	35.2	56.1	83.6	284.8	300.0	100	100	5.8	10.0	16.0	28.3	90.5	300.0
d_{18}	70	70	1.0	31.0	55.8	104.9	288.1	300.0	100	100	7.4	15.1	27.4	69.2	267.7	300.0
d_{19}	70	70	1.0	42.5	276.8	287.8	295.1	300.0	100	100	1.0	31.1	63.1	110.7	286.9	300.0
d_{20}	70	80	1.0	4.0	10.0	20.0	150.0	300.0	90	100	1.0	4.0	10.0	20.0	150.0	300.0
d_{21}	70	80	1.0	60.8	120.6	180.4	240.2	300.0	90	100	1.0	60.8	120.6	180.4	240.2	300.0

For reference in later tables, the prior distributions for each respective mean or correlation parameter θ_i is labelled $\pi(\theta_i)$, so that, for example, the prior $E_a \sim N(4.91 \times 10^4, 4000^2)$ is labelled $\pi(E_a)$. The three priors on λ are labelled, $\pi_1(\lambda)$: $\lambda \sim B(4,1)$, $\pi_2(\lambda)$: $\lambda \sim U(0,1)$ and $\pi_3(\lambda)$: $\lambda \sim B(2, 12.8571)$.

Several scenarios, shown in Table 4.21, were considered

1. Scenarios where the parameter values are drawn from the prior distributions used when finding designs $d_{17} - d_{19}$. This set of scenarios was chosen to investigate the impact of mis-specifying the prior distribution for λ while designing, and to confirm that the theoretically optimal designs gave the most accurate parameter estimates when considering finite data sets.

2. Edge scenarios, where the mean and correlation parameters were misspecified, and the value of λ was drawn from $\pi_1(\lambda)$, $\pi_2(\lambda)$ or $\pi_3(\lambda)$. The edge scenarios were chosen to investigate how effective the designs found would be at estimating the mean parameters when the prior distributions on the mean and correlation parameters were misspecified. The parameter values for the edge scenarios were chosen from the edge scenarios in the larger simulation study (s_{14}-s_{21}), with two combinations of E_a and k_r chosen, representing a very fast reaction and a very slow reaction respectively.

Table 4.21: Scenarios used in the simulation study, to investigate the impact of error heteroscedasticity on accuracy of parameter estimation

	Prior distribution for each parameter value						
Scenario	E_a	k_r	τ	ρ	r	q	λ
s_{25}	$\pi(E_a)$	$\pi(k_r)$	$\pi(\tau)$	$\pi(\rho)$	$\pi(r)$	$\pi(q)$	$\pi_1(\lambda)$
s_{26}	$\pi(E_a)$	$\pi(k_r)$	$\pi(\tau)$	$\pi(\rho)$	$\pi(r)$	$\pi(q)$	$\pi_2(\lambda)$
s_{27}	$\pi(E_a)$	$\pi(k_r)$	$\pi(\tau)$	$\pi(\rho)$	$\pi(r)$	$\pi(q)$	$\pi_3(\lambda)$
s_{28}	2×10^4	10×10^{-4}	0.1	1	1	1	$\pi_1(\lambda)$
s_{29}	7×10^4	1×10^{-4}	0.1	1	1	1	$\pi_1(\lambda)$
s_{30}	2×10^4	10×10^{-4}	0.1	1	100	1	$\pi_1(\lambda)$
s_{31}	7×10^4	1×10^{-4}	0.1	1	100	1	$\pi_1(\lambda)$
s_{32}	2×10^4	10×10^{-4}	0.1	1	1	1	$\pi_2(\lambda)$
s_{33}	7×10^4	1×10^{-4}	0.1	1	1	1	$\pi_2(\lambda)$
s_{34}	2×10^4	10×10^{-4}	0.1	1	100	1	$\pi_2(\lambda)$
s_{35}	7×10^4	1×10^{-4}	0.1	1	100	1	$\pi_2(\lambda)$
s_{36}	2×10^4	10×10^{-4}	0.1	1	1	1	$\pi_3(\lambda)$
s_{37}	7×10^4	1×10^{-4}	0.1	1	1	1	$\pi_3(\lambda)$
s_{38}	2×10^4	10×10^{-4}	0.1	1	100	1	$\pi_3(\lambda)$
s_{39}	7×10^4	1×10^{-4}	0.1	1	100	1	$\pi_3(\lambda)$

4.10.2 Methodology

Simulation to assess the impact of error heteroscedasticity on parameter accuracy is difficult using the transformation approach taken in this thesis. Errors simulated from a multivariate normal distribution with variance proportional to the expected response raised to a power of $2(1-\lambda)$ will no longer be normally distributed if the response is transformed by a power of λ, and so the designs found may no longer be optimal. The approach given by Atkinson assumes that the error distribution is also skewed: to see this in practice would usually require real data.

Under least squares regression, when λ is assumed known, D-optimal designs found assuming the response and expected response will be transformed by a power of λ are also D-optimal for a statistical model with a response with variance proportional to the expected response raised to a power of $2(1-\lambda)$. The optimal designs were assessed under this assumption.

Data were generated as in Section 4.4, with errors drawn from a MVN

distribution with mean $\mathbf{0}$ and variance

$$\Delta Q(\boldsymbol{\theta}_c)\Delta\sigma^2,$$

where Δ is an $Nm \times Nm$ matrix whose diagonal entries are $[\boldsymbol{\eta}(\boldsymbol{t}, \boldsymbol{T}, \boldsymbol{\theta}_m)]^{1-\lambda}$ and all other entries are zero. The constant term σ was assumed equal to 1×10^{-4}.

Data attained by simulation was analysed as in Section 4.4 using two different approaches in order to investigate how much impact ignoring heteroscedasticity during analysis has on the accuracy of parameter estimation. Data obtained was analysed by pre-multiplying the response and expected response by the transpose of either

$$Q(\boldsymbol{\theta}_c)^{-1/2} \tag{4.4}$$

or

$$\left[\hat{\Delta} Q(\boldsymbol{\theta}_c)\hat{\Delta}\right]^{-1/2} \tag{4.5}$$

where $\hat{\Delta}$ is an $Nm \times Nm$ matrix whose diagonal entries are $\left[\boldsymbol{\eta}(\boldsymbol{t}, \boldsymbol{T}, \hat{\boldsymbol{\theta}}_m)\right]^{1-\lambda}$.

Note that while the model parameters are estimated when calculating $\hat{\Delta}$, λ was assumed known. This is an unrealistic assumption, as in practice both the correlation matrix and λ would need to be estimated, but to allow ease of simulation, this simplifying assumption was made.

4.10.3 Results

We first examine the results of scenarios $s_{25} - s_{27}$, to investigate the performance of Bayesian D-optimal designs when only the prior distributions on λ are mis-specified. Table 4.22 gives the DSE relative to the minimum DSE for each parameter and scenario, when the response and expected response were premultiplied by Expression (4.4), i.e. when heteroscedasticity was ignored during analysis.

When heteroscedasticity is not considered during analysis, the optimal designs

Table 4.22: DSE relative to minimum DSE for scenarios $s_{25} - s_{27}$, when the data were analysed ignoring heteroscedasticity

		Designs						
Scenario	Parameter	d_9	d_{16}	d_{17}	d_{18}	d_{19}	d_{20}	d_{21}
s_{25}	E_a	1.00	1.00	0.97	0.93	0.46	0.19	0.42
	k_r	1.00	1.00	0.91	0.93	0.55	0.45	0.58
s_{26}	E_a	0.97	0.98	0.98	1.00	0.41	0.15	0.37
	k_r	0.95	0.99	0.87	1.00	0.52	0.42	0.48
s_{27}	E_a	0.87	0.91	0.90	1.00	0.36	0.14	0.29
	k_r	0.85	0.90	0.80	1.00	0.43	0.38	0.36

which assumed the "correct" prior distribution for λ while designing do not always provide the most accurate parameter estimates. In particular, d_{16} and d_9 provide more accurate parameter estimates than d_{17} in s_{25}, and d_{19} never performs well. The accuracy of d_{16} is always higher than that of d_9, implying that extending the time interval for taking observation was beneficial. Note that d_{18}, which found optimal designs for when a uniform distribution was assumed on λ, performs the best in two scenarios, and reasonably well in s_{25}, so it would seem a reasonable suggestion that practitioners should use d_{18} for experiments, even if it was not certain that heteroscedasticity will be accounted for during analysis. The ad hoc designs, d_{20} and d_{21}, never achieve a high relative accuracy of parameter estimation, so are not preferred for any scenario.

Table 4.23: DSE relative to minimum DSE for scenarios $s_{25} - s_{27}$ when the data were analysed considering heteroscedasticity

		Designs						
Scenario	Parameter	d_9	d_{16}	d_{17}	d_{18}	d_{19}	d_{20}	d_{21}
s_{25}	E_a	0.96	1.00	0.99	0.95	0.51	0.21	0.43
	k_r	0.97	1.00	0.99	0.97	0.63	0.51	0.63
s_{26}	E_a	0.84	0.91	0.96	1.00	0.75	0.34	0.44
	k_r	0.82	0.89	0.93	1.00	0.81	0.72	0.65
s_{27}	E_a	0.47	0.58	0.72	0.79	1.00	0.53	0.40
	k_r	0.42	0.56	0.71	0.79	1.00	0.80	0.61

Table 4.23 gives the DSE relative to the minimum DSE for each parameter and scenarios $s_{25} - s_{27}$, when the response and expected response were premultiplied by the transpose of $\left[\hat{\Delta}Q(\boldsymbol{\theta}_c)\hat{\Delta}\right]^{-1/2}$. Analysis allowing for heteroscedasticity gives anticipated results, with the Bayesian D-optimal designs performing the best for the scenario they were designed for, with the exception of s_{25}, where d_{16} slightly outperforms d_{17}. This is a somewhat surprising result, but d_{17} still has a very high relative performance under this scenario, so this is not too concerning. Design d_{18}, which was found assuming a uniform distribution over λ maintains a high relative accuracy in all three scenarios. We would thus recommend using d_{18} in practice assuming practitioners had no knowledge of the likely value of λ. Once heteroscedasticity is included in the analysis, d_{16} has an even higher relative accuracy than d_9 than in Table 4.22.

The DSE obtained when designs were analysed without considering heteroscedasticity relative to the DSE obtained when designs where analysed assuming heteroscedasticity was known and present was calculated, to investigate the gain in parameter accuracy obtained by considering heteroscedasticity while analysing. The results are in Table 4.24.

Table 4.24: DSE obtained when the data were analysed ignoring heteroscedasticity relative to the DSE obtained when the data were analysed considering heteroscedasticity for scenarios $s_{25} - s_{27}$

		Designs						
Scenario	Parameter	d_9	d_{16}	d_{17}	d_{18}	d_{19}	d_{20}	d_{21}
s_{25}	E_a	1.30	1.35	1.37	1.37	1.51	1.55	1.37
	k_r	1.16	1.20	1.30	1.25	1.36	1.37	1.30
s_{26}	E_a	2.15	2.30	2.43	2.46	4.53	5.37	2.96
	k_r	1.87	1.97	2.33	2.18	3.39	3.72	2.95
s_{27}	E_a	8.89	10.50	13.25	13.09	45.49	61.93	22.94
	k_r	6.57	8.19	11.81	10.45	30.89	27.82	22.40

In all cases, including heteroscedasticity during analysis gives a marked improvement in parameter estimates obtained. As the mean of λ in the scenario reduces, the gains in accuracy from including heteroscedasticity during analysis

become much larger. For example, in scenario s_{27}, where $\pi_3(\lambda)$ has a mean of 0.14, the ratio of DSE can become as large as 61.93. This is intuitive, as lower values of λ correspond to the data becoming more heteroscedastic. The ad hoc designs lose even more accuracy from not including heteroscedasticity in analysis than the Bayesian D-optimal designs. Henceforth, we will analyse the simulated data considering heteroscedasticity during analysis. We present the results for the edge scenarios when heteroscedasticity has been included during analysis. Table 4.25 gives the results for $s_{28} - s_{31}$, the edge scenarios when $\lambda \sim$Beta(4,1).

Table 4.25: DSE relative to minimum DSE for scenarios $s_{28} - s_{31}$. Speed of the reaction, determined by the value of E_a and k_r, is given in brackets next to the scenario label

		Designs						
Scenario	Parameter	d_9	d_{16}	d_{17}	d_{18}	d_{19}	d_{20}	d_{21}
s_{28} (fast)	E_a	1.00	0.98	1.00	0.78	0.28	0.05	0.54
	k_r	1.00	0.99	1.00	0.66	0.35	0.10	0.84
s_{29}(slow)	E_a	0.66	0.69	0.80	0.93	1.00	0.45	0.37
	k_r	0.63	0.67	0.77	0.89	1.00	0.89	0.63
s_{30}(fast)	E_a	1.00	1.00	1.00	0.87	0.33	0.06	0.61
	k_r	0.99	0.99	1.00	0.75	0.41	0.12	0.92
s_{31}(slow)	E_a	0.57	0.64	0.62	0.88	1.00	0.42	0.40
	k_r	0.66	0.71	0.64	0.89	1.00	0.75	0.73

The designs with the most accurate parameter estimates for fast reactions, as in scenarios s_{28} and s_{30} are d_9, d_{16} and d_{17}, which achieve similar relative accuracy of parameter estimation. All three designs take their observations earlier in time than the other designs included in this study. In the slow reaction scenarios, d_{19} gives the most accurate parameter estimates, and takes many observations much later in time. These results imply that mis-specifying the mean parameters during design is having more impact on the accuracy of parameter estimation than mis-specifying the prior distribution of λ.

Table 4.26 gives the results for $s_{32} - s_{35}$, the edge scenarios when $\lambda \sim$U(0,1).

For the slower scenarios d_{19} provides the most accurate parameter estimates, and

Table 4.26: DSE relative to minimum DSE for scenarios $s_{32} - s_{35}$. Speed of the reaction, determined by the value of E_a and k_r, is given in brackets next to the scenario label

		Designs						
Scenario	Parameter	d_9	d_{16}	d_{17}	d_{18}	d_{19}	d_{20}	d_{21}
s_{32} (fast)	E_a	0.96	0.99	1.00	0.98	0.54	0.14	0.41
	k_r	0.82	0.90	0.90	1.00	0.64	0.16	0.44
s_{33} (slow)	E_a	0.51	0.62	0.74	0.79	1.00	0.53	0.37
	k_r	0.51	0.54	0.77	0.74	1.00	0.78	0.53
s_{34} (fast)	E_a	0.82	0.90	0.90	1.00	0.64	0.16	0.44
	k_r	0.95	1.00	0.93	1.00	0.64	0.37	0.61
s_{35} (slow)	E_a	0.51	0.66	0.66	0.77	1.00	0.48	0.41
	k_r	0.49	0.71	0.73	0.87	1.00	0.80	0.81

when the reaction is occurring more quickly the design providing the most accurate parameter estimates is d_{18}, for which it was assumed that $\lambda \sim$ U(0,1).

Table 4.27 gives the results for $s_{36} - s_{39}$, the edge scenarios when $\lambda \sim$Beta(2,12.8571).

Table 4.27: DSE relative to minimum DSE for scenarios $s_{36} - s_{39}$. Speed of the reaction, determined by the value of E_a and k_r, is given in brackets next to the scenario label

		Designs						
Scenario	Parameter	d_9	d_{16}	d_{17}	d_{18}	d_{19}	d_{20}	d_{21}
s_{36} (fast)	E_a	0.87	0.83	0.87	0.96	1.00	0.69	0.44
	k_r	0.81	0.80	0.80	0.95	0.84	1.00	0.65
s_{37} (slow)	E_a	0.34	0.58	0.77	0.82	1.00	0.81	0.63
	k_r	0.28	0.49	0.68	0.71	0.88	1.00	0.79
s_{38} (fast)	E_a	0.87	0.87	0.70	0.91	1.00	0.64	0.53
	k_r	0.91	0.89	0.61	1.00	0.88	0.95	0.85
s_{39} (slow)	E_a	0.39	0.71	0.84	0.90	1.00	0.74	0.74
	k_r	0.34	0.65	0.77	0.84	0.92	0.97	1.00

The design d_{19} provides the most accurate estimates of E_a throughout, but for all scenarios except s_{38}, the ad hoc designs provide the most accurate estimates of k_r. The relative accuracy of d_{19} is fairly high in all scenarios, and the estimates of

E_a provided by the ad hoc designs often have a low relative accuracy, so the ad hoc designs would not be preferred.

When $\lambda = 1$ in the larger simulation study, d_9 outperformed the ad hoc designs in all edge scenarios, but once λ is allowed to vary this is no longer always true, especially for when $\lambda \sim$Beta(2,12.8571). This study demonstrates that specifying λ correctly does have an impact on which design provides the most accurate parameter estimates.

4.11 Conclusion

In this chapter two aspects of the performance of the designs obtained in Chapter 3 were investigated: the extent to which D-optimal designs are optimal for finite samples, and the robustness of the optimal designs to the assumptions under which they were designed being incorrect. There is little evidence of bias in the estimated parameters, but there is evidence that the estimates of σ are biased when correlation is ignored when the data are analysed. The designs studied were found to produce estimates of σ with similar levels of accuracy across all designs and scenarios. In addition, estimation assuming that correlation is not present reduced the accuracy of parameter estimates obtained from the D-optimal designs so that they were no longer optimal. This result is only apparent from simulation, as theoretical measures such as D-efficiency assume that correlation is known.

If a practitioner is not going to account for correlation in analysing the experimental data, then this study appears to show that a reasonable strategy would be to use an optimal design which remained robust to serial correlation, but ignores run-to-run error (assumed $r = 0$). The best course would be to include correlation in the analysis, as this gives better estimates of the mean parameters and ensures that variance estimates are not biased. The tools for analysing non-linear models when there is correlation between observations from the same run are not well established, but approaches to analysing data with correlation can be found in Chapter 7, Pinheiro and Bates (2004).

The designs which provided the most accurate parameter estimates on average were the Bayesian D-optimal designs. If the "true" parameters during experimentation are close to those expected before experimentation (as in s_1-s_{13}) then finding designs which assume a uniform prior distribution for the mean parameters gives accurate parameter estimates, while if the "true" parameter values were unexpectedly large or small, assuming a normal prior distribution for E_a and a gamma prior distribution for k_r gave the most accurate parameter estimates.

On average, the best prior distribution for τ appears to be a gamma distribution, for ρ a uniform distribution and the best prior distributions on r and q are log uniform prior distributions.

A study on the impact of heteroscedasticity on accuracy of parameter estimation was also conducted, which found that the Bayesian D-optimal designs were the best choice for remaining robust to the impacts of heteroscedasticity, in particular assuming a uniform distribution for λ, the parameter controlling the strength of heteroscedasticity, gave a design robust over several different scenarios. It was also discovered that not including heteroscedasticity during analysis could lead to a large drop in accuracy of parameter estimation.

This study is fairly comprehensive, but, as in previous chapters, assumed that the model given for the expected response and the correlation structure was correct. The impact of the assumed expected response being misspecified could be very large on the accuracy of parameter estimates. In particular, ad hoc designs, which were outperformed throughout this study, might outperform the Bayesian D-optimal designs in this case. Throughout this chapter it was assumed that the correlation and heteroscedasticity parameters were known in the model employed in the analysis, which will not be true in practice. This was done to allow simulation to be performed on a larger scale, but a more realistic simulation study would include the estimation of the correlation parameters and λ in the analysis of the simulated data.

Chapter 5

Cost considerations in design selection

5.1 Introduction

Cost of experimentation is an important aspect of experimental design. Not all observations will cost the same amount, and a design should use resource as efficiently as possible to attain accurate parameter estimates.

In this chapter cost considerations are incorporated into design selection via a method introduced by Elfving (1952). Suppose that the total budget for an experiment is c_T and the cost of making an observation at a support point $(T_i, \boldsymbol{t}_i)$ $(i = 1, \ldots, n)$ of an approximate design ξ is $c(T_i, \boldsymbol{t}_i)$. Then the total cost is constrained by

$$\sum_{i=1}^{n} n_i c(T_i, \boldsymbol{t}_i) \leq c_T,$$

where n_i is the number of observations taken at the support point $(T_i, \boldsymbol{t}_i)$. In this method, we normalise the information matrix for ξ by using:

$$M_{c_T}(\xi) = \sum_{i=1}^{n} w_i \frac{F_i' Q_i^{-1}(\boldsymbol{\theta}_c) F_i}{c(T_i, \boldsymbol{t}_i)} \tag{5.1}$$

where F is the sensitivity matrix and Q_i the matrix holding the correlation between observations from the ith run, as defined in Chapter 1. For a fixed cost c_T, the number of observations to be taken at each support point is obtained by rounding the value of $w_i c_T / c(T_i, \boldsymbol{t}_i)$.

This method can be used to attain approximate designs which control for the cost of observation; designs are found for model (1.7) in which errors are assumed to be homoscedastic. Throughout, we use the D-optimality criterion (given by Equation (1.14)) with the information matrix M replaced by M_{c_T}.

The chapter begins by briefly reviewing the literature on designs with cost considerations for non-linear models in Section 5.2. Section 5.3 introduces two cost functions, as well as the assumptions made on the mean and correlation parameter values when the experiment is designed. Section 5.4 explores the designs found for the first cost function considered, and the optimal m for given ratios of cost of an observation to cost of a run. Section 5.5 explores the designs found for the second cost function which includes the cost of the length of a run of a process. In Section 5.6 the difficulties of converting approximate designs into exact designs are considered, and some possible solutions proposed. Section 5.7 investigates the impact of cost considerations on the selection of Bayesian D-optimal designs through an example and, in the final section, some conclusions are given.

5.2 Literature review

There has been some investigation into finding approximate designs for non-linear models where the designs are weighted by the cost of observation.

Mentré et al. (1997) investigated optimal design for random effects regression models, and pharmacokinetic examples in particular. They found approximate locally D-optimal designs, using the method given in Section 5.1. They considered

four cost functions, which take account of one of the following: the cost of the number of samples taken on an individual, a penalty for taking more than one sample from an individual, the cost of taking new individuals into a study, and the time taken to sample from an individual. They adopted a model with two parameters, each varying at subject level. It was assumed that the values of the distribution of the random effects were known. They also assumed that an approximate design could be rounded without too much loss. Retout et al. (2009), using the same model and cost functions, created a modification of the Federov-Wynn algorithm which includes M_{c_T} in the objective function to find optimal designs.

Gagnon and Leonov (2005) investigated cost constraints in the design of a pharmacokinetic experiment for a non-linear model with random effects. They found locally D-optimal designs consisting of a sequence of observations (blood samples) where the frequency of sampling is controlled. The cost of an experiment was determined by the number of samples taken from each patient enrolled which had two forms: one where the cost per patient increased linearly with the number of samples taken, and one where the cost of an additional sample from a patient is proportional to the number of samples already taken. As might be anticipated, the higher the cost of an additional observation, the less frequently observations were taken. Fedorov and Leonov (2007) considered a very similar problem using the same model and cost function and obtaining similar results.

Dragalin et al. (2008) investigated the design for a dose response model, which is a bivariate probit model. The cost was based on the efficacy of the drug at the point of observation and hence was ethical, not financial. They found locally D-optimal designs and Bayesian D-optimal designs subject to these cost constraints. Unlike other examples considered in the literature and in this thesis, observations were assumed to be uncorrelated. Pronzato (2010) also applied cost functions in finding designs for dose response studies and a similar model. He approached the problem by solving it as a constrained design problem, rather than the standard D-optimal design problem with normalised information matrix as in Equation (5.1). He argued that the advantage of this approach is that in the unconstrained approach, finding

designs which do not have undesirable properties (low efficiency, or costs that are still too high) involves essentially ad hoc alteration of the cost function. He proposed that formulating the problem in the constrained manner makes clear the balance between gaining information and controlling cost. For the example considered in the thesis, the cost functions have a reasonably straight forward interpretation, so we did not investigate his approach.

Wright et al. (2010) investigated exact D-optimal designs for an example from pharmacokinetics of a dynamic non-linear model, without random effects. The authors argued that approximate designs are usually not practically useful and found exact designs instead. They developed a design search algorithm for the case with and without constraints on sampling cost. They found that the designs obtained were reasonably robust to the assumptions made on the values of the mean parameters.

5.3 Cost functions, criteria and parameter assumptions

We define two cost functions and the assumptions on the parameter values which will be used to find designs in later sections. In Equation (5.1), each support point $(T_i, \boldsymbol{t}_i)$, where $\boldsymbol{t}_i$ has length m, is weighted by a cost function, $c_j = c_j(\boldsymbol{t}_i, m)$, $j = 1, 2$, where:

$$c_1 = 1 + \frac{m-1}{\alpha}, \tag{5.2}$$

$$c_2 = 1 + \frac{m-1}{\alpha} + \beta\frac{\max(\boldsymbol{t}_i)}{200}. \tag{5.3}$$

where $\alpha \geq 1$ and $\beta > 0$ are constants to be specified, and $\max(\boldsymbol{t}_i)$ denotes the maximum observation time in the vector $\boldsymbol{t}_i$.

Cost function c_1 imposes a penalty for taking additional observations on each run. It is scaled by $\alpha \geq 1$ to allow for differences in cost between a run of the

process and observations taken within that run. Typically, it will cost less resource to take, for example, six observations per run than to conduct six runs, each with one observation. For this reason we have $\alpha \geq 1$ in Equations (5.2) and (5.3).

As we have seen in Chapter 3, the optimal designs for many scenarios tend to have observations at the maximum permitted time. However, running a process for a long time may be undesirable, and should incur a cost. The cost function c_2 extends c_1 to include the cost of the run time of the process. The cost is defined as the maximum time of observation in each run (the length of time that the process will be conducted) divided by the maximum permitted time, 200 here, scaled by $\beta > 0$.

Six different sets of assumptions on the model and correlation parameter values are used in later investigations of designs, i.e. in the optimal allocation of resource. For each set it was assumed that $E_a = 4.91 \times 10^4$, $k_r = 4.59 \times 10^{-4}$ and $\rho = 1$. The values assumed for τ, r and q are given in Table 5.1.

Table 5.1: Values for τ, r and q used to find designs

Set	τ	r	q
1	0.1	0	0
2	1	0	0
3	0.1	1	1
4	1	1	1
5	0.1	10	1
6	1	10	1

The values of τ, r and q were selected for the following reasons. In Sets 1 and 2 there is no run-to-run error or measurement error: serial correlation is present at low and high levels, $\tau = 1$ and $\tau = 0.1$ respectively. This allows an investigation into the impact of serial correlation on the optimal allocation of resource. Sets 3-6 have non-zero run-to-run error and measurement error. Sets 3-4 have r=1 so that the variance component for run-to-run error is at a comparable level to each of the other variance components. Sets 5-6 have $r = 10$ so that there is a relatively large variance component for run-to-run error.

5.4 Optimal designs for c_1

In this section we apply the cost function c_1 and explore the impact of different assumptions on the parameter values (Sets 1-6) and different costs of runs relative to number of observations taken within runs on the choice of optimal design.

We first investigate the optimal number of observations per run, by using the cost function c_1. Under this assumption, and given that m has the same value for each run of the process, then it is straightforward to find a design that is D-optimal for c_1 for a given value of α. This is because the value of α does not affect the determinant of the information matrix M_{c_T} for fixed m, and hence does not influence the optimal design obtained. Optimal designs can be found for a variety of m values and then the cost function used to determine which value of m is optimal for a given α.

The optimal designs for $m = 1, \ldots, 10$ were found for each of Sets 1-6 in Table 5.1. Figure 5.1 shows the optimal number of observations per run $0 < m \leq 10$ for $\alpha = 1, 2, \ldots, \alpha_{max}$, where α_{max} is the smallest value of α for which $m \geq 10$.

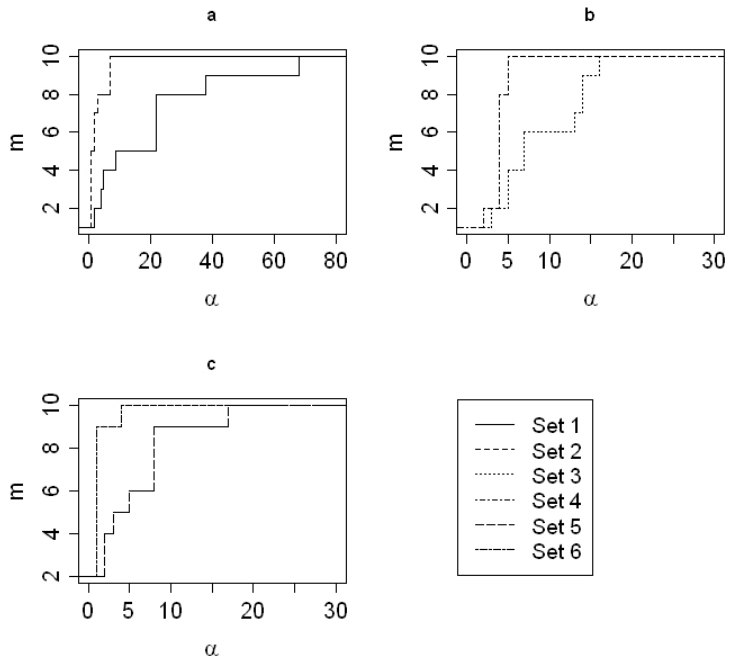


Figure 5.1: The optimal value for m (vertical axis) for a given α for each set of parameter assumptions: (a)-Sets 1,2 (b)-Sets 3,4 (c)-Sets 5,6

When only serial correlation is included in the statistical model (Sets 1 and 2),

the optimal number of observations per run increases as the cost of each additional observation per run is reduced as might be expected. This occurs much more quickly when the serial correlation between observations is weaker, as for Set 2, for which $m = 10$ is optimal for $\alpha \geq 5$. When there is strong serial correlation between observations on the same run (Set 1), less information is gained by increasing m and $m = 10$ only becomes optimal when $\alpha > 70$.

Similar results are obtained when run-to-run error and measurement error are included in the statistical model (Sets 3 and 4). For each set, the optimal m for the designs obtained becomes 10 for much lower values of α, due to the impact of run-to-run error. In contrast to serial correlation, which, as it increases, reduces the optimal value of m for a fixed α, increasing the value of $r = \sigma_r^2/\sigma_m^2$ increases the optimal value of m. For sufficiently large r the optimal value of m for $\alpha = 1$ is greater than one. This can be seen for Sets 5 and 6 in Figure 5.1 (c) where the optimal value for m is 2 when $\alpha = 1$. This result has been observed for linear models by Ju and Lucas (2002) and Goos and Vandebroek (2004) and was also explored in Subsection 3.6.1

5.5 Locally D-optimal designs for c_2

In this section we apply the cost function c_2 and explore designs obtained for a variety of values of α and β and assumptions on the parameter values.

As the cost of the length of a run of the process is included in c_2 (Equation (5.3)), the optimal value for m can no longer be obtained by finding the optimal design for each m, as the impact of m on the information matrix will depend on the maximum value of t in each support point.

In the following study, locally optimal designs are found using an algorithm which begins by assuming m=1, then finds the locally D-optimal design for m. The value of m is then increased by one and the same process is conducted until increasing m does not lead to an increase in the determinant of M_{c_T} obtained for the optimal design. Locally D-optimal designs were found for $\alpha = 1, 4, 8$ and $\beta = 0.5, 1, 2$, for

each set of assumptions (1-6). The optimal designs obtained for Set 1 are given in Table 5.2.

Table 5.2: Designs for cost function c_2 and Set 1 of parameter assumptions

α	β	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)			
1	0.5	38.3			
1	1	36.7			
1	2	34.4			
4	0.5	30.3	55.7		
4	1	29.2	53.3		
4	2	27.5	49.8		
8	0.5	25.6	45.8	68.2	
8	1	21.7	38.1	55.0	74.0
8	2	20.5	35.8	51.2	68.1
α	β	$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)			
1	0.5	10.0			
1	1	9.8			
1	2	9.6			
4	0.5	7.9	16.8		
4	1	7.8	16.2		
4	2	7.5	15.1		
8	0.5	1.0	5.9	14.7	
8	1	1.0	4.4	9.6	17.8
8	2	1.0	4.2	8.9	16.1

For Set 1 of parameter assumptions, the impact of adding a cost to the length of a run of the process is that the optimal design obtained makes its observations slightly earlier in time than the 54 designs found (but not shown) in the previous section. For higher values of β, the final observation becomes earlier in time.

Table 5.3 gives the designs for Set 2 which are similar to those for Set 1 except that when $\alpha = 1$, and $\beta \geq 1$ the optimal design has $m = 2$, one more observation than the design obtained for Set 1.

Tables 5.4 and 5.5 give the designs obtained for Set 3 and Set 4 of the parameter assumptions. We see that including a cost for length of a run has a large impact on the optimal designs obtained. The results in Chapter 3 indicated that the locally D-optimal design for $r = q = 1$ would be expected to have observations at the

Table 5.3: Designs for cost function c_2 and Set 2 of parameter assumptions

α	β	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)											
1	0.5	38.3											
1	1	35.6	41.5										
1	2	34.2	39.8										
4	0.5	26.2	30.6	34.7	38.6	42.4	46.3	50.3	54.5	59.1			
4	1	25.7	30.1	34.1	37.9	41.7	45.4	49.3	53.4	57.7			
4	2	24.9	29.2	33.1	36.8	40.4	44.0	47.7	51.5	55.5			
8	0.5	22.6	26.7	30.4	33.8	37.1	40.4	43.7	47.0	50.4	53.9	57.6	61.5	65.7
8	1	22.1	26.2	29.7	33.1	36.4	39.6	42.8	46.0	49.3	52.7	56.2	59.9	63.8
8	2	24.5	27.9	31.2	34.3	37.3	40.3	43.3	46.3	49.4	52.5	55.7	59.1	62.6
α	β	$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)											
1	0.5	9.97											
1	1	8.4	12.4										
1	2	8.3	12.2										
4	0.5	4.5	6.8	8.9	10.9	13.0	15.2	17.5	20.1	23.2			
4	1	4.5	6.8	8.8	10.9	12.9	15.0	17.3	19.9	22.9			
4	2	4.4	6.7	8.7	10.7	12.7	14.8	17.0	19.5	22.3			
8	0.5	3.7	5.7	7.4	9.0	10.7	12.3	14.0	15.8	17.6	19.6	21.7	24.2	27.1
8	1	3.7	5.6	7.3	8.9	10.6	12.2	13.9	15.6	17.4	19.3	21.4	23.8	26.5
8	2	3.4	5.3	6.9	8.5	11.5	13.0	14.6	16.2	18.0	20.6	21.8	23.9	26.4

allowed end points for time (1 and 200). However, the optimal design only includes $t = 1$ for $\alpha = 1$ and $\beta = 0.5$ and no observations are made at 200 in any optimal design. This is probably due to the cost of running a process until $t = 200$ being greater than the gain in information from having observations at that point.

Tables 5.6 and 5.7 hold the designs found for Sets 5 and 6 which are similar to those for Sets 3 and 4. However, r is larger than for Sets 3 and 4 and when $\beta = 0.5$ the optimal design does have an observation at 200 for the first support point. This implies that the loss from the cost of making observations at 200 is compensated by the information gained from the last observation.

Locally D-optimal designs found without cost considerations which have later observations times will be more affected by increased cost of a longer process run. The value assigned to β is important for determining whether these late observations are worthwhile or not.

Table 5.4: Designs for cost function c_2 and Set 3 of parameter assumptions

α	β	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)			
1	0.5	38.3			
1	1	36.7			
1	2	34.4			
4	0.5	32.6	47.9		
4	1	31.6	45.6		
4	2	30.1	42.2		
8	0.5	29.6	40.8	52.6	
8	1	28.8	39.2	49.4	
8	2	26.2	33.9	41.5	47.4
α	β	$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)			
1	0.5	10.0			
1	1	9.8			
1	2	9.6			
4	0.5	8.6	12.0		
4	1	8.6	11.8		
4	2	8.5	11.3		
8	0.5	1.0	7.8	11.6	
8	1	1.0	7.7	11.3	
8	2	1.0	6.9	9.2	11.8

Table 5.5: Designs for cost function c_2 and Set 4 of parameter assumptions

α	β	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)						
1	0.5	38.3						
1	1	36.7						
1	2	34.4						
4	0.5	36.5	41.2					
4	1	35.3	39.7					
4	2	32.5	36.2	39.6				
8	0.5	1.0	32.0	35.4	38.4	41.3	44.3	47.5
8	1	32.9	36.4	39.6	42.8			
8	2	31.3	34.6	37.5	40.3			
α	β	$\boldsymbol{t}_2$ ($T_2 = 100$, $w_1 = 0.5$)						
1	0.5	10.0						
1	1	9.8						
1	2	9.6						
4	0.5	8.8	11.7					
4	1	8.7	11.5					
4	2	7.9	10.1	12.3				
8	0.5	1.0	7.1	9.0	10.8	12.7	14.9	73.3
8	1	7.6	9.5	11.3	13.3			
8	2	7.4	9.3	11.0	12.9			

Table 5.6: Designs for cost function c_2 and Set 5 of parameter assumptions

α	β	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)						
1	0.5	1.0	38.5					
1	1	1.0	37.5					
1	2	1.0	31.2	43.8				
4	0.5	1.0	29.3	39.3	49.9			
4	1	1.0	28.6	38.2	47.4			
4	2	1.0	29.9	40.8				
8	0.5	1.0	27.0	37.3	48.0	61.1	182.2	200.0
8	1	1.0	26.0	32.1	38.1	44.4	49.3	
8	2	1.0	25.1	30.9	36.5	42.2	44.7	
α	β	$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)						
1	0.5	10.1	91.0					
1	1	1.0	9.0					
1	2	1.0	7.7	10.1				
4	0.5	8.6	12.4	82.6	110.7			
4	1	8.6	12.4	74.5	97.2			
4	2	1.0	7.6	9.8				
8	0.5	1.0	7.0	9.7	13.8	75.3	97.0	117.1
8	1	1.0	7.0	9.5	13.2	69.4	88.3	
8	2	1.0	7.0	9.4	13.1	61.8	75.5	

Table 5.7: Designs for cost function c_2 and Set 6 of parameter assumptions

α	β	$\boldsymbol{t}_1$ ($T_1 = 70$, $w_1 = 0.5$)													
1	0.5	40.4	200.0												
1	1	1.0	36.9	41.4											
1	2	1.0	35.0	38.8	42.3										
4	0.5	1.0	2.4	35.6	39.7	43.7									
4	1	1.0	2.5	33.9	37.5	40.8	44.2								
4	2	1.0	2.0	3.7	31.0	34.0	36.6	39.0	41.4	43.8	46.3				
8	0.5	1.0	31.7	35.2	38.4	41.5	44.6	47.9	51.7	185.8	190.0	193.7	197.0	200.0	
8	1	1.0	31.6	35.2	38.4	41.5	44.7	48.0	51.9	183.5	187.6	191.1	194.3	197.3	200.0
8	2	1.0	2.6	31.5	34.6	37.3	39.9	42.4							
α	β	$\boldsymbol{t}_2$ ($T_2 = 100$, $w_2 = 0.5$)													
1	0.5	10.1	91.2												
1	1	8.9	11.7	87.6											
1	2	8.9	11.9	81.6	86.8										
4	0.5	8.8	12.0	88.4	93.8	98.8									
4	1	8.0	10.5	13.2	81.6	86.8	91.4								
4	2	7.0	9.0	10.9	12.9	15.2	75.3	80.1	84.2	88.2	91.9				
8	0.5	6.4	8.1	9.6	11.1	12.7	14.5	16.6	86.8	92.3	97.3	102.1	106.8	111.3	
8	1	6.3	8.1	9.6	11.2	12.8	14.6	16.8	78.6	83.4	87.8	91.9	95.9	99.7	103.5
8	2	8.0	10.6	13.3	70.1	74.3	78.0	81.5							

5.6 Exact locally optimal designs

In practice, exact designs are used, as defined in Section 1.3. When cost considerations are included, rounding the weights of approximate designs to obtain exact designs is more difficult. This section explores methods of obtaining exact designs from approximate designs.

Given a fixed cost c_T, the number of runs of the process taken at a support point $(T_i, \boldsymbol{t}_i)$ should be $w(T_i, \boldsymbol{t}_i)c_T/c(T_i, \boldsymbol{t}_i)$, where $w(T_i, \boldsymbol{t}_i)$ is the weight of that support point, and $c(T_i, \boldsymbol{t}_i)$ is the cost of that support point. Due to the nature of the cost function, it may simply not be possible to round in a way that exactly adds up to c_T, leading to an over or under spend of resource. Given an approximate design, a practitioner might be reluctant to spend computational resources to determine what the exact design is, and might be interested in how much efficiency would be lost or gained by under or over spending and following the approximate design.

To investigate this for c_1, exact designs were found for $c_T = 6$, with $\alpha = 1, 2, 4, 6$. Table 5.8 gives the optimal values of m found for the approximate designs for each set of assumptions made on the parameters values. Table 5.9 gives the optimal values of N and m found for the exact designs for each value of α.

Table 5.8: Optimal values of m for the approximate designs for α=1, 2, 4, 6 and Sets $1, \ldots, 6$

	Set					
α	1	2	3	4	5	6
1	1	1	1	1	2	2
2	1	5	1	1	2	9
4	2	8	2	2	5	9
6	4	8	4	10	6	10

Table 5.10 gives the relative efficiency of the approximate design compared to the exact design, both for underspend and overspend, compared to the amount of under and overspend of resource, for each assumption made on the mean and correlation parameter values, for $c_T = 6$. Note that as all approximate designs have two support points, rounding up or down might lead to an odd number of runs. In those cases

Table 5.9: Optimal values of N and m for α=1, 2, 4, 6, when $c_T = 6$ and Sets $1, \ldots, 6$

		Set					
α		1	2	3	4	5	6
1	N	6	6	6	6	3	3
	m	1	1	1	1	2	2
2	N	6	2	6	6	4	2
	m	1	5	1	1	2	5
4	N	4	2	6	2	3	2
	m	3	9	1	9	5	9
6	N	4	2	4	2	3	2
	m	4	13	4	13	7	13

either support point can be chosen, as they have equal weighting. For parameter assumption Set six, $\alpha = 2$, there is no value given for underspend. This is due to the optimal m for the approximate design being nine. When $c_T = 6$, this means that for two runs of the process an experiment is over budget, and an experiment with only one run of the process would not give enough information for a practitioner to estimate E_a.

Table 5.10: Under/over efficiency, with ratio of under/over spend given in brackets of rounded approximate designs for all six sets of parameter assumptions. All values given to 2 d.p.

	Set					
α	1	2	3	4	5	6
1	1/1	1/1	1/1	1/1	1/1	1/1
	(1/1)	(1/1)	(1/1)	(1/1)	(1/1)	(1/1)
2	1/1	1/1	1/1	1/1	1/1	/1.68
	(1/1)	(1/1)	(1/1)	(1/1)	(1/1)	(/1.67)
4	0.85/1.04	0.92/1.30	0.87/1.07	0.87/1.07	1/1	1/1
	(0.83/1.04)	(0.92/1.38)	(0.83/1.04)	(0.83/1.04)	(1/1)	(1/1)
6	1/1	0.72/1.01	1/1	0.86/1.21	0.93/1.32	0.81/1.14
	(1/1)	(0.83/1.08)	(1/1)	(0.833/1.25)	(0.92/1.22)	(0.833/1.25)

Table 5.10 demonstrates that the gains or losses in efficiency by using a rounded approximate design are usually roughly equal to the increase or decrease in respective

costs. This means a practitioner who wishes to use approximate designs can weigh up the increase or decrease in cost by the same amount of loss or gain of accuracy in parameter estimation.

Another alternative to finding exact designs would be to use the following procedure:

- For a given resource c_T, round the approximate design to the 'under spending' exact design
- Append extra observation times to each run until all available resource is used

For $\alpha = 4$ and 6 the under spending exact designs were appended with random time points 1,000 times, and the mean D-efficiency across all designs was calculated. This is by no means an optimal method for obtaining designs, but gives the efficiency a practitioner might make by an ad hoc choice. Table 5.11 gives the averaged values.

Table 5.11: Relative efficiency of 'under spending' exact designs with randomly appended observation times, compared to the optimal exact design for all six sets. Improvement over underspend in Table 5.10 given in brackets

	Set					
α	1	2	3	4	5	6
4	0.90 (0.05)	0.94 (0.02)	0.95 (0.08)	0.97 (0.1)	1	1
6	1	0.80 (0.08)	1	0.94 (0.08)	0.98 (0.05)	0.90 (0.09)

This method improves the attained efficiency, and while the design will not be optimal, in nearly all cases the relative efficiency is greater than 0.9, so this method is not unreasonable for a practitioner who does not wish to have to search for an optimal exact design. Note that for $\alpha = 2$, Set 6, this method is not applicable, as the optimal approximate design has a higher value of m than there is resource available.

If the practitioner did not wish to simply append time points at random, they could instead follow a strategy of appending points at $\min(\boldsymbol{t})$-5 or $\max(\boldsymbol{t})$+5 for runs taken at $T = 70$ or 100 respectively, until all resource was used up. This gave the efficiencies in Table 5.12.

Table 5.12: Relative efficiency of 'under spending' exact designs with appended observation times, compared to the optimal exact design for all six sets of parameter assumptions. Improvement over Table 5.11 given in brackets (note that the improvement is rounded)

	Set					
α	1	2	3	4	5	6
4	0.92 (0.02)	0.99 (0.04)	0.96 (0.01)	0.99 (0.01)	0.89	NA
6	1	0.85 (0.05)	1	0.94 (-0.00)	0.98 (0.01)	0.91 (0.01)

In most cases this is an improvement over randomly appending observations to the under spending exact design, although for $\alpha = 6$ and Set 4 of the assumptions on the parameter values, appending the points is slightly less effective (the difference being -0.003) than adding random observations, so in general this is not a significant improvement over random appending of the designs.

Finding exact designs directly could be time consuming as c_T and α increases, and practitioners might lack the expertise to be able to do so. Appending approximate designs seems a reasonable alternative if that is the case.

5.7 Cost considerations for a Bayesian D-optimal design

In this section cost considerations are applied to finding Bayesian D-optimal designs, and an example of a Bayesian D-optimal design is found for c_1.

A Bayesian D-optimal design for c_1 was found for the prior distributions $E_a \sim N(4.91 \times 10^4, 4000^2)$, $k_r \sim \text{G}(m_k^2/v_k, v_k/m_k)$, with a mean of $m_k = 4.59 \times 10^{-4}$ and variance of $v_k = (4 \times 10^{-4})^2$, $\tau \sim G(1, 2)$, $\rho \sim U(0, 2)$, $\log(r) \sim U(-2.5, 2.5)$ and $\log(q) \sim U(-2.5, 2.5)$. This prior was chosen as it was found to be robust to a large range of mean and correlation parameters in Chapters 3 and 4. The optimal design was found for $m = 1, \ldots, 10$, and the optimal value of m for α was found, displayed in Figure 5.2.

The results are similar to those observed for the optimal designs obtained for

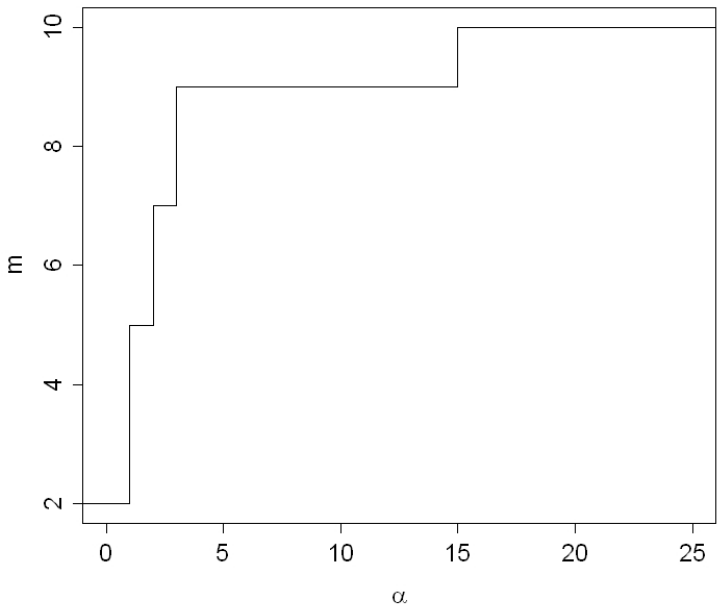


Figure 5.2: The optimal value of m for a Bayesian D-optimal design for given α and prior distributions given in Section 5.7

the sixth set of assumptions on the parameter values, as $m = 2$ when $\alpha = 1$, but it takes until $\alpha = 16$ for the optimal $m = 10$.

Locally D-optimal designs are found based on the unrealistic assumption that the parameter values are known before design, which Bayesian optimal design compensates for. However, Bayesian D-optimal designs will take longer to find due to computational issues, and including a cost constraint will increase the amount of time required to find the optimal design. For c_1, we needed to find 9 optimal designs (for $m = 2, \ldots, 10$), and for c_2 it becomes much harder to obtain the Bayesian D-optimal design. This is an area on which future research could focus.

5.8 Conclusion

In this chapter cost considerations were included while finding locally D-optimal and Bayesian D-optimal designs. The optimal number of observations that should be taken per run of the process was shown to depend on the cost function used

and the strength and type of correlation. When serial correlation is present and strong, designs with fewer observations per run are preferred, while if run-to-run error is present designs with more observations per experimental run are preferred. As observations per run become cheaper with respect to the cost of a run of the process, the optimal number of observations per run increases, as would be expected.

Including a cost for length of experimentation was shown to strongly affect designs when run-to-run error was believed to be present, as the optimal designs would normally take observations as late as possible in time. It was also demonstrated that, while approximate designs cannot always be directly converted into exact designs which use all available resource, several methods, such as appending additional random or uniformly spread observation times to the runs of the under-spending exact design obtained from the approximate design had high D-efficiencies.

Cost considerations were also considered for a Bayesian D-optimal design, and the results obtained were similar to those for the locally D-optimal designs.

This study into including cost considerations while designing has some short comings. Although the impact on the optimal allocation of cost for a range of different assumptions on the mean and correlation parameters was explored, the impact of mis-specifying these assumptions on the mis-allocation of resource was not investigated, and could provide an avenue for future research. Methods for rounding approximate designs were explored, and though no simple method of doing so was provided, it is unlikely that one exists. Two cost functions were investigated, which probably cover most of the cost concerns experimenters might have, although methods of obtaining exact designs from approximate designs for the second cost function, c_2, were not investigated, and doing so may prove difficult in practice. The errors were also assumed to be homoscedastic, and error heteroscedasticity could have a large impact on the optimal allocation of resource.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis a variety of issues have been explored which have been raised by finding optimal designs for experiments to estimate parameters in the mean function of non-linear models. These models arise in chemical kinetics, and we have studied an example provided by chemists at GlaxoSmithKline. The goals enumerated in Section 1.4 were investigated in full.

Analytical results were obtained for finding locally D-optimal designs for the considered example when independent, normally distributed errors were assumed, and the response and expected response were transformed to reduce heteroscedasticity. It was discovered that the optimal temperatures at which to conduct experiments were at the maximum and minimum permitted values, with the optimal times of observation given by Equation (2.5). Additionally, as errors became more heteroscedastic, the optimal design took observations later in time.

These results were applied to assist in obtaining locally D-optimal designs when correlation between observations from the same run was present, where the statistical model included both serial correlation and run-to-run error. It was discovered that the stronger serial correlation was, the further apart in time observations on the same run were made. When run-to-run error was present and the corresponding variance component was large, optimal designs made observations at the maximum

and minimum permitted times as well as observations close to those specified by Equation (2.5).

Mis-specifying the mean parameters, correlation parameters or transformation parameter was found to have a large impact on D-efficiency. For the mean parameters, mis-specifying k_r had more of an impact than mis-specifying E_a over the range of parameter values considered.

Bayesian D-optimal designs, were found for prior distributions specified across the mean, correlation and transformation parameters. The robustness of the design's performance to mis-specification of the parameter distributions was then assessed via a large simulation study, which demonstrated that the Bayesian D-optimal designs performed well over a range of scenarios, out-performing the locally D-optimal designs and ad hoc designs used by experimenters in the past. The simulation study also demonstrated that failing to include correlation and heteroscedasticity during analysis when either was present would lead to a large drop in the accuracy of the parameter estimates obtained.

Cost considerations were also incorporated into the design criterion. Two cost functions were considered. Both included the cost of taking multiple observations during a run of a process. The second cost function included a cost penalty for the length of time that a process was conducted. For both cost functions, the magnitude of serial correlation between observations from the same run was found to strongly affect the optimal number of observations per run: the stronger the correlation, the fewer observations per run required in the optimal design. The second cost function led to optimal designs with observations taken earlier in time.

An investigation was also conducted into how a practitioner, given an approximate design, could find exact designs to use in an experiment. It was demonstrated that methods such as appending additional random time points to runs from under spending designs could give a reasonable approximation to an optimal exact design.

6.2 Recommendations for practitioners

The investigation into optimal design for non-linear dynamic models conducted in this thesis allows us to give recommendations to practitioners working with similar non-linear dynamic models, when their goal during experimentation is to obtain accurate estimates of the mean parameters of the expected response.

We recommend finding Bayesian D-optimal designs, which have been shown to provide accurate parameter estimates of the mean parameters for a wide range of potential parameter values. The following process to determine the prior distributions for each parameter is recommended:

- If previous experimentation is available:
  - Mean parameters should be assigned prior probability distributions which best reflect information gathered from previous experimentation
  - Correlation parameters can be assigned prior distributions using a similar method as for the mean parameters if experimental data are judged to be of sufficient quality to accurately estimate the correlation parameters. Alternatively the following prior distributions, found to provide designs robust to correlation parameter values in this thesis, can be used:
    - $\tau \sim G(1, 2)$
    - $\rho \sim U(0, 10)$
    - $\log(r) \sim U(-2.5, 2.5)$
    - $\log(m) \sim U(-2.5, 2.5)$
  - If there is evidence of heteroscedasticity in the data, we would recommend using $\lambda \sim U(0, 1)$ as the prior distribution on λ
- If previous experimentation is not available:
  - Mean parameters can be given prior distributions specified by the best scientific knowledge of practitioners: probably uniform distributions over a specified interval

– Unless there is strong reason to believe that correlation or heteroscedasticity will not be present, we would recommend using the same prior distributions on those parameters as given above

For the first cost function presented in Chapter 5, which gave a ratio of cost between conducting additional runs of a process and taking more observations on each run, approximate designs should be found for a range of values of m for the assumed prior distributions on the mean and correlation parameters. These will determine, along with the cost ratio, the optimal number of observations per run which should be taken. When exact designs are required, if the exact design obtained from the approximate design does not spend all available resource, then it can be appended with random observation times, as outlined in Chapter 5. Finding optimal designs for the second cost function, which assigned a penalty to the length of experiment will be more difficult in practice, and we would recommend finding exact designs for the specific problem, rather than finding many approximate designs in advance, as the latter may prove difficult.

6.3 Future work

There are some extensions which could be applied to each chapter. The analytic results in Chapter 2 might be extended to a wider range of values of h, b and λ, as empirical investigation implied. It also might be possible to obtain analytic results for when experimenters also control the starting concentrations, a_0 and b_0, which were assumed fixed in this thesis.

There was some investigation into the relationship between the magnitude of the variance component for run-to-run error and when the optimal times of observation include the extreme values for time in Chapter 3. This work could be extended by an even larger empirical study to investigate the precise impact of the different values of mean parameters and number of observations per run as was explored briefly in Subsection 3.7.2. Another extension would be to model the impact of separate runs on the statistical errors through assuming non-linear random effects

rather than additive random effects. This is the approach frequently taken with pharmacokinetic models, as explored in the literature review in Section 3.2, and could have a substantial impact on the optimal designs obtained.

The simulation study in Chapter 4 assumed that the values of the correlation parameters and λ were known: a more comprehensive study might investigate the effect of estimating these parameters during analysis. We could assess how accurate the estimates of correlation would be in practice, and what impact this would have on the accuracy of mean parameter estimates.

In Chapter 5 the cost study could be extended to include more cost functions, such as making higher temperatures cost more. Better methods could also be explored for obtaining exact designs from approximate designs; the computational efficiency of a design algorithm which uses the approximate design plus appended points as starting points could be studied. Another extension could be to allow different numbers of observations to be taken in different runs, as there is no guarantee that the optimal design has equal number of observations in each run. Such an exploration could be computationally burdensome, as multiple experimental designs would need to be explored to find the optimum.

The methods used to obtain designs in this thesis could be applied to different examples in chemistry. In particular, the approach could be applied to a process where an analytic form of the expected response is not available; for example, where the expected response is available in the form of a system of differential equations with no closed form solution. The "direct method" of Bauer et al. (2000), popularised in the statistical literature by Atkinson and Bogacka (2002), could be applied. This approach combines the differential equations for the response model and the sensitivity equations (as given in Equation (1.13)) into simultaneous equations which can be solved numerically to find optimal designs. Finding optimal designs when the analytical form of the expected response is not available is likely to be more computationally intensive, and may require more efficient algorithms and methods for finding optimal designs than employed here.

During this thesis the primary focus has been on parameter estimation, while

assuming that the specified model is correct. This may not be a realistic assumption for some applications. In practice, the chemists may believe that several possible models could describe the reaction, and experimenters would need to discriminate between them before conducting experiments.

There is a large literature on model discrimination for dynamic models. The current focus of most frequentist model discrimination is on T-optimal designs, as given by Atkinson and Fedorov (1975). These include Ucinski and Bogacka (2004), Ucinski and Bogacka (2005) and López-Fidalgo et al. (2008). However, T-optimality is difficult to apply to more than two models at once, and this could be an area for more research. Some investigation of this has been conducted by Schwabb et al. (2006) who proposed a method for compromise between T-optimality and an approach given by Buzzi-Ferraris et al. (1984), for model discrimination.

Another problem with T-optimal designs is that the designs produced do not necessarily provide very accurate parameter estimation. In some cases, not all the parameters in the models considered can even be estimated for some T-optimal designs. One way to overcome this issue might be to design using a compound criterion, as investigated by Waterhouse et al. (2008), McGree et al. (2008), Dette et al. (2008) and Atkinson (2008a).

Appendix A

Proof of Lemmas in Chapter 2

A.1 Proof of Lemma 2.3.3

Lemma 2.3.3 *The function g(x) defined in Equation* (2.4) *has exactly one root,* x_0. *In addition,* $g(x) < 0\ \forall x < x_0$ *and* $g(x) > 0\ \forall x > x_0$.

Proof

Note $g(0) = h - b < 0$ and as $x \to \infty\ g(x) \to \infty$ so a root x_0 exists.

1: If $\lambda \geq 1$ and $h \geq 0$ then

$$g'(x) = b(x\lambda + \lambda - 1)e^x + h > 0.$$

So g(x) is monotonically increasing $\forall x$ and the lemma is proved.

2: Consider $\lambda < 1$ or $h < 0$. We have

$$g'(0) = b(\lambda - 1) + h.$$

As $x \to \infty$, $g'(x) \to \infty$. If $g'(0) < 0$, then $\exists x^* > 0$ s.t. $g'(x^*) = 0$.

If $g'(0) \geq 0$ then either $g'(x) > 0\ \forall x > 0$ (in which case, $g(x)$ is monotonically

increasing as before) or $\exists x^* > 0$ st $g'(x^*) = 0$. Then

$$g''(x) = b(x\lambda + 2\lambda - 1)e^x,$$

and inserting x* we obtain

$$g''(x^*) = b(x^*\lambda + 2\lambda - 1)e^{x^*}.$$

2a: When $\lambda \geq 1/2$

$$g''(x^*) \geq b(x^*\lambda)e^{x^*} > 0,$$

this implies that $g(x^*)$ is a unique local minimum for $g(x)$, as all roots x_r such that $g'(x_r) = 0$ must be minima, which implies there can be at most one. $g(0) < 0$, so $g(x^*) < 0$, so $x_0 > x^*$, and $g'(x) > 0$ for $x > x^*$, which proves the lemma.

2b: Suppose $\lambda < 1/2$ and $\lambda < \frac{b-h}{b}$, then

$$g'(0) = b(1 - \lambda) + h < 0.$$

We observe that $g''(x) = 0$ for

$$x_1 = \frac{1 - 2\lambda}{\lambda} > 0,$$

as $\lambda < 1/2$. In addition $g''(x) < 0 \ \forall x < x_1$ and $g''(x) > 0 \ \forall x > x_1$. This implies that $g'(x)$ is decreasing $\forall\ x < x_1$ and $g'(x)$ is increasing $\forall\ x > x_1$.

From previous argument a root x^* s.t. $g'(x^*) = 0$ exists, and as $g'(0) < 0$, $g'(x_1) < 0$, so $x^* > x_1$ and $g''(x^*) > 0$, which implies that $g(x^*)$ is a unique minimum for $g(x)$. As before, this is sufficient to demonstrate that the lemma is true.

2c: Suppose $\lambda < 1/2$ and $\lambda \exp(1) > h/b$. Then define x_r to be s.t. $g'(x_r) = 0$. As before, there exists at least one such x_r or $g'(x) > 0 \ \forall x$.

If $g(x_r) < 0 \ \forall\ x_r$ then $g(x)$ has only one root (since $g(0) < 0$ and x_0 exists by

previous argument) and $g(x) < 0\ \forall x < x_0$ and $g(x) > 0\ \forall x > x_0$. Note that

$$\begin{aligned} g(x_r) =& b(\lambda x_r + \lambda - 1)e^{x_r} - b\lambda e^{x_r} + hx_r + h \\ =& g'(x_r) + hx_r - b\lambda e^{x_r} = hx_r - b\lambda e^{x_r} \\ \leq& hx_r - b\lambda e^1 x_r \text{ (as } e^x \geq e^1 x \text{ for } x \geq 0) \\ <& hx_r - hx_r = 0, \end{aligned}$$

which is sufficient to demonstrate that the lemma is true. Thus the lemma is true for all conditions assumed for λ and h/b.

A.2 Proof of Lemma 2.3.4

Lemma 2.3.4 *For a two point design $d(T_1, t_1; T_2, t_2)$, with fixed positive values for T_1, T_2 and t_i $(i = 1, 2)$, the value of t_j, $j \neq i$ $(j = 1, 2)$, which maximises the determinant of the information matrix $M(\xi, \boldsymbol{\theta})$ is*

$$t_{T_j} = \frac{x_0}{c\theta_1 \exp(-l\theta_2/T_j)},$$

where x_0 is the root of the function

$$g(x) = b(\lambda x - 1)\exp(x) + h(x + 1).$$

Proof

WLOG, let $i = 2$ and $j = 1$. The proof proceeds by showing $\frac{\partial |M|}{\partial t_1} = 0$ at t_{T_1}, and that this is a maximum.

The information matrix for this model is

$$M = 1/2 \begin{pmatrix} M_1(T_1, t_1; T_2, t_2) & M_2(T_1, t_1; T_2, t_2) \\ M_2(T_1, t_1; T_2, t_2) & M_3(T_1, t_1; T_2, t_2) \end{pmatrix},$$

where

$$\begin{aligned}
M_1(T_1,t_1;T_2,t_2) =&\lambda^2\eta(T_1,t_1)^{2\lambda-2}\left(\frac{\partial\eta(T_1,t_1)}{\partial\theta_1}\right)^2\\
&+\lambda^2\eta(T_2,t_2)^{2\lambda-2}\left(\frac{\partial\eta(T_2,t_2)}{\partial\theta_1}\right)^2,\\
M_2(T_1,t_1;T_2,t_2) =&\lambda^2\eta(T_1,t_1)^{2\lambda-2}\frac{\partial\eta(T_1,t_1)}{\partial\theta_1}\frac{\partial\eta(T_1,t_1)}{\partial\theta_2}\\
&+\lambda^2\eta(T_2,t_2)^{2\lambda-2}\frac{\partial\eta(T_2,t_2)}{\partial\theta_1}\frac{\partial\eta(T_2,t_2)}{\partial\theta_2},\\
M_3(T_1,t_1;T_2,t_2) =&\lambda^2\eta(T_1,t_1)^{2\lambda-2}\left(\frac{\partial\eta(T_1,t_1)}{\partial\theta_2}\right)^2\\
&+\lambda^2\eta(T_2,t_2)^{2\lambda-2}\left(\frac{\partial\eta(T_2,t_2)}{\partial\theta_2}\right)^2.
\end{aligned}$$

This gives a determinant for M of

$$\begin{aligned}
|M| =&1/4\lambda^4\eta(T_1,t_1)^{2\lambda-2}\eta(T_2,t_2)^{2\lambda-2}\\
&\times\left(\frac{\partial\eta(T_1,t_1)}{\partial\theta_1}\frac{\partial\eta(T_2,t_2)}{\partial\theta_2}-\frac{\partial\eta(T_1,t_1)}{\partial\theta_2}\frac{\partial\eta(T_2,t_2)}{\partial\theta_1}\right)^2.
\end{aligned}$$

We define $A=A(T,t)=ct\theta_1\exp(-l\theta_2/T)$. The definition of A means that there is a bijection between t and A and that $A>0$. This gives

$$\begin{aligned}
\frac{\partial\eta(T_1,t_1)}{\partial\theta_1} &= \frac{-baA(T_1,t_1)\exp(A(T_1,t_1))}{\theta_1(b\exp(A(T_1,t_1))-h)^2},\\
\frac{\partial\eta(T_1,t_1)}{\partial\theta_2} &= -\frac{\partial\eta(T_1,t_1)}{\partial\theta_1}\left(\frac{l\theta_1}{T_1}\right),
\end{aligned}\tag{A.1}$$

which allows $|M|$ to be rewritten as

$$\begin{aligned}
|M| =&\frac{(l\theta_1)^2}{4}\lambda^4\eta(T_1,t_1)^{2\lambda-2}\eta(T_2,t_2)^{2\lambda-2}\\
&\times\left(\frac{\partial\eta(T_1,t_1)}{\partial\theta_1}\right)^2\left(\frac{\partial\eta(T_2,t_2)}{\partial\theta_1}\right)^2\left(\frac{1}{T_1}-\frac{1}{T_2}\right)^2.
\end{aligned}$$

The derivative of this function with respect to t_1 is then

$$\begin{aligned}\frac{\partial|M|}{\partial t_1} =&|M|\left[2\left(\frac{\partial\eta(T_1,t_1)}{\partial\theta_1}\right)^{-1}\left(\frac{\partial^2\eta(T_1,t_1)}{\partial\theta_1\partial t_1}\right)\right.\\ &\left.+(2\lambda-2)\left(\frac{\partial\eta(T_1,t_1)}{\partial t_1}\right)\eta(T_1,t_1)^{-1}\right],\end{aligned}$$

where

$$\begin{aligned}\frac{\partial^2\eta(T_1,t_1)}{\partial\theta_1\partial t_1} &= \left(\frac{\partial\eta(T_1,t_1)}{\partial\theta_1}\right)\left[\frac{1}{t_1}+\frac{A(T_1,t_1)}{t_1}-2\frac{bA(T_1,t_1)\exp(A(T_1,t_1))}{t_1(b\exp(A(T_1,t_1))-h)}\right].\\ &= -\left(\frac{1}{t_1(b\exp(A(T_1,t_1))-h)}\right)\left(\frac{\partial\eta(T_1,t_1)}{\partial\theta_1}\right)\\ &\quad\times\left[b\exp(A(T_1,t_1))(A(T_1,t_1)-1)+h(1+A(T_1,t_1))\right],\end{aligned}$$

and

$$\frac{\partial\eta(T_1,t_1)}{\partial t_1} = -\eta(T_1,t_1)\frac{b\exp(A(T_1,t_1))A(T_1,t_1)}{t_1(b\exp(A(T_1,t_1))-h)}.$$

Let $g^*(A) = [b(A-1)\exp(A)+h(1+A)]$, and
$g(A) = [b(A\lambda-1)\exp(A)+h(1+A)]$, then

$$\begin{aligned}\frac{\partial|M|}{\partial t_1} =&\frac{-2|M|\left[g^*(A(T_1,t_1))+b(\lambda-1)(A)\exp(A)\right]}{t_1(b\exp(A(T_1,t_1))-h)}\\ =&\frac{-2|M|g(A(T_1,t_1))}{t_1(b\exp(A(T_1,t_1))-h)}.\end{aligned}\tag{A.2}$$

Any $t_1 \in (t_{min}, t_{max})$ which maximises the determinant will lead to the derivative with respect to t_1 becoming 0. Clearly, if t_1 is chosen such that $g(A(T_1,t_1)) = 0$ then the derivative is 0. Clearly all other factors in (A.2) are nonzero for all values of t_1 except $t_1 = 0$ and $t_1{=}\infty$, neither of which would maximise the determinant.

By definition, x_0 is the value of A s.t. $g(x_0) = 0$. As $A > 0$, it follows from Lemma 2.3.3 that x_0 exists, and $g(A) < 0$ for all $A < x_0$ and $g(A) > 0$ for all $A > x_0$. We know $b > h$, so $be^A - h > 0$ which implies that $\frac{\partial|M|}{\partial t_1} > 0$ for all $A < x_0$

and $\frac{\partial|M|}{\partial t_1} < 0$ for all $A > x_0$. This implies that x_0 does maximise $|M|$.

A.3 Proof of Lemma 2.3.5

Lemma 2.3.5 *If the time t_{T_i} is in the interval $[t_{min}, t_{max}]$ for all T_i $(i = 1, 2)$, then the D-optimal two point design is $d(T_{max}, t_{T_{max}}; T_{min}, t_{T_{min}})$.*

Proof

If $T_1 = T_2$ then $|M| = 0$, so the optimal design will have unequal values for T_1 and T_2. WLOG assume that $T_1 > T_2$. Then

$$\begin{aligned}\frac{\partial|M|}{\partial T_1} =& 2|M|\bigg(\frac{1}{T_1 - T_2} - \frac{1}{T_1} + \left(\frac{\partial\eta(T_1, t_1)}{\partial\theta_1}\right)^{-1} \frac{\partial^2\eta(T_1, t_1)}{\partial\theta_1\partial T_1} \\ &+ \frac{\lambda - 1}{\eta(T_1, t_1)}\frac{\partial\eta(T_1, t_1)}{\partial T_1}\bigg). \end{aligned} \tag{A.3}$$

We define $A' = \frac{\partial A(T_1,t_1)}{\partial T_1}$ and obtain

$$\begin{aligned}\frac{\partial^2\eta(T_1, t_1)}{\partial\theta_1\partial T_1} =& -\frac{abA'\exp(A(T_1, t_1))}{\theta_1(b\exp(A(T_1, t_1)) - h)^2} \\ &\times \left[1 + A(T_1, t_1) - \frac{2bA(T_1, t_1)\exp(A(T_1, t_1))}{b\exp(A(T_1, t_1)) - h}\right], \\ =& -\frac{\partial\eta(T_1, t_1)}{\partial\theta_1}\frac{A'/A(T_1, t_1)}{(b\exp(A(T_1, t_1)) - h)}g^*(A),\end{aligned}$$

and

$$\frac{\partial\eta(T_1, t_1)}{\partial T_1} = -\eta(T_1, t_1)\frac{b\exp(A(T_1, t_1))A'}{b\exp(A(T_1, t_1)) - h}.$$

For any given value of T_1, the value of t_1 which maximises $|M|$ is t_{T_1}. Inserting this value into (A.3) gives

$$\begin{aligned}\frac{\partial|M|}{\partial T_1} &= 2|M|\left(\frac{1}{T_1 - T_2} - \frac{1}{T_1}\right), \\ &= 2|M|\left(\frac{T_2}{T_1(T_1 - T_2)}\right).\end{aligned}$$

This is positive for all T_1, so the optimal value for T_1 is T_{max}.

An analogous argument can be applied to show that the optimal value for T_2 is T_{min}.

A.4 Proof of Lemma 2.3.6

Lemma 2.3.6 *The D-optimal design for model* (2.1) *is the two support point design* $d(T_{max}, t_{T_{max}}; T_{min}, t_{T_{min}})$.

Proof

The equivalence theorem is applied to the design $d(T_{max}, t_{T_{max}}; T_{min}, t_{T_{min}})$. To save space, we write $T_{max} = T_x$, and $T_{min} = T_n$. Then

$$M^{-1} = \frac{1}{2|M|} \begin{pmatrix} M_3(T_x, t_{T_x}, T_n, t_{T_n}) & -M_2(T_x, t_{T_x}, T_n, t_{T_n}) \\ -M_2(T_x, t_{T_x}, T_n, t_{T_n}) & M_1(T_x, t_{T_x}, T_n, t_{Tn}) \end{pmatrix},$$

and obtain

$$\begin{aligned} v(T, t, \boldsymbol{\theta}, \xi) = & \frac{\lambda^2 \eta(T,t)^{2\lambda-2}}{2|M|} \left[\left(\frac{\partial \eta(T,t)}{\partial \theta_1} \right)^2 M_3(T_x, t_{T_x}, T_n, t_{T_n}) \right. \\ & + \left(\frac{\partial \eta(T,t)}{\partial \theta_2} \right)^2 M_1(T_x, t_{T_x}, T_n, t_{T_n}) \\ & \left. - 2 \frac{\partial \eta(T,t)}{\partial \theta_1} \frac{\partial \eta(T,t)}{\partial \theta_2} M_2(T_x, t_{T_x}, T_n, t_{T_n}) \right]. \end{aligned}$$

Using the identity (A.1), this expression is simplified to obtain:

$$\begin{aligned} v(T,t,\boldsymbol{\theta},\xi) =& \frac{V(T,t)}{2|M|}\Bigg[\left(\frac{\partial\eta(T_x,t_{T_x})}{\partial\theta_1}\right)^2 \eta(T_x,t_{T_x})^{2\lambda-2}\left(\frac{1}{T_x^2}+\frac{1}{T^2}-\frac{2}{T_xT}\right) \\ &+\left(\frac{\partial\eta(T_n,t_{T_n})}{\partial\theta_1}\right)^2 \eta(T_n,t_{T_n})^{2\lambda-2}\left(\frac{1}{T_n^2}+\frac{1}{T^2}-\frac{2}{T_nT}\right)\Bigg], \\ =& \frac{V(T,t)}{2|M|}\Bigg[\left(\frac{\partial\eta(T_x,t_{T_x})}{\partial\theta_1}\right)^2 \eta(T_x,t_{T_x})^{2\lambda-2}\left(\frac{1}{T_x}-\frac{1}{T}\right)^2 \\ &+\left(\frac{\partial\eta(T_n,t_{T_n})}{\partial\theta_1}\right)^2 \eta(T_n,t_{T_n})^{2\lambda-2}\left(\frac{1}{T_n}-\frac{1}{T}\right)^2\Bigg], \end{aligned}$$

where

$$V(T,t) = \lambda^4\eta(T,t)^{2\lambda-2}(l\theta_1)^2\left(\frac{\partial\eta(T,t)}{\partial\theta_1}\right)^2.$$

When $T = T_x$ and $t = t_{T_x}$ or $T = T_n$ and $t = t_{T_n}$, this expression simplifies to

$$\frac{4|M|}{2|M|} = 2 = p_m.$$

Now it must be shown that the support points are the maxima of $v(T,t,\boldsymbol{\theta},\xi)$. A fixed T is assumed and the expression $v(T,t,\boldsymbol{\theta},\xi)$ differentiated with respect to t

$$\begin{aligned} \frac{\partial v(T,t,\boldsymbol{\theta},\xi)}{\partial t} =& 2v(T,t,\boldsymbol{\theta},\xi)\Bigg[\left(\frac{\partial\eta(T_1,t_1)}{\partial\theta_1}\right)^{-1}\left(\frac{\partial^2\eta(T_1,t_1)}{\partial\theta_1\partial t_1}\right) \\ &+(\lambda-1)\left(\frac{\partial\eta(T_1,t_1)}{\partial t_1}\right)\eta(T_1,t_1)^{-1}\Bigg]. \end{aligned}$$

By the same argument used in the proof of Lemma 2.3.4, the derivative has a root at $t = t_T$ and $v(T,t,\boldsymbol{\theta},\xi)$ has a maximum at $t = t_T$. Note that for any $t = t_T$, $A = x_0$ which is a constant that does not depend on T or t. Hence $\frac{\partial\eta(T_x,t_{T_x})}{\partial\theta_1} = \frac{\partial\eta(T_n,t_{T_n})}{\partial\theta_1}$ and $\eta(T_n,t_{T_n}) = \eta(T_x,t_{T_x})$. This simplifies the expression to:

$$v(T,t_T,\boldsymbol{\theta},\xi) = \frac{V(T,t_T)W(T_x,t_{T_x})}{|M|}\left[\left(\frac{1}{T_x}-\frac{1}{T}\right)^2+\left(\frac{1}{T_n}-\frac{1}{T}\right)^2\right],$$

where

$$W(T_x, t_{T_x}) = \left(\frac{\partial \eta(T_x, t_{T_x})}{\partial \theta_1}\right)^2 (\eta(T_x, t_{T_x}))^{2\lambda-2}.$$

The derivative with respect to T is found.

$$\begin{aligned}\frac{\partial v(T, t_T, \boldsymbol{\theta}, \xi)}{\partial T} = \frac{V(T, t_T) W(T_x, t_{T_x})}{2|M|} \left[\frac{2}{T^2}\left(\frac{1}{T_x} - \frac{1}{T}\right) + \frac{2}{T^2}\left(\frac{1}{T_n} - \frac{1}{T}\right)\right] \\ + 2v(T, t, \boldsymbol{\theta}, \xi)\left[\left(\frac{\partial \eta(T_1, t_1)}{\partial \theta_1}\right)^{-1} \frac{\partial^2 \eta(T_1, t_1)}{\partial \theta_1 \partial T_1} \right. \\ \left. + \frac{\lambda - 1}{\eta(T_1, t_1)} \frac{\partial \eta(T_1, t_1)}{\partial T_1}\right].\end{aligned}$$

From the proof of Lemma 2.3.5 it can be seen that the second term in this sum is equal to 0 for $t = t_T$, which leaves:

$$\frac{\partial v(T, t_T, \boldsymbol{\theta}, \xi)}{\partial T} = \frac{V(T, t_T) W(T_x, t_{T_x})}{2|M|} \left[\frac{2}{T^2}\left(\frac{1}{T_x} - \frac{1}{T}\right) + \frac{2}{T^2}\left(\frac{1}{T_n} - \frac{1}{T}\right)\right].$$

This expression is equal to 0 only at

$$T^* = \frac{2T_x T_n}{T_x + T_n}.$$

When $T > T^*$ then

$$\frac{\partial v(T, t, \boldsymbol{\theta}, \xi)}{\partial T} > 0,$$

and for $T < T^*$ then

$$\frac{\partial v(T, t, \boldsymbol{\theta}, \xi)}{\partial T} < 0,$$

which implies that at T^* $v(T, t_T, \boldsymbol{\theta}, \xi)$ has a minimum, and thus $v(T, t_T, \boldsymbol{\theta}, \xi)$ is

maximal on the boundary of $[T_n, T_x]$. As

$$v(T_x, t_{T_x}, \boldsymbol{\theta}, \xi) = v(T_n, t_{T_n}, \boldsymbol{\theta}, \xi),$$

this expression obtains its maximum at its support points, where it is equal to p_m.

Thus the conditions for optimality as expressed in the equivalence theorem are satisfied, and the design is indeed optimal.

A.5 Proof of Corollary 2.3.7

Corollary 2.3.7 *For a two point design $d(T_1, t_1; T_2, t_2)$, with fixed positive values for T_1, T_2 and t_i $(i = 1, 2)$, if the value of t_j $(j = 1, 2)$ $j \neq i$ is constrained to be within $[t_{min}, t_{max}]$, then the value of $t_j = t_{opt}$ that maximises the determinant of the information matrix M is*

$$t_{opt} = \begin{cases} t_{min} & \text{if } t_{T_j} < t_{min} \\ t_{max} & \text{if } t_{T_j} > t_{max}. \end{cases}$$

Proof

WLOG let $j = 1$ and $i = 2$. The proof follows from the proof of Lemma 2.3.4, as $A = ct\theta_1 \exp(-l\theta_2/T)$. Thus if $t_{T_1} > t_{max}$ then $\frac{\partial |M|}{\partial t_1} > 0$ and t_{max} is the optimal choice. If $t_{T_1} < t_{min}$ then $\frac{\partial |M|}{\partial t_1} < 0$ and t_{min} is the optimal choice.

Bibliography

Asprey, S. P. and Macchietto, S. (2002) Designing robust optimal dynamic experiments. *Journal of Process Control*, **12**, 545–556.

Atkinson, A. C. (2003) Horwitz's rule, transforming both sides and the design of experiments for mechanistic models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **52**, 261–278.

— (2004) Some Bayesian optimum designs for response transformation in nonlinear models with nonconstant variance. *MODA 7 - Advances in Model-Oriented Design and Analysis, Heidelberg: Physica-Verlag*, 13–22.

— (2005a) Efficiencies for optimum designs when transforming the response in nonlinear models with nonconstant variance. *Metrika*, **62**, 127–138.

— (2005b) Robust optimum designs for transformation of the responses in a multivariate chemical kinetic model. *Technometrics*, **47**, 478–487.

— (2008a) DT-optimum designs for model discrimination and parameter estimation. *Journal of Statistical Planning and Inference*, **138**, 56–64.

— (2008b) Examples of the use of an equivalence theorem in constructing optimum experimental designs for random-effects nonlinear regression models. *Journal of Statistical Planning and Inference*, **138**, 2595–2606.

Atkinson, A. C. and Bogacka, B. (1997) Compound D- and D_S-optimum designs for determining the order of a chemical reaction. *Technometrics*, **39**, 347–356.

— (2002) Compound and other optimum designs for systems of non-linear differential equations arising in chemical kinetics. *Chemometrics and Intelligent Laboratory Systems*, **61**, 17–33.

Atkinson, A. C., Chaloner, K., Herzberg, A. M. and Juritz, J. (1993) Optimum experimental designs for properties of a compartmental model. *Biometrics*, **49**, 325–337.

Atkinson, A. C., Donev, A. N. and Tobias, R. D. S. (2007) *Optimum Experimental Designs, with SAS*. Oxford University Press.

Atkinson, A. C. and Fedorov, V. V. (1975) The design of experiments for discriminating between two rival models. *Biometrika*, **62**, 57–70.

Barz, T., Arellano-Garcia, H. and Wozny, G. (2010) Handling uncertainty in model-based optimal experimental design. *Industrial and Engineering Chemistry Research*, **49**, 5702–5713.

Bates, D. and Watts, D. (1988) *Nonlinear Regression Analysis and its Applications*. John Wiley & Sons.

Bauer, I., Bock, H., Körkel, H. and Schlöder, J. (2000) Numerical methods for optimum experimental design in DAE systems. *Journal of Computational and Applied Mathematics*, **120**, 1–25.

Bogacka, B., Patan, M., Johnson, P. J., Youdim, K. and Atkinson, A. C. (2011) Optimum design of experiments for enzyme inhibition kinetic models. *Journal of Biopharmaceutical Statistics*, **21**, 555–572.

Bogacka, B. and Wright, F. J. (2005) Non-linear design problem in a chemical kinetic model with non-constant error variance. *Journal of Statistical Planning and Inference*, **128**, 633–648.

Braess, D. and Dette, H. (2007) On the number of support points of maximin and Bayesian optimal designs. *Annals of Statistics*, **35**, 772–792.

Buzzi-Ferraris, G., Forzatti, P., Emig, G. and Hofmann, H. (1984) Sequential experimental design procedure for model discrimination in the case of multiple responses. *Chemical Engineering Science*, **39**, 81–85.

Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995) A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing*, **16**, 1190–1208.

Cassity, C. (1965) Abscissas, coefficients, and error term for the generalized gauss-laguerre quadrature formula using the zero ordinate. *Mathematics of Computation*, **19**, 287–296.

Chaloner, K. and Verdinelli, I. (1995) Bayesian experimental design: a review. *Statistical Science*, **10**, 273–304.

Cheng, C. (1995) Optimal regression designs under random block-effects models. *Statistica Sinica*, **5**, 485–497.

Chu, Y. and Hahn, J. (2008) Integrating parameter selection with experimental design under uncertainty for nonlinear dynamic systems. *AIChE*, **54**, 2310–2320.

Dette, H. and Biedermann, S. (2003) Robust and efficient designs for the Michaelis-Menten model. *Journal of the American Statistical Association*, **98**, 679–686.

Dette, H., Bretz, F., Pepelyshev, A. and Pinheiro, J. (2008) Optimal designs for dose-finding studies. *Journal of the American Statistical Association*, **103**, 1225–1237.

Dette, H., Melas, V. B., Pepelyshev, A. and Strigul, N. (2005) Robust and efficient design of experiments for the Monod model. *Journal of Theoretical Biology*, **234**, 537–550.

Dette, H., Pepelyshev, A. and Holland-Letz, T. (2010) Optimal designs for random effect models with correlated errors with applications in population pharmacokinetics. *The Annals of Applied Statistics*, **4**, 1430–1450.

Dette, H., Pepelyshev, A., Melas, V. B. and Strigul, N. (2003) Efficient design of experiments in the Monod model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 725–742.

Dragalin, V., Fedorov, V. V. and Wu, Y. (2008) Adaptive designs for selecting drug combinations based on efficacy-toxicity response. *Journal of Statistical Planning and Inference*, **138**, 352–373.

Dufful, S. B., Mentré, F. and Aarons, L. (2001) Optimal design of a population pharmacodynamic experiment for ivabradine. *Pharmaceutical Research*, **18**, 83–89.

Elfving, G. (1952) Optimum allocation in linear regression theory. *Annals of Mathematical Statistics*, **23**, 255–262.

Fang, K. T., Li, R. and Sudjianto, A. (2006) *Design and Modelling for Computer Experiments*. Chapman and Hall.

Fedorov, V. V. and Leonov, S. (2007) Population pharmacokinetic measures, their estimation and selection of sampling times. *Journal of Biopharmaceutical Statistics*, **17**, 919–941.

Firth, D. and Hinde, J. P. (1997) On Bayesian D-optimum design criteria and the equivalence theorem in non- linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, **59**, 793–797.

Franceschini, G. and Macchietto, S. (2008a) Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, **63**, 4846–4872.

— (2008b) Novel anticorrelation criteria for model-based experiment design: theory and formulations. *AIChE*, **54**, 1010–1024.

Gagnon, R. and Leonov, S. (2005) Optimal population designs for PK models with serial sampling. *Journal of Biopharmaceutical statistics*, **15**, 143–163.

Geudj, J., Bazzoli, C., Neumann, A. U. and Mentré, F. (2011) Design evaluation and optimization for models of hepatitis C viral dynamics. *Statistics in Medicine*, **30**, 1045–1056.

Goos, P. and Vandebroek, M. (2004) Outperforming completely randomized designs. *Journal of Quality Technology*, **36**, 12–26.

Gotwalt, C. M., Jones, B. A. and Steinberg, D. M. (2009) Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics*, **51**, 88–95.

Gregson, J. (2009) *Pseudo-Bayesian Design of Experiments*. Master's thesis, School of Mathematics, University of Southampton.

Gueorguieva, I., Aarons, L., Ogungbenro, K., Joga, K. M., Rodgers, T. and Rowland, M. (2006) Optimal design for multivariate response pharmacokinetic models. *Journal of Pharmacokinetics and Pharmacodynamics*, **33**, 97–124.

Hamada, M., Martz, H. F., Reese, C. S. and Wilson, A. G. (2001) Finding near-optimal Bayesian experimental designs via genetic algorithms. *The American Statistician*, **55**, 175–181.

Hughes-Oliver (1998) Optimal designs for nonlinear models with correlated errors. *IMS Lecture Notes- Monograph Series*, **34**, 163–174.

Ju, H. L. and Lucas, J. M. (2002) L^k factorial experiments with hard-to-change and easy-to-change factors. *Journal of Quality Technology*, **34**, 411–421.

Kiefer, J. and Wolfowitz, J. (1960) The equivalence of two extremum problems. *Canadian Journal of Mathematics*, **12**, 363–366.

Laidler, K. J. (1984) The development of the Arrhenius equation. *Journal of Chemical Education*, **61**, 494.

Lischer, P. (1999) Good statistical practice in analytical chemistry. In *Probability Theory and Mathematical Statistics*, 1–12. Dordrecht: VSP.

López-Fidalgo, J., Tommasi, C. and Trandafir, P. C. (2008) Optimal designs for discriminating between some extensions of the Michaelis-Menten model. *Journal of Statistical Planning and Inference*, **138**, 3797–3804.

Mannaswamy, A., Munson-McGee, S. H. and Anersen, P. K. (2010) D-optimal designs for the cross viscosity model applied to guar gum mixtures. *Journal of Food Engineering*, **97**, 403–409.

Matthews, J. N. S. and Allcock, G. C. (2004) Optimal designs for Michaelis-Menten kinetic studies. *Statistics in Medicine*, **23**, 477–491.

Matthews, J. N. S. and James, P. W. (2005) Restricted optimal design in the measurement of cerebral blood flow using the Kety-Schmidt technique. In *Applied Optimal Designs*, chap. 8. John Wiley & Sons.

McGree, J. M., Eccleston, J. A. and Duffull, S. B. (2008) Compound optimal design criteria for nonlinear models. *Journal of Biopharmaceutical Statistics*, **18**, 646–661.

McKay, M. D., Conover, W. and Beckman, R. (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**, 239–240.

Mentré, F., Mallet, A. and Baccar, D. (1997) Optimal design in random-effects regression models. *Biometrika*, **84**, 429–442.

Monahan, J. and Genz, A. (1997) Spherical-radial integration rules for Bayesian computation. *Journal of the American Statistical Association*, **92**, 664–674.

Ogungbenro, K. and Aarons, L. (2008) Optimisation of sampling windows design for population pharmacokinetic experiments. *Journal of Pharmacokinetic Pharmacodynamics*, **35**, 465–482.

Owen, M., Luscombe, C., Lai, L., Godbert, S. Crookes, D. and Emiabata-Smith, D. (2001) Efficiency by design: Optimisation in process research. *Organic Process Research & Development*, **5**, 308–323.

Patan, M. and Bogacka, B. (2007) Optimum experimental designs for dynamic systems in the presence of correlated errors. *Computational Statistics and Data Analysis*, **51**, 6544–5661.

Pinheiro, J. and Bates, D. (2004) *Mixed Effects Models in S and S-Plus*. Springer.

Press, W., Teuklosky, S., Vetterling, W. and Flannery, B. (1992) *Numerical Recipes in C (2nd edition)*. New York: Cambridge University Press.

Pronzato, L. (2010) Penalised optimal designs for dose finding. *Journal of Statistical Planning and Inference*, **140**, 283–296.

Pukelsheim, F. and Rieder, S. (1992) Efficient rounding of approximate designs. *Biometrika*, **79**, 763–770.

Retout, S., Comets, E., Bazzoli, C. and Mentré, F. (2009) Design optimization in nonlinear mixed effects models using cost functions: application to a joint model of infliximab and methotrexate pharmacokinetics. *Communications in Statistics-Theory and Methods*, **38**, 3351–3368.

Retout, S., Comets, E., Samson, A. and Mentré, F. (2007) Design in nonlinear mixed effects models: Optimization using the Fedorov-Wynn algorithm and power of the Wald test for binary covariates. *Statistics in Medicine*, **26**, 5162–5179.

Retout, S. and Mentré, F. (2003) Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics*, **13**, 209–227.

Rodriguez-Aragón, L. J. and López-Fidalgo, J. (2005) Optimal designs for the Arrhenius equation. *Chemometrics and Intelligent Laboratory Systems*, **77**, 131–138.

Rodríguez-Díaz, J. M. and Santos-Martín, M. T. (2009) Study of the best designs for modifications of the Arrhenius equation. *Chemometrics and Intelligent Laboratory Systems*, **95**, 199–208.

Schwabb, M., Silva, F. M., Queipo, C. A., Barreto Jr., A. G., Nele, M. and Pinto, J. C. (2006) A new approach for sequential experimental design for model discrimination. *Chemical Engineering Science*, **61**, 5791–5806.

Silvey, S. D. (1980) *Optimum Design*. Chapman and Hall, London.

Song, D. and Wong, W. K. (1998) Optimal two-point designs for the Michaelis-Menten model with heteroscedastic errors. *Communications in Statistics- Theory and Methods*, **27**, 1503–1516.

Stehlík, M., Rodríguez-Díaz, J. M., Müller, W. and López-Fidalgo, J. (2008) Optimal allocation of bioassays in the case of parameterized covariance functions: an application to Lung's retention of radioactive particles. *Test*, **17**, 56–68.

Tang, B. (1993) Orthogonal array-based latin hypercubes. *Journal of the American Statistical Association*, **88**, 1392–1397.

Ucinski, D. and Atkinson, A. C. (2004) Experimental design for time-dependent models with correlated observations. *Studies in Nonlinear Dynamics & Econometrics*, **8**, 206–219.

Ucinski, D. and Bogacka, B. (2004) Heteroscedastic T-optimum designs for multiresponse dynamic models. *MODA 7 - Advances in Model-Oriented Design and Analysis, Heidelberg: Physica-Verlag*, 191–199.

— (2005) T-optimum designs for discrimination between two multiresponse dynamic models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**, 3–18.

Wang, Y., Ekridge, K. M. and Nadarajah, S. (2012) Optimal design of mixed-effects PK/PD models based on differential equations. *Journal of Biopharmaceutical Statistics*, **22**, 180–205.

Waterhouse, T. H., Woods, D. C., Eccleston, J. A. and Lewis, S. M. (2008) Design selection criteria for discrimination/estimation for nested models and a binomial response. *Journal of Statistical Planning and Inference*, **138**, 132–144.

Woods, D. C. and van de Ven, P. (2011) Blocked designs for experiments with correlated non-normal response. *Technometrics*, **53**, 173–182.

Wright, S. E., Sigal, B. M. and Bailer, A. J. (2010) Workweek optimization of experimental designs: exact designs for variable sampling costs. *Journal of Agricultural, Biological, and Environmental Statistics*, **15**, 491–509.