

Bayesian inference about outputs of computationally expensive algorithms with uncertainty on the inputs

by Richard G.E. Haylock, MSc

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy,
May, 1997

CONTENTS

1	Introduction	7
1.1	Motivation for this research	7
1.2	Human exposure to radiation	8
1.3	Internal dosimetry	13
2	Introduction to Uncertainty analysis	16
2.1	What is uncertainty analysis?	16
2.2	Parameter value uncertainty analysis	17
2.2.1	Classical design selection	20
2.2.2	Classical parameter uncertainty analysis	21
2.2.3	Alternative methodology for parameter uncertainty analysis	23
2.3	Predicting an algorithm's output using stochastic process techniques	24
2.3.1	The classical predictor	27
2.3.2	Examples of the use of the classical predictor	29
2.3.3	The Bayesian predictor	30
2.3.4	Examples of the use of the Bayesian predictor . . .	31
2.3.5	Criterion based design selection	33
2.4	The specific objectives of this research	35
2.5	Overview of the remaining chapters	36
3	General methodology for Bayesian uncertainty analysis	38
3.1	Introduction	38
3.2	Development of the Bayesian model	39
3.3	Predicting the output of the algorithm	40

3.3.1	Calculating the probability that the true output of the algorithm will exceed a critical value	42
3.4	Estimating the mean of the algorithm's uncertainty distribution	43
3.5	Estimating the variance of the algorithm's uncertainty distribution	44
3.5.1	Calculation of the posterior expectation of L	45
3.5.2	Calculation of the posterior variance of L	46
3.6	Selection of optimum design points	48
4	Methodology for Bayesian uncertainty analysis of uncertain parameters with normal prior distributions	53
4.1	Selection of the design points	56
4.2	Updating the smoothing parameter values	57
4.3	Estimation of $\eta(\mathbf{x}_0)$ for normally distributed uncertain parameters	59
4.4	Estimation of K the expected value of $\eta(\mathbf{X})$ for normally distributed uncertain parameters	59
4.5	Estimation of L the variance of $\eta(\mathbf{X})$ for normally distributed uncertain parameters	62
4.5.1	Calculation of the posterior expectation of L	62
4.5.2	Calculation of the posterior variance of L	64
5	Comparison of Bayesian and classical uncertainty analysis methodologies using an algorithm for the calculation of dose due to ingestion of ^{131}I	72
5.1	Introduction	72
5.2	A simplified recycling algorithm for ^{131}I	73
5.3	A classical analysis of the uncertainties associated with the ^{131}I algorithm	75
5.3.1	The uncertain parameters of the simplified recycling algorithm	75
5.3.2	The methodology of the classical uncertainty analysis	76

5.4	Calculation of the ‘true uncertainty’ associated with the ^{131}I algorithm	77
5.5	Bayesian analysis of the ^{131}I algorithm	78
5.5.1	Selection of the design points	79
5.5.2	Updating the smoothing parameter values	80
5.5.3	Estimation of the value of the ^{131}I algorithm for specified values of lw and lf	80
5.5.4	Estimation of the mean and variance of the ^{131}I algorithm over the distributions of the uncertain parameters	81
5.6	Results	81
5.6.1	The ‘true’ uncertainty	81
5.6.2	The NRPB Monte Carlo analysis	82
5.6.3	The Bayesian analysis	84
5.7	Discussion	88
6	Comparison of Bayesian and classical uncertainty analysis methodologies using the ICRP 67 model for plutonium metabolism	92
6.1	Introduction	92
6.2	The development of the ICRP 67 biokinetic model for plutonium metabolism	93
6.3	Why is an uncertainty analysis of the Pu model required? .	98
6.4	Classical uncertainty analysis of the Pu model	98
6.4.1	Selection of the parameters on which to perform the classical analysis	98
6.4.2	Details of the analysis performed	99
6.5	Bayesian uncertainty analysis of the Pu model	101
6.5.1	Transforming the parameters	101
6.5.2	Selection of the design points	102
6.5.3	Updating the smoothing parameter values	104
6.5.4	Details of the analyses performed	105
6.6	Results of the classical uncertainty analysis	105
6.7	Results of the Bayesian uncertainty analysis	106

6.8	Discussion	110
7	Comments, discussion and future work	113
7.1	Future work	126
8	Appendix : the design selection and analysis programs	127
8.1	Introduction	127
8.2	The Design selection program	128
8.2.1	Reloading an old problem or defining a new problem	129
8.2.2	Defining/editing the parameters of the problem . .	130
8.2.3	Selecting a design	132
8.3	The Analysis program	137
8.3.1	Loading the details of the problem	138
8.3.2	Analysis of the selected problem	139
8.3.3	Updating the smoothing parameters	142
8.3.4	Sensitivity Plots	143
8.4	Discussion	145

LIST OF FIGURES

1.1	Breakdown of the average annual radiation dose to the population of the UK.	9
5.1	A graphical representation of the ^{131}I algorithm.	74
5.2	Frequency histogram of the sample output of the Monte Carlo analysis	83
5.3	Contour plot of the probability that the output of the algorithm for particular w and f values will exceed the predicted mean output value	89
7.1	Evaluation of the ^{131}I algorithm, extending to approximately 2.5 standard deviations either side of the mean values . . .	118
7.2	Surface representation of the percentage errors in the point estimates generated by the Bayesian analysis using design y_{15} .	120
7.3	Surface representation of the percentage errors in the point estimates generated by the Bayesian analysis using design y_{20} .	121
7.4	Uncertainty plot for the parameter ‘Loss from testes’ based on a 50 point design.	124
7.5	Uncertainty plot for the parameter ‘Loss from testes’ based on a 150 point design.	125

LIST OF TABLES

5.1	Parameters of the lognormal prior distributions	76
5.2	Measures of the ‘true’ uncertainty	81
5.3	Results of the Monte Carlo analysis	83
5.4	The improved estimates for δ_{lw} and δ_{lf}	84
5.5	Summary statistics for the quality of the estimates of CEDE	85
5.6	Parameters of the posterior t distribution for K , the mean of the CEDE	86
5.7	Mean and Variance of the posterior distribution for L , the variance of the CEDE	87
6.1	Standard transfer rates per day for the Plutonium model .	97
6.2	The parameters (day^{-1}) and their subjective prior distribu- tions to which the plutonium model is sensitive. ($\#$ = upper and lower bounds $\#\#$ = shape parameters)	100
6.3	Results of the Monte Carlo analysis	106
6.4	Mean and Variance of the posterior distribution for K , the mean of the CEDE (Prior5 : mean values, Prior8 : median values)	108
6.5	Mean and Variance of the posterior distribution for L , the variance of the CEDE	109
6.6	Consistency checks on the Bayesian analyses (values in the table are percentages).	111

Abstract

In the field of radiation protection, complex computationally expensive algorithms are used to predict radiation doses, to organs in the human body from exposure to internally deposited radionuclides. These algorithms contain many inputs, the true values of which are uncertain. Current methods for assessing the effects of the input uncertainties on the output of the algorithms are based on Monte Carlo analyses, i.e. sampling from subjective prior distributions that represent the uncertainty on each input, evaluating the output of the model and calculating sample statistics. For complex computationally expensive algorithms, it is often not possible to get a large enough sample for a meaningful uncertainty analysis. This thesis presents an alternative general theory for uncertainty analysis, based on the use of stochastic process models, in a Bayesian context. The measures provided by the Monte Carlo analysis are obtained, plus extra more informative measures, but using a far smaller sample. The theory is initially developed in a general form and then specifically for algorithms with inputs whose uncertainty can be characterised by independent normal distributions.

The Monte Carlo and Bayesian methodologies are then compared using two practical examples. The first example, is based on a simple model developed to calculate doses due to radioactive iodine. This model has two normally distributed uncertain parameters and due to its simplicity an independent measurement of the true uncertainty on the output is available for comparison. This exercise appears to show that the Bayesian methodology is superior in this simple case. The purpose of the second example is to determine if the methodology is practical in a ‘real-life’ situation and to compare it with a Monte Carlo analysis. A model for calculating doses due to plutonium contamination is used. This model is computationally expensive and has fourteen uncertain inputs. The Bayesian analysis compared favourably to the Monte Carlo, indicating that it has the potential to provide more accurate uncertainty analyses for the parameters of computationally expensive algorithms.

1. INTRODUCTION

1.1. Motivation for this research

The aim of this thesis is to develop and test a general theory for performing uncertainty analysis. The motivation for this project is related to my background as a statistician working for the National Radiological Protection Board (NRPB). This body was established by the Radiological Protection Act, 1970, to provide advice to the UK government on matters regarding the protection of the public from all forms of radiation. At NRPB, complex computationally expensive mathematical algorithms are used to simulate the activity of radioactive substances taken into the human body. These algorithms are mainly used in the calculation of doses due to exposures from radioactive substances carried within the body. This field of work is known as internal dosimetry. Over the last decade analyses have been performed at NRPB on a selection of these algorithms, in an attempt to quantify the effects of various sources of uncertainty on the reliability and accuracy of the predictions obtained from these algorithms. These uncertainty analyses made inferences about an algorithm based on a sample distribution of its output. As a general rule, the larger the sample the more accurate was the picture of the associated uncertainty obtained.

The algorithms used in the field of internal dosimetry tend to be complex and computationally expensive. Consequently, the accuracy of these analyses is limited by the number of evaluations of the algorithm that can be performed within the available time scale. An examination of some stochastic process methodology, used in the related field of estimating the output of algorithms, led me to believe that from this methodology a general theory might be developed, from a Bayesian perspective, to perform more informative and efficient uncertainty analyses, in terms of the num-

ber of evaluations of the algorithm required, for an accurate uncertainty analysis to be performed.

In this thesis, a general theory for uncertainty analysis, based on stochastic process methodology, will be developed and tested on examples of internal dosimetry algorithms. Although these examples come from one specific field, the theory is of a general nature and could be applied more widely to other algorithms.

The next section of this chapter will introduce the field of radiation protection and explain why the monitoring of human exposure to ionising radiation is so important. The final section will introduce and explain the development of the internal dosimetry algorithm and will indicate why it is important to obtain accurate measures of the uncertainty about their outputs.

In the first part of chapter two, the current classical methodology for uncertainty analysis will be described along with its drawbacks. Following this, the stochastic process methodology referred to above will be detailed to provide a background to the development of the new uncertainty analysis theory. Finally, the specific objectives of this thesis will be identified along with a description of the remaining chapters.

1.2. Human exposure to radiation

Throughout history man has always been exposed to radiation from natural sources. In 1895 William Rontgen [Ron95] observed that certain crystals gave off light when placed near to an electrical discharge taking place in a partially evacuated tube. He had discovered a way of producing X-rays artificially. This breakthrough led to a number of diagnostic and therapeutic applications for X-rays in the field of medicine that resulted in the first human exposures to radiation generated from an artificial source.

Today in the UK, an average member of the population receives a radiation dose of 2.6mSv. Exposure to all forms of natural background radiation contributes 85.5% of the total dose (50% Radon gas, 14% Gamma rays, 10% Cosmic rays and 11.5% food and drink) while the largest artificial source of radiation exposure, which constitutes 14% of the total,

is attributed to medical procedures, e.g. X-ray examinations and cancer treatments. Other artificial sources such as nuclear discharges, fallout and occupational exposure make up only 0.5% of the total dose. Figure 1.1 details the full breakdown of the sources of exposure to a typical member of the UK population.

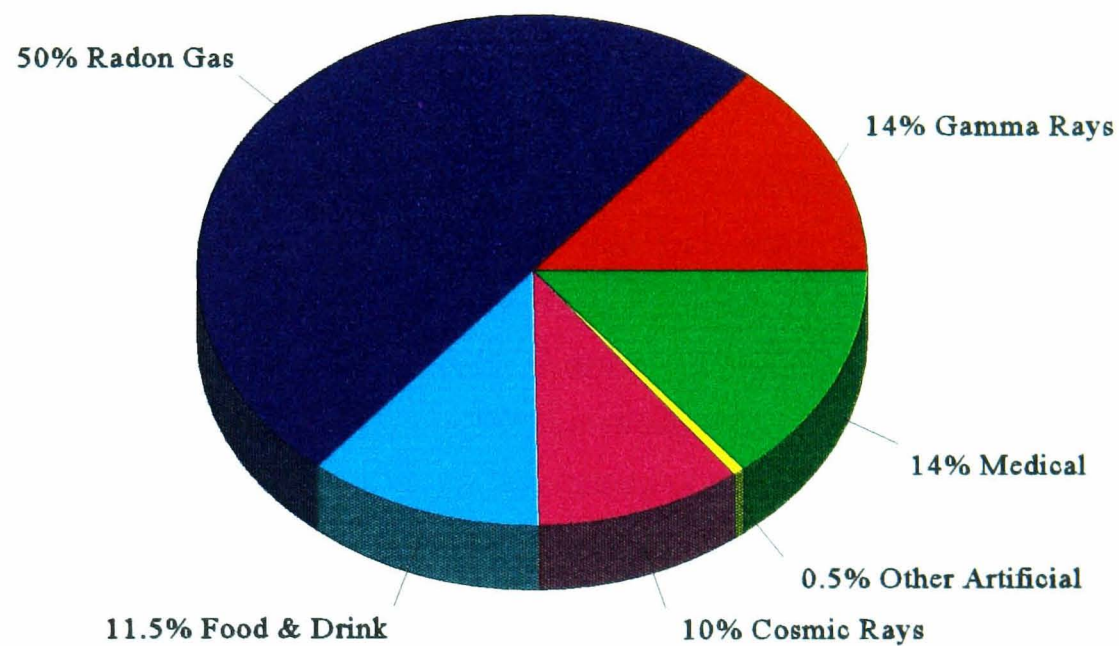


Figure 1.1: Breakdown of the average annual radiation dose to the population of the UK.

Unexpected illnesses and deaths amongst the first people to work with artificially produced radiation quickly led researchers to the conclusion that over exposure to radiation was harmful. Since then considerable effort has been devoted to developing radiation protection strategies to enable its safe use.

The detrimental effects of radiation exposure can be divided into two categories:

1) Early effects: these occur when the body is exposed to an extremely large dose over a short period of time. For example, an instantaneous uniform dose to the whole body of 5Gy will result in severe damage to

the bone marrow and gastrointestinal tract that is almost certain to cause death within weeks, even with medical treatment.

2) Late effects: these occur following exposures that are too small to cause early effects. The most important late effect of exposure to ionising radiation is cancer induction. Groups of people exposed to radiation in sub-lethal doses were found in subsequent years to have higher incidence and mortality rates from various cancers than comparable unexposed groups. The other important late effect of radiation exposure is an increase in the risk of hereditary disease in the offspring of exposed individuals.

Protecting people from the early effects of radiation, also referred to as deterministic effects, was found to be relatively easy. The first signs of these effects such as reddening of the skin, irritation of the eyes could be recognised and with the use of early dosimeters the high exposures that cause these and other more serious life threatening effects could be avoided.

Guarding against the risk of the late effects of radiation, also referred to as stochastic effects, was found to be more complicated. The first problem is that these effects are not easily quantifiable. For early effects, exposure of say an area of skin to a certain dose can be predicted with reasonable accuracy to cause a particular observable effect. For late effects, i.e. cancer induction, the consequences of an exposure can only be expressed in terms of an elevated risk of contracting a cancer at some point later in life. It is not possible to define a cut-off dose level above which the risk of cancer is increased and below which it is not. Thus the quantification of an individual's risk of a late effect is difficult.

The precise method by which an exposure causes cancer is not yet understood. This means that only empirical evidence may be used to quantify the risks of contracting, for instance, radiation-induced leukaemia. Current evidence regarding the late effects of radiation exposure comes principally from three sources:

1) Animal experiments. A number of different types of animals have been and are still used as experimental subjects. The advantage of using animals is that it is considered 'acceptable' to perform experiments in which

potentially dangerous doses are given to subjects such as rats or mice, whereas it would be ethically unacceptable to perform such experiments on humans. The main disadvantage of using animals is that even when using mammals such as rats, differences have been found in the physiological responses of these animals to radiation exposures compared to those of man.

2) The Life Span study. In 1945, the United States dropped two atomic bombs on the Japanese cities of Hiroshima and Nagasaki. This resulted in the killing of a large number of people who were close to the explosions. However, a considerable number of people, within these cities, were far enough from the explosions to escape instant death or a lethal dose but who still received a measurable exposure. In 1950, a cohort of 120,000 of these survivors was created. The various criteria used for the selection of this cohort are detailed by Preston, Kato, Kopecky and Fujita [PKKF87]. The health status of this cohort has been followed since that time with particular reference to the incidence of cancer. The Life Span study still represents the largest and most important body of data on the exposure of man to radiation currently available.

3) Human experiments. A small number of experiments have been carried out on humans using volunteers, mainly terminally ill patients, who received low doses. Most of these experiments looked at the behaviour of radioactive substances taken into the body either by inhalation or ingestion.

The data collected from the three sources listed above have highlighted a number of features concerning the pattern of cancer risk. Firstly, the increased risk of cancer does not begin immediately following the exposure. In fact, the evidence suggests that radiation-induced cancers do not start to appear until a number of years following the exposure. This delay is called the latency period and has been observed to differ between cancer types. Solid cancers, such as lung or liver cancer, seem to take a minimum of about ten years to develop following exposure. In contrast, for leukaemia the period is much shorter at around two years.

Empirical evidence also shows that a radiation-induced cancer does not always appear directly following the end of the latency period but can occur a number of years later. The period during which a radiation induced cancer can occur is known as the expression period. As with the latency period the expression period has been seen to vary with cancer type from as short as forty years with leukaemia to the end of life for solid cancers. Furthermore, there is some evidence that the cancer risk during the expression period does not remain constant although current evidence available on this point is limited.

Different cancer types also exhibit varying sensitivity to induction by radiation. For example, evidence suggests that in general leukaemia is the most sensitive form of cancer to induction by radiation exposure, based on the proportional increase in risk. However, the incidence of the leukaemia subtype known as chronic lymphatic leukaemia does not appear to be altered by radiation exposure. Other less extreme differences occur between the excess incidence rates of various solid cancers.

Finally, the empirical evidence suggests that there is variability between people in their susceptibility to contracting radiation induced cancer. A good analogy to this is the effect of smoking on lung cancer risk. There is a generally accepted rule that as a person's rate of smoking increases then so does their risk of lung cancer. However, some people can smoke heavily and still not contract lung cancer while others who smoke relatively little in comparison do succumb to the disease. A current theory, that has been proposed to explain this anomaly, is that a person's genetic characteristics affect their sensitivity to cancer induction due to environmental exposures such as smoking and radiation.

Clearly, the important task of deriving a relationship between the risk of cancer and the radiation dose received is not a simple one. Various national and international bodies have attempted to quantifying this relationship. They have produced a range of mathematical models which are used to calculate various dose constraints that are intended to define acceptable levels of risk to both the general public and to radiation workers. Thus, given a person's dose profile, these models can be used to predict their

radiation induced excess risk of contracting cancer and ensure that this does not exceed current ‘acceptable’ levels of risk.

An essential component of such risk calculations is the person’s dose profile, but obtaining this information can be a difficult problem. Exposure to the three types of radiation, alpha, beta and gamma can occur from two types of source; those external to the body and those taken into the body. Measurement of the dose received from external sources is relatively straightforward. Levels of radiation entering the body from external sources can be measured using personal dosimeters worn about a person’s body or can be derived from environmental measurements in some cases. These can provide a relatively accurate picture of a person’s external exposure. Considerably more difficult, and potentially more important, is the measurement and consideration of exposures from internal radiation sources. The problems concerning the measurement of internal exposures will be discussed in the next section.

1.3. Internal dosimetry

Internal contamination from radioactive substances can be extremely dangerous since they can remain in the body for extended periods of time and result in high doses to parts of the body. Radioactive substances can enter the body through three main pathways; by ingestion, by inhalation, and directly through the skin, for example, at the site of a wound. Once in the body, radioactive substances behave in various different ways that make dose calculations extremely difficult. In common with a lot of materials, radioactive substances tend to become concentrated in particular parts of the body. Also, certain types of radioactive substance have an affinity for specific organs or systems within the body.

These features can make internal exposures potentially dangerous. For example, a quantity of a radioactive substance when evenly dispersed throughout the body might give a small relatively inconsequential dose to the whole body. However, the same quantity of the material, concentrated in one organ, may give a larger and more hazardous dose to that particular organ and surrounding tissue with only a negligible dose to the

rest of the body. Thus, when considering the effects of internal exposures it is not only the type and quantity of the radioactive substance that is important but also its location in the body and the time spent at that location.

The uneven distribution of radioactive materials in the body causes a large problem for the assessment of internally generated doses. To accurately quantify such doses it is desirable to measure the amount of the radioactive substance in each of the organs in which it concentrates. Additionally, information is needed on how the amount in each organ changes with time following the initial exposure. In general, obtaining these measurements directly is not possible for a living subject so various mathematical models have been created to simulate the time-dependent distribution of radioactive substances in the various organs of the body.

The first generation of these models consisted of simple mathematical expressions that attempted to predict the time-dependent distribution of various radioactive substances in a few major organs. The next development was the introduction of mechanistic compartmental models. These were so called because they could be thought of as consisting of a number of boxes or compartments with pathways between them along which the radioactive material could move at a predetermined-determined rate.

These models introduced an additional level of complexity because they allowed for the possibility of recycling. Recycling is the process by which material excreted from an organ is not removed from the body, but is returned to the blood stream where it can be reabsorbed by the same organ or another part of the body. Recycling has been shown to be a critical factor in the calculation of internal doses. Further, the term ‘mechanistic’ was applied to these models as they were the first to attempt to simulate the actual pattern of movement of radioactive material in the body.

An integral part of this type of model is the specification of pathways between the model’s compartments. These pathways represent the routes that the radioactive material can take round the body. Each path has an associated coefficient that determines the transfer rate along that path. In many cases the true values of these rate coefficients are not known. Often

they are not directly (physically) measurable, nor, owing to the complexity of the human body is it possible to obtain definitive experimental data from which accurate estimates of these coefficients can be derived. Thus, the coefficients have to be estimated by ‘experts’ using any available subjective information and their expert knowledge.

Often such models are applied to a reference subject. That is, a fictitious person whose characteristics are selected to be in some way representative of an average member of the population. However, problems occur when a model is applied to a specific individual as opposed to the reference subject. Suppose, for example, that a model requires, as an input, the mass of an internal organ. For the reference subject a suitable representative value can be chosen but for a specific individual obtaining the mass of an internal organ without resorting to invasive actions is very difficult. Therefore, subjective information, based for example on the individual’s body size, must be used to derive an appropriate estimate.

This lack of knowledge about a model’s parameters or inputs introduces uncertainty into the model’s output since the output will, to an unknown degree, vary according to the values of the subjectively defined parameters. For a model to be used with confidence in the radiological protection environment, it is important to quantify this uncertainty.

2. INTRODUCTION TO UNCERTAINTY ANALYSIS

2.1. What is uncertainty analysis?

Uncertainty analysis, in the context of this work, is the name given to any technique that investigates the accuracy with which a computer-based mathematical algorithm or model can represent a complex (possibly physical) system.

Under this definition it is possible to define two sub-types of uncertainty analysis:

1) Analysis of model inadequacy : A computer algorithm that is used to provide a mathematical representation of a complex system will usually also be a simplification of the system. As a result of this simplification there are likely to be systematic differences between the output of the algorithm and the true value of the system. If it is intended to use the outputs of such an algorithm in place of measurements from the complex system then it is desirable to quantify these differences.

This can be a very difficult task since an exact definition of the complex system is not usually available. A good example of such a scenario is the problem of the measurement of internal radiation doses described in chapter one. In this case the exact nature of the movement of radionuclides about the body is unknown so that the computer algorithm represents our current understanding of the system. Thus, without a clear definition of the complex system it is difficult to assess how good a computer algorithm is at representing it.

Currently, it is usual to define the structure of an algorithm on the basis of the currently available information and then to ignore this source

of uncertainty. This type of uncertainty will not be considered further.

2) Analysis of parameter value uncertainty : A computer algorithm that represents a large complex system, although a simplification of reality, can also be complex with many parameters and inputs that represent underlying features of the system. In many cases it is not possible to define precisely the true value of each input or parameter. To overcome this problem estimates are used in place of the true, but unknown, values.

This lack of knowledge about the true values of the parameters propagates through the algorithm and results in uncertainty about the true value of the output. Again, if outputs of the algorithm are to be used in place of measurements from the underlying system then it is important that the effects of the choice of particular estimates of the uncertain parameters on the output of the algorithm are known. The aim of parameter value uncertainty analysis is to quantify these effects.

Unlike the problem of model inadequacy, classical statistical techniques have been developed to investigate this problem.

2.2. Parameter value uncertainty analysis

As discussed above, the aim of this analysis is to assess how the lack of knowledge about the true values of particular parameters will affect the reliability of the output of a mathematical algorithm. All forms of parameter uncertainty analysis techniques have the same basic structure.

The first task in the analysis is to select which parameters are going to be considered as uncertain. This may seem a trivial task: either the true value is known or it is unknown. However, for large and computationally expensive algorithms, and especially when using classical techniques, it may not be possible to analyse all of the uncertain parameters simultaneously. It is then necessary to select the most important uncertain parameters upon which to perform an analysis. A good way of resolving this selection problem is on the basis of how sensitive the output of the model is to changes in each of the uncertain parameters. Regardless of how much uncertainty exists about the value of a particular parameter, no useful information about the resulting uncertainty in the algorithm's out-

put will be obtained if this output is insensitive to the chosen value of the parameter. A wide range of sensitivity analysis techniques exist, [Ham95].

The next task is to quantify the uncertainty about the true value of each of the parameters selected for the analysis. The ideal way of doing this is to associate a joint probability distribution with all the uncertain parameters. This requires knowledge of the correlation structure between the parameters which is often not available. Thus it is usual to associate a single probability distribution with each parameter. The selection of these distributions forms a large topic of research on its own. In some cases, data can be used to construct a sample distribution from which a suitable distribution can be selected and the necessary parameters estimated. In other cases, if no data are available or if the parameter is not a physically measurable quantity then ‘expert’ judgment can be used. The process of defining parameter distributions using expert judgment has been approached in various ways, [Coo91].

Having selected the uncertain parameters and their associated distributions, upon which to perform the analysis, the distribution of the output induced by the input distributions can now be determined.

The ideal method would be to replace the estimates of all the uncertain parameter values with distributions, possibly multivariate for correlated parameters, and then to analytically determine the resulting distribution on the output of the model. In all but the simplest cases this is not possible due to mathematical complexity. An alternative method would be to evaluate the output of the model for all possible values of the uncertain parameters and then to construct the output distribution from the resulting values. This is theoretically impossible for continuous distributions and unrealistic for discrete distributions due to its computational expense. Instead, it is usual to perform a statistical analysis which avoids these mathematical complexities. All such statistical analyses have the same basic format.

The first step is to obtain data about the computer algorithm. One value is selected from each of the distributions associated with the uncertain parameters. The output of the algorithm is then evaluated using this set

of input parameter values. This process is repeated n times. The whole group of n sets of inputs will be referred to as the ‘design set’. Each set of parameter values used to obtain an output value from the algorithm will be referred to as an element of the design set.

Clearly, the more evaluations of the algorithm’s output that are made the more data will be obtained and hence the more accurate will be the subsequent uncertainty analysis. However, if an algorithm is computationally expensive to evaluate then the size of the design set will be limited by available computer time. This means it is important to maximise the quality of information about the parameter value uncertainty derived from each evaluation of the algorithm. The positions in the parameter space of the elements of the design set are the most important factor in maximising the quality of the information obtained. For example, if the uncertainty distribution of a parameter is uniform between two values a and b then a sample of points evenly spread between a and b would be more appropriate than one in which the values were irregularly spaced. In contrast, for model input that has an associated normal uncertainty distribution it would be preferable to select more of the values around the mean with fewer values in the tails. This means that more information will be gathered about the output of the algorithm from areas of the distribution that are considered to have a high probability of containing the true value of the uncertain parameter.

The second step of the analysis then consists of obtaining suitable statistical measures from the data to make inference about the effect of the parameter value uncertainty on that of the algorithm’s output.

Clearly, the methods by which both the design set is selected and the resulting data analysed will greatly affect the quality of the uncertainty analysis and thus the inferences made about the suitability of the algorithm to represent the underlying system.

A number of different design selection techniques have been proposed. In classical uncertainty analysis these are mainly based on Monte-Carlo selection methods. However, other design selection methods have been proposed based on the idea of selecting a design that maximises a criterion

that in some way measures the predicted quality of the information that will be obtained.

Having obtained the data, classical methods of deriving measures of the parameter uncertainty are mainly based around an analysis of the characteristics of the sample output distribution.

2.2.1. Classical design selection

As noted above, the quality of the results of an uncertainty analysis will be affected by the positions of the design points in the parameter space. The selection of the design set should be made so as to maximise the quality of the information obtained from each element. Classical techniques for the selection of designs are based on various forms of Monte-Carlo selection.

In the most basic form of Monte-Carlo selection, each of the distributions of the uncertain parameters is sampled, at random, the required number of times. This method can result in a poor coverage of the parameter space and hence a sample that does not maximise the information provided. To overcome this problem a structured random selection method called Latin Hypercube sampling, LHS, has been devised by McKay, Conover and Beckman [MCB79]. It is a widely used and popular method.

Suppose a design set, of size n , is required for a model with uncertain parameters, X_1, \dots, X_k . A single sample from this set will contain k values, one value from each of the distributions of the uncertain parameters.

To select a design set using LHS, the distribution of each uncertain parameter, X_i , ($i = 1, \dots, k$) is first divided into n nonoverlapping intervals on the basis of equal probability. One value is then selected, at random, from each interval to obtain the set $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,n})$.

To obtain the first sample of the design set, a value is selected at random from each of the sets \mathbf{x}_i . The next $n - 2$, samples are obtained similarly by sampling without replacement from the remaining values. The final sample in the design set consists of the remaining n values.

The use of this method ensures that the whole range of the parameter space is covered by the sample and thus provides more information than a simple unstructured sample. An empirical investigation, Rose [Ros83],

demonstrated that for a variety of statistical measures a simple random sample of size 400 was equivalent, in terms of the information it provided, to an LHS sample of size 200. The relative efficiency of LHS compared to MC sampling does not appear to have been considered theoretically.

2.2.2. Classical parameter uncertainty analysis

The measures obtained from a classical parameter uncertainty analysis are usually estimates of the mean and variance of the output distribution, as well as a 95% confidence interval for the mean. These measures are derived from the sample output distribution. The sample estimates of the mean and variance are then considered to be good estimates of the true population values, i.e. those values that would be obtained if all possible values of the algorithm's output were available.

A histogram of the output distribution is also usually constructed in order to examine the shape of the sample distribution. This is usually the extent of the analyses performed.

Three examples of classical analyses are Crick, Hofer, Jones and Haywood [CHJH88], Haywood and Smith [HS93] and Helton, Garner, McCurley and Rudeen [HGMR91].

This simple form of uncertainty analysis suffers from one major problem. As the number of uncertain parameters being analysed increases then so does the size of the sample of the outputs required to perform a meaningful uncertainty analysis.

Consider the following examples that demonstrate this problem using simple additive and multiplicative models. The additive model, as defined in 2.1, will be examined first. Let y be the output of the model with x and z as inputs.

$$y = x + z \tag{2.1}$$

Suppose that the true value of z is unknown and that we assign a distribution $N(\mu_z, \sigma_z^2)$ to represent our lack of knowledge about it. The uncertainty in z will propagate through the model and generate uncertainty in the value of y . For instance, if a sample of size n selected from the distribution of z is used to generate, from the model, a set of y values then

the variance of \bar{y} , the sample mean of the outputs, can be written as

$$Var[\bar{y}] = \frac{\sigma_z^2}{n}. \quad (2.2)$$

Now suppose the true value of x is also unknown and that its uncertainty can be represented independently of z by the distribution $N(\mu_x, \sigma_x^2)$. If we also take a sample of size n from this distribution and evaluate y for each of the n pairs of inputs then the variance of \bar{y} becomes

$$Var[\bar{y}] = \frac{\sigma_x^2 + \sigma_z^2}{n}. \quad (2.3)$$

We can determine m , the number of extra samples needed from each input distribution, for the two sample variances to be equal as

$$m = \frac{n\sigma_x^2}{\sigma_z^2}. \quad (2.4)$$

Thus for the variance of \bar{y} to be the same for the model with two uncertain inputs as that of the model with one uncertain input the number of samples from each must be increased by m . In the case where σ_x^2 and σ_z^2 are equal this will mean doubling the size of the sample. The number of extra samples required will increase further if the distributions of x and z are positively correlated since the expression for $Var[\bar{y}]$ will contain an added covariance term.

Now, using a simple multiplicative model, as defined in (2.5), the same calculations can be performed.

$$y = x \times z \quad (2.5)$$

Initially, the value of x will be considered known and z unknown with uncertainty distribution, $N(\mu_z, \sigma_z^2)$, as above. The variance of \bar{y} is then obtained as

$$Var[\bar{y}] = \frac{x^2 \sigma_z^2}{n} \quad (2.6)$$

where n is again defined as the sample size. If the value of x is now also considered unknown with uncertainty distribution, $N(\mu_x, \sigma_x^2)$, independent of z , then the variance of the sample mean becomes

$$Var[\bar{y}] = \frac{\sigma_x^2 \sigma_z^2 + \mu_x^2 \sigma_z^2 + \mu_z^2 \sigma_x^2}{n}. \quad (2.7)$$

Thus, m , the number of extra samples needed from each input distribution, for the two sample variances to be equal is

$$m = \frac{n(\sigma_x^2\sigma_z^2 + \mu_x^2\sigma_z^2 + \mu_z^2\sigma_x^2 - x^2\sigma_z^2)}{x^2\sigma_z^2}. \quad (2.8)$$

In the case where the means and variances of the uncertainty distributions are identical m becomes

$$m = n \left(\frac{\sigma^2 + 2\mu^2}{x^2} - 1 \right). \quad (2.9)$$

Further, if the known value of x is assumed to be μ then m becomes

$$m = n \left(\frac{\sigma^2}{\mu^2} + 1 \right). \quad (2.10)$$

Thus, for the variance of \bar{y} to be the same for the model with two uncertain inputs as that of the model with one uncertain input the number of samples from each must be at least doubled. If the magnitude of x is smaller than μ then the number of extra points will increase, while for values of x with a greater magnitude the number of extra points will decrease.

These simple examples illustrate that the sample size required to achieve a particular level of accuracy for \bar{y} will be dependent on the number and distributional assumptions of the uncertain parameters in the model. It can also be seen that the form of the model itself will affect the exact number of extra samples required. For algorithms with a large number of uncertain parameters this means that potentially a large number of evaluations could be required or that the uncertainty ranges for the measures of the analysis could be wide.

It is quite possible that for an algorithm with many uncertain parameters, it would not be possible to perform enough evaluations, using Monte Carlo methods, to obtain a reliable uncertainty analysis.

2.2.3. Alternative methodology for parameter uncertainty analysis

Little attention has been given to the problem of developing alternative, more efficient methods of performing parameter uncertainty analysis. A

method is required that makes more efficient use of the information provided by each data point. In addition it would be advantageous to develop a better method of design selection, that is, one that maximises the information provided by each point in the sample.

In recent years, considerable work has been done on the closely related problem of predicting the output of computationally expensive algorithms for specified inputs.

These methods are based on the use of a stochastic process (random function) to model departures of the algorithm's output from simple functions. An important assumption made in this methodology is that the output of the algorithm is, to a certain degree, smooth. That is, evaluating the output of the algorithm at a value x of an uncertain parameter will provide information about the value of the output at other points in the local neighbourhood of x .

The classical development of these methods still requires the sample to be selected using Monte Carlo based methods as described in **2.2.1**. However, if they are considered from a Bayesian point of view, the sample of the algorithm's output does not have to be selected at random. Instead, it can be selected in such a way as to maximise the information provided by each element of the design set, using the smoothness assumption. In the next section we will describe the development of the stochastic process prediction techniques from both the classical and Bayesian viewpoints.

2.3. Predicting an algorithm's output using stochastic process techniques

As introduced above, the principle around which these methods are based is of modelling the departures of the algorithm's output from a constant or a regression function as a realisation of a stochastic process. Let the known output of a deterministic algorithm be represented by $y(\mathbf{x})$ where \mathbf{x} represents a vector of values from the distributions of the uncertain parameters.

The unevaluated value of the algorithm for the general input vector \mathbf{x} ,

represented by $Y(\mathbf{x})$, is considered as a realisation of a stochastic process that includes either a constant term, i.e.

$$Y(\mathbf{x}) = \mu + Z(\mathbf{x}) \quad (2.11)$$

or a regression function

$$Y(\mathbf{x}) = \sum_k \beta_k f_k(\mathbf{x}) + Z(\mathbf{x}) \quad (2.12)$$

where $Z(\cdot)$ is a stationary stochastic process with mean zero. It is also usual to assume for computational convenience that $Z(\cdot)$ takes a Gaussian form. For two sets of inputs \mathbf{w} and \mathbf{x} the correlation between $Z(\mathbf{x})$ and $Z(\mathbf{w})$ is given by the covariance function $\sigma^2 R(\mathbf{w}, \mathbf{x})$, where σ^2 is the variance of the process and $R(\mathbf{w}, \mathbf{x})$ is a suitably defined correlation function. That is, it must be a positive function that decreases with increasing distance between the two points.

Two papers by Welch, Buck, Sacks, Wynn, Mitchell and Morris [WBS⁺92] and Sacks, Welch, Mitchell and Wynn [SWMW89] both cite one rationale for this use of stochastic processes: that departures of the complex function from μ or $\sum_k \beta_k f_k(\mathbf{x})$, though deterministic, may resemble a sample path of a (suitably chosen) stochastic process $Z(\cdot)$. The second form of the model allows for the inclusion of a trend with the uncertain parameters if the behaviour of the algorithm indicates such a trend. However, Sacks, Welch, Mitchell and Wynn [SWMW89] found that the use of the simple model (2.11) did not affect the performance of their predictor.

The most crucial part of this methodology is the correlation function $R(\cdot, \cdot)$ which defines the extent to which information about the output of the algorithm at \mathbf{w} is useful for predicting the value at \mathbf{x} . For algorithms with p uncertain parameters it is usual and computationally convenient to apply a product correlation rule that takes the form

$$R(\cdot, \cdot) = \prod_{j=1}^p R_j(\cdot, \cdot). \quad (2.13)$$

This implies that *a priori* the effect of each uncertain parameter is considered independent of the other uncertain parameters. A number of different

correlation functions have been used. All are based on two elements: firstly, a distance function component relating \mathbf{w} and \mathbf{x} ; and secondly, a scale parameter. The scale parameter represents a measure of the smoothness of the algorithm with respect to an uncertain parameter, i.e. the extent to which knowledge about the value of the algorithm at one value of an uncertain parameter will be of use in the prediction of the algorithm at another value. One commonly used example of a correlation function is the Gaussian correlation function,

$$\text{corr}(Z(\mathbf{w}), Z(\mathbf{x})) = \exp \left(- \sum_{j=1}^p \theta_j (x_j - w_j)^2 \right). \quad (2.14)$$

The actual choice of the function $R(.,.)$ is determined individually for each application so that it best represents the predicted form of the correlation.

For situations where the output of the algorithm cannot be measured deterministically, for instance, when the uncertain parameters represent inputs to some physical system then the model is extended to include a stochastic process, $\varepsilon(\mathbf{x})$ that represents this measurement error, i.e.

$$Y(\mathbf{x}) = \sum_k \beta_k f_k(\mathbf{x}) + Z(\mathbf{x}) + \varepsilon(\mathbf{x}). \quad (2.15)$$

This form of the model will not be pursued further since our interest lies mainly with the prediction of deterministic mathematical algorithms.

The way in which a predictor for the algorithm is obtained from this model differs depending on whether a classical or Bayesian analysis is being performed. In both scenarios, however, the data used to construct the predictor takes the form of a sample of evaluations of the algorithm over the uncertain parameter space. The values of the uncertain parameters at which the algorithm is evaluated are referred to as the design set $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_n]$ while the vector of the algorithm's output at the design points is defined as $\mathbf{y}_\mathbf{S} = [y(\mathbf{s}_1), \dots, y(\mathbf{s}_n)]^T$.

2.3.1. The classical predictor

The classical method for obtaining estimates using these models is to construct a linear predictor for $y(\mathbf{x})$ i.e.

$$\hat{y}(\mathbf{x}) = \mathbf{c}(\mathbf{x})^T \mathbf{y}_{\mathbf{S}}. \quad (2.16)$$

To obtain $\mathbf{c}(\mathbf{x})$ the estimate $\hat{y}(\mathbf{x})$ is considered as a random quantity and its mean square error is then minimised subject to an unbiasedness constraint to obtain the best linear unbiased predictor (BLUP). The calculation of the BLUP is clearly defined by Sacks, Welch, Mitchell and Wynn [SWMW89] as follows. First, define $\mathbf{Y}_{\mathbf{S}}$ as the random vector $\mathbf{Y}_{\mathbf{S}} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]^T$.

Further, let

$$\mathbf{f}(\mathbf{x})^T = [f_1(\mathbf{x}), \dots, f_k(\mathbf{x})], \quad (2.17)$$

$$\mathbf{R} = [R(Y(\mathbf{s}_i), Y(\mathbf{s}_j))]_{1 \leq i, j \leq n}, \quad (2.18)$$

$$\mathbf{r}(\mathbf{x})^T = [R(\mathbf{s}_1, \mathbf{x}), \dots, R(\mathbf{s}_n, \mathbf{x})], \quad (2.19)$$

$$\mathbf{F} = [f_l(\mathbf{s}_i)]_{1 \leq i \leq n, 1 \leq l \leq k}, \quad (2.20)$$

where σ^2 is considered known.

Now if $\mathbf{c}(\mathbf{x})^T \mathbf{y}_{\mathbf{S}}$ is a linear predictor of $y(\mathbf{x})$ then it has mean square error (MSE) defined as

$$\begin{aligned} E[\mathbf{c}^T \mathbf{Y}_{\mathbf{S}} - Y(\mathbf{x})]^2 &= E \left[\sum_{i=1}^n \left(c_i \left[\sum_{j=1}^k \beta_j f_j(s_i) + Z(s_i) \right] \right) \right]^2, \\ &= E \left[\sum_{i=1}^n \left[\sum_{m=1}^n c_i c_m Z(s_i) Z(s_m) \right] \right. \\ &\quad \left. + Z(x_i)^2 - 2c_i Z(s_i) Z(x_i) \right], \\ &= \sigma^2 [\mathbf{c}(\mathbf{x})^T \mathbf{R} \mathbf{c}(\mathbf{x}) + 1 - 2\mathbf{c}(\mathbf{x})^T \mathbf{r}(\mathbf{x})], \end{aligned} \quad (2.21)$$

where the summations i and m are over the set of n design points and the summation j is over the k estimating functions.

To obtain the BLUP the MSE must be minimised subject to the unbiasedness condition, $\mathbf{F}^T \mathbf{c}(\mathbf{x}) = \mathbf{f}(\mathbf{x})$. This constraint on the predictor

ensures that it will interpolate the data points. The constrained minimisation can be performed using Lagrange multipliers. Thus, the coefficient $\mathbf{c}(\mathbf{x})$ of the BLUP must satisfy the matrix equation

$$\begin{bmatrix} 0 & \mathbf{F}^T \\ \mathbf{F} & \mathbf{R} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}(\mathbf{x}) \\ \mathbf{c}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbf{f}(\mathbf{x}) \\ \mathbf{r}(\mathbf{x}) \end{bmatrix}. \quad (2.22)$$

where $\boldsymbol{\lambda}(\mathbf{x})$ is a Lagrange multiplier. Now, by inverting the initial partitioned matrix, the following matrix equation is obtained,

$$\begin{bmatrix} \boldsymbol{\lambda}(\mathbf{x}) \\ \mathbf{c}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} -(\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} & (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{R}^{-1} \\ \mathbf{R}^{-1} \mathbf{F} (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} & \mathbf{R}^{-1} (\mathbf{I} - \mathbf{F} (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{R}^{-1}) \end{bmatrix} \begin{bmatrix} \mathbf{f}(\mathbf{x}) \\ \mathbf{r}(\mathbf{x}) \end{bmatrix}. \quad (2.23)$$

Thus $\mathbf{c}(\mathbf{x})$ equals

$$\mathbf{c}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) \mathbf{R}^{-1} \mathbf{F} (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} + \mathbf{r}(\mathbf{x}) \mathbf{R}^{-1} (\mathbf{I} - \mathbf{F} (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{R}^{-1}). \quad (2.24)$$

The resulting predictor can be written as

$$\hat{y}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \hat{\boldsymbol{\beta}} + \mathbf{r}(\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y}_S - \mathbf{F} \hat{\boldsymbol{\beta}}), \quad (2.25)$$

where

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{y}_S, \quad (2.26)$$

is the standard generalised least squares estimate of $\boldsymbol{\beta}$. It is now clear why the correlation function $R(.,.)$, must be selected with care since to obtain the predictor $\hat{y}(\mathbf{x})$, it is necessary to invert the matrix \mathbf{R} of correlations between the elements of the design set.

This form of prediction is very similar to the method of ‘kriging’ used in geostatistics. In kriging, the stochastic process $Y(\mathbf{x})$ is also taken to be second order stationary, as in (2.11) and (2.12). That is, for given \mathbf{x} , the mean of the process is constant. The main difference is in the specification of the spatial correlation, which, for kriging, is obtained using a variogram. To estimate a variogram from data, half the mean value of the squared distance between all data points separated by a distance h is plotted against h or $|h|$. The parameters c_1 and a of a selected variogram function are then obtained by fitting the function to these points using

non-linear regression. Typical isotropic variogram functions, i.e. those in which the value of the function depends only on the distance h between two points, include the exponential function,

$$\gamma(h) = c_1 [1 - \exp(-|h|/a)], \quad (2.27)$$

and the Gaussian function,

$$\gamma(h) = c_1 [1 - \exp(-|h|^2/a^2)]. \quad (2.28)$$

An extra constant c_0 is also sometimes included in the function, e.g., $\gamma(h) = c_0 + c_1 [1 - \exp(-|h|^2/a^2)]$ to represent non-spatial variability and/or spatial variability below the sampling density. This constant is estimated at the same time as c_1 and a . This fitted variogram function is then used in place of the correlation function $R(.,.)$ to define a BLUP.

2.3.2. Examples of the use of the classical predictor

Two typical examples of the use of the stochastic process model with the Gaussian correlation function are provided by Sacks, Welch, Mitchell and Wynn [SWMW89] and Sacks, Schiller and Welch [SSW85]. The first of these papers shows, that in a circuit simulator example in 6 dimensions the BLUP is a more accurate predictor than a polynomial regression model. The BLUP, in this case, is based on the basic form of the model (2.11). It is suggested by Sacks that from past experience this simplification will not affect the performance of the predictor. In both studies the selection of the smoothing parameters, constant, μ , and the process variance, σ^2 , for the calculation of the BLUP is by maximum likelihood.

Other classical applications of the basic stochastic model (2.11) are provided by Welch, Buck, Sacks, Wynn, Mitchell and Morris [WBS⁺92] and Bowman, Sacks and Chang [BSC93]. Welch applies the model to two examples, both having 20 uncertain inputs, using the full exponential correlation function $R(d) = \prod_{j=1}^k \exp(-\theta_j |d_j|^{P_j})$. Due to the size of the examples the values of θ_j and P_j are estimated by a constrained maximum likelihood technique to reduce the computational burden of maximising the likelihood over 42 parameters. A screening exercise to eliminate the parameters to which the output of the algorithm is insensitive is suggested as an

alternative way of reducing the computational burden of high dimensional problems.

Bowman uses exactly the same methodology as described by Sacks, Welch, Mitchell and Wynn [SWMW89] with the same correlation functions as Welch, Buck, Sacks, Wynn, Mitchell and Morris [WBS⁺92] but with the simpler formulation of the model. It is applied to a two dimensional problem. The values of θ_j and P_j are also estimated by maximum likelihood using the design point evaluations.

2.3.3. The Bayesian predictor

We now turn to consider the development of a predictor, from the stochastic process model, in a Bayesian framework. The stochastic process (2.11) or (2.12) is now considered as a prior process that represents our knowledge about the algorithm before any evaluations are made. The assumption that the process, $Z(\cdot)$, is stationary and that it takes a Gaussian form is retained. The data vector \mathbf{y}_S is taken to be multivariate normally distributed such that

$$\mathbf{y}_S | \sigma^2 \sim N(\boldsymbol{\mu}, \mathbf{R}) \quad (2.29)$$

or

$$y_S | \sigma^2, \boldsymbol{\beta} \sim N(\mathbf{F}\boldsymbol{\beta}, \mathbf{R}) \quad (2.30)$$

for models (2.11) and (2.12) respectively.

The mean of the posterior process obtained using the data and the prior is used as a point predictor for the algorithm. Since the prior takes a Gaussian form and the data are considered as multivariate normal then the posterior will also take a Gaussian form and it is possible, using standard techniques, to obtain the posterior process. The Bayesian predictor will be taken as the mean of this posterior process and can, conditional on both $\sigma^2, \boldsymbol{\beta}$ or $\boldsymbol{\mu}$ be written as

$$\boldsymbol{\mu} + \mathbf{r}(\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}_S - \boldsymbol{\mu}^T \mathbf{1}) \quad (2.31)$$

when the basic model (2.11) is used, or as

$$\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} + \mathbf{r}(\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}_S - \mathbf{F}\boldsymbol{\beta}) \quad (2.32)$$

when the model including a regression function, (2.12) is used. In the former predictor, $\boldsymbol{\mu}$ is a vector of length k with elements μ . If a standard non-informative prior is placed on the vector $\boldsymbol{\beta}$ in the latter expression then the mean of the posterior process becomes

$$\mathbf{f}(\mathbf{x})^T \hat{\boldsymbol{\beta}} + \mathbf{r}(\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}_S - \mathbf{F}\hat{\boldsymbol{\beta}}) \quad (2.33)$$

where $\hat{\boldsymbol{\beta}}$ is as defined in (2.26).

This last predictor is the same as the BLUP defined in the classical analysis. However, for informative priors we would not expect the two estimators to be equal. In the Bayesian analysis the posterior variance of the process may also be obtained to provide a measure of the uncertainty associated with the point estimate.

In both the classical and Bayesian frameworks it can be seen that the choice of the correlation function is not arbitrary, The function must be selected so that the \mathbf{R} matrix is invertible else it is not possible to calculate the BLUP/posterior mean. This indicates that the choice of correlation functions is restricted to those for which the \mathbf{R} matrix has non-zero determinant. Further, it is assumed that the process variance, σ^2 , and the smoothness parameters, θ , one for each uncertain parameter, are known. Unfortunately these assumptions are not usually realistic and, as a result, a number of different ways of estimating these parameters have been advocated.

Both of these forms of prediction have been used with complex mathematical algorithms in a number of different scenarios.

2.3.4. Examples of the use of the Bayesian predictor

Probably the first use of the stochastic model in a Bayesian form to predict the output of an unknown function is by Kimeldorf and Wahba [KW70]. O'Hagan [O'H78], develops the same model but using a non-stationary stochastic process, also in the Bayesian framework, in order to find a new approach to the theory of optimal design selection. O'Hagan describes the fitting of a 'localised regression model' to a set of values, analogous to \mathbf{y}_S , drawn from a normal distribution. The general theory for the model is de-

tailed for the situation where the data are obtained subject to measurement error.

To illustrate the theory, a number of one dimensional examples are given in which the simpler form of the theory is used in which the points \mathbf{y}_s are assumed to be obtained without measurement error. In these examples, a Gaussian correlation function of the form $\exp(-\frac{1}{2}h^2/\theta^2)$ is used and a linear regression function, $\boldsymbol{\beta}^T \mathbf{f}(\mathbf{x})$ where $\mathbf{f}(\mathbf{x}) = (1, \mathbf{x})$. The problem of having to estimate the coefficients of the regression function is overcome by placing a prior multivariate normal distribution on the coefficients, i.e. $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, k\mathbf{B})$ and letting k tend to infinity. The effects on the predictive function of estimating the values of θ and \mathbf{B} are considered by comparing the quality of the predictions for two different values for θ and various values for the elements of \mathbf{B} . O'Hagan [O'H78] shows that a value of θ that overestimates the smoothness of the algorithm will cause the posterior predictive process to be overconfident, while one that underestimates it will result in the predictive process being pessimistic in assessing its accuracy. The effects of varying the elements of \mathbf{B} are shown to be a change in the smoothness of the posterior predictive process.

In a technical report, [CMMY88], and subsequent paper, [CMMY91], Currin, Mitchell, Morris and Ylvisaker use the basic model, (2.11), in a Bayesian framework. Here it is suggested that a fully specified prior distribution for the smoothing parameters, θ , the mean value, μ , and the process variance σ^2 would be optimal. However, this is not pursued further. Instead three optimization criteria, the 'leave one out' predictive density, the 'leave one out' bias and the maximum likelihood are proposed with which to select suitable values. In the multidimensional examples described by Currin, an *ad hoc* method that uses all three criteria is applied to obtain estimates of the parameter values.

A number of different correlation functions were also examined by Currin; linear, non-negative linear, cubic, non-negative cubic and an exponential function due to Sacks, Welch, Mitchell and Wynn [SWMW89] of the form $R(d) = \prod_{j=1}^k \exp(-\theta_j |d_j|^P)$ (the case of $P = 2$ gives the Gaussian form used by O'Hagan [O'H78]). For the examples provided, no one cor-

relation function performed consistently better than the others when they were used to predict over a set of randomly selected test points at which the true value of each example function was known.

Morris, Mitchell and Ylvisaker [MMY93] also used the Gaussian form of the exponential correlation function i.e. with $P = 2$ in a Bayesian framework. However, Morris also assumes that information about the first partial derivatives of the function with respect to each uncertain parameter is available at the design points. The method of maximum likelihood is used to estimate the unknown parameters of the model but it is noted that this can be a computationally expensive operation for problems with a large number of dimensions.

2.3.5. Criterion based design selection

The LHS sampling scheme, described in 2.2.1, has been used extensively to provide ‘good’ designs for the stochastic process models, e.g. Welch, Buck, Sacks, Wynn, Mitchell and Morris [WBS⁺92], and Bowman, Sacks and Chang [BSC93]. However, the LHS design selection technique does not make the assumption of the smoothness of the output of the algorithm over the input parameter space discussed above. Thus, by using this assumption, the LHS methodology could potentially be improved upon.

Extensive research has been devoted to the definition of suitable criteria for design selection that do take account of the smoothness assumption. A range of such criteria based on minimising the expected variance or expected error of the predictor have been widely explored. The first use of such a criterion was by Box and Draper [BD59], who selected designs that minimised the average of the mean squared error over the parameter space (and normalised with respect to the variance of the function and the number of design points).

Three further criteria that have been used extensively are the G, A, and D optimisation measures. Mitchell and Morris and Ylvisaker [MMY94] define these as follows. A design selected to be asymptotically G (global) optimum is obtained by minimising the maximum posterior variance of an unspecified point, \mathbf{x}_0 , on the parameter space. The A (average) criterion

is similar but in this case the average posterior variance is minimised. The D (determinant) criterion, on the other hand, is based on maximising the determinant of the correlation matrix of the design set. Other recent references to the use of these criteria are Johnson, Moore and Ylvisaker [JMY90] and Mitchell, Sacks and Ylvisaker [MSY94].

Another widely used criterion is called entropy. This criterion, proposed by Lindley [Lin56], is based on Shannon's entropy measure which quantifies the 'amount of information' provided by each element in the design. Currin, Mitchell, Morris and Ylvisaker [CMMY88] show that when the prior stochastic process has a Gaussian form and the coefficients β (as in equation (2.12)) are considered known then this criterion reduces to maximising the determinant of the matrix of prior correlations between the design points, and is the same as the D-optimal criterion.

Additionally, the minimisation of both the integrated mean squared error and the maximum mean squared error have been used by Sacks, Schiller and Welch, [SSW85] and Sacks, Welch, Mitchell and Wynn [SWMW89]. In both these studies, the inclusion of a weight function in the integrated mean squared error criterion is advocated but not implemented. O'Hagan [O'H78] defines a loss function using the mean squared error of prediction. The design which minimises the integral of this loss with respect to a weight function is considered optimal. O'Hagan defines the weight function to be the normal prior density of the unknown parameter indicating that the prediction of the algorithm close to the mean value of the uncertain parameter is most important.

There is one major problem with using these criteria. They are all functions of the correlation function 2.14, defined in **2.3**, which contains one smoothing parameter for each uncertain input in the algorithm. As stated above, these describe the smoothness of the algorithm's output with respect to each uncertain input. In order to use the criteria above to select 'good' designs it is necessary to specify the values of the smoothing parameters. However, until the algorithm has been evaluated for a design set no objective information is available to estimate these values.

Sacks, Schiller and Welch [SSW85] performed a robustness study in

order to find a suitable estimate of the smoothing value that would perform well over a range of true values. A value of one was found to be most efficient in terms of minimising the integrated mean square error of the predictor when the true value of the smoothness parameter ranged between 0.25 and 100. Selection of a large estimated value, 100, was found to protect against very large errors in the predictor but also resulted in a predictor that was uniformly poor over the input parameter space. Another study by Sacks, Welch, Mitchell and Wynn [SWMW89] implements a two stage selection procedure in which the first 16 points of a 32 point design in six dimensions is selected based on robust estimates of the smoothing parameters. This data is then used to update the estimates using MLE. The remaining 16 points are then selected sequentially using the new estimates.

The remaining problem with this form of design selection is that it is usually computationally expensive to optimise these criteria in all but the smallest problems. Ideally, one would like to perform an unconstrained search over the input parameter space. It is more usual to find designs selected from a grid of potential points using a variety of search procedures. The use of a grid brings in the extra problem of determining the best grid spacing. Too fine a grid and the computational burden is large but too coarse a grid and the design obtained may not be optimal.

2.4. The specific objectives of this research

The stochastic process methodology described above provides a powerful way to predict the value of an algorithm for particular values of the uncertain parameters. The critical link in using these methods for uncertainty analysis is made by O'Hagan [O'H91] who uses these stochastic process methods to perform not just point estimation but also the prediction of the integral of a complex algorithm. Thus, as well as estimating $y(\mathbf{x})$, O'Hagan details the estimation of

$$\mathbf{k} = \int \mathbf{r}(\mathbf{x})y(\mathbf{x}) dG(\mathbf{x}) \quad (2.34)$$

where $\mathbf{r}(\cdot)$ is a known vector of functions of \mathbf{x} and $G(\mathbf{x})$ is a measure over \mathbf{x} . In the current context, $dG(\mathbf{x})$ will be the distributions of the uncertain

parameters. If the value of $r(\mathbf{x})$ is set to 1 then k becomes the expectation of $y(\mathbf{x})$, one of the measures obtained in a classical uncertainty analysis. The methods for uncertainty analysis that will be developed will expand on these techniques.

The objectives of this research can be described as follows:

1) to develop a general approach to parameter uncertainty analysis based on a Bayesian interpretation of stochastic process models that improve on the currently available classical methods in three ways. Firstly, to obtain more accurate estimates of the mean and variance of the algorithms output using fewer evaluations of the algorithm. Further, to develop other measures of parameter uncertainty not available in the classical analysis, and to define an efficient design selection criterion.

2) to objectively demonstrate any improvement in the Bayesian methodology over the classical methodology

3) to demonstrate that the Bayesian methodology is useable in a ‘real-life’ uncertainty analysis.

2.5. Overview of the remaining chapters

In chapter three, the basic theory underlying Bayesian uncertainty analysis will be detailed. This will consist of the development of the estimates of the mean, variance and other measures of the algorithm’s output, and the derivation of a design selection criterion. In chapter four, the general theory will be reworked specifically for the analysis of algorithms with normally distributed uncertain parameters. The following two chapters, five and six, detail the application of this theory to two internal dosimetry algorithms.

In chapter five, the Bayesian and classical methodologies are compared objectively using a simplified recycling algorithm for the calculation of committed effective dose following ingestion of radioactive iodine, ^{131}I .

This model was selected for a number of reasons. First, a classical uncertainty analysis had already been performed on this model at NRPB. For this NRPB analysis only two of the model’s parameters had been considered as uncertain, so it provided a simple two-dimensional problem on

which to test the performance of the Bayesian methodology. Further, Fortran code was available, at NRPB, to run the algorithm with user-defined values for the uncertain parameters. This code took a relatively short period of time to run on the NRPB mainframe computers and so estimates of the true uncertainty in the algorithm's output could also be derived using a very large set of model evaluations. Thus, the classical and Bayesian estimates of the uncertainty could be compared not only to each other but also objectively to the true uncertainty.

The iodine algorithm provided a suitable test of the potential of the Bayesian methodology. However, for the size and complexity of the ^{131}I algorithm it would be possible to produce results from a classical analysis to a high degree of accuracy by simply using a suitably large sample. To be of any use the Bayesian methodology must be shown to perform better than the classical methodology when applied to higher dimensional and more computationally expensive models.

Chapter six describes such a 'real life' uncertainty analysis problem also from the field of internal dosimetry. The algorithm used is for the metabolism of plutonium in the human body and was considered, after some transformations were carried out, to have fourteen uncertain parameters. In parallel to the Bayesian analysis a classical analysis was being performed by staff at NRPB. It was thus possible to compare the results of the Bayesian and classical uncertainty analyses. Since this was a 'real life' problem for which the true uncertainty was not obtainable, other means of validating the Bayesian methodology were also considered.

Finally, in chapter seven the merits and pitfalls of the Bayesian methodology are discussed and compared to those of the classical methodology and further areas of potential research are identified.

3. GENERAL METHODOLOGY FOR BAYESIAN UNCERTAINTY ANALYSIS

3.1. Introduction

In this chapter the general theory underlying the Bayesian approach to uncertainty analysis, based on the use of stochastic process models, is detailed. Initially, a hierarchical Bayesian stochastic model is defined. From this model a number of measures are derived to quantify the uncertainty in the output of a computer algorithm.

The development of the uncertainty measures falls into four sections:

- a) construction of the Bayesian model,
- b) predicting the output of the algorithm for specific values of the uncertain parameters,
- c) estimating the expected value of the algorithm's output over the range of possible values of the uncertain parameters,
- d) estimating the variance of the algorithm's output over the range of possible values of the uncertain parameters

The purpose of parts a and b is mainly to lay the foundations of the Bayesian methodology that will be required for the development of the uncertainty measures in sections c and d. However, in part b, the development of the Bayesian predictor as described in **2.3.3** is illustrated. This predictor, although not a measure of uncertainty has two relevant uses. First, it can be used as a means of quantifying the accuracy of the Bayesian uncertainty analysis methodology. Secondly, it can be adapted to provide a probability that the true output of the algorithm does not exceed a predefined critical value. Both of these uses will be demonstrated

in the following chapters.

3.2. Development of the Bayesian model

Let $\eta(\cdot)$ represent the complex computer algorithm to which the uncertainty analysis is to be applied. In general, such algorithms will be computationally expensive and often too complex to represent as a single explicit mathematical expression. Models used in the field of radiation protection are frequently defined by large sets of differential equations.

Let $\eta(\mathbf{x})$, $\eta(\mathbf{z})$ represent the known output of the algorithm when the p uncertain inputs/parameters are given by $\mathbf{x} = (x_1, x_2, \dots, x_p)$ and $\mathbf{z} = (z_1, z_2, \dots, z_p)$ respectively. Following the Bayesian philosophy, our prior knowledge about $\eta(\cdot)$, at a point prior to evaluation, will be expressed as a hierarchical stochastic model. This model, which has four key elements, was first proposed in the context of Bayesian quadrature [O'H91].

We first make the assumption that the function $\eta(\cdot)$ can be approximated, to a reasonable degree, as a linear combination of k simple functions $h_j(\cdot)$, $j = 1, 2, \dots, k$. Then, we define the expectation and variance of $\eta(\mathbf{x})$ for all \mathbf{x} in the space χ of possible input values as

$$E(\eta(\mathbf{x}) \mid \boldsymbol{\beta}, \sigma^2) = \mathbf{h}(\mathbf{x})^T \boldsymbol{\beta} \quad (3.1)$$

$$Var(\eta(\mathbf{x}) \mid \boldsymbol{\beta}, \sigma^2) = \sigma^2, \quad (3.2)$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)$ is a vector of unknown 'regression coefficients', σ^2 , an unknown parameter that quantifies the variability of $\eta(\cdot)$ about its expected value, and where $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_k(\mathbf{x}))$ defines the vector of chosen 'regressor functions'.

The next part of the model, defines the covariance between $\eta(\mathbf{x})$ and $\eta(\mathbf{z})$ as

$$Cov(\eta(\mathbf{x}), \eta(\mathbf{z})) = \sigma^2 C(\mathbf{x}, \mathbf{z}) \quad (3.3)$$

where $C(\cdot, \cdot)$ is a correlation function, such that $C(\mathbf{x}, \mathbf{x}) = 1$ and where $C(\mathbf{x}, \mathbf{z})$ is a monotonically decreasing function with respect to a selected measure of the distance between \mathbf{x} and \mathbf{z} . This covariance function is the most important part of the Bayesian model since it represents our belief in the smoothness of $\eta(\cdot)$ with respect to \mathbf{x} .

The third component of the model combines these elements by assuming that conditional on the vector $\boldsymbol{\beta}$ and the parameter σ^2 the joint prior distribution of $\eta(\cdot)$ at a finite set of n points $(\eta(\mathbf{x}_1), \eta(\mathbf{x}_2), \dots, \eta(\mathbf{x}_n))$ is multivariate normal.

Formally, the prior distribution of $\eta(\cdot)$ conditional on $(\boldsymbol{\beta}, \sigma^2)$ is a Gaussian process and is written

$$\eta(\cdot) \mid \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{h}(\cdot)^T \boldsymbol{\beta}, \sigma^2 C(\cdot, \cdot)) \quad (3.4)$$

where $\mathbf{h}(\cdot)^T \boldsymbol{\beta}$ is the mean function and $\sigma^2 C(\cdot, \cdot)$ the covariance function.

Finally, to complete the hierarchical model we must specify prior distributions for the hyper-parameters $\boldsymbol{\beta}$ and σ^2 . If suitable prior information is available a conjugate prior can be specified as a Normal inverse Gamma distribution such that

$$f(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{2-(d+p+2)/2} \exp\left(\frac{-(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{V}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + a}{2\sigma^2}\right). \quad (3.5)$$

This implies that the conditional prior distribution of $\boldsymbol{\beta} \mid \sigma^2$ is multivariate normally distributed and that the prior marginal distribution of σ^2 has an inverse gamma form. Often, little prior information is available and so a noninformative prior

$$f(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2} \quad (3.6)$$

is usually substituted. Formally it can be obtained from (3.5) by letting the elements of \mathbf{V} tend to ∞ and by setting $a = 0$ and $d = -p$.

A comparison with chapter two will show that (3.4) is in fact the same as the stochastic process defined in (2.12).

3.3. Predicting the output of the algorithm

Suppose the computer algorithm is run for n different sets of uncertain parameter values, then define $\mathbf{y}^T = [\eta(\mathbf{x}_1), \eta(\mathbf{x}_2), \dots, \eta(\mathbf{x}_n)]$ as a vector containing the n outputs of $\eta(\cdot)$ at $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. These n sets of points will in future be known as the set of ‘design points’, $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$, as described in 2.3. The vector \mathbf{y} represents the objective data about the output of the algorithm and will be used to derive a posterior distribution for $\eta(\cdot)$.

The vector of observations \mathbf{y} , conditional on $\boldsymbol{\beta}$ and σ^2 will be considered multivariate normally distributed

$$\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{H}\boldsymbol{\beta}, \sigma^2 \mathbf{A}), \quad (3.7)$$

where

$$\mathbf{H} = [\mathbf{h}(\mathbf{x}_1), \dots, \mathbf{h}(\mathbf{x}_n)]^T, \quad (3.8)$$

and \mathbf{A} is the $n \times n$ symmetric matrix with the $[i, j]$ -th element $C(\mathbf{x}_i, \mathbf{x}_j)$.

From this distribution a likelihood function for $(\boldsymbol{\beta}, \sigma^2)$ can be obtained. The application of Bayes theorem to this function and the prior distribution of $(\boldsymbol{\beta}, \sigma^2)$ will enable posterior distributions to be obtained. The conditional posterior distribution of $\boldsymbol{\beta} \mid \mathbf{y}, \sigma^2$ and the marginal posterior distribution of σ^2 are obtained in [O'H91] as

$$\boldsymbol{\beta} \mid \sigma^2, \mathbf{y} \sim N(\hat{\boldsymbol{\beta}}, \sigma^2 (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H})^{-1}), \quad (3.9)$$

and

$$\frac{(n - q - 2)\sigma^2}{\hat{\sigma}^2} \mid \mathbf{y} \sim \chi_{n-q}^{-2}, \quad (3.10)$$

where

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{A}^{-1} \mathbf{y}, \quad (3.11)$$

$$\hat{\sigma}^2 = \frac{\mathbf{y}^T (\mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{H} (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H})^{-1} \mathbf{A}^{-1}) \mathbf{y}}{n - q - 2}, \quad (3.12)$$

and where n is the length of the vector \mathbf{y} and q is the rank of \mathbf{H} .

The posterior distribution of $\eta(\cdot)$ is obtained in three stages. First the posterior distribution of $\eta(\cdot) \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}$ must be derived from the joint distribution of $\eta(\cdot) \mid \boldsymbol{\beta}, \sigma^2$ and $\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2$. Next, the product of $\eta(\cdot) \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}$ with that of $\boldsymbol{\beta} \mid \sigma^2, \mathbf{y}$ must be integrated over $\boldsymbol{\beta}$ to obtain the Gaussian process

$$\eta(\cdot) \mid \sigma^2, \mathbf{y} \sim N(m^*(\cdot), \sigma^2 C^*(\cdot, \cdot)), \quad (3.13)$$

where

$$m^*(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T \hat{\boldsymbol{\beta}} + \mathbf{t}(\mathbf{x})^T \mathbf{A}^{-1} (\mathbf{y} - \mathbf{H} \hat{\boldsymbol{\beta}}), \quad (3.14)$$

and

$$\begin{aligned} C^*(\mathbf{x}, \mathbf{z}) &= C(\mathbf{x}, \mathbf{z}) - \mathbf{t}(\mathbf{x})^T \mathbf{A}^{-1} \mathbf{t}(\mathbf{z}) \\ &\quad + (\mathbf{h}(\mathbf{x})^T - \mathbf{t}(\mathbf{x})^T \mathbf{A}^{-1} \mathbf{H}) (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H})^{-1} (\mathbf{h}(\mathbf{z})^T - \mathbf{t}(\mathbf{z})^T \mathbf{A}^{-1} \mathbf{H})^T, \end{aligned} \quad (3.15)$$

and

$$\mathbf{t}(\mathbf{x}) = [C(\mathbf{x}, \mathbf{x}_1), C(\mathbf{x}, \mathbf{x}_2), \dots, C(\mathbf{x}, \mathbf{x}_n)]^T. \quad (3.16)$$

Finally, integrating the product of $\eta(\cdot) \mid \sigma^2, \mathbf{y}$ (3.13) and $\sigma^2 \mid \mathbf{y}$ (3.10) with respect to σ^2 yields a posterior distribution for $\eta(\cdot) \mid \mathbf{y}$ which for a given \mathbf{x} can be written as

$$\frac{\eta(\mathbf{x}) - m^*(\mathbf{x})}{\hat{\sigma} \sqrt{C^*(\mathbf{x}, \mathbf{x})}} \mid \mathbf{y} \sim t_{n-q}. \quad (3.17)$$

This distribution is a generalisation of the multivariate t distribution in the same way that a Gaussian process generalises a multivariate normal distribution. Thus given a set of inputs \mathbf{x}_0 the output of the model can be predicted. For example, a point estimate is provided by the posterior expectation, $m^*(\mathbf{x}_0)$, the same function as noted in **2.3.1** and **2.3.3**. The posterior variance of this point estimate is obtained as $\hat{\sigma}^2 C^*(\mathbf{x}_0, \mathbf{x}_0)$. At the design points, X , the point estimate will equal the true value and the variance will be zero. For all other points, $m^*(\mathbf{x})$ represents a smooth interpolation of the objective data. Further, at all points, X , the estimates are also unbiased.

3.3.1. Calculating the probability that the true output of the algorithm will exceed a critical value

A useful measure that can be obtained from the previous result with little effort is the probability that the true value of the algorithm's output exceeds a specified value for given input values. Let $P(\mathbf{x})$ be the probability that the unknown true value of the output of the algorithm using a particular selection of the uncertain parameters exceeds some predefined value. Let \mathbf{x}_0 be the vector of selected parameter values and let c be the predefined critical value. The t distribution (3.17) gives $P(\mathbf{x}_0)$ as

$$P(\mathbf{x}_0) = t_\nu \left(\frac{c - m^*(\mathbf{x}_0)}{\hat{\sigma} \sqrt{C^*(\mathbf{x}_0, \mathbf{x}_0)}} \right) \quad (3.18)$$

where t_ν is the complementary cumulative distribution function for the t distribution on ν degrees of freedom and where $\nu = n - q$.

3.4. Estimating the mean of the algorithm's uncertainty distribution

In this section we will derive a distribution to represent our information about the mean value of the algorithm's output. Let G be the distribution function for the uncertain inputs \mathbf{x} . Then for the purposes of uncertainty analysis we are interested in the distribution of the random variable $\eta(X)$ where X has the distribution G . The most important feature of this distribution to obtain is a measure of its location. In this section a posterior distribution for the mean of the uncertainty distribution associated with $\eta(\mathbf{x})$ will be derived. We will denote the mean by K . Therefore,

$$K = \int_{\mathcal{X}} \eta(\mathbf{x}) dG(\mathbf{x}). \quad (3.19)$$

If the form of the algorithm were known and found to be analytically integrable then in principle the value of K could be obtained immediately from (3.19). However, in the scenario considered here our knowledge about $\eta(\cdot)$ is restricted to a posterior distribution (3.13) consequently it is not possible to obtain the exact value of K although, since K is a linear functional of $\eta(\cdot)$, we can derive a posterior distribution for it. This situation represents a particular form of the general Bayesian quadrature problem described in [O'H91]. By following the same approach we obtain

$$K \mid \sigma^2, \mathbf{y} \sim N(\hat{k}, \sigma^2 W). \quad (3.20)$$

The mean and the variance are obtained as

$$\hat{k} = \int_{\mathcal{X}} m^*(\mathbf{x}) dG(\mathbf{x}) = \mathbf{R}\hat{\boldsymbol{\beta}} + \mathbf{T}\mathbf{A}^{-1}(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}}), \quad (3.21)$$

and

$$\begin{aligned} W &= \int_{\mathcal{X}} \int_{\mathcal{X}} C^*(\mathbf{x}, \mathbf{z}) dG(\mathbf{x}) dG(\mathbf{z}) \\ &= U - \mathbf{T}\mathbf{A}^{-1}\mathbf{T}^T + (\mathbf{R} - \mathbf{T}\mathbf{A}^{-1}\mathbf{H})(\mathbf{H}^T\mathbf{A}^{-1}\mathbf{H})^{-1}(\mathbf{R} - \mathbf{T}\mathbf{A}^{-1}\mathbf{H})^T, \end{aligned} \quad (3.22)$$

in which

$$\mathbf{R} = \int_{\mathcal{X}} \mathbf{h}(\mathbf{x})^T dG(\mathbf{x}), \quad (3.23)$$

$$\mathbf{T} = \int_{\chi} \mathbf{t}(\mathbf{x}) dG(\mathbf{x}), \quad (3.24)$$

$$U = \int_{\chi} \int_{\chi} C(\mathbf{x}, \mathbf{z}) dG(\mathbf{x}) dG(\mathbf{z}). \quad (3.25)$$

The distribution for K , (3.20) is conditional not only on \mathbf{y} but also σ^2 . This is not desirable since σ^2 will rarely be known. The conditioning on σ^2 can be removed by integrating the product of the distribution function for K with the marginal distribution of σ^2 , (3.10), giving a posterior t distribution for K of the form

$$\frac{K - \hat{k}}{\hat{\sigma}\sqrt{W}} \mid \mathbf{y} \sim t_{n-q} \quad (3.26)$$

where n and q are as previously defined (3.10). A point estimate and variance for the value of K can be obtained from this distribution.

This theory has been already published as Haylock & O'Hagan [HO96].

3.5. Estimating the variance of the algorithm's uncertainty distribution

Let L be defined as the variance of $\eta(\mathbf{X})$, where

$$L = K_2 - K^2 \quad (3.27)$$

and

$$K_2 = \int_{\chi} \eta^2(\mathbf{x}) dG(\mathbf{x}). \quad (3.28)$$

K_2 is a p-dimensional integral taken over G , the joint distribution function of the uncertain parameters. As with the quantity K , we would like to derive the posterior distribution for L . Due to the form of K_2 (3.28), this is not a mathematically tractable problem since the posterior distribution for $\eta^2(\cdot)$ would take the form of a non-central F distribution and it would then be difficult (if not impossible) to obtain distribution for K_2 in closed form. It is, however, possible to obtain posterior moments of L . The first two moments of L about the mean will now be derived.

3.5.1. Calculation of the posterior expectation of L

The expectation of $L \mid \mathbf{y}$, that is, the mean of the unknown posterior distribution for the variance of $\eta(X)$ can be expressed as

$$E[L \mid \mathbf{y}] = E[K_2 - K^2 \mid \mathbf{y}] = E[K_2 \mid \mathbf{y}] - (Var[K \mid \mathbf{y}] + E[K \mid \mathbf{y}]^2). \quad (3.29)$$

It is only the component, $E[K_2 \mid \mathbf{y}]$, of this formula that is unknown (the other components can be extracted from (3.26)). To derive $E[K_2 \mid \mathbf{y}]$ we first obtain $E[K_2 \mid \sigma^2, \mathbf{y}]$ and then marginalise this expectation with respect to σ^2 .

Now,

$$E[K_2 \mid \sigma^2, \mathbf{y}] = E \left[\int_{\chi} \eta^2(\mathbf{x}) dG(\mathbf{x}) \mid \sigma^2, \mathbf{y} \right] = \int_{\chi} E[\eta^2(\mathbf{x}) \mid \sigma^2, \mathbf{y}] dG(\mathbf{x}), \quad (3.30)$$

where

$$E[\eta^2(\mathbf{x}) \mid \sigma^2, \mathbf{y}] = m^*(\mathbf{x})^2 + \sigma^2 C^*(\mathbf{x}, \mathbf{x}). \quad (3.31)$$

Substituting (3.31) into (3.30) and expanding the expressions for $m^*(\mathbf{x})$, (3.14), and $C^*(\mathbf{x}, \mathbf{x})$, (3.15), gives

$$\begin{aligned} E[K_2 \mid \sigma^2, \mathbf{y}] &= \hat{\beta}^T \mathbf{Q} \hat{\beta} + (\mathbf{y} - \mathbf{H} \hat{\beta})^T \mathbf{A}^{-1} \mathbf{P} \mathbf{A}^{-1} (\mathbf{y} - \mathbf{H} \hat{\beta}) \\ &\quad + 2(\mathbf{y} - \mathbf{H} \hat{\beta})^T \mathbf{A}^{-1} \mathbf{S} \hat{\beta} + \sigma^2 [1 - \text{tr} \{ \mathbf{A}^{-1} \mathbf{P} \} \\ &\quad + \text{tr} \{ (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H})^{-1} \mathbf{Q} \} - 2 \text{tr} \{ \mathbf{A}^{-1} \mathbf{H} (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H})^{-1} \mathbf{S} \} \\ &\quad + \text{tr} \{ \mathbf{A}^{-1} \mathbf{H} (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{A}^{-1} \mathbf{P} \}], \end{aligned} \quad (3.32)$$

where

$$\mathbf{P} = \int_{\chi} \mathbf{t}(\mathbf{x}) \mathbf{t}(\mathbf{x})^T dG(\mathbf{x}), \quad (3.33)$$

$$\mathbf{Q} = \int_{\chi} \mathbf{h}(\mathbf{x}) \mathbf{h}(\mathbf{x})^T dG(\mathbf{x}), \quad (3.34)$$

$$\mathbf{S} = \int_{\chi} \mathbf{t}(\mathbf{x}) \mathbf{h}(\mathbf{x})^T dG(\mathbf{x}), \quad (3.35)$$

with \mathbf{R} and \mathbf{T} defined as in (3.23), (3.24) respectively.

The conditioning of $E[K_2 \mid \sigma^2, \mathbf{y}]$ on σ^2 can now be removed by taking its expectation with respect to the posterior distribution of $\sigma^2 \mid \mathbf{y}$, (3.10).

This will result in an expression identical to $E[K_2 \mid \sigma^2, \mathbf{y}]$ but with σ^2 replaced by $\hat{\sigma}^2$, the posterior expectation of $\sigma^2 \mid \mathbf{y}$ given in (3.12).

Now, all the components of (3.29) are obtained and $E[L \mid \mathbf{y}]$ can be calculated.

3.5.2. Calculation of the posterior variance of L

The variance of $L \mid \mathbf{y}$, that is, the variance of the posterior distribution for the variance of $\eta(X)$ is derived using the standard formula

$$\text{Var}[L \mid \mathbf{y}] = E \left[(K_2 - K^2)^2 \mid \mathbf{y} \right] - E \left[K_2 - K^2 \mid \mathbf{y} \right]^2. \quad (3.36)$$

The component $E \left[(K_2 - K^2)^2 \mid \mathbf{y} \right]$ has already been obtained in the previous section, so it only remains to calculate

$$E \left[(K_2 - K^2)^2 \mid \mathbf{y} \right] = E \left[E \left[(K_2 - K^2)^2 \mid \sigma^2, \mathbf{y} \right] \mid \sigma^2 \right] \quad (3.37)$$

using

$$\begin{aligned} E \left[(K_2 - K^2)^2 \mid \sigma^2, \mathbf{y} \right] &= E \left[K_2^2 \mid \sigma^2, \mathbf{y} \right] - 2E \left[K_2 K^2 \mid \sigma^2, \mathbf{y} \right] \\ &\quad + E \left[K^4 \mid \sigma^2, \mathbf{y} \right]. \end{aligned} \quad (3.38)$$

We now quote a general result. The normal moment generating function is defined as

$$M_X(t) = \exp(0.5t'\Omega t + \mu't) \quad (3.39)$$

and the fourth partial derivative of this function, evaluated at $t = 0$, gives

$$\begin{aligned} E[X_i X_j X_k X_l] &= \mu_i \omega_{jk} \mu_l + \omega_{jk} \omega_{il} + \mu_j \omega_{ik} \mu_l + \omega_{ik} \omega_{jl} + \mu_k \omega_{ij} \mu_l + \omega_{ij} \omega_{kl} \\ &\quad + \mu_i \mu_j \mu_k \mu_l + \mu_i \omega_{kl} \mu_j + \mu_i \omega_{jl} \mu_k + \mu_j \omega_{il} \mu_k. \end{aligned} \quad (3.40)$$

where the ω represents the various elements of the variance-covariance matrix Ω .

Using this result, the various components of (3.38) can be expanded and evaluated. Thus, the first element becomes

$$\begin{aligned} E \left[K_2^2 \mid \sigma^2, \mathbf{y} \right] &= E \left[\int_{\chi} \eta^2(\mathbf{x}) dG(\mathbf{x}) \int_{\chi} \eta^2(\mathbf{z}) dG(\mathbf{z}) \mid \sigma^2, \mathbf{y} \right], \quad (3.41) \\ &= \int_{\chi} \int_{\chi} E \left[\eta^2(\mathbf{x}) \eta^2(\mathbf{z}) \mid \sigma^2, \mathbf{y} \right] dG(\mathbf{x}) dG(\mathbf{z}). \end{aligned}$$

Now using (3.40) with $i = j$ and $k = l$ we obtain

$$\begin{aligned}
E [K_2^2 \mid \sigma^2, \mathbf{y}] &= \int_{\chi} \int_{\chi} 4m^*(\mathbf{x})m^*(C^*\mathbf{z}) \\
&\quad + (m^*(\mathbf{x})^2 + C^*(\mathbf{x}, \mathbf{x})\sigma^2) (m^*(\mathbf{z})^2 + C^*(\mathbf{z}, \mathbf{z})\sigma^2) \\
&\quad + 2C^*(\mathbf{x}, \mathbf{z})^2\sigma^4 dG(\mathbf{x}) dG(\mathbf{z}), \\
&= 4[MMC]\sigma^2 + 2[C^2]\sigma^4 + ([M^2] + [V]\sigma^2)^2. \quad (3.42)
\end{aligned}$$

The second element of (3.38) can be expanded as

$$\begin{aligned}
E [K_2 K^2 \mid \sigma^2, \mathbf{y}] &= E \left[\int_{\chi} \eta^2(\mathbf{x}) dG(\mathbf{x}) \int_{\chi} \eta(\mathbf{v}) dG(\mathbf{v}) \int_{\chi} \eta(\mathbf{z}) dG(\mathbf{z}) \mid \sigma^2, \mathbf{y} \right], \\
&= \int_{\chi} \int_{\chi} \int_{\chi} E[\eta^2(\mathbf{x})\eta(\mathbf{v})\eta(\mathbf{z}) \mid \sigma^2, \mathbf{y}] dG(\mathbf{x}) dG(\mathbf{v}) dG(\mathbf{z}), \quad (3.43)
\end{aligned}$$

again using (3.40) this can be expressed as

$$\begin{aligned}
E [K_2 K^2 \mid \sigma^2, \mathbf{y}] &= \int_{\chi} \int_{\chi} \int_{\chi} 4m^*(\mathbf{x})C^*(\mathbf{x}, \mathbf{v})\sigma^2 m^*(\mathbf{z}) \\
&\quad + 2C^*(\mathbf{x}, \mathbf{v})C^*(\mathbf{x}, \mathbf{z})\sigma^4 \\
&\quad + m^*(\mathbf{x})^2 m^*(\mathbf{v})m^*(\mathbf{z}) \\
&\quad + C^*(\mathbf{x}, \mathbf{x})C^*(\mathbf{v}, \mathbf{z})\sigma^4 \\
&\quad + m^*(\mathbf{v})C^*(\mathbf{x}, \mathbf{x})\sigma^2 m^*(\mathbf{z}) \\
&\quad + m^*(\mathbf{x})C^*(\mathbf{v}, \mathbf{z})\sigma^2 m^*(\mathbf{z}) dG(\mathbf{x})dG(\mathbf{v})dG(\mathbf{z}), \\
&= 4[M][MC]\sigma^2 + 2[CC]\sigma^4 + [M^2][M]^2 \\
&\quad + [V][C]\sigma^4 + [M]^2[V]\sigma^2 + [M^2][C]\sigma^2. \quad (3.44)
\end{aligned}$$

Finally,

$$\begin{aligned}
E [K^4 \mid \sigma^2, \mathbf{y}] &= E \left[\int_{\chi} \eta(\mathbf{x}) dG(\mathbf{x}) \int_{\chi} \eta(\mathbf{v}) dG(\mathbf{v}) \int_{\chi} \eta(\mathbf{z}) dG(\mathbf{z}) \int_{\chi} \eta(\mathbf{q}) dG(\mathbf{q}) \mid \sigma^2, \mathbf{y} \right], \\
&= \iiint\limits_{\chi} E [\eta(\mathbf{x})\eta(\mathbf{v})\eta(\mathbf{z})\eta(\mathbf{q}) \mid \sigma^2, \mathbf{y}] dG(\mathbf{x})dG(\mathbf{v})dG(\mathbf{z})dG(\mathbf{q}), \\
&= \iiint\limits_{\chi} 3C^*(\mathbf{x}, \mathbf{v})C^*(\mathbf{q}, \mathbf{z})\sigma^4 + 6m^*(\mathbf{x})C^*(\mathbf{q}, \mathbf{z})\sigma^2 m^*(\mathbf{v}) \\
&\quad + m^*(\mathbf{x})m^*(\mathbf{v})m^*(\mathbf{z})m^*(\mathbf{q}) dG(\mathbf{x}) dG(\mathbf{v}) dG(\mathbf{z}) dG(\mathbf{q}), \\
&= 3[C]^2\sigma^4 + 6[M]^2[C]\sigma^2 + [M]^4, \quad (3.45)
\end{aligned}$$

where

$$[M] = \int_{\chi} m^*(\mathbf{x}) dG(\mathbf{x}), \quad (3.46)$$

$$[MC] = \int_{\chi} \int_{\chi} m^*(\mathbf{x}) C^*(\mathbf{x}, \mathbf{v}) dG(\mathbf{x}) dG(\mathbf{v}), \quad (3.47)$$

$$[MMC] = \int_{\chi} \int_{\chi} m^*(\mathbf{x}) m^*(\mathbf{v}) C^*(\mathbf{x}, \mathbf{v}) dG(\mathbf{x}) dG(\mathbf{v}), \quad (3.48)$$

$$[CC] = \int_{\chi} \int_{\chi} \int_{\chi} C^*(\mathbf{x}, \mathbf{v}) C^*(\mathbf{x}, \mathbf{z}) dG(\mathbf{x}) dG(\mathbf{v}) dG(\mathbf{z}), \quad (3.49)$$

$$[C^2] = \int_{\chi} \int_{\chi} C^*(\mathbf{x}, \mathbf{v})^2 dG(\mathbf{x}) dG(\mathbf{v}), \quad (3.50)$$

$$[M^2] = \int_{\chi} m^*(\mathbf{x})^2 dG(\mathbf{x}), \quad (3.51)$$

$$[V] = \int_{\chi} C^*(\mathbf{x}, \mathbf{x}) dG(\mathbf{x}), \quad (3.52)$$

$$[C] = \int_{\chi} \int_{\chi} C^*(\mathbf{x}, \mathbf{v}) dG(\mathbf{x}) dG(\mathbf{v}). \quad (3.53)$$

Each of the expressions (3.46) to (3.53) can now be evaluated following expansion. In fact, the evaluation of $[M^2]$ and $[V]$ was implicit in the derivation of the expectation of L . The resulting expression for $E[(K_2 - K^2)^2]$ given σ^2 and \mathbf{y} is large and so is not listed.

To remove the conditioning of this expectation on σ^2 a further expectation must be taken with respect to the posterior distribution for $\sigma^2 \mid \mathbf{y}$ (3.10). The result of this action will be to cause instances of σ^2 and σ^4 to be replaced by their posterior expected values $\hat{\sigma}^2$ and $\hat{\sigma}^4$, where $\hat{\sigma}^4$ is calculated as $Var(\sigma^2) + E[\sigma^2]^2$. Thus the variance of the unknown distribution estimating L conditional on \mathbf{y} is obtained.

This development of the estimate of L has been already published as Haylock & O'Hagan [HO].

3.6. Selection of optimum design points

When performing a classical uncertainty analysis the points at which the computer algorithm is evaluated must be selected according to some random selection procedure, such as the Latin Hypercube sampling scheme, as discussed in 2.2.1.

This is necessary to ensure that the sample statistics can be used to make inference about the population values. In contrast, the method of selection of the design points for a Bayesian analysis is not subject to this constraint. In fact the Bayesian sample is specifically selected to provide the highest quality information about the hypothesis in question.

For this reason we consider the Bayesian methodology to be most advantageous when inference is required about computationally expensive functions. In such situations each element of \mathbf{y} , the vector of function evaluations, will be costly to obtain but by selecting the positions of the design points to provide the highest quality information, the total number of points required will be minimised.

The selection of optimum designs based on the optimisation of a criterion has been discussed in **2.3.5**. A criterion for the selection of optimum design points based on a simple squared distance loss function will now be derived as follows.

Initially, the loss function is defined as

$$L[\eta(\mathbf{x}_0), d(\mathbf{x}_0), \mathbf{x}_0, \mathbf{y}, X] = [\eta(\mathbf{x}_0) - d(\mathbf{x}_0)]^2 \quad (3.54)$$

where

X is the selected design,

\mathbf{y} is the vector of true observations evaluated at the design points X ,

\mathbf{x}_0 is the point on the parameter space at which it is required to estimate the value of the algorithm $\eta(\cdot)$,

$d(\mathbf{x}_0)$ is the estimate of the algorithm at the point \mathbf{x}_0 ,

$\eta(\mathbf{x}_0)$ is the true value of the algorithm at the point \mathbf{x}_0 .

Clearly, this function is not suitable since it depends on $\eta(\mathbf{x}_0)$, $d(\mathbf{x}_0)$, \mathbf{x}_0 , \mathbf{y} as well as X , the design. A criterion with which to select best designs must only depend on X since when selecting the design the other parameters will be unknown. The first step to obtaining from $L[\eta(\mathbf{x}_0), d(\mathbf{x}_0), \mathbf{x}_0, \mathbf{y}, X]$ such a criterion is to take the expectation of the loss function over $\eta(\mathbf{x}_0)$, the true value of the function, giving

$$L[d(\mathbf{x}_0), \mathbf{x}_0, \mathbf{y}, X] = E [(\eta(\mathbf{x}_0) - d(\mathbf{x}_0))^2 \mid d(\mathbf{x}_0), \mathbf{x}_0, \mathbf{y}, X]. \quad (3.55)$$

Moving to the next parameter, we now minimise the loss function over $d(\mathbf{x}_0)$ by replacing $d(\mathbf{x}_0)$ with the expected value of $\eta(\mathbf{x}_0)$, i.e., substituting $E[\eta(\mathbf{x}_0) \mid \mathbf{x}_0, \mathbf{y}, X]$ for $d(\mathbf{x}_0)$. This further modifies the loss function to reveal

$$\begin{aligned} L[\mathbf{x}_0, \mathbf{y}, X] &= E [(\eta(\mathbf{x}_0) - E[\eta(\mathbf{x}_0) \mid \mathbf{x}_0, \mathbf{y}, X])^2 \mid \mathbf{x}_0, \mathbf{y}, X] \\ &= Var[\eta(\mathbf{x}_0) \mid \mathbf{x}_0, \mathbf{y}, X] \end{aligned} \quad (3.56)$$

since $E[A - E[A]]^2 = Var[A]$.

Now in general,

$$Var[A \mid C] = E[Var[A \mid B, C] \mid C] + Var[E[A \mid B, C] \mid C] \quad (3.57)$$

thus

$$\begin{aligned} L[\mathbf{x}_0, \mathbf{y}, X] &= E[Var[\eta(\mathbf{x}_0) \mid \mathbf{x}_0, \mathbf{y}, X, \sigma^2] \mid \mathbf{x}_0, \mathbf{y}, X] \\ &\quad + Var[E[\eta(\mathbf{x}_0) \mid \mathbf{x}_0, \mathbf{y}, X, \sigma^2] \mid \mathbf{x}_0, \mathbf{y}, X]. \end{aligned} \quad (3.58)$$

Further,

$$E[\eta(\mathbf{x}_0) \mid \mathbf{x}_0, \mathbf{y}, X, \sigma^2] = m^*(\mathbf{x}_0) \quad (3.59)$$

which is independent of σ^2 , hence

$$Var[m^*(\mathbf{x}_0) \mid \sigma^2, \mathbf{x}_0, \mathbf{y}, X] = 0. \quad (3.60)$$

Also,

$$Var[\eta(\mathbf{x}_0) \mid \mathbf{x}_0, \mathbf{y}, X, \sigma^2] = \sigma^2 C^*(\mathbf{x}_0, \mathbf{x}_0) \quad (3.61)$$

so

$$\begin{aligned} E[Var[\eta(\mathbf{x}_0) \mid \mathbf{x}_0, \mathbf{y}, X, \sigma^2] \mid \mathbf{x}_0, \mathbf{y}, X] &= E[\sigma^2 C^*(\mathbf{x}_0, \mathbf{x}_0) \mid \mathbf{x}_0, \mathbf{y}, X] \\ &= E[\sigma^2 \mid \mathbf{x}_0, \mathbf{y}, X] C^*(\mathbf{x}_0, \mathbf{x}_0) \end{aligned} \quad (3.62)$$

and hence

$$L[\mathbf{x}_0, \mathbf{y}, X] = E[\sigma^2 \mid \mathbf{x}_0, \mathbf{y}, X] C^*(\mathbf{x}_0, \mathbf{x}_0). \quad (3.63)$$

Next, we must take the expectation of $L[\mathbf{x}_0, \mathbf{y}, X]$ with respect to \mathbf{x}_0 . In taking this expectation we must decide at which points on the parameter

space it is most important to best estimate the function. If all points on the parameter space are equally important then we might take the expectation with respect to a uniform distribution over the uncertain parameters. However, in this case the expectation of $L[\mathbf{x}_0, \mathbf{y}, X]$ will be taken with respect to the joint distribution of the uncertain parameters, denoted by $G(\mathbf{x})$. This means that point estimates of the function will be most accurately estimated for values of the uncertain parameters that, according to their prior distributions, are most likely to occur.

Thus,

$$L[\mathbf{y}, X] = \int_{\mathbf{x}} E[\sigma^2 | \mathbf{x}, \mathbf{y}, X] C^*(\mathbf{x}, \mathbf{x}) dG(\mathbf{x}). \quad (3.64)$$

This expression can be simplified further by observing that

$$E[\sigma^2 | \mathbf{x}_0, \mathbf{y}, X] = E[\sigma^2 | \mathbf{y}, X] \quad (3.65)$$

since the calculation of the expected value of σ^2 from its distribution (3.10,3.12), only involves functions of \mathbf{y} and X , thus

$$L[\mathbf{y}, X] = E[\sigma^2 | \mathbf{y}, X] \int_{\mathbf{x}} C^*(\mathbf{x}, \mathbf{x}) dG(\mathbf{x}). \quad (3.66)$$

Finally, to obtain the desired criterion the expectation of the loss function over \mathbf{y} must be taken.

Now considering the first part,

$$E[E[\sigma^2 | \mathbf{y}, X] | X] = E[\sigma^2 | X] \quad (3.67)$$

using the general formula

$$E[A | C] = E[E[A | B, C] | C] \quad (3.68)$$

and since the second part, $\int_{\mathbf{x}} C^*(\mathbf{x}, \mathbf{x}) dG(\mathbf{x})$, does not depend on \mathbf{y} it remains unchanged when its expectation with respect to \mathbf{y} is taken. Hence,

$$\begin{aligned} L[X] &= E[\sigma^2 | X] \int_{\mathbf{x}} C^*(\mathbf{x}, \mathbf{x}) dG(\mathbf{x}) \\ &= E[\sigma^2 | X] \int_{\mathbf{x}} [C(\mathbf{x}, \mathbf{x}) - \mathbf{t}(\mathbf{x})^T \mathbf{A}^{-1} \mathbf{t}(\mathbf{x})] dG(\mathbf{x}). \end{aligned} \quad (3.69)$$

Two further steps can be taken to simplify the selection criterion. Firstly, considering the expectation of σ^2 ,

$$E [\sigma^2 \mid X] = E [\sigma^2]$$

since knowledge about the positions of the design points without also obtaining the values of \mathbf{y} will provide no extra information about the expected value of σ^2 . Thus this expectation will remain constant for all designs and so constitutes only a scaling factor in the design selection criterion and may be excluded. Hence,

$$L[X] \propto \int_{\chi} [C(\mathbf{x}, \mathbf{x}) - \mathbf{t}(\mathbf{x})^T \mathbf{A}^{-1} \mathbf{t}(\mathbf{x})] dG(\mathbf{x}). \quad (3.70)$$

The other simplification of the criterion is based on the prior assumption that $C(\mathbf{x}, \mathbf{x}) = 1 \forall \mathbf{x} \in \chi$, see **3.2**, so this component of $L(X)$ can also be excluded since it will remain constant for all designs .

Therefore, the best design, of size n , independent of any knowledge about \mathbf{y} , \mathbf{x}_0 , $d(\mathbf{x})$ or $\eta(\mathbf{x})$, will be that which **maximises**

$$\int_{\chi} \mathbf{t}(\mathbf{x})^T \mathbf{A}^{-1} \mathbf{t}(\mathbf{x}) dG(\mathbf{x}). \quad (3.71)$$

This can be thought of, in general terms, as the design that is predicted to give the maximum reduction in the average posterior variance of the predictor, weighted by $G(\mathbf{x})$, over the parameter space.

4. METHODOLOGY FOR BAYESIAN UNCERTAINTY ANALYSIS OF UNCERTAIN PARAMETERS WITH NORMAL PRIOR DISTRIBUTIONS

In this chapter, the methodology appropriate to performing a Bayesian uncertainty analysis in which the uncertain parameters are associated with independent normal prior distributions will be examined.

We will first outline some definitions. Let $\eta(\mathbf{x})$ represent the algorithm, which has p uncertain inputs/parameters where $\mathbf{x} = (x_1, x_2, \dots, x_p)$. The symbol x_α , where $1 \leq \alpha \leq p$, will be used to represent a general element of \mathbf{x} . We will assume that the prior knowledge about each of these parameters is best expressed as independent normal prior distributions. The mean and variance of these distributions will be represented by the vectors

$$\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p] \quad (4.1)$$

and

$$\boldsymbol{\sigma}^2 = [\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2]. \quad (4.2)$$

It would in theory be possible to use multivariate normal distributions to represent groups of correlated uncertain parameters. However, this would increase considerably the complexity of the development of the uncertainty measures. In addition, it would be necessary, as part of the prior information, to define the values of the covariances between the various correlated parameters. In cases where little information is available about the uncertain parameters it would be difficult to provide such values with any degree of confidence. Further, in many cases it will be possible to transform correlated uncertain parameters to obtain a set of independent parameters.

Thus, the Bayesian theory will be developed assuming prior independence between the uncertain parameters.

Next, we will represent a set of n design points as $X = [\mathbf{x}_1, \dots, \mathbf{x}_p]^T$ where $\mathbf{x}_\alpha = [x_{\alpha,1}, \dots, x_{\alpha,n}]^T$. Thus, $x_{\alpha,i}$ represents the value of uncertain parameter x_α in the i^{th} design point.

Vector $\mathbf{h}(\cdot)$ will be defined using the assumption that the function $\eta(\cdot)$ is approximately a linear additive function of the uncertain parameters, thus

$$\mathbf{h}(\mathbf{x}) = [1, x_1, x_2, \dots, x_p]. \quad (4.3)$$

It would be possible to develop the Bayesian uncertainty measures using a more complicated form of $\mathbf{h}(\cdot)$, eg, with non-linear components. However, in general the true form of the relationship will be unknown and there is no reason to suppose that in this situation more complicated functions in $\mathbf{h}(\cdot)$ would be any better at approximating the true relationship than would linear functions. Further, the inclusion of more complicated functions would greatly increase the complexity of the calculation of the uncertainty measures.

Finally, we will also define the correlation function as

$$C(\mathbf{x}, \mathbf{x}') = \exp [-(\mathbf{x} - \mathbf{x}')^T \mathbf{Z} (\mathbf{x} - \mathbf{x}')] \quad (4.4)$$

where

$$\mathbf{Z} = \text{diag}(\boldsymbol{\delta}) \quad (4.5)$$

and

$$\boldsymbol{\delta} = \left[\frac{1}{\delta_1}, \frac{1}{\delta_2}, \dots, \frac{1}{\delta_p} \right]. \quad (4.6)$$

By assuming this exponential form for $C(\cdot, \cdot)$ we indicate a belief that the function $\eta(\cdot)$ is locally smooth with respect to each of its uncertain parameters and further that it is infinitely differentiable.

The elements of \mathbf{Z} describe our prior information about the smoothness of $\eta(\cdot)$ with respect to each of the uncertain parameters and about the extent of correlations between the uncertain parameters. In this general scenario, the off-diagonal elements of \mathbf{Z} are set to zero to indicate an initial belief that the uncertain parameters are independent. That is, we

do not believe that the smoothness of $\eta(\cdot)$ with respect to one parameter is influenced by the value of any other parameter.

The actual size of the elements $\delta_1, \dots, \delta_p$ are determined by our beliefs about the smoothness of $\eta(\cdot)$ with respect to each of the uncertain parameters. Assigning a large value to an element implies a belief that $\eta(\cdot)$ is smooth with respect to that parameter, a small value implies a belief that $\eta(\cdot)$ is rough.

The assumption that the uncertain parameters are independent, as indicated by the diagonal nature of \mathbf{Z} , is very useful as it enables us to consider this analysis not as a p -dimensional problem but as the product of p one-dimensional problems. This will greatly simplify the Bayesian uncertainty analysis.

Thus, consider a single dimension, α , from the p dimensions of the problem. If $x_{\alpha,i}$ represents the coordinate of design point, i in this dimension and $x_{\alpha,j}$ that of point j then we will represent the contribution to the correlation between $\eta(\mathbf{x}_i)$ and $\eta(\mathbf{x}_j)$ from this dimension by

$$C(x_{\alpha,i}, x_{\alpha,j}) = \exp \left[\frac{-(x_{\alpha,i} - x_{\alpha,j})^2}{\delta_\alpha} \right]. \quad (4.7)$$

Thus, the total correlation over all dimensions $C(\mathbf{x}_i, \mathbf{x}_j)$ is given by

$$C(\mathbf{x}_i, \mathbf{x}_j) = \prod_{\theta=1, \dots, p} \exp \left[\frac{-(x_{\theta,i} - x_{\theta,j})^2}{\delta_\theta} \right]. \quad (4.8)$$

This technique of rewriting p -dimensional expressions as the product of p ‘1-dimensional’ expressions will be used throughout this chapter.

Prior to observing at least two values of $\eta(\cdot)$ there is no way of objectively estimating the smoothness of $\eta(\cdot)$ with respect to each parameter. The best that can be done is to make estimates based on prior subjective information. Smoothness is not an absolute quantity it depends on scale of the function. Thus two different δ values can represent the same degree of smoothness depending on the range of possible values of the parameters.

As in the previous chapter, however, the problem of selecting the design points will be considered first.

4.1. Selection of the design points

In **3.2** a criterion for the selection of the design points was obtained. In its most general form it states that for a particular design X the value of the criterion is defined as $\int_{\chi} \mathbf{t}(\mathbf{x})^T \mathbf{A}^{-1} \mathbf{t}(\mathbf{x}) dG(\mathbf{x})$, (3.71).

For the case of p independent uncertain parameters considered here this criterion can be re-written as

$$\int_{\chi} \mathbf{t}(\mathbf{x})^T \mathbf{A}^{-1} \mathbf{t}(\mathbf{x}) dG(\mathbf{x}) = \prod_{\theta=1, \dots, p} \left[\int_{\chi_{\theta}} \mathbf{t}(\mathbf{x}_{\theta})^T \mathbf{A}_{\theta}^{-1} \mathbf{t}(\mathbf{x}_{\theta}) dG(x_{\theta}) \right] \quad (4.9)$$

where $G(x_{\theta})$ is the normal density function of the prior distribution for uncertain parameter θ and χ_{θ} is the parameter space of x_{θ} . Now $\mathbf{t}(\mathbf{x}_{\theta})^T \mathbf{A}_{\theta}^{-1} \mathbf{t}(\mathbf{x}_{\theta})$ can be re-written as $tr(\mathbf{A}_{\theta}^{-1} \mathbf{t}(\mathbf{x}_{\theta}) \mathbf{t}(\mathbf{x}_{\theta})^T)$, because

$$\begin{aligned} \int_{\chi} \mathbf{t}(\mathbf{x})^T \mathbf{A}^{-1} \mathbf{t}(\mathbf{x}) dG(\mathbf{x}) &= \prod_{\theta=1, \dots, p} \left[\int_{\chi_{\theta}} tr(\mathbf{A}_{\theta}^{-1} \mathbf{t}(\mathbf{x}_{\theta}) \mathbf{t}(\mathbf{x}_{\theta})^T) dG(x_{\theta}) \right] \\ &= \prod_{\theta=1, \dots, p} \left[tr \left(\int_{\chi_{\theta}} \mathbf{A}_{\theta}^{-1} \mathbf{t}(\mathbf{x}_{\theta}) \mathbf{t}(\mathbf{x}_{\theta})^T dG(x_{\theta}) \right) \right] \end{aligned} \quad (4.10)$$

since the integration and calculation of the trace of a matrix are commutative operations. Next, since the matrix \mathbf{A}_{θ}^{-1} is independent of \mathbf{x}_{θ} the criterion can be expressed as

$$\begin{aligned} \int_{\chi} \mathbf{t}(\mathbf{x})^T \mathbf{A}^{-1} \mathbf{t}(\mathbf{x}) dG(\mathbf{x}) &= \prod_{\theta=1, \dots, p} \left[tr \left(\mathbf{A}_{\theta}^{-1} \int_{\chi_{\theta}} \mathbf{t}(\mathbf{x}_{\theta}) \mathbf{t}(\mathbf{x}_{\theta})^T dG(x_{\theta}) \right) \right] \\ &= \prod_{\theta=1, \dots, p} [tr(\mathbf{A}_{\theta}^{-1} \mathbf{P}_{\theta})] \end{aligned} \quad (4.11)$$

where \mathbf{P}_{θ} is as defined in (3.33).

In order to evaluate this criterion we need to evaluate \mathbf{A}_{θ} and \mathbf{P}_{θ} . We obtain the $[i, j]^{th}$ element of the $n \times n$ matrix \mathbf{P}_{θ} as

$$\begin{aligned} \mathbf{P}_{\theta}[i, j] &= \int_{\chi_{\theta}} C(x_{\theta}, x_{\theta, i}) C(x_{\theta}, x_{\theta, j}) dG(x_{\theta}), \\ &= \int_{\chi_{\theta}} \frac{1}{\sqrt{2\pi\sigma_{\theta}^2}} \exp \left[\frac{-(x_{\theta} - x_{\theta, i})^2 - (x_{\theta} - x_{\theta, j})^2}{\delta_{\theta}} \right] \\ &\quad \times \exp \left[\frac{-(x_{\theta} - \mu_{\theta})^2}{2\sigma_{\theta}^2} \right] dx_{\theta}, \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \frac{x_{\theta}}{\sqrt{2\pi\sigma_{\theta}^2}} \\
&\quad \times \exp \left[\frac{-(4\sigma_{\theta}^2 + \delta_{\theta})}{2\sigma_{\theta}^2\delta_{\theta}} \left[x_{\theta} - \frac{2x_{\theta,i}\sigma_{\theta}^2 + 2x_{\theta,j}\sigma_{\theta}^2 + \mu_{\theta}\delta_{\theta}}{4\sigma_{\theta}^2 + \delta_{\theta}} \right]^2 \right] dx_{\theta} \\
&\quad \times \exp \left[\frac{-\delta_{\theta}^2 [(\mu_{\theta} - x_{\theta,i})^2 + (\mu_{\theta} - x_{\theta,j})^2] - 2\sigma_{\theta}^2(x_{\theta,i} - x_{\theta,j})^2}{(\delta_{\theta}^2 + 4\sigma_{\theta}^2)\delta_{\theta}^2} \right], \\
&= \exp \left[\frac{-\delta_{\theta}^2 [(\mu_{\theta} - x_{\theta,i})^2 + (\mu_{\theta} - x_{\theta,j})^2] - 2\sigma_{\theta}^2(x_{\theta,i} - x_{\theta,j})^2}{(\delta_{\theta}^2 + 4\sigma_{\theta}^2)\delta_{\theta}^2} \right] \\
&\quad \times \frac{\delta_{\theta}^2}{\sqrt{\delta_{\theta}^2 + 4\sigma_{\theta}^2}}, \tag{4.12}
\end{aligned}$$

and the $[i, j]^{th}$ element of \mathbf{A}_{θ} as

$$\mathbf{A}_{\theta}[i, j] = C(x_{\theta,i}, x_{\theta,j}) = \exp \left[\frac{-(x_{\theta,i} - x_{\theta,j})^2}{\delta_{\theta}} \right]. \tag{4.13}$$

In order to use the criterion it is necessary to supply the values of δ . It would be an unusual situation to know the true values of these parameters when selecting a design, so it is necessary to estimate them. Before obtaining at least two evaluations of the algorithm for different values of the uncertain parameters it is not possible to objectively estimate the true values of the smoothing parameters. Thus, subjective prior knowledge will be used to provide initial estimates for δ .

4.2. Updating the smoothing parameter values

The selection of a design and the evaluation of the output of the algorithm at each point in the design, \mathbf{y} , provides objective information that can be used to improve on the initial estimates of the smoothing parameters used to select the design.

The multivariate normal distribution for \mathbf{y} (3.7) contains an implicit conditioning on δ through the \mathbf{A} matrix. Its density function can be rewritten explicitly conditional on δ as,

$$f(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \delta) = \frac{|\mathbf{A}|^{-\frac{1}{2}}}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp \left[-(\mathbf{y} - \mathbf{H}\boldsymbol{\beta})^T \frac{\mathbf{A}^{-1}}{2\sigma^2} (\mathbf{y} - \mathbf{H}\boldsymbol{\beta}) \right]. \tag{4.14}$$

This density function can be thought of as a likelihood function for $\boldsymbol{\beta}$, σ^2 and δ . Now by taking the product of this likelihood with the conjugate joint

prior distribution for β and σ^2 (3.6), and independently an uninformative uniform prior on δ , the joint posterior distribution can be obtained as

$$f(\beta, \sigma^2, \delta | \mathbf{y}) \propto \frac{|\mathbf{A}|^{-\frac{1}{2}}}{\sigma^{n+2}(2\pi)^{\frac{n}{2}}} \exp \left[-(\mathbf{y} - \mathbf{H}\beta)^T \frac{\mathbf{A}^{-1}}{2\sigma^2} (\mathbf{y} - \mathbf{H}\beta) \right]. \quad (4.15)$$

Ideally, the parameters β, σ^2 would now be integrated out of this joint distribution to obtain a marginal posterior distribution for $\delta | \mathbf{y}$. Now, the first integration to remove β can be performed to reveal

$$f(\sigma^2, \delta | \mathbf{y}) \propto \frac{|\mathbf{A}|^{-\frac{1}{2}} |\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H}|^{-\frac{1}{2}}}{\sigma^2 (2\pi)^{\frac{n}{2}}} \exp \left[-(\mathbf{y} - \mathbf{H}\hat{\beta})^T \frac{\mathbf{A}^{-1}}{2\sigma^2} (\mathbf{y} - \mathbf{H}\hat{\beta}) \right] \quad (4.16)$$

where $\hat{\beta} = (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H}^T)^{-1} \mathbf{H}^T \mathbf{A}^{-1} \mathbf{y}$ (3.11). However, it is not possible to further integrate this distribution with respect to σ^2 . The integral that must be performed takes the general form

$$\int_0^\infty \frac{c}{\sigma^2} \exp \left[-\frac{k}{2\sigma^2} \right] d\sigma^2 \quad (4.17)$$

where c and k are constants with respect to σ^2 . If a change of variable is performed, using $x = \frac{k}{2\sigma^2}$ and $d\sigma^2 = -\frac{k}{2x^2} dx$, then this integral becomes

$$\int_0^\infty \frac{c}{x} \exp[-x] dx \quad (4.18)$$

which does not converge.

Thus posterior estimates of δ were obtained by calculating the posterior mode of either the full joint distribution of $\beta, \sigma^2, \delta | \mathbf{y}$, (4.15) or the marginal distribution of $\sigma^2, \delta | \mathbf{y}$, (4.16). In both cases the point estimates of δ were calculated by partially differentiating the distribution function with respect to each parameter, setting each of the resulting expressions to zero and solving them simultaneously.

In the rest of the chapter all expressions that contain the correlation function will be considered to have an implicit conditioning on the value of δ .

4.3. Estimation of $\eta(\mathbf{x}_0)$ for normally distributed uncertain parameters

In **3.3** we obtained a t -distribution (3.17) to represent our knowledge about $\eta(\cdot)$ at a general point \mathbf{x}_0 . From this distribution, a predictor and a measure of the accuracy of the predictor, the mean and variance, were obtained as

$$E[\eta(\mathbf{x}_0) | \mathbf{y}] = m^*(\mathbf{x}_0) \quad (4.19)$$

and

$$Var[\eta(\mathbf{x}_0) | \mathbf{y}] = C^*(\mathbf{x}_0, \mathbf{x}_0) \frac{\mathbf{y}^T \left(\mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{H} (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H})^{-1} \mathbf{A}^{-1} \right) \mathbf{y}}{n - q - 2}. \quad (4.20)$$

Now, by substituting into these expressions the specific forms of $\mathbf{h}(\mathbf{x}_0)$ and $C(\mathbf{x}_0, \mathbf{x}_0)$ defined for parameters with independent normally distributed priors in (4.3) and (4.4) respectively, estimates of $\eta(\mathbf{x}_0)$ for all $\mathbf{x} \in \chi$ can be obtained.

4.4. Estimation of K the expected value of $\eta(\mathbf{X})$ for normally distributed uncertain parameters

The general theory relating to this question is detailed in **3.4**. There a t distribution (3.26) was described that contained all the subjective and objective information available about the expected value of the algorithm, defined as K . From the t distribution generic expressions for $E[K | \mathbf{y}]$ and $Var[K | \mathbf{y}]$ were obtained as

$$E[K | \mathbf{y}] = \mathbf{R}\hat{\boldsymbol{\beta}} + \mathbf{T}\mathbf{A}^{-1} \left(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}} \right) \quad (4.21)$$

and

$$Var[K | \mathbf{y}] = W \frac{\mathbf{y}^T \left(\mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{H} (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H})^{-1} \mathbf{A}^{-1} \right) \mathbf{y}}{n - q - 2}, \quad (4.22)$$

where W is defined in (3.22).

To derive $E[K | \mathbf{y}]$ and $Var[K | \mathbf{y}]$ specific to parameters with normally distributed priors we need to evaluate the expressions \mathbf{R} , \mathbf{T} and U defined in equations (3.23), (3.24), (3.25).

Consider the evaluation of vector expression \mathbf{R} ,

$$\begin{aligned}\mathbf{R} &= \int_{\chi} \mathbf{h}(\mathbf{x})^T dG(\mathbf{x}), \\ &= \left[1, \int_{\chi_1} x_1 dG(x_1), \dots, \int_{\chi_\alpha} x_\alpha dG(x_\alpha), \dots, \int_{\chi_p} x_p dG(x_p) \right] \quad (4.23)\end{aligned}$$

A general element of this vector, excluding the first, can be expanded in the form

$$\int_{-\infty}^{\infty} \frac{x_\alpha}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left[\frac{-(x - \mu_\alpha)^2}{2\sigma_\alpha^2}\right] dx_\alpha = \mu_\alpha, \quad (4.24)$$

thus,

$$\mathbf{R} = [1, \mu_1, \dots, \mu_p]. \quad (4.25)$$

The evaluation of \mathbf{T} is as follows,

$$\begin{aligned}\mathbf{T} &= \int_{\chi} \mathbf{t}(\mathbf{x}) dG(\mathbf{x}), \\ &= \left[\int_{\chi} C(\mathbf{x}, \mathbf{x}_1) dG(\mathbf{x}), \dots, \int_{\chi} C(\mathbf{x}, \mathbf{x}_n) dG(\mathbf{x}) \right]. \quad (4.26)\end{aligned}$$

Now expanding the element relating to the i^{th} design point we obtain

$$\int_{\chi} C(\mathbf{x}, \mathbf{x}_i) dG(\mathbf{x}) = \prod_{\theta=1, \dots, p} \int_{\chi_\theta} C(x_\theta, x_{\theta,i}) dG(x_\theta), \quad (4.27)$$

$$\begin{aligned}&= \prod_{\theta=1, \dots, p} \int_{-\infty}^{\infty} \exp\left[\frac{-(x_\theta - x_{\theta,i})^2}{\delta_\theta}\right] \\ &\quad \times \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left[\frac{-(x_\theta - \mu_\theta)^2}{2\sigma_\theta^2}\right] dx_\theta, \\ &= \prod_{\theta=1, \dots, p} \exp\left[\frac{-(x_\theta - \mu_\theta)^2}{\delta_\theta + 2\sigma_\theta^2}\right] \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \\ &\quad \times \exp\left[\frac{-(2\sigma_\theta^2 + \delta_\theta^2)}{2\sigma_\theta^2\delta_\theta^2} \left[x_\theta - \left(\frac{2x_{\theta,i}\sigma_\theta^2 + \mu_\theta\sigma_\theta^2}{2\sigma_\theta^2 + \delta_\theta^2}\right)\right]^2\right] dx_\theta. \\ &= \prod_{\theta=1, \dots, p} \exp\left[\frac{-(x_\theta - \mu_\theta)^2}{\delta_\theta + 2\sigma_\theta^2}\right] \sqrt{\frac{\delta_\theta}{\delta_\theta + 2\sigma_\theta^2}} \quad (4.28) \\ &\quad \times \int_{-\infty}^{\infty} \sqrt{\frac{\delta_\theta + 2\sigma_\theta^2}{\sigma_\theta^2\delta_\theta^2}} \frac{1}{\sqrt{2\pi}} \\ &\quad \times \exp\left[\frac{-(2\sigma_\theta^2 + \delta_\theta^2)}{2\sigma_\theta^2\delta_\theta^2} \left[x_\theta - \left(\frac{2x_{\theta,i}\sigma_\theta^2 + \mu_\theta\sigma_\theta^2}{2\sigma_\theta^2 + \delta_\theta^2}\right)\right]^2\right] dx_\theta.\end{aligned}$$

$$= \prod_{\theta=1,\dots,p} \sqrt{\frac{\delta_\theta}{\delta_\theta + 2\sigma_\theta^2}} \exp \left[\frac{-(x_{\theta,i} - \mu_\theta)^2}{\delta_\theta + 2\sigma_\theta^2} \right], \quad (4.29)$$

where (4.29) is obtained from (4.28) using the fact that the integral is that of a normal density function, which always integrates to one. The elements of \mathbf{T} can now be calculated.

The derivation of U is slightly more complicated in that it involves a double integral, over $dG(\mathbf{x})$ and $dG(\mathbf{z})$.

$$\begin{aligned} U &= \int_{\chi} \int_{\chi} C(\mathbf{x}, \mathbf{z}) dG(\mathbf{x}) dG(\mathbf{z}), \\ &= \prod_{\theta=1,\dots,p} \int_{\chi_\theta} \int_{\chi_\theta} C(x_\theta, z_\theta) dG(x_\theta) dG(z_\theta), \end{aligned} \quad (4.30)$$

$$= \prod_{\theta=1,\dots,p} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left[\frac{-(x_\theta - z_\theta)^2}{\delta_\theta} \right] dG(x_\theta) dG(z_\theta). \quad (4.31)$$

Now, consider the integral with respect to $G(x_\theta)$. A similar integration has already been performed in the evaluation of an element of \mathbf{T} above, thus

$$\begin{aligned} \int_{-\infty}^{\infty} \exp \left[\frac{-(x_\theta - z_\theta)^2}{\delta_\theta} \right] dG(x_\theta) &= \int_{-\infty}^{\infty} \exp \left[\frac{-(x_\theta - z_\theta)^2}{\delta_\theta} \right] \\ &\quad \times \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp \left[\frac{-(x_\theta - \mu_\theta)^2}{2\sigma_\theta^2} \right] dx_\theta, \\ &= \sqrt{\frac{\delta_\theta}{\delta_\theta + 2\sigma_\theta^2}} \exp \left[\frac{-(z_\theta - \mu_\theta)^2}{\delta_\theta + 2\sigma_\theta^2} \right]. \end{aligned} \quad (4.32)$$

The expression for U can now be written as

$$\begin{aligned} U &= \prod_{\theta=1,\dots,p} \int_{-\infty}^{\infty} \sqrt{\frac{\delta_\theta}{\delta_\theta + 2\sigma_\theta^2}} \exp \left[\frac{-(z_\theta - \mu_\theta)^2}{\delta_\theta + 2\sigma_\theta^2} \right] dG(z_\theta), \\ &= \prod_{\theta=1,\dots,p} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp \left[\frac{-(z_\theta - \mu_\theta)^2}{2\sigma_\theta^2} \right] \\ &\quad \times \sqrt{\frac{\delta_\theta}{\delta_\theta + 2\sigma_\theta^2}} \exp \left[\frac{-(z_\theta - \mu_\theta)^2}{\delta_\theta + 2\sigma_\theta^2} \right] dz_\theta, \\ &= \prod_{\theta=1,\dots,p} \sqrt{\frac{\delta_\theta}{\delta_\theta + 2\sigma_\theta^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp \left[\frac{-(4\sigma_\theta^2 + \delta_\theta)(z_\theta - \mu_\theta)^2}{\sigma_\theta^2(2\sigma_\theta^2 + \delta_\theta)} \right] dz_\theta. \end{aligned}$$

$$\begin{aligned}
&= \prod_{\theta=1,\dots,p} \sqrt{\frac{\delta_\theta}{\delta_\theta + 2\sigma_\theta^2}} \sqrt{\frac{\delta_\theta + 2\sigma_\theta^2}{\delta_\theta + 4\sigma_\theta^2}} \\
&\quad \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \sqrt{\frac{\delta_\theta + 4\sigma_\theta^2}{\delta_\theta + 2\sigma_\theta^2}} \exp \left[\frac{-(4\sigma_\theta^2 + \delta_\theta)(z_\theta - \mu_\theta)^2}{\sigma_\theta^2(2\sigma_\theta^2 + \delta_\theta)} \right] dz_\theta, \\
&= \prod_{\theta=1,\dots,p} \sqrt{\frac{\delta_\theta}{\delta_\theta + 2\sigma_\theta^2}} \sqrt{\frac{\delta_\theta + 2\sigma_\theta^2}{\delta_\theta + 4\sigma_\theta^2}}, \\
&= \prod_{\theta=1,\dots,p} \sqrt{\frac{\delta_\theta}{\delta_\theta + 4\sigma_\theta^2}}. \tag{4.33}
\end{aligned}$$

Having evaluated the above integrals, the specific t distribution representing the available information about $K \mid \mathbf{y}$ along with $E[K \mid \mathbf{y}]$ and $Var[K \mid \mathbf{y}]$ can be calculated.

4.5. Estimation of L the variance of $\eta(\mathbf{X})$ for normally distributed uncertain parameters

The general theory relating to this problem is described in 3.5 where the first two moments of the distribution describing the subjective and objective information about $L = Var[\eta(\mathbf{X})]$ are obtained. All that remains is to interpret the general theory and obtain the formulae specific to uncertain parameters with normal prior distributions.

4.5.1. Calculation of the posterior expectation of L

Now as described in 3.5.1 the expected value of $L \mid \mathbf{y}$ can be written as

$$E[L \mid \mathbf{y}] = E[K_2 \mid \mathbf{y}] - (Var[K \mid \mathbf{y}] + E[K \mid \mathbf{y}]^2). \tag{4.34}$$

The t distribution obtained in the previous section, 4.4, provides values of $E[K \mid \mathbf{y}]$ and $Var[K \mid \mathbf{y}]$. It only remains to evaluate $E[K_2 \mid \mathbf{y}]$.

This expectation is calculated using the formula (3.32) with σ^2 replaced by $\hat{\sigma}^2$ derived from (3.10) and where \mathbf{P} , \mathbf{Q} , and \mathbf{S} , the unknown components, are defined by (3.33), (3.34) and (3.35).

Now,

$$\mathbf{P} = \int_{\chi} \mathbf{t}(\mathbf{x})\mathbf{t}(\mathbf{x})^T dG(\mathbf{x}). \tag{4.35}$$

is an $n \times n$ matrix that has elements of the form

$$\mathbf{P}[i, j] = \prod_{\theta=1, \dots, p} \mathbf{P}_\theta[i, j] \quad (4.36)$$

where $\mathbf{P}_\theta[i, j]$ is defined in (4.12). Further, \mathbf{Q} is a $(p+1) \times (p+1)$ matrix defined as

$$\mathbf{Q} = \int_{\chi} \mathbf{h}(\mathbf{x}) \mathbf{h}(\mathbf{x})^T dG(\mathbf{x}) \quad (4.37)$$

such that

$$\mathbf{Q}[1, 1] = 1, \quad (4.38)$$

$$\mathbf{Q}[1, \alpha + 1] = \mathbf{Q}[\alpha + 1, 1] = \int_{\chi_\alpha} \mathbf{h}[\alpha] dG(x_\alpha) = \mu_\alpha, \quad 1 \leq \alpha \leq p, \quad (4.39)$$

and

$$\begin{aligned} \mathbf{Q}[\phi, \alpha] &= \int_{\chi_\phi} \mathbf{h}[\phi] dG(x_\phi) \times \int_{\chi_\alpha} \mathbf{h}[\alpha] dG(x_\alpha) \\ &= \int_{-\infty}^{\infty} \frac{x_\phi}{\sqrt{2\pi\sigma_\phi^2}} \exp\left[\frac{-(x - \mu_\phi)^2}{2\sigma_\phi^2}\right] dx_\phi \\ &\quad \times \int_{-\infty}^{\infty} \frac{x_\alpha}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left[\frac{-(x - \mu_\alpha)^2}{2\sigma_\alpha^2}\right] dx_\alpha \\ &= \mu_\phi \mu_\alpha, \quad 1 \leq \phi, \alpha \leq p. \end{aligned} \quad (4.40)$$

Finally, \mathbf{S} is an $n \times (p+1)$ matrix defined as

$$\mathbf{S} = \int_{\chi} \mathbf{t}(\mathbf{x}) \mathbf{h}(\mathbf{x})^T dG(\mathbf{x}). \quad (4.41)$$

The first column of this matrix is the vector \mathbf{T} , and of the other p columns the i^{th} element of column α will take the form

$$\int_{\chi_\alpha} x_\alpha C(x_\alpha, x_{\alpha,i}) dG(x_\alpha) \times \prod_{\theta=1, \dots, p, \theta \neq \alpha} \int_{\chi_\theta} C(x_\theta, x_{\theta,i}) dG(x_\theta) \quad (4.42)$$

which can be expanded as

$$\begin{aligned} &\int_{-\infty}^{\infty} \exp\left[\frac{-(x_\alpha - x_{\alpha,i})^2}{\delta_\alpha}\right] \frac{x_\alpha}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left[\frac{-(x_\alpha - \mu_\alpha)^2}{2\sigma_\alpha^2}\right] dx_\alpha \\ &\times \prod_{\theta=1, \dots, p, \theta \neq \alpha} \int_{-\infty}^{\infty} \exp\left[\frac{-(x_\theta - x_{\theta,i})^2}{\delta_\theta}\right] \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left[\frac{-(x_\theta - \mu_\theta)^2}{2\sigma_\theta^2}\right] dx_\theta, \\ &= \exp\left[\frac{-(x_{\alpha,i} - \mu_\alpha)^2}{\delta_\alpha + 2\sigma_\alpha^2}\right] \times \end{aligned}$$

$$\begin{aligned}
& \int_{-\infty}^{\infty} \frac{x_{\alpha}}{\sqrt{2\pi\sigma_{\alpha}^2}} \exp \left[\frac{-(2\sigma_{\alpha}^2 + \delta_{\alpha})}{2\sigma_{\alpha}^2\delta_{\alpha}} \left[x_{\alpha} - \frac{2x_{\alpha,i}\sigma_{\alpha}^2 + \mu_{\alpha}\delta_{\alpha}}{2\sigma_{\alpha}^2 + \delta_{\alpha}} \right]^2 \right] dx_{\alpha} \\
& \times \prod_{\theta=1, \dots, p, \theta \neq \alpha} \exp \left[\frac{-(x_{\theta,i} - \mu_{\theta})^2}{\delta_{\theta} + 2\sigma_{\theta}^2} \right] \\
& \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{\theta}^2}} \exp \left[\frac{-(2\sigma_{\theta}^2 + \delta_{\theta})}{2\sigma_{\theta}^2\delta_{\theta}} \left[x_{\theta} - \frac{2x_{\theta,i}\sigma_{\theta}^2 + \mu_{\theta}\delta_{\theta}}{2\sigma_{\theta}^2 + \delta_{\theta}} \right]^2 \right] dx_{\theta}, \\
& = \left(\frac{2x_{\alpha,i}\sigma_{\alpha}^2 + \delta_{\alpha}\mu_{\alpha}}{\delta_{\alpha} + 2\sigma_{\alpha}^2} \right) \times \prod_{\theta=1, \dots, p} \sqrt{\frac{\delta_{\theta}}{\delta_{\theta} + 2\sigma_{\theta}^2}} \exp \left[\frac{-(x_{\theta,i} - \mu_{\theta})^2}{\delta_{\theta} + 2\sigma_{\theta}^2} \right]. \quad (4.43)
\end{aligned}$$

Having evaluated the above integrals, all the components of $E[L_2 | \mathbf{y}]$ are now available and posterior expectation of L can be calculated.

4.5.2. Calculation of the posterior variance of L

The general theory for this problem is described in 3.5.2 where $Var[L | \mathbf{y}]$ is given by (3.36) as

$$Var[L | \mathbf{y}] = E[(K_2 - K^2)^2 | \mathbf{y}] - E[K_2 - K^2 | \mathbf{y}]^2. \quad (4.44)$$

The component $E[(K_2 - K^2)^2 | \mathbf{y}]^2$ is the square of the expectation of $L | \mathbf{y}$ derived in the previous section. Thus it only remains to calculate the specific form of $E[(K_2 - K^2)^2 | \mathbf{y}]$ for the case of uncertain parameters with normally distributed priors. We recall that

$$E[(K_2 - K^2)^2 | \sigma^2, \mathbf{y}] = E[K_2^2 | \mathbf{y}] - 2E[K_2 K^2 | \mathbf{y}] + E[K^4 | \mathbf{y}], \quad (4.45)$$

where each component is defined by (3.41), (3.43) and (3.45) respectively.

To calculate $E[(K_2 - K^2)^2 | \mathbf{y}]$ for this specific case we need to evaluate $[M]$, $[MC]$, $[MMC]$, $[CC]$, $[C^2]$, $[M^2]$, $[V]$ and $[C]$, as defined by equations (3.46) to (3.53).

The expression $[M]$ is equal to \hat{k} , (3.21) and the calculation of $[M^2]$ and $[V]$ were implicit in that of the expectation of L above. Thus, if we let

$$\mathbf{G} = (\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H})^{-1}, \mathbf{F} = \mathbf{y} - \mathbf{H}\hat{\beta}, \mathbf{D} = \mathbf{A}^{-1} \mathbf{H} \mathbf{G} \mathbf{H}^T \mathbf{A}^{-1} \quad (4.46)$$

and

$$\mathbf{V} = (\mathbf{A} \mathbf{H} \mathbf{G})^T \quad (4.47)$$

then

$$[M^2] = \hat{\beta}^T \mathbf{Q} \hat{\beta} + \mathbf{F}^T \mathbf{A}^{-1} \mathbf{P} \mathbf{A}^{-1} \mathbf{F} + 2\mathbf{F}^T \mathbf{A}^{-1} \mathbf{S} \hat{\beta}, \quad (4.48)$$

and

$$[V] = 1 - \text{tr} \{ \mathbf{A}^{-1} \mathbf{P} \} + \text{tr} \{ \mathbf{G} \mathbf{P} \} + \text{tr} \{ \mathbf{A}^{-1} \mathbf{H} \mathbf{G} \mathbf{H}^T \mathbf{A}^{-1} \mathbf{P} \}. \quad (4.49)$$

The remaining expressions are evaluated as follows,

$$\begin{aligned} [MC] = & MC_1 \beta + \hat{\beta} \mathbf{Q} \mathbf{G} \mathbf{R} - \hat{\beta} \mathbf{Q} \mathbf{V} \mathbf{T} - \hat{\beta} \mathbf{Q} \hat{\beta} \mathbf{T} + MC_2 \mathbf{A} \mathbf{F} + \mathbf{F} \mathbf{A}^{-1} \mathbf{S} \mathbf{G} \mathbf{R} \\ & - \mathbf{F} \mathbf{A}^{-1} \mathbf{S} \mathbf{V} \mathbf{T} - \mathbf{F} \mathbf{A}^{-1} \mathbf{P} \hat{\beta} \mathbf{T} - \hat{\beta} \mathbf{S} \mathbf{V} \mathbf{R} - \mathbf{F} \mathbf{A}^{-1} \mathbf{P} \mathbf{V} \mathbf{R}, \end{aligned} \quad (4.50)$$

where MC_1 is a vector of length $(p+1)$ obtained by integrating,

$$\int_{\chi} \int_{\chi} \mathbf{h}(\mathbf{x}) C(\mathbf{x}, \mathbf{z}) dG(\mathbf{x}) dG(\mathbf{z}). \quad (4.51)$$

Thus,

$$MC_1[1] = \int_{\chi} \int_{\chi} 1 C(\mathbf{x}, \mathbf{z}) dG(\mathbf{x}) dG(\mathbf{z}), \quad (4.52)$$

which equals U as illustrated in (4.31), so

$$MC_1[1] = \prod_{\theta=1, \dots, p} \sqrt{\frac{\delta_{\theta}}{\delta_{\theta} + 4\sigma_{\theta}^2}}. \quad (4.53)$$

The other p elements of the vector have the general form, such that

$$\begin{aligned} MC_1[\alpha + 1] &= \int_{\chi} \int_{\chi} \mathbf{h}_x[\alpha] C(\mathbf{x}, \mathbf{z}) dG(\mathbf{x}) dG(\mathbf{z}) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{\alpha} C(x_{\alpha}, z_{\alpha}) dG(x_{\alpha}) dG(z_{\alpha}) \end{aligned} \quad (4.54)$$

$$\times \prod_{\theta=1, \dots, p, \theta \neq \alpha} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(x_{\theta}, z_{\theta}) dG(x_{\theta}) dG(z_{\theta}), \quad (4.55)$$

$$= \mu_{\alpha} MC_1[1]. \quad (4.56)$$

Now, consider the evaluation of (4.54). To perform the integration of this expression with respect to $G(z_{\alpha})$, reveals

$$\int_{-\infty}^{\infty} C(x_{\alpha}, z_{\alpha}) dG(z_{\alpha}) = x_{\alpha} \sqrt{\frac{\delta_{\alpha}}{\delta_{\alpha} + 2\sigma_{\alpha}^2}} \exp \left[\frac{-(x_{\alpha} - \mu_{\alpha})^2}{\delta_{\alpha} + 2\sigma_{\alpha}^2} \right]. \quad (4.57)$$

If the component x_{α} is taken outside the integral sign the remaining expression has already been evaluated as part of \mathbf{T} , (4.27). The further integration of this expression with respect to $G(x_{\alpha})$ can be written as

$$\begin{aligned}
& \int_{-\infty}^{\infty} \sqrt{\frac{\delta_{\alpha}}{\delta_{\alpha} + 2\sigma_{\alpha}^2}} \frac{x_{\alpha}}{\sqrt{2\pi\sigma_{\alpha}^2}} \exp \left[\frac{-(x_{\alpha} - \mu_{\alpha})^2}{\delta_{\alpha} + 2\sigma_{\alpha}^2} \right] \exp \left[\frac{-(x_{\alpha} - \mu_{\alpha})^2}{2\sigma_{\alpha}^2} \right] dx_{\alpha} \\
&= \sqrt{\frac{\delta_{\alpha}}{\delta_{\alpha} + 2\sigma_{\alpha}^2}} \int_{-\infty}^{\infty} \frac{x_{\alpha}}{\sqrt{2\pi\sigma_{\alpha}^2}} \exp \left[\frac{-(4\sigma_{\alpha}^2 + \delta_{\alpha})(x_{\alpha} - \mu_{\alpha})^2}{2\sigma_{\alpha}^2(\delta_{\alpha} + 2\sigma_{\alpha}^2)} \right] dx_{\alpha}, \\
&= \mu_{\alpha} \sqrt{\frac{\delta_{\alpha}}{\delta_{\alpha} + 4\sigma_{\alpha}^2}}.
\end{aligned}$$

The expression, (4.55), has the same form as $MC_1[1]$, thus

$$MC_1[\alpha + 1] = \mu_{\alpha} MC_1[1]. \quad (4.58)$$

The next element of $[MC]$, MC_2 is a vector of length n with elements of the form

$$MC_2[i] = \prod_{\theta=1, \dots, p} \int_{\chi_{\theta}} \int_{\chi_{\theta}} C(x_{\theta}, z_{\theta}) C(z_{\theta}, x_{\theta,i}) dG(x_{\theta}) dG(z_{\theta}), \quad (4.59)$$

Now,

$$\begin{aligned}
& \int_{\chi_{\theta}} C(x_{\theta}, z_{\theta}) C(z_{\theta}, x_{\theta,i}) dG(x_{\theta}) \\
&= \int_{\chi_{\theta}} \frac{1}{\sqrt{2\pi\sigma_{\theta}^2}} \exp \left[\frac{-(x_{\theta} - z_{\theta})^2}{\delta_{\theta}} \right] \exp \left[\frac{-(z_{\theta} - x_{\theta,i})^2}{\delta_{\theta}} \right] \exp \left[\frac{-(x_{\theta} - \mu_{\theta})^2}{2\sigma_{\theta}^2} \right] dx_{\theta}, \\
&= \exp \left[\frac{-(z_{\theta} - \mu_{\theta})^2}{\delta_{\theta} + 2\sigma_{\theta}^2} \right] \exp \left[\frac{-(z_{\theta} - x_{\theta,i})^2}{\delta_{\theta}} \right] \\
&\quad \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{\theta}^2}} \exp \left[\frac{-(2\sigma_{\theta}^2 + \delta_{\theta})}{2\sigma_{\theta}^2\delta_{\theta}} \left(x_{\theta} - \frac{2z_{\theta}\sigma_{\theta}^2 + \mu_{\theta}\delta_{\theta}}{2\sigma_{\theta}^2 + \delta_{\theta}} \right)^2 \right] dx_{\theta}, \\
&= \sqrt{\frac{\delta_{\theta}}{\delta_{\theta} + 2\sigma_{\theta}^2}} \exp \left[\frac{-(z_{\theta} - \mu_{\theta})^2}{\delta_{\theta} + 2\sigma_{\theta}^2} \right] \exp \left[\frac{-(z_{\theta} - x_{\theta,i})^2}{\delta_{\theta}} \right]. \quad (4.60)
\end{aligned}$$

The second integral is evaluated in the same way as the first, by completing the square with respect to z_{θ} , to construct a normal density function which integrates to one. Thus,

$$MC_2[i] = \prod_{\theta=1, \dots, p} \exp - \left[\frac{8\sigma_{\theta}^4 (x_{\theta,i}^2 + \mu_{\theta}^2) + 6\sigma_{\theta}^2\delta_{\theta} (x_{\theta,i}^2 + \mu_{\theta}^2)}{(4\sigma_{\theta}^2 + 6\delta_{\theta}\sigma_{\theta}^2 + \delta_{\theta}^2) (\delta_{\theta} + 2\sigma_{\theta}^2)} \right]$$

$$\begin{aligned}
& \times \exp - \left[\frac{(x_{\theta,i}^2 + \mu_{\theta}^2) \delta_{\theta}^2 - 16x_{\theta,i}\sigma_{\theta}^4\mu_{\theta} - 12x_{\theta,i}\sigma_{\theta}^2\mu_{\theta}\delta_{\theta} - 2x_{\theta,i}\sigma_{\theta}^2\mu_{\theta}}{(4\sigma_{\theta}^2 + 6\delta_{\theta}\sigma_{\theta}^2 + \delta_{\theta}^2)(\delta_{\theta} + 2\sigma_{\theta}^2)} \right] \\
& \times \sqrt{\frac{\delta_{\theta}}{\delta_{\theta} + 2\sigma_{\theta}^2}} \times \sqrt{\frac{\delta_{\theta}(\delta_{\theta} + 2\sigma_{\theta}^2)}{4\sigma_{\theta}^2 + 6\delta_{\theta}\sigma_{\theta}^2 + \delta_{\theta}^2}}.
\end{aligned} \tag{4.61}$$

Next,

$$\begin{aligned}
[C] &= \int_{\chi} \int_{\chi} C(\mathbf{x}, \mathbf{z}) dG(\mathbf{x})dG(\mathbf{z}) - \mathbf{T}\hat{\beta}\mathbf{T} + \mathbf{RGR} - 2\mathbf{RVT}, \\
&= \left(\prod_{\theta=1,\dots,p} \sqrt{\frac{\delta_{\theta}}{\delta_{\theta} + 4\sigma_{\theta}^2}} \right) - \mathbf{T}\hat{\beta}\mathbf{T} + \mathbf{RGR} - 2\mathbf{RVT}
\end{aligned} \tag{4.62}$$

since, $\int_{\chi} \int_{\chi} C(\mathbf{x}, \mathbf{z}) dG(\mathbf{x})dG(\mathbf{z})$ equals $MMC_1[1]$ (4.53).

Also,

$$\begin{aligned}
[MMC] &= \hat{\beta} MMC_1 \hat{\beta} + \hat{\beta} \mathbf{Q} \mathbf{G} \mathbf{Q} \hat{\beta} - \hat{\beta} \mathbf{Q} \mathbf{V} \mathbf{S} \hat{\beta} + 2\hat{\beta} MMC_2 \mathbf{A}^{-1} \mathbf{F} \\
&+ \hat{\beta} \mathbf{Q} \mathbf{G} \mathbf{S} \mathbf{A}^{-1} \mathbf{F} - \hat{\beta} \mathbf{S} \mathbf{V} \mathbf{S} \mathbf{A}^{-1} \mathbf{F} - \hat{\beta} \mathbf{S} (\mathbf{A}^{-1} - \mathbf{D}) \mathbf{S} \hat{\beta} \\
&- \hat{\beta} \mathbf{S} (\mathbf{A}^{-1} - \mathbf{D}) \mathbf{P} \mathbf{A}^{-1} \mathbf{F} + \hat{\beta} \mathbf{Q} \mathbf{G} \mathbf{S} \mathbf{A}^{-1} \mathbf{F} - \hat{\beta} \mathbf{Q} \mathbf{V} \mathbf{P} \mathbf{A}^{-1} \mathbf{F} \\
&+ \mathbf{F} \mathbf{A}^{-1} MMC_3 \mathbf{A}^{-1} \mathbf{F} + \mathbf{F} \mathbf{A}^{-1} \mathbf{S} \mathbf{G} \mathbf{S} \mathbf{A}^{-1} \mathbf{F} \\
&- \mathbf{F} \mathbf{A}^{-1} \mathbf{S} \mathbf{V} \mathbf{P} \mathbf{A}^{-1} \mathbf{F} - \hat{\beta} \mathbf{S} (\mathbf{A}^{-1} - \mathbf{D}) \mathbf{P} \mathbf{A}^{-1} \mathbf{F} \\
&- \mathbf{F} \mathbf{A}^{-1} \mathbf{P} (\mathbf{A}^{-1} - \mathbf{D}) \mathbf{P} \mathbf{A}^{-1} \mathbf{F} - \hat{\beta} \mathbf{S} \mathbf{V} \mathbf{Q} \hat{\beta} - \mathbf{F} \mathbf{A}^{-1} \mathbf{P} \mathbf{V} \mathbf{Q} \hat{\beta} \\
&- \hat{\beta} \mathbf{S} \mathbf{V} \mathbf{S} \mathbf{A}^{-1} \mathbf{F} - \mathbf{F} \mathbf{A}^{-1} \mathbf{P} \mathbf{V} \mathbf{S} \mathbf{A}^{-1} \mathbf{F},
\end{aligned} \tag{4.63}$$

where MMC_1 is the $(p+1) \times (p+1)$ symmetric matrix that results from performing the integral

$$\int_{\chi} \int_{\chi} \mathbf{h}(\mathbf{x}) C(\mathbf{x}, \mathbf{z}) \mathbf{h}(\mathbf{z})^T dG(\mathbf{x})dG(\mathbf{z}) \tag{4.64}$$

such that

$$MMC_1[1, 1] = \int_{\chi} \int_{\chi} C(\mathbf{x}, \mathbf{z}) \mathbf{h}(\mathbf{z})^T dG(\mathbf{x})dG(\mathbf{z}) \tag{4.65}$$

which equals U , (4.30).

The details of the evaluations of the remaining expressions in this chapter will be kept brief since they are all performed in a similar way to those already detailed. That is, by converting each integral into the product of an expression and a normal density function, the second of which will always integrate to one.

Thus,

$$\begin{aligned}
MMC_1[1, (\alpha + 1)] &= \int_{\chi} \int_{\chi} \mathbf{h}_{\mathbf{x}}[\alpha] C(\mathbf{x}, \mathbf{z}) dG(\mathbf{x}) dG(\mathbf{z}), \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{\alpha} C(x_{\alpha}, z_{\alpha}) dG(x_{\alpha}) dG(z_{\alpha}) \\
&\quad \times \prod_{\theta=1, \dots, p, \theta \neq \alpha} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(x_{\theta}, z_{\theta}) dG(x_{\theta}) dG(z_{\theta}), \\
&= \mu_{\theta} U.
\end{aligned} \tag{4.66}$$

Similarly,

$$\begin{aligned}
MMC_1[(\alpha + 1), (\phi + 1)] &= \int_{\chi} \int_{\chi} \mathbf{h}_{\mathbf{x}}[\alpha] C(\mathbf{x}, \mathbf{z}) \mathbf{h}_{\mathbf{z}}[\phi] dG(\mathbf{x}) dG(\mathbf{z}), \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{\phi} C(x_{\phi}, z_{\phi}) dG(x_{\phi}) dG(z_{\phi}) \\
&\quad \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{\alpha} C(x_{\alpha}, z_{\alpha}) dG(x_{\alpha}) dG(z_{\alpha}) \\
&\quad \times \prod_{\theta} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(x_{\theta}, z_{\theta}) dG(x_{\theta}) dG(z_{\theta}), \\
&= \mu_{\phi} \mu_{\theta} U.
\end{aligned} \tag{4.67}$$

where $\theta = 1, \dots, p$, $\theta \neq \alpha$, $\theta \neq \phi$ and where $1 \leq \phi, \alpha \leq p$ excluding the case where $\alpha = \phi$. Finally, the diagonal of this matrix has elements of the form

$$\begin{aligned}
MMC_1[(\alpha + 1), (\alpha + 1)] &= \int_{\chi} \int_{\chi} \mathbf{h}_{\mathbf{x}}[\alpha] C(\mathbf{x}, \mathbf{z}) \mathbf{h}_{\mathbf{z}}[\alpha] dG(\mathbf{x}) dG(\mathbf{z}), \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{\alpha} C(x_{\alpha}, z_{\alpha}) z_{\alpha} dG(x_{\alpha}) dG(z_{\alpha}) \\
&\quad \times \prod_{\theta} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(x_{\theta}, z_{\theta}) dG(x_{\theta}) dG(z_{\theta}), \\
&= U \left(\mu_{\alpha}^2 + \frac{2\sigma_{\alpha}^4}{\delta_{\alpha} + 2\sigma_{\alpha}^2} \right),
\end{aligned} \tag{4.68}$$

where $\theta = 1, \dots, p$, $\theta \neq \alpha$ for $1 \leq \alpha \leq p$.

Next, MMC_2 is an $n \times (p+1)$ matrix obtained by evaluating the integral

$$\int_{\chi} \int_{\chi} \mathbf{h}(\mathbf{x}) C(\mathbf{x}, \mathbf{z}) \mathbf{t}(\mathbf{z})^T dG(\mathbf{x}) dG(\mathbf{z}), \tag{4.69}$$

the i^{th} row of this matrix will consist of $p + 1$ elements such that

$$\begin{aligned} MMC_2[i, 1] &= \prod_{\theta=1, \dots, p} \int_{\chi_\theta} \int_{\chi_\theta} C(x_\theta, z_\theta) C(x_\theta, z_{\theta,i}) dG(x_\theta) dG(z_\theta), \\ &= MC_2[i], \end{aligned} \quad (4.70)$$

as defined in (4.61), and

$$\begin{aligned} MMC_2[i, \alpha + 1] &= \prod_{\theta=1, \dots, p} \int_{\chi_\theta} \int_{\chi_\theta} \mathbf{h}_x[\alpha] C(x_\theta, z_\theta) C(x_\theta, z_{\theta,i}) dG(x_\theta) dG(z_\theta), \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_\alpha C(x_\alpha, z_\alpha) C(x_\alpha, z_{\alpha,i}) dG(x_\alpha) dG(z_\alpha) \\ &\quad \times \prod_{\theta=1, \dots, p} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(x_\theta, z_\theta) C(x_\theta, z_{\theta,i}) dG(x_\theta) dG(z_\theta), \\ &= \left[\frac{4\mu_\alpha \delta_\alpha \sigma_\alpha^2 + 4x_{\alpha,i} \sigma_\alpha^4 + 2x_{\alpha,i} \sigma_\alpha^2 \delta_\alpha + \mu_\alpha \delta_\alpha^2}{4\sigma_\alpha^2 + 6\delta_\alpha \sigma_\alpha^2 + \delta_\alpha^2} \right] \\ &\quad \times \left[\frac{2\sigma_\alpha^2}{2\sigma_\alpha^2 + \delta_\alpha} \right] \times MC_2[i] \\ &\quad + \left[\frac{2\mu_\alpha \delta_\alpha}{2\sigma_\alpha^2 + \delta_\alpha} \right] \times MC_2[i]. \end{aligned} \quad (4.71)$$

Finally, MMC_3 is an $n \times n$ matrix obtained by evaluating the integral

$$\int_{\chi} \int_{\chi} \mathbf{t}(\mathbf{x}) C(\mathbf{x}, \mathbf{z}) \mathbf{t}(\mathbf{z})^T dG(\mathbf{x}) dG(\mathbf{z}), \quad (4.72)$$

a general element $MMC_3[i, j]$ from this matrix takes the form

$$\begin{aligned} MMC_3[i, j] &= \prod_{\theta=1, \dots, p} \int_{\chi_\theta} \int_{\chi_\theta} C(x_\theta, x_{\theta,i}) C(x_\theta, z_\theta) C(z_\theta, z_{\theta,j}) dG(x_\theta) dG(z_\theta), \\ &= \exp - \left[\frac{-32x_{\theta,i} \sigma_\theta^6 z_{\theta,j} + 20\mu_\theta \delta_\theta^2 \sigma_\theta^2 + 2\mu_\theta \delta_\theta^3}{(12\sigma_\theta^4 + 8\delta_\theta \sigma_\theta^2 + \delta_\theta^2) \delta_\theta (\delta_\theta + 4\sigma_\theta)} \right] \\ &\quad \times \exp - \left[\frac{-8x_{\theta,i} \sigma_\theta^4 z_{\theta,j} \delta + 48\mu_\theta^2 \delta_\theta \sigma_\theta^4}{(12\sigma_\theta^4 + 8\delta_\theta \sigma_\theta^2 + \delta_\theta^2) \delta_\theta (\delta_\theta + 4\sigma_\theta)} \right] \\ &\quad \times \exp - \left[\frac{(x_{\theta,i}^2 + z_{\theta,j}^2) (28\sigma_\theta^4 \delta_\theta + 10\sigma_\theta^2 \delta_\theta^2 + 16\sigma_\theta^6 + \delta_\theta^3)}{(12\sigma_\theta^4 + 8\delta_\theta \sigma_\theta^2 + \delta_\theta^2) \delta_\theta (\delta_\theta + 4\sigma_\theta)} \right] \\ &\quad \times \exp - \left[\frac{-(x_{\theta,i} + z_{\theta,j}) (48\sigma_\theta^4 \mu_\theta \sigma_\theta + 20\sigma_\theta^2 \mu_\theta \delta_\theta^2 + 2\delta_\theta^3 \mu_\theta)}{(12\sigma_\theta^4 + 8\delta_\theta \sigma_\theta^2 + \delta_\theta^2) \delta_\theta (\delta_\theta + 4\sigma_\theta)} \right] \\ &\quad \times \sqrt{\frac{\delta_\theta^2}{12\sigma_\theta^4 + 8\delta_\theta \sigma_\theta^2 + \delta_\theta^2}}. \end{aligned} \quad (4.73)$$

Next,

$$\begin{aligned}
[CC] &= \mathbf{RGQGR} - 2\mathbf{RGQVT} + \mathbf{TVQVT} \\
&\quad - 2\mathbf{T}(\mathbf{A}^{-1} - \mathbf{D})\mathbf{S}(\mathbf{GR} - \mathbf{VT}) + \mathbf{T}(\mathbf{A}^{-1} - \mathbf{D})\mathbf{P}(\mathbf{A}^{-1} - \mathbf{D})\mathbf{T} \\
&\quad - 2\mathbf{RVSGR} + 2\mathbf{RVSVT} + \mathbf{T}(\mathbf{A}^{-1} - \mathbf{D})\mathbf{PVR} + \mathbf{RVPVR} \\
&\quad + CC_1 + 2MC_1\mathbf{GR} + MC_1\mathbf{VT} - MC_2(\mathbf{A}^{-1} - \mathbf{D})\mathbf{T} \\
&\quad + MC_2\mathbf{VR},
\end{aligned} \tag{4.74}$$

where

$$\begin{aligned}
CC_1 &= \int_{\chi} \int_{\chi} \int_{\chi} C(\mathbf{v}, \mathbf{x}) C(\mathbf{x}, \mathbf{z}) dG(\mathbf{v}) dG(\mathbf{x}) dG(\mathbf{z}), \\
&= \prod_{\theta=1, \dots, p} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(v_{\theta}, x_{\theta}) C(x_{\theta}, z_{\theta}) dG(v_{\theta}) dG(x_{\theta}) dG(z_{\theta}), \\
&= \prod_{\theta=1, \dots, p} \frac{\delta_{\theta}}{\sqrt{(\delta_{\theta} + 2\sigma_{\theta}^2)(\delta_{\theta} + 6\sigma_{\theta}^2)}}.
\end{aligned} \tag{4.75}$$

In conclusion,

$$\begin{aligned}
[C^2] &= \text{tr} \{ \mathbf{GQGQ} \} - 2\text{tr} \{ \mathbf{GQVS} \} - \text{tr} \{ \mathbf{GS}(\mathbf{A}^{-1} - \mathbf{D})\mathbf{S} \} \\
&\quad + \text{tr} \{ \mathbf{VQVP} \} + 2\text{tr} \{ \mathbf{VS}(\mathbf{A}^{-1} - \mathbf{D})\mathbf{P} \} - 2\text{tr} \{ \mathbf{GSVQ} \} \\
&\quad + \text{tr} \{ (\mathbf{A}^{-1} - \mathbf{D})\mathbf{P}(\mathbf{A}^{-1} - \mathbf{D})\mathbf{P} \} + 2\text{tr} \{ \mathbf{SVP}(\mathbf{A}^{-1} - \mathbf{D}) \} \\
&\quad + 2\text{tr} \{ \mathbf{VSVS} \} + \text{tr} \{ \mathbf{VPVQ} \} + C_1^2 + 2\text{tr} \{ \mathbf{G} \mathbf{M} \mathbf{M} C_1 \} \\
&\quad - 4\text{tr} \{ \mathbf{V} \mathbf{M} \mathbf{M} C_2 \} - 2\text{tr} \{ (\mathbf{A}^{-1} - \mathbf{D}) \mathbf{M} \mathbf{M} C_2 \},
\end{aligned} \tag{4.76}$$

where

$$\begin{aligned}
C_1^2 &= \int_{\chi} \int_{\chi} C(\mathbf{x}, \mathbf{z}) C(\mathbf{x}, \mathbf{z}) dG(\mathbf{x}) dG(\mathbf{z}), \\
&= \prod_{\theta=1, \dots, p} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(x_{\theta}, z_{\theta}) C(x_{\theta}, z_{\theta}) dG(x_{\theta}) dG(z_{\theta}), \\
&= \prod_{\theta=1, \dots, p} \sqrt{\frac{\delta_{\theta}}{\delta_{\theta} + 8\sigma_{\theta}^2}}.
\end{aligned} \tag{4.77}$$

Using the above results the specific form of $E[(K_2 - K^2)^2 | y]$ for uncertain parameters with normally distributed priors is obtained and thus $\text{Var}[L | y]$.

The expressions evaluated above demonstrate that the Bayesian uncertainty analysis is analytically tractable for the case of independent uncertain parameters that have normal prior distributions and in which the function to be analysed is assumed to be approximately linear in each of the uncertain parameters. The justification for these assumptions was considered at the beginning of this chapter. Although they seem rather constricting, it is often possible to obtain, through transformations, parameters with normal prior distributions, eg. the log transformation for parameters with lognormal prior distributions. Further, in most cases the form of the relationship between an uncertain parameter and the output of the function will not be known, so the assumption of a linear relationship is not unreasonable.

5. COMPARISON OF BAYESIAN AND CLASSICAL UNCERTAINTY ANALYSIS METHODOLOGIES USING AN ALGORITHM FOR THE CALCULATION OF DOSE DUE TO INGESTION OF ^{131}I .

5.1. Introduction

Iodine, in its stable form occurs naturally in the human body where it concentrates in the thyroid gland. This organ uses iodine to produce two hormones called thyroxine and triiodothyrosine which the body uses to regulate cell metabolism. If the radioactive form of iodine, ^{131}I , is taken into the human body, it will also concentrate in the thyroid where it will undergo radioactive decay. The exposure of this gland to ionising radiation may lead to an increased risk of thyroid cancer [doc93].

It is important to quantify the size of this risk so that if a potential dose, as measured by the committed effective dose equivalent (CEDE)¹, is predicted to result in an unacceptably high level of risk then remedial action can be initiated. An effective form of remedial action involves the prompt administration of a large dose of stable iodine which will be taken into the thyroid, saturating it and blocking the absorption of the radioactive element causing it to be excreted from the body more quickly.

Contamination of the environment with ^{131}I is one potential effect of a nuclear reactor accident. Consequently, the accurate prediction of a

¹This is a widely used measure that quantifies detriment from an exposure as the dose accumulates over the following 50 years.

person's CEDE value due to an intake of ^{131}I is important. CEDE cannot be measured directly; its value, amongst other things, is a function of the amount of ^{131}I remaining in the body. It is also a function of the rate at which the ^{131}I is removed from the body (referred to as its biological half-life) and the radioactive decay half-life. The quantity and dispersion of ^{131}I in the body is very difficult to measure directly. Consequently, its behaviour is usually inferred from a mathematical algorithm or model designed to predict the movement and retention, over time, of this element in the human body.

A number of algorithms exist for performing this task. All contain many parameters for which an exact value is not known, owing either to lack of knowledge or because a value cannot be assigned until a specific situation is envisaged. To illustrate this second point, in some models the subject's age is defined as a parameter. Until the subject for measurement is selected their age is unknown. It can thus be considered as a random variable. Quantification of the uncertainty induced in the output of these algorithms by their uncertain parameters is vital if CEDE values obtained using them are to be reliable for the purposes of radiation protection.

In this chapter, one such algorithm will be used to demonstrate, in a simple low-dimensional scenario, the viability of the Bayesian uncertainty analysis methodology. The algorithm selected is a simplified recycling algorithm for ^{131}I which was developed at NRPB during the 1980's. The Bayesian methodology will be applied to the same uncertain parameters that were investigated in a classical uncertainty analysis previously performed at NRPB. This will enable the relative quality of the two methods to be compared directly.

In addition, because the chosen algorithm is relatively inexpensive to evaluate, two measures of the 'true' uncertainty will also be derived to provide an absolute measure of the effectiveness of the Bayesian methodology.

5.2. A simplified recycling algorithm for ^{131}I

The algorithm that will be examined is a simplified recycling algorithm proposed by Adams and Fell [AF88]. A graphical representation of the

algorithm is presented in Figure 5.1.

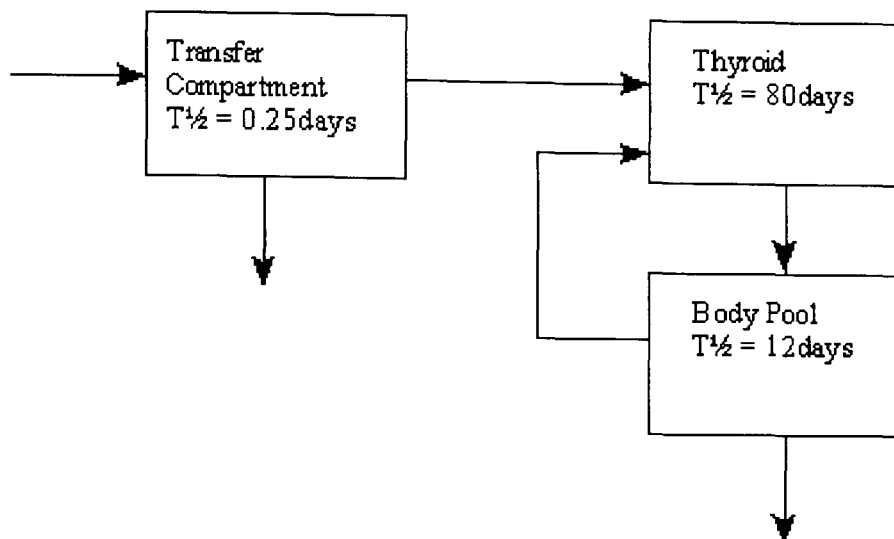


Figure 5.1: A graphical representation of the ^{131}I algorithm.

This is known as a compartmental algorithm or box model. Different parts of the body are represented by the three boxes or compartments. Once a quantity of ^{131}I has been ingested it moves down through the oesophagus to the stomach. From there the iodine can take one of two routes. Either it passes on down the digestive tract and is then excreted, or it is absorbed through the stomach wall into the blood stream and is transferred to the thyroid gland. The time taken for both these processes to remove half of the iodine from the stomach, $T_{\frac{1}{2}}$ is estimated at a quarter of a day. They are represented in the algorithm by the transfer compartment.

The iodine component transferred to the thyroid will remain there for considerably longer than a quarter of a day. In fact, the algorithm estimates that it takes 80 days for half of any quantity of iodine to be removed from the thyroid back to the blood. Once there it is supposed that it will be retained in the rest of the body, known as the ‘body pool’, for a certain period, $T_{\frac{1}{2}} = 12$ days. On leaving the body pool a proportion will be returned to the thyroid, a process known as recycling, while the remaining iodine will be excreted.

Each of the arrowed lines in the figure has a transfer rate associated with it which quantifies the flow rate from one compartment to another. In addition to the transfers and losses described above, the radioactive iodine

will also be decaying naturally to become stable iodine. The time taken for half a unit quantity of this radioactive form of iodine to decay to its stable form (i.e., its radioactive half life) is approximately eight days.

At time zero a unit quantity of ^{131}I is considered ingested. The algorithm provides as its output the quantity of iodine in each of the three compartments at a specified time t following ingestion. These values can be used to calculate an appropriate measure of exposure, the CEDE. At NRPB a program called Pedal [KP90] was developed which evaluates the algorithm and calculates the CEDE from a unit ingestion of ^{131}I . This program is written in Fortran 77 and, using the fastest main frame computer available at NRPB at the time the program was developed, it took approximately 4.5 seconds to calculate one CEDE value.

In the remainder of this chapter we will denote the mathematical form of the ^{131}I algorithm by $\eta(\cdot)$.

5.3. A classical analysis of the uncertainties associated with the ^{131}I algorithm

Soon after the publication of the simplified iodine algorithm by Adams and Fell [AF88], as described above, the NRPB decided to perform an uncertainty analysis in order to determine the variability in the CEDE caused by the uncertainties in the parameters of the algorithm. Sensitivity calculations were first performed and the two parameters whose values most affected the algorithm's output were selected for the uncertainty analysis.

5.3.1. The uncertain parameters of the simplified recycling algorithm

The classical uncertainty analysis performed at NRPB considered two parameters of the algorithm to be uncertain. They were

- a) the mass of the thyroid, w
- b) the fraction of iodine, f , contained in a unit quantity of blood that is taken up by the thyroid.

The value of the parameter w is uncertain because thyroid mass varies

from person to person. It would be most accurate to use an individual's thyroid mass in a calculation of CEDE. However, it is difficult to obtain a safe, cheap and accurate estimate of an individual's thyroid dose, so an estimate is usually made.

The other uncertain parameter, the uptake fraction f , is a quantity that is usually considered to be the same for everybody. Unfortunately, methods have not yet been devised to measure it accurately, so the uncertainty associated with this parameter is caused by a lack of knowledge about how iodine behaves in the human body.

5.3.2. The methodology of the classical uncertainty analysis

The first step in the analysis involved the selection of prior distributions to represent the uncertainties about the values of w and f . Dunning and Schwarz [DSS81] have collected and analysed data from various sources about both the fractional uptake and the thyroid mass. The data show a good fit to the log-normal distribution in both cases. Actual plots of the data can be found in Dunning and Schwarz [DSS81]. Other information concerning these parameters was obtained from ICRP Publication 23 [ICR75].

As a result of combining the information from both these sources log-normal prior distributions were selected for both w and f . The parameters of these prior distributions are given in Table 5.1.

Thyroid mass, w (grams)		Fractional uptake, f	
μ	σ	μ	σ
2.889	0.463	-1.315	0.355

Table 5.1: Parameters of the lognormal prior distributions

The next step in the classical analysis was to choose 1000 values from each of these distributions using simple Monte Carlo random selection. The PEDAL program [KP90] was then run with each of the 1000 pairs of w and f values to generate a set of 1000 CEDE values.

This set of values was then used to create a frequency histogram. Summary measures were obtained in the form of the sample mean and variance.

A confidence interval for the predicted value of the CEDE was also derived. These were the only measures obtained from this analysis and are detailed in 5.6.2.

5.4. Calculation of the ‘true uncertainty’ associated with the ^{131}I algorithm

One of the main reasons that the iodine algorithm was selected for testing the viability of the Bayesian uncertainty analysis methodology was that independent estimates of the true uncertainty could be obtained for the purposes of comparison.

The ^{131}I algorithm is relatively simple and the computer processing power available at the NRPB has increased considerably since the classical analysis of this algorithm was carried out in 1988. The 1000 evaluations required for that analysis originally took 75 minutes CPU time to calculate, whereas now they take 2 minutes. As a result it has been possible to perform many more than 1000 evaluations of the CEDE in a reasonably short time. This has enabled estimates to be made of the true uncertainty in the algorithm independent of either the classical or Bayesian methodologies.

A 1000×1000 regular grid of points was selected from the two dimensional parameter space of w and f , the original lognormally distributed parameters. The grid was determined by selecting the points so that the it extended 2.5 standard deviations, in each direction, either side of each parameter’s prior mean value as defined in Table 5.1. The CEDE was evaluated using the PEDAL program [KP90] at each grid point. This took approximately 33 hours CPU time. Using the computers available at the time the classical analysis was performed, in 1988, this would have taken about 51 days. From this large set of values a sample mean and a sample variance for the CEDE, weighted by the distributions of the uncertain parameters, were obtained as measures of the true uncertainty.

Further, to determine if the restriction of the grid to within 2.5 standard deviations of the mean values had an effect on the values of the sample mean or variance, a 40×40 regular grid of CEDE values was selected as

a subset of the larger grid of values constructed above. The best fitting regression function, to estimate the value of the CEDE, was then selected for this data using GLIM [AAFH90]. The values for the mean and variance of the CEDE were obtained analytically by integrating the regression function over the parameter space, χ .

The results of both calculations of the true uncertainty on ^{131}I algorithm are detailed in section 5.6.1.

5.5. Bayesian analysis of the ^{131}I algorithm

This analysis was based on the theory detailed in chapter four. The first step was to define the uncertain parameters and their prior distributions. The theory detailed in the previous chapter requires that the uncertain parameters of the problem have normal prior distributions. Hence this Bayesian analysis was performed using the log of the thyroid mass and fractional uptake as the uncertain parameters which were considered to have normally distributed priors. The transformed parameters log thyroid mass and log fractional uptake are denoted by x_{lw} and x_{lf} respectively.

Although in the five years since the NRPB classical uncertainty analysis was performed new data has become available, the same prior distributions were used for the Bayesian analysis in order to ensure the closest comparison of the two methodologies. The means, μ_w and μ_f , and the variances, σ_w^2 and σ_f^2 , of these priors are detailed in Table 5.1.

The general elements of the analysis specified at the start of chapter four will now be detailed specifically for this analysis. Hence, p which represents the number of uncertain parameters in the problem will therefore become 2. A general point on the uncertain parameter space was defined by the vector $\mathbf{x} = [x_{lw}, x_{lf}]$. The vector of regression functions, $\mathbf{h}(\mathbf{x})$, was set to

$$\mathbf{h}(\mathbf{x}) = [1, x_{lw}, x_{lf}], \quad (5.1)$$

while

$$\boldsymbol{\delta} = [\delta_{lw}, \delta_{lf}] \quad (5.2)$$

and

$$C(\mathbf{x}, \mathbf{z}) = C(x_{lw}, z_{lw}) \times C(x_{lf}, z_{lf})$$

$$= \exp \left[\frac{-(x_{lw} - z_{lw})^2}{\delta_{lw}} \right] \times \exp \left[\frac{-(x_{lf} - z_{lf})^2}{\delta_{lf}} \right]. \quad (5.3)$$

The next step in the Bayesian analysis was to define the set of design points.

5.5.1. Selection of the design points

It would have been perfectly legitimate to select points at random and use these as a set of design points as is required for the classical analysis. However, this would not have maximised the potential of the Bayesian methodology. In this analysis the design points were selected using the criterion defined in (4.11). This criterion can be simplified from the general p -dimensional case to this two dimensional problem as

$$[tr(\mathbf{A}_{lw}^{-1} \mathbf{P}_{lw})] \times [tr(\mathbf{A}_{lf}^{-1} \mathbf{P}_{lf})] \quad (5.4)$$

where the general forms of \mathbf{P}_θ and \mathbf{A}_θ are defined in (4.12) and (4.13) respectively.

Now, the parameter space for the normally distributed uncertain parameters is continuous. This meant that it was not practical to perform an unrestricted search for the best design (of a given size) as this would take a prohibitively large amount of computer processing time. To overcome this problem a 20×20 regular grid was defined over the parameter space of w and f , with the largest and smallest values in each dimension of the grid 2.5 standard deviations either side of the parameter's expected value as defined by its prior distribution (i.e. at the 99th percentiles).

The best n point designs were selected from those points defined by the grid using a one-step replacement algorithm. This can be described as follows. First, a design of the required size is selected at random. The value of the criterion is then evaluated using prior estimates for δ . A new point is then selected from the grid and the value of the criterion calculated n times with the new point replacing each one of the design's original points in turn. The criterion values are then compared and the point whose replacement by the new point leads to the largest increase in the criterion value over that of the original design is discarded in favour of the new

point. If the substitution of this point does not result in any improvement in the criterion value then the original design is retained. This process is repeated many times until the potential for further improvement in the criterion is thought to be small, at which time the current design is selected and used as the ‘best design’. A computer program was written in APL to perform this search

5.5.2. Updating the smoothing parameter values

Once a design had been selected and the output of the algorithm evaluated for each point, to give the vector \mathbf{y} , this objective data could be used to update the prior estimate of δ as described in 4.2.

An APL program was written to numerically determine the values of δ that had the highest posterior probability. The program was designed to obtain the modal values of δ using both the full joint posterior distribution and the distribution that has been marginalised with respect to β . The program calculates both of these estimates using a downhill simplex algorithm due to Nelder and Mead [NM65].

5.5.3. Estimation of the value of the ^{131}I algorithm for specified values of lw and lf

This calculation is not strictly part of the Bayesian uncertainty analysis. However, the ability to measure the accuracy with which the Bayesian methodology can predict the output of the algorithm for particular values of the uncertain parameters will be useful. The exact value of the ^{131}I algorithm has already been calculated over a 1000×1000 grid of points on the parameter space, in order to obtain independent estimates of the mean and variance of the algorithm 5.4.

By comparing the predictions over this grid to the true values it can be determined how well the Bayesian methodology estimates individual values of the algorithm. The effect on the quality of the predictions of increasing the design size and updating the estimates of δ can also be examined. This should provide valuable information as to the most useful allocation of resources in order to obtain the most accurate uncertainty analysis.

Strictly speaking, this analysis will only give information about how to optimise the estimation of individual values of the algorithm. However, we hope to demonstrate that measures which maximise the accuracy of the point predictions will also have a beneficial effect on the estimates of the mean and variance.

5.5.4. Estimation of the mean and variance of the ^{131}I algorithm over the distributions of the uncertain parameters

The generic form of the expressions for the mean and variance of an algorithm for which the uncertain parameters have normally distributed priors is detailed in 4.4 and 4.5. These can be directly applied to the ^{131}I algorithm. The results of the Bayesian analysis are detailed in 5.6.3.

5.6. Results

5.6.1. The ‘true’ uncertainty

The two sets of ‘true value’ estimates of K and L , the mean and variance of the ^{131}I algorithm, are detailed in Table 5.2.

Weighted Sample method		Model based method	
Mean, K	Variance, L	Mean, K	Variance, L
2.532×10^{-8}	2.286×10^{-16}	2.578×10^{-8}	3.213×10^{-16}

Table 5.2: Measures of the ‘true’ uncertainty

The first set of ‘true value’ estimates was calculated as weighted averages over the 1000×1000 grid of ^{131}I values, each observation being weighted by the probability of selecting the associated parameter values when sampling from the distributions of the uncertain parameters.

The second set of estimates was obtained from the linear regression model

$$\eta(w, f) = \alpha + \gamma.w + \xi.f + \epsilon.w.f + \theta.w^2 + \lambda.f^2 \quad (5.5)$$

which was found to be the best-fitting, most parsimonious model. using GLIM [AAFH90], based on a subset of 1600 points selected from the grid

of true values of $\eta(\cdot)$. The fitted function, which takes a linear-quadratic form in both w and f , was then integrated to obtain analytical expressions for the expected value and the expected value squared of $\eta(\cdot)$, from which the values for the ‘true’ mean and variance were calculated. i.e.

$$\begin{aligned} K &= E[\eta(w, f)] \\ &= \int_0^\infty \int_0^\infty (\alpha + \gamma.w + \xi.f + \epsilon.w.f + \theta.w^2 + \lambda.f^2) dG(w) dG(f) \end{aligned} \quad (5.6)$$

The difference between the two estimates of K is small, less than 2%, suggesting that both estimates are close to the true value of K . For the estimate of L , the weighted sample method gives a smaller value than the model based method. A plausible explanation for this difference is that the grid of CEDE values from which the weighted sample variance is calculated only extends in each direction to the 99th percentile of the distributions of w and f . This could result in some variation in the extreme tails of the distributions being lost and thus causing this estimate of the variance to be smaller than the correct value. The model based method will not suffer from this problem since the generalised linear model extends over the whole w, f parameter space. However, it may be subject to other errors since the simplified model may not be a good enough representation of the true function to enable a sufficiently accurate estimate of the variance of $\eta(\cdot)$ to be obtained. Hence, the value of L is likely to be greater than the weighted sample mean but there is no way of determining if the model based estimate is too large or small.

Of course, in a ‘real-life’ analysis the above calculations would not be possible but in this case they will provide a valuable independent comparison of the Bayesian and classical methodologies.

5.6.2. The NRPB Monte Carlo analysis

The results of the Monte Carlo (MC) analysis consist of 1000 output values from the ^{131}I algorithm. From this set of values the frequency distribution was constructed, as shown in Figure 5.2, along with a sample mean, a sample variance, and a 95% confidence interval, whose bounds were defined

as the 2.5 and 97.5 percentiles. These values are provided in Table 5.3.

Sample Mean	Sample Variance	95% confidence interval for the prediction of a future single value
2.34×10^{-8}	4.21×10^{-16}	$3.88 \times 10^{-9} - 8.18 \times 10^{-8}$

Table 5.3: Results of the Monte Carlo analysis

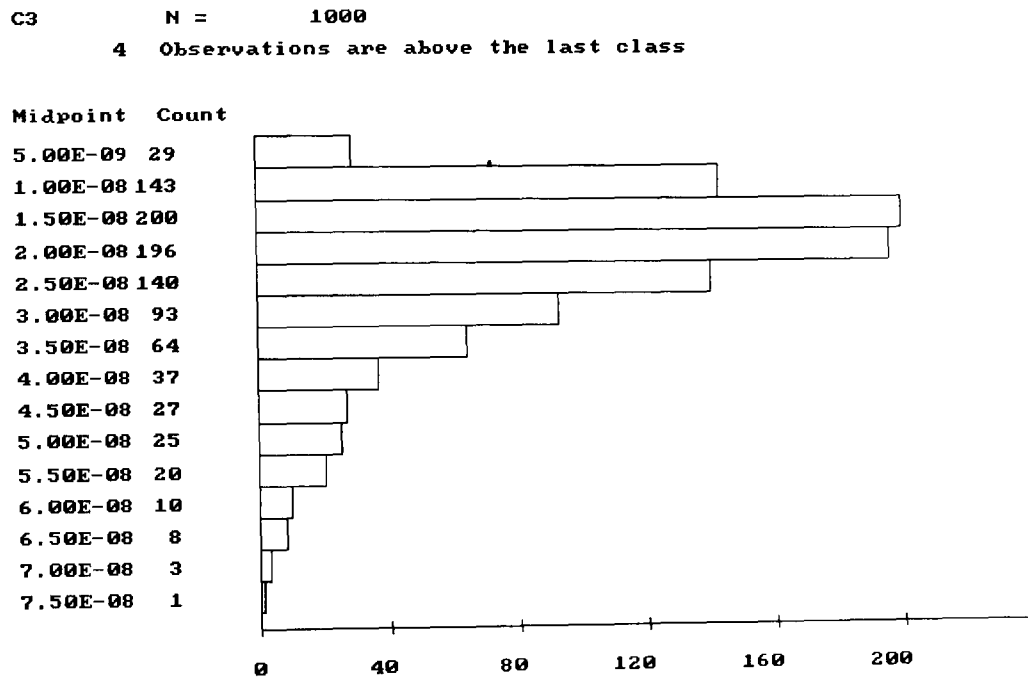


Figure 5.2: Frequency histogram of the sample output of the Monte Carlo analysis

From the frequency histogram it can be seen that the sample distribution is approximately lognormal. The sample mean represents the MC estimate of K . Its value differs from the lower of the two ‘true value’ estimates by just over 7.5%. Similarly the sample variance represents an estimate of L which differs from the upper of the two ‘true values’ for L by almost 24%. In addition, the confidence interval provides bounds for the value of a future single evaluation of the CEDE. These are the only results provided by the MC analysis.

However, a 95% confidence interval of the sample mean, based on the central limit theorem, can be calculated to represent a confidence interval

for K based on the MC sample. This interval is obtained as $2.21 \times 10^{-8} - 2.47 \times 10^{-8}$. Clearly, the interval does not contain either of the estimates of the true value of K derived above.

5.6.3. The Bayesian analysis

To enable the Bayesian methodology to be studied in detail, three different sets of design points were selected, one each of 10, 15 and 20 points, all using $\delta = [1, 1]$. As has been discussed previously there is no practical way of selecting the correct smoothing parameter values, at the design selection stage, so the value one was arbitrarily selected as a default value for the point selection process. The three designs will be referred to as d_{10} , d_{15} and d_{20} respectively. For each of the designs, the set of n outputs from the ^{131}I algorithm was obtained using the Pedal program [KP90]. These will be referred to as y_{10} , y_{15} and y_{20} .

The next step in the analysis was to update the estimates of the smoothing parameters, $\delta = [\delta_{lw}, \delta_{lf}]$ using each of the selected designs in turn. Initially, estimates were obtained using the posterior distribution $f(\delta, \sigma^2 | \mathbf{y})$ (4.16). Unfortunately, the surface of this function was very flat and by starting the search at different positions on the parameter space greatly different answers were obtained for the estimates. In many cases the values of δ_{lw} , δ_{lf} tended to zero or infinity indicating that no non-trivial modal values existed. Consequently, these estimates were rejected and new estimates were calculated using the full joint posterior distribution $f(\beta, \sigma^2, \delta, | \mathbf{y})$ (4.15). The original and corresponding updated δ values are shown in Table 5.4.

Design	Size	Original δ_{lw}, δ_{lf}	Improved estimates ² for δ_{lw}, δ_{lf}
y_{10}	10	1, 1	1.35, 1.15
y_{15}	15	1, 1	3.62, 2.34
y_{20}	20	1, 1	5.46, 3.74

Table 5.4: The improved estimates for δ_{lw} and δ_{lf}

In all cases the improved estimates of δ are larger than the original val-

ues. This indicates that the original values underestimated the smoothness of the algorithm with respect to the uncertain parameters.

The next step of this analysis was to examine how well the Bayesian methodology estimated individual values of the CEDE for given values of the parameters w and f (see 5.5.3). Using designs d_{10} , d_{15} and d_{20} with both their original δ values and the improved estimates for the δ parameters, the value of the CEDE was estimated over a 30×30 regularly spaced grid of points selected from the 1000×1000 grid that was used to calculate the ‘true uncertainty’ and at which the true value of the CEDE had already been calculated. The Bayesian point estimates were compared with the true values to assess their quality. Summary measures for this comparison are shown in Table 5.5.

Design y	δ	Indicators of percentage error			
	δ_{lw}, δ_{lf}	Max error	Min error	Absolute mean	Absolute weighted mean
y_{10}	1, 1	386	-20.4	12.8	3.95
y_{10}	1.35, 1.15	353	-17.8	11.1	3.36
y_{15}	1, 1	116	-22.4	3.48	1.20
y_{15}	3.62, 2.34	3.63	-48.1	2.24	0.69
y_{20}	1, 1	53.1	-23.4	3.11	1.28
y_{20}	5.46, 3.74	8.60	-7.40	0.90	0.19

Table 5.5: Summary statistics for the quality of the estimates of CEDE

A number of observations can be made from this table. Firstly, as the number of points in the design increased the errors in the Bayesian estimates decreased. Secondly, for a particular set of design points the magnitude of the percentage errors was smaller when the improved estimates of the smoothing parameters were used in place of the original values. The importance of using the best estimates of the smoothing parameters is demonstrated by the observation that the estimates, generated using the 15 point design with the improved δ estimates, are more accurate than those generated using the 20 point design and the original δ values. This indicates that it may be more efficient to optimise δ than to increase the

number of points in the design. In this example, it is possible to see the effects of increasing the design size and optimising the estimates of δ on the estimates of the mean and variance. Unfortunately, in a ‘real-life’ analysis the effectiveness of these optimisations would be difficult to quantify since the construction of Table 5.5 would not be feasible.

Next, using each set of design points with both their original δ values and the improved estimates, the mean and variance of the posterior t distribution for K , the expected value of the CEDE was calculated. From this distribution a 95% probability interval for K was derived. These results are shown in Table 5.6.

Design	Size	δ_{lw}, δ_{lf}	Mean	Variance	95% probability
y_{10}	10	1, 1	2.549×10^{-8}	6.748×10^{-21}	$2.529 - 2.568 \times 10^{-8}$
y_{10}	10	1.35, 1.15	2.550×10^{-8}	4.933×10^{-21}	$2.533 - 2.566 \times 10^{-8}$
y_{15}	15	1, 1	2.559×10^{-8}	2.750×10^{-20}	$2.522 - 2.595 \times 10^{-8}$
y_{15}	15	3.62, 2.34	2.565×10^{-8}	1.699×10^{-21}	$2.536 - 2.593 \times 10^{-8}$
y_{20}	20	1, 1	2.565×10^{-8}	4.983×10^{-21}	$2.550 - 2.579 \times 10^{-8}$
y_{20}	20	5.46, 3.74	2.566×10^{-8}	1.333×10^{-22}	$2.563 - 2.568 \times 10^{-8}$

Table 5.6: Parameters of the posterior t distribution for K , the mean of the CEDE

There is a high level of consistency between the Bayesian point estimates of K ; the largest and smallest differ by less than 1%. Also, all of these estimates lie between the two ‘true value’ estimates suggesting that the Bayesian results are reliable. Further, as either the design size is increased or the estimates of δ are updated then the variance associated with the point estimate for K decreases. Again it can be seen that a smaller variance can be obtained for the point estimate by updating δ than can be obtained by increasing the design size by 5. The variance can be interpreted as a self-assessment of how accurate the Bayesian methodology is at estimating K . All but the 20 point design using the updated δ values provide a 95% probability interval that contains at least one of the ‘true value’ estimates. The interval for the 20 point design lies between the ‘true value’ estimates which could indicate that either the Bayesian methodol-

ogy is being too optimistic in assessing its accuracy or that the ‘true value’ estimates could be improved.

Finally, the mean and variance of the posterior distribution for L , the variance of the CEDE, was derived for each design and with both the original and improved estimates of the δ parameters. These values are given in Table 5.7.

Design	Size	δ_{lw}, δ_{lf}	Mean	Variance
y ₁₀	10	1, 1	2.448×10^{-16}	5.992×10^{-33}
y ₁₀	10	1.35, 1.15	2.451×10^{-16}	6.505×10^{-33}
y ₁₅	15	1, 1	2.571×10^{-16}	2.099×10^{-32}
y ₁₅	15	3.62, 2.34	2.629×10^{-16}	7.068×10^{-32}
y ₂₀	20	1, 1	2.586×10^{-16}	1.562×10^{-32}
y ₂₀	20	5.46, 3.74	2.664×10^{-16}	2.874×10^{-31}

Table 5.7: Mean and Variance of the posterior distribution for L , the variance of the CEDE

This table again shows a high level of consistency between the Bayesian point estimates of the variance, L . All the estimates lie between the ‘true value’ estimates of the variance. The classical methodology does not provide a measure to predict the accuracy of its estimate of the variance of the output that can be compared with the Bayesian estimates. The Bayesian estimates run counter to what might be expected. As the number of design points increases the variance of the unknown distribution also increases. This indicates that the methodology is less confident in its estimate of the variance of the CEDE as the number of design points increases. One possible explanation for this is that little variability is observed when very few design points are used, but as more objective data are provided further variability is revealed. If this were the case then one would expect the variance to continue to increase as the number of design points increases up to a certain number of points and then fall.

5.7. Discussion

The object of this analysis was to demonstrate that, in a simple low dimensional problem, the Bayesian uncertainty analysis could achieve greater accuracy than the classical MC based methods. The ^{131}I algorithm provided an ideal subject for this purpose as it was possible to estimate the true uncertainty in the model's output with respect to the parameters w and f , independently of the choice of methodology.

The Bayesian estimates of K were all in close agreement with the 'true values', while the MC estimate was more than an order of magnitude further from the 'true value' than any of the Bayesian estimates.

The close proximity of the Bayesian estimates to the 'true values' was also seen in the estimation of the variance of the algorithm's output. Again, all the Bayesian point estimates lie between the two 'true value' estimates whereas the MC estimate exceeds both of these values considerably. Another point in favour of the Bayesian analysis is that it provides a measure of uncertainty for the point estimate of the variance, in contrast to the MC methodology.

In each of the Bayesian analyses, the improvements in the accuracy of the estimates of the mean, K , and the variance, L , were associated with an increase in the size of the set of design points used and with the updating of the estimates of δ . It could be seen, in a number of cases, that improving the estimates of the smoothing parameters was more beneficial than increasing the size of the design. This has important implications for the use of this methodology when performing uncertainty analyses on large algorithms with many uncertain parameters in which evaluating extra points could be more computationally expensive than updating the estimates of the smoothing parameter values.

A useful and informative addition provided by the Bayesian methodology is the ability to provide a P-value, representing the probability that the true value of the model's output, for particular parameter values, exceeds some predefined critical value, see **3.3.1**. Using a grid of these P-values a contour plot can be drawn. This provides a clear picture of how sensitive the value of the algorithm is to uncertainty in the values of the parameters.

An example of such a contour plot is shown in Figure 5.3 for the 10 point design.

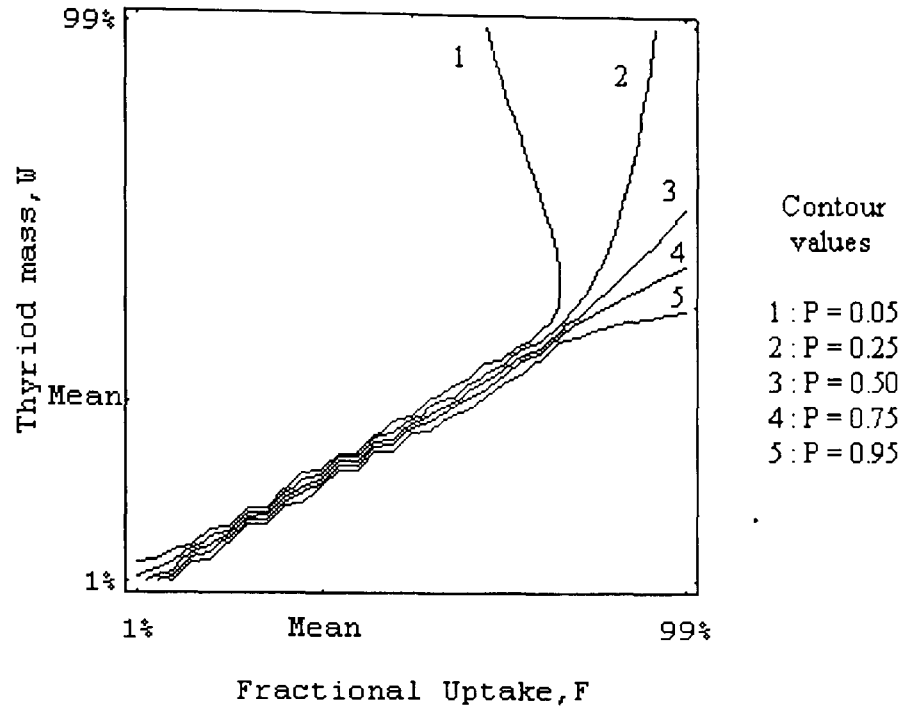


Figure 5.3: Contour plot of the probability that the output of the algorithm for particular w and f values will exceed the predicted mean output value

The contours on this plot indicate the various levels of probability that an output will exceed the ‘true value’ of the mean for the particular parameter value combinations. It can be seen that for most of the parameter space the probability is close to either one or zero. For only a small region is there any sizeable doubt as to which side of the true mean the actual value would fall.

Taking these results at face value, it is easy to conclude that the reason the Bayesian measures were closer to the ‘true values’ was that this was a more accurate form of uncertainty analysis, compared to the MC analysis. However, the following factors should be taken into consideration before this conclusion is accepted.

The 1000 points of the MC dataset were randomly selected from the full distributions of the uncertain parameters. In contrast, the data for the ‘true’ value and Bayesian estimates were selected from within a range of 2.5 standard deviations either side of the mean value of the uncertain

distributions. We would expect the MC dataset to contain about 10 observations outside the 2.5 standard deviation range. Thus the ‘true values’ and the Bayesian estimates are really measures of the uncertainty in output of the algorithm based on uncertain input distributions truncated to approximately the 99th percentiles whereas the MC analysis provides estimates of the algorithms uncertainty based on the full input distributions.

As a result, the closeness of the ‘true values’ and Bayesian estimates might be expected since both are using data from the same bounded parameter space to measure the same thing. Furthermore, a difference between these values and the results of the MC analysis may occur not because the MC analysis is less accurate but because it is in fact estimating something slightly different.

The degree to which these differences in the input parameter space affects the analyses depends on the behaviour of the algorithm for values of the uncertain parameters outside the 2.5 standard deviation range, i.e. in the tails of the distributions. This behaviour will be dependent on the type of distribution assigned to each uncertain parameter. For a reliable comparison of the methodologies the bounds of the sample space of the input parameters, from which the Bayesian and ‘true value’ datasets are selected, should be sufficiently wide to ensure that the effects of excluding the tails of the input parameter distributions are negligible.

In this example, normal distributions were selected to represent the uncertainty in the thyroid mass and uptake fraction. These distributions are symmetric with short tails and thus it was considered that the exclusion of the sample space outside the 99th percentile range of these distributions should not adversely affect the validity of the comparison of the MC and Bayesian uncertainty analysis methodologies.

In conclusion, taking into account the factors highlighted above, the Bayesian methodology has, in the case of this small example, demonstrated its ability to provide a wider range of measures about the uncertainties in the output of an algorithm resulting from uncertainties about its parameters. Further, these measures are based on a smaller quantity of data than required by an MC based analysis to obtain a similar or higher degree of

accuracy. Thus, the results of these analyses support the testing of the Bayesian methodology on a ‘real-life’ problem. A paper based on a subset of the above results has been published [HO96].

6. COMPARISON OF BAYESIAN AND CLASSICAL UNCERTAINTY ANALYSIS METHODOLOGIES USING THE ICRP 67 MODEL FOR PLUTONIUM METABOLISM

6.1. Introduction

In the previous chapter it was established that Bayesian uncertainty analysis was feasible in a simple two-dimensional problem. The object of this chapter is to demonstrate the application of Bayesian uncertainty analysis to a ‘real-life’ multi-dimensional problem. In contrast to the analysis of the iodine algorithm it will not be possible to check the accuracy of this analysis by comparing it to an independently calculated set of ‘true’ answers. This being a ‘real-life’ problem, if it were possible to obtain the true uncertainties then performing either a Bayesian or classical uncertainty analysis would be pointless. However, it will be possible to examine any differences between the various Bayesian analyses and make a comparison with the results generated by the classical analysis.

The model selected for these analyses is that used currently at NRPB for the estimation of radiation doses due to plutonium (Pu) absorption into the human body. Due to its very long biological and decay half-lives Pu is an important source of long-term radiation exposure since a large proportion of any inhaled or ingested Pu will persist in the body for the remainder of a person’s life. The selected model is also that currently recommended by the International Commission on Radiological Protection (ICRP) for the measurement of plutonium metabolism and so is widely used in the radiation protection community for the prediction of doses. It

is a large and computationally expensive algorithm upon which it is hoped to demonstrate that when compared to a classical analysis a Bayesian uncertainty analysis is more informative and accurate using fewer data and less computational effort.

6.2. The development of the ICRP 67 biokinetic model for plutonium metabolism

The International Commission on Radiological Protection (ICRP) was founded in 1928 and, since 1950, has provided guidance on the wide-spread use of radiation sources caused by developments in the field of nuclear energy production. Part of this guidance has been the provision of intake limits for radiation workers to ensure that dose limits are not exceeded.

To perform this task it was necessary to develop models to determine the dose received from a unit intake of a radioactive substance. These models differ between elements, to reflect the wide range of behaviour of radioactive substances in the human body.

Plutonium has been recognised by ICRP as a potential health hazard since it was first produced in the 1940's. Biomedical experiments to examine the behaviour of Pu in the human body were begun in 1944. Initially, data were obtained from studies using the rat as a surrogate subject for man. These data indicated that only a small fraction of ingested Pu entered the blood stream but that a much larger fraction of inhaled Pu was absorbed through the wall of the lung due to the prolonged retention in the lung as compared to that in the gut. The early experiments also indicated that once Pu entered the blood stream, it was quickly deposited around the body, mainly in the skeleton and the liver. Pu was also found to have a long biological half-life. This means that once taken up it remains in the body for a prolonged period of time. In fact a large proportion will remain to the end of life. As a result the accurate measurement of the quantity and location of Pu in the human body, for the purposes of dose calculations, is very important.

Between 1945 and 1946 an experiment was carried out, on 18 human

volunteers, [LBHC50], [Lan59], all of whom were seriously ill and had short life expectancies. Each subject was injected with trace quantities of Pu, either in a nitrate or citrate form. The aim of the experiment was to determine the relationship between urinary and faecal excretion rates and the amount of Pu remaining in the body. Quantities of Pu in the excreta of the subjects were measured in the weeks following the injection of the Pu and measurements of tissue concentrations were taken after the deaths of some of the subjects. Tissue measurements were not possible for the all subjects, within the time frame of the experiment, since not all the subjects died in the short period expected; eight lived longer than eight years and four were still alive 30 years later. This experiment still represents a large proportion of the direct human evidence concerning the metabolism of Pu.

From this study and data on the accidental contamination of radiation workers Langham [LBHC50] derived simple urinary and faecal excretion curves that related the quantity of the injected Pu in the excreta to the time since administration. These equations were for many years used by ICRP as the basis for the quantification of Pu intakes and hence the calculation of CEDE, the quantity used to set dose limits.

Since the 1980's a number of studies have indicated that Langham's equations may overestimate the CEDE received for a given intake. Langham's equations are based on data covering a relatively short period following first exposure and assume that the removal of Pu from the various organs and structures of the body continues at a constant rate over the lifetime of the subject with no possibility of recycling between organs. This has been found to be an unrealistic assumption since Pu deposited on the surface of the skeleton will be buried by new bone growth and therefore be immobile until uncovered again in later years by the natural process of bone loss in old age.

As a result, Langham's equations become less accurate with increasing age and time since exposure. It has been shown by Norwood and Newton [NN75], and McInroy [McI76] that at times more than five years following exposure these equations can overestimate the actual amount of Pu in the body, as determined by autopsy, by as much as 10 times.

The inaccuracies in Langham's equations cause a problem in radiological protection scenarios, since intake limits are set with reference to the CEDE which estimates a subject's predicted dose over the 50 years following exposure.

In recent years, further work has been done on the metabolism of Pu and other similarly behaving radioactive elements such as Americium and Neptunium in the human body. These new data were used by Leggett [Leg85] to derive a mechanistic model to predict the behaviour of these elements. The term mechanistic is used since the model attempts to describe, at least in a relatively simplistic way, the physiological processes involved in the actual movement of Pu about the body.

The ICRP saw this type of model as an advance in the drive towards improving standards in radiological protection and, in its Publication 56 [ICR89], adopted this form of model for the calculation of doses from intakes of Plutonium, Americium and Neptunium. The mechanistic models are much better able to describe the time-dependent behaviour and thus provide more accurate predictions of committed effective doses. They do, however, have the disadvantage of requiring considerably more computing power than their simpler predecessors to implement.

The current model recommended by ICRP for dose calculations of Pu intakes is a modification of the mechanistic model described in Publication 56. It is identical in structure to that recently published by Leggett [Leg92] except that a second liver compartment has been included. This latest model is detailed in Appendix B of ICRP 67 [ICR93].

As with the iodine algorithm, the model is defined in terms of body compartments and transfer coefficients between the compartments. It comprises 19 compartments, with 29 transfer rates detailing the patterns of movement of Pu between the compartments. One restriction is that it only describes the behaviour of Pu once it has entered the bloodstream; extra components can be added to simulate the method of entry to the blood stream, namely inhalation, ingestion or directly through a wound.

A pictorial representation of the model is given in Figure 6.1, and the standard values of the transfer coefficients between the various compart-

ments are listed in Table 6.1.

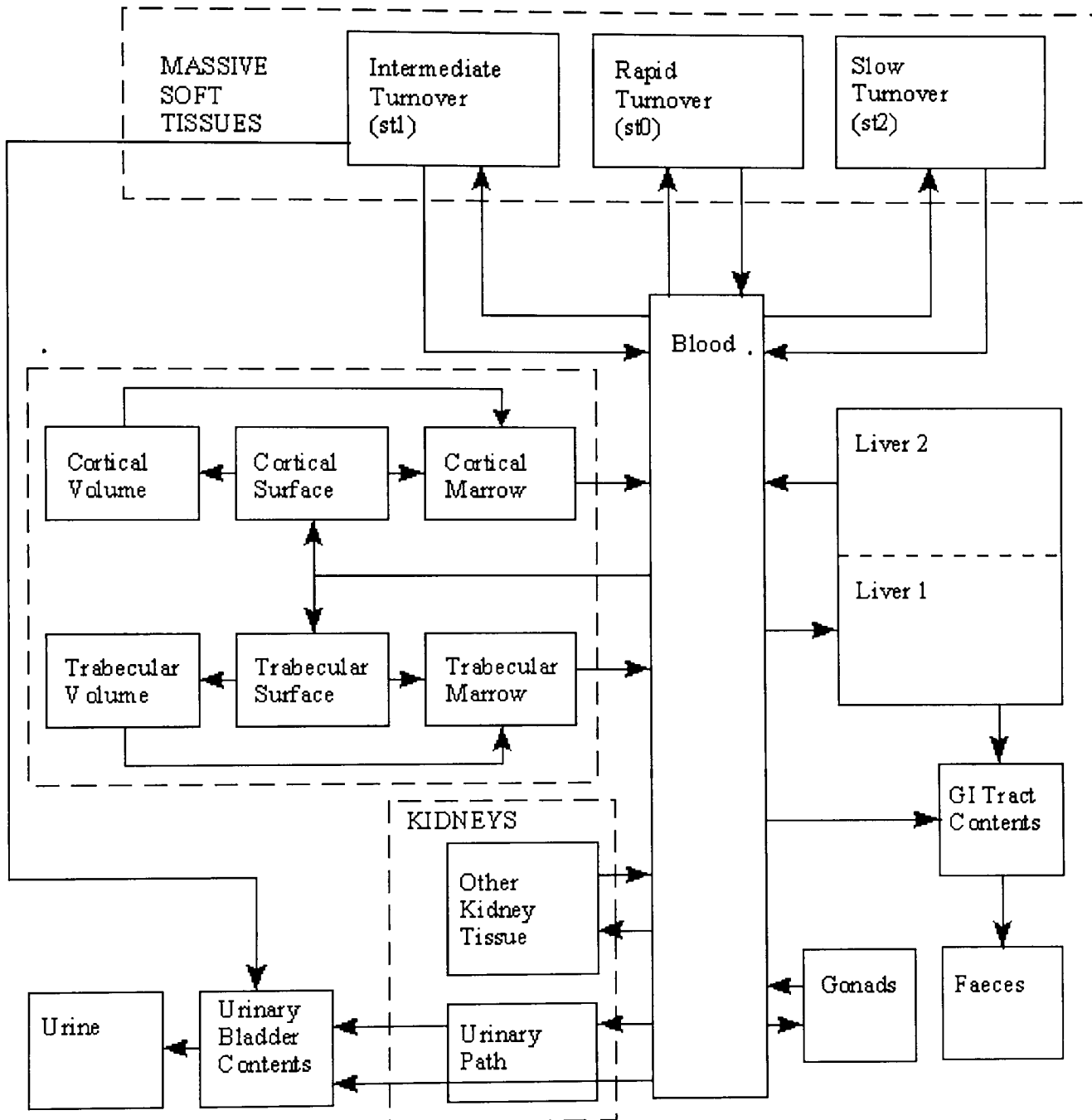


Figure 6.1: The ICRP biokinetic model for plutonium metabolism

The illustration shows that this is a large and complex model incorporating many recycling pathways. Again the output is in the form of a CEDE value. In contrast to the iodine example, no ingestion pathway will be specified for this uncertainty analysis. It will be assumed that a unit

Transfer coefficient	Standard value
blood to liver 1	0.1941
blood to cortical surface	0.1294
blood to trabecular surface	0.1941
blood to urinary bladder contents	0.0129
blood to kidney (urinary path)	0.00647
blood to other kidney tissue	0.00323
blood to upper large intestine contents	0.0129
blood to testes	0.00023
blood to ovaries	0.000071
blood to soft tissue 0	0.2773
blood to soft tissue 1	0.0806
blood to soft tissue 2	0.0129
soft tissue 0 to blood	0.693
kidneys (urinary path) to bladder	0.01386
other kidney tissue to blood	0.0139
soft tissue1 to blood	0.000475
soft tissue1 to urinary bladder contents	0.000475
soft tissue 2 to blood	0.000019
trabecular surface to volume	0.000247
trabecular surface to marrow	0.000493
cortical surface to volume	0.0000411
cortical surface to marrow	0.0000821
trabecular surface to volume	0.000493
cortical volume to marrow	0.0000821
cortical/trabecular marrow to blood	0.00076
liver 1 to liver 2	0.00177
liver 1 to small intestine	0.000133
liver 2 to blood	0.000211
gonads to blood	0.00019

Table 6.1: Standard transfer rates per day for the Plutonium model

quantity of Pu is injected directly into the blood stream, thus avoiding the need to complicate the model further by specifying an intake pathway.

6.3. Why is an uncertainty analysis of the Pu model required?

The true values of most of the parameters of the model listed in Table 6.1 are not known. They are not directly measurable, neither is it possible to obtain definitive experimental data from which to estimate these values. Thus these parameters must be estimated by ‘experts’ using available subjective information.

This lack of knowledge causes parameter value uncertainty to be introduced into the model. Hence, the output of the model will vary in some unknown way with the expert’s choice of coefficient values. For the model to be used with confidence in the radiological protection environment, it is important to quantify this uncertainty.

This means that a reliable analysis of the model’s parameter value uncertainty is of great importance. Ideally, this analysis should provide an estimate of the expected value of the model’s output for a unit intake and also estimate the variability on both a single predicted value of the model’s output and the expected output.

6.4. Classical uncertainty analysis of the Pu model

6.4.1. Selection of the parameters on which to perform the classical analysis

To perform a classical uncertainty analysis encompassing all 29 estimated transfer coefficients would be a large task and in fact this may not actually be necessary. It is likely that the value of the model’s output is only sensitive to variation in the values of a subset of these coefficients. Hence, at NRPB, a sensitivity analysis was performed which identified a subset of the coefficients to which the model was most sensitive. The members of this subset were not all independent and to perform a classical uncertainty

analysis on this set of dependent coefficients would be difficult. Therefore a substitute set of independent parameters was created. This process is illustrated by the following example.

The output of the model was found to be sensitive to the rate of transfer of Pu from the blood to the cortical bone surface and from the blood to the trabecular bone surface. These are the only two bone compartments of the model into which Pu can be transferred from the blood; hence the sum of these two transfer rates must equal the total rate of transfer from the blood to the skeleton. However, the total transfer to the skeleton is also an uncertain parameter to which the model is sensitive. In order to retain coherence, the setting of two of these parameters should automatically define the value of the third. Thus the cortical and trabecular transfer coefficients were replaced, for the purposes of the uncertainty analysis, by a single parameter representing the fraction of the total Pu transferred to the skeleton that is deposited on the cortical bone surface. The remainder of the Pu is then assumed deposited on the trabecular bone surface.

Similar procedures were carried out on other parameters of the subset, at the end of which a new set of 14 independent and coherent parameters (in terms of our prior beliefs about them) had been defined to describe the coefficients to which the model output was sensitive. An ‘expert’ was then asked to provide distributions and 95% ranges for the uncertainty associated with these parameters, as shown in Table 6.2.

6.4.2. Details of the analysis performed

Five hundred sets of the fourteen parameters defined in Table 6.2 were selected using a Latin Hypercube sampling procedure which was discussed in **2.2.1**. For each of these sets of 14 values and using those fixed coefficients to which the model was found to be insensitive (set to their standard values as in Table 6.1), a full complement of the 29 coefficients for the model was calculated. The model was then run with each set of coefficients and assuming a standard unit of Pu injected into the bloodstream. This produced 500 values forming a sample distribution of the output which was then used to obtain the same descriptive statistics as in the analysis of the

Parameter	Distribution	Lower 95% bound	Upper 95% bound
Rate of clearance from blood	Lognormal	0.1085	7.870
Rate of clearance from liver 2 to blood	Lognormal	3.744×10^{-5}	1.189×10^{-5}
Rate of clearance from soft tissue1	Lognormal	1.11568×10^{-4}	8.0892×10^{-3}
Rate of clearance from cortical surface	Lognormal	3.520×10^{-5}	4.313×10^{-4}
Rate of clearance from trabecular surface	Lognormal	2.114×10^{-4}	2.590×10^{-3}
Ratio of fraction soft tissue 1-> ubc to soft tissue 1 -> blood	Beta	0.00, 1.5 [#]	2,2 ^{##}
Fraction of initial deposition to skeleton	Beta	0.25, 0.75 [#]	2,2 ^{##}
Fraction of initial deposition to trabecular/cortical surface	Beta	0.3, 0.9 [#]	2,2 ^{##}
Fraction of initial deposition in the testes	Lognormal	4.11×10^{-5}	2.98×10^{-3}
Rate of clearance from testes	Lognormal	4.11×10^{-5}	2.98×10^{-3}
Rate of clearance from small intestine to upper large intestine	Lognormal	0.7046	51.09
Rate of clearance from upper large intestine to lower large intestine	Lognormal	0.2114	15.33
Rate of clearance from lower large intestine to faeces	Lognormal	0.1174	8.515

Table 6.2: The parameters (day^{-1}) and their subjective prior distributions to which the plutonium model is sensitive. (# = upper and lower bounds ## = shape parameters)

iodine model. The results of this analysis are described in 6.6.

6.5. Bayesian uncertainty analysis of the Pu model

For the iodine example, software was written specifically to perform the calculations required. For instance, the dimensions of the problem and the parameters of the prior distributions were hard-wired into the software. However, the Bayesian methodology which was found to be superior in the simple iodine example would be of little use if new software had to be written to analyse each new problem. Thus, to perform the calculations for this ‘real-life’ problem, two general programs were written that allowed the user to specify the details of the specific uncertainty analysis under consideration. The first was designed to select optimum design sets while the second provided the means to update the smoothing parameters and also to obtain the summary measures of the analysis. Both were conceived to be user-friendly and adaptable to a wide range of problems. All the results of the Bayesian analyses in this chapter have been obtained using this software. A detailed description of the programs is given in the appendix.

6.5.1. Transforming the parameters

The first problem that had to be overcome, in order to perform a Bayesian uncertainty analysis, was that the 14 independent parameters used in the classical analysis have a number of non-normal uncertainty distributions associated with them, i.e. Beta or Lognormal as defined in Table 6.2. The Bayesian methodology as described in chapter four was constructed to be applied only to parameters with independent normal uncertainty distributions. To overcome this difficulty, for each of the non-normal uncertain parameters, transformations were applied to each of these parameters to normalise them.

Lognormal: For these parameters the standard log transformation was applied.

Beta: For these parameters the logistic transformation was applied. Thus for example if a parameter θ was distributed as $\text{Beta}(\alpha, \beta)$ with

upper and lower bounds of A and B respectively then the parameter

$$\Theta = \log \left[\frac{\theta - A}{B - \theta} \right]$$

would be considered as normally distributed with mean and variance given by

$$\begin{aligned} \mu &= 2 \log Z - \left(\frac{(Z_2 - Z^2)}{2} \right) + Z^2, \\ \sigma^2 &= (Z_2 - Z^2) + Z^2 - 2 \log Z, \end{aligned}$$

respectively, where

$$\begin{aligned} Z &= \frac{\alpha}{\beta - 1}, \\ Z_2 &= \frac{\alpha(\alpha + 1)}{(\beta - 1)(\beta - 2)}. \end{aligned}$$

Gamma: For these parameters the uncertainty distribution was initially replaced by a lognormal distribution with the same mean and variance. The standard log transformation was then applied as described above.

Using these transformations a set of 14 normally distributed parameters was obtained on which to perform a Bayesian analysis.

6.5.2. Selection of the design points

The sets of design points were chosen using the criterion defined in 4.1. To recap, the ‘best’ design of a certain size was defined to be that which maximised

$$\int_{\chi} \mathbf{t}(\mathbf{x}) \mathbf{A}^{-1} \mathbf{t}(\mathbf{x}) dG(\mathbf{x}) \quad (6.1)$$

which for the Pu model, having only independent uncertain parameters, can be written as

$$\prod_i \left[\int_{\chi_i} \mathbf{t}(x_i) \mathbf{A}_{x_i}^{-1} \mathbf{t}(x_i) dG(x_i) \right] = \prod_i [\text{tr}(\mathbf{A}_{x_i}^{-1} \mathbf{P}_{x_i})] \quad (6.2)$$

where the subscript i runs over the 14 dimensions of the problem and the matrices \mathbf{A} and \mathbf{P} take the same form as described in (4.13) and (4.12). For the iodine problem it was sufficient to simply define a grid of points. and then choose each design by starting with a random selection and updating

it with randomly drawn selections from the remaining points of the grid. In a 'real-life' problem this method of optimising the set of design points is not practical.

For the iodine problem the 20×20 grid equated to only 400 points in total. However, to create a comparable grid for the current problem the grid would have to contain approximately 1.6×10^{18} points which is not practical to maintain in electronic format. Furthermore, while the method of updating the design, by selecting points at random from the grid, is effective in the initial stages of the selection process, its efficiency falls rapidly as the design gets close to the optimum. Hence, the selection procedure was modified in two ways.

First, the initial random design was selected, then new points with which to update the design were also obtained by random selection from each of the parameter's uncertainty distributions and tested for inclusion in the design. It was noted that as the number of iterations increased, the frequency with which the new points were incorporated into the design decreased. Thus a second stepwise procedure was implemented in order to improve the speed and efficiency of the selection process as the design approached the 'best' design.

The principle behind this stepwise procedure is that, instead of testing for the improvement in the design when a point in the design is replaced by a new point, each existing point in the design is moved in turn a random number of steps of a predefined size in a random direction. The size of a step is user-defined. At each move the design criterion is recalculated and, if it has improved, the new position of the point is retained. As the design gets closer to its optimum the step size can be decreased, in order to make the procedure more sensitive. Hence, the most efficient method for selecting designs is to start with the random selection procedure and, when this becomes inefficient as the design approaches the 'best' design, to switch to the stepwise procedure.

The general design selection program detailed in the appendix has the option to use either the random or stepwise procedure. It allows the user to monitor the progress of the optimisation and to switch between the

procedures at an appropriate time.

A number of different sized designs were selected for comparison. As with the iodine model, it was necessary to select appropriate initial values for the smoothing parameters. For the current problem an attempt was made to select suitable values according to a more objective method.

The initial estimates of the smoothing parameters were derived as follows. For each transformed uncertain parameter the associated normal uncertainty distribution was used to determine the bounds of a 99% probability interval for the parameter's values. Using these upper and lower bound values the two smoothing parameter values that result in correlations of 0.99 and 0.01 respectively between the bound values were derived. These two smoothing parameter values were then taken to be the upper and lower 99% probability bounds of a lognormal prior distribution for the true value of the smoothing parameter. The mean and median values for this lognormal distribution were then obtained and used as initial estimates of the smoothing parameters.

6.5.3. Updating the smoothing parameter values

In the Bayesian uncertainty analysis of the iodine model it was observed that updating the smoothing parameter values could improve the accuracy of the analysis more than the addition of another five points to the size of the design. It was also found that the use of the posterior distribution $f(\sigma^2, \delta \mid \mathbf{y})$ to obtain a modal point estimate for δ was impractical in that example, owing to instability of the estimates, and was discarded in favour of optimisation using the full joint posterior distribution $f(\beta, \sigma^2, \delta \mid \mathbf{y})$.

In recognition of the fact that for 'real-life' problems the updating of these parameters could well be less computationally expensive than selecting a larger design, this updating procedure was extended to include the choice of another criterion. This criterion is based on the same likelihood function, (4.14), but the independent uniform prior distributions of δ were replaced by the lognormal distributions derived as part of the definition of the initial values of the smoothing parameters, 6.5.2.

The main advantage of this new criterion is that because of the influ-

ence of the prior distributions it tends to produce posterior values that are not very large or small compared to the size of the prior estimates. Using the uniform priors, each prior value was given equal probability whereas with the lognormal priors, very large or small values were given a small prior probability. Hence the risk of the updated values becoming unrealistically large or small was reduced. The general program written to perform Bayesian analyses allows the user to optimise the smoothing parameters using either of these criteria. The appendix contains the details of this program.

6.5.4. Details of the analyses performed

For this problem a number of different analyses were performed. Three different sizes of design were selected with 50, 100 and 150 points. The Bayesian uncertainty measures were then calculated based on both the original and improved smoothing parameter estimates for each design, using the general program detailed in the appendix. In order to provide a direct comparison between the two uncertainty analysis methodologies, a further Bayesian analysis was carried out using the Latin Hypercube design selected for the classical analysis. This design was also used to provide a means of objectively estimating the relative performance of the various Bayesian analyses. The results of these analyses are detailed in **6.7**.

6.6. Results of the classical uncertainty analysis

As described above, the output of the classical analysis consisted of a sample of 500 values. Figure 6.2 shows a frequency histogram of these values.

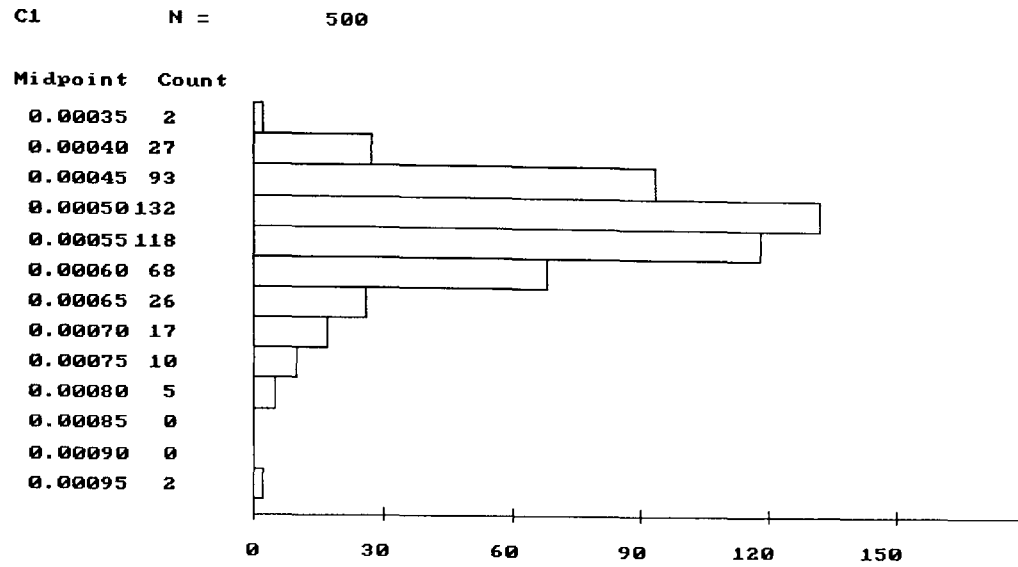


Figure 6.2: Frequency histogram of the sample output of the classical analysis

From these 500 values the sample mean and variance were derived as shown in Table 6.3, along with a 95% confidence interval for the mean.

Sample Mean	Sample Variance	Sample Variance of the Mean
5.35×10^{-4}	7.20×10^{-9}	1.44×10^{-11}

Table 6.3: Results of the Monte Carlo analysis

A 95 % confidence interval for the prediction of a future value was also obtained as $5.27 \times 10^{-4}, 5.43 \times 10^{-4}$.

6.7. Results of the Bayesian uncertainty analysis

As in the previous example, a number of different analyses were performed based on designs of different sizes and using various scenarios. Three different sizes of design were selected, consisting of 50, 100, and 150 points respectively. Two different sets of smoothing parameter values were used to select a design of each size. The set of initial smoothing parameter values called Prior5 are the means of the derived uncertainty distributions while the set labelled Prior8 are the medians of these distributions.

Once the six designs had been selected, improved estimates of the smoothing parameters were calculated using both criteria for each design and each set of initial smoothing parameter values.

The mean and variance of the t distribution for K and unknown distribution for L were then calculated for each design and smoothing parameter combination defined above, (see Tables 6.4 and 6.5). In addition, two extra analyses were performed using the classical Latin Hypercube sample of 500 points. The first analysis used smoothing parameters all set to a default value of 1. The other used a set of optimised values derived using one of the 150 point designs.

It can be seen from Table 6.4 that the difference between the largest and smallest values for the mean of K is only 1.3%. The size of the point estimates increases slightly as the number of points in the design increases from 50 to 150 but the change is only in the second decimal place.

A greater difference is, however, noticeable in the estimates of the variance of the distribution for K. For the calculations based on the original smoothing parameter values, Prior5 and Prior8, the variance of the distribution decreases by around an order of magnitude between the 50 and 150 point designs. Looking within each design, in each case the use of the updated smoothing parameters also reduced the size of the variance. The improvement was similar irrespective of which method was used to optimise the smoothing parameters.

The estimates of the mean and variance of the unknown distribution for L, (i.e. the variance of the model's output) follow a similar pattern to those for K. The mean values do, however, show less of a consistent increase in size with that of the design. Further, for the variance of L there appears to be no consistent trend in the values as either the size of the design increases or when the optimised smoothing parameters are used.

As discussed at the beginning of this chapter it was not possible to obtain an objective measure of the accuracy of these analyses since, in a real-life problem, this would defeat the object of the uncertainty analysis. However, it was possible to provide two forms of consistency check for the different Bayesian analyses. The first of these is based on the

Estimation of K				
Design	Size	Smoothing Parameters	Mean K	Variance K
g3	50	Original : Prior5	5.43×10^{-4}	1.44×10^{-11}
g3	50	Improved : Lognormal	5.42×10^{-4}	6.51×10^{-12}
g3	50	Improved : Uniform	5.42×10^{-4}	6.82×10^{-12}
g7	50	Original : Prior8	5.47×10^{-4}	2.34×10^{-11}
g7	50	Improved : Lognormal	5.47×10^{-4}	9.48×10^{-12}
g7	50	Improved : Uniform	5.49×10^{-4}	9.67×10^{-12}
g8	100	Original : Prior5	5.48×10^{-4}	3.32×10^{-12}
g8	100	Improved : Lognormal	5.45×10^{-4}	1.78×10^{-12}
g8	100	Improved : Uniform	5.46×10^{-4}	1.58×10^{-12}
g4	100	Original : Prior8	5.47×10^{-4}	4.51×10^{-12}
g4	100	Improved : Lognormal	5.47×10^{-4}	3.27×10^{-12}
g4	100	Improved : Uniform	5.46×10^{-4}	3.14×10^{-12}
g6	150	Original : Prior5	5.48×10^{-4}	2.32×10^{-12}
g6	150	Improved : Lognormal	5.46×10^{-4}	1.07×10^{-12}
g6	150	Improved : Uniform	5.47×10^{-4}	8.80×10^{-13}
g9	150	Original : Prior8	5.46×10^{-4}	2.94×10^{-12}
g9	150	Improved : Lognormal	5.46×10^{-4}	1.15×10^{-12}
g9	150	Improved : Uniform	5.46×10^{-4}	1.05×10^{-12}
LHS	500	All 1	5.28×10^{-4}	4.15×10^{-12}
LHS	500	g9 : Improved : Uniform	5.44×10^{-4}	1.97×10^{-13}

Table 6.4: Mean and Variance of the posterior distribution for K , the mean of the CEDE (Prior5 : mean values, Prior8 : median values)

Estimation of L				
Design	Size	Smoothing Parameters	Mean L	Variance L
g3	50	Original : Prior5	5.50×10^{-9}	5.33×10^{-19}
g3	50	Improved : Lognormal	5.36×10^{-9}	7.68×10^{-19}
g3	50	Improved : Uniform	5.71×10^{-9}	9.11×10^{-19}
g7	50	Original : Prior8	6.30×10^{-9}	8.00×10^{-19}
g7	50	Improved : Lognormal	6.14×10^{-9}	7.84×10^{-19}
g7	50	Improved : Uniform	6.63×10^{-9}	5.40×10^{-19}
g8	100	Original : Prior5	5.96×10^{-9}	8.16×10^{-19}
g8	100	Improved : Lognormal	5.52×10^{-9}	3.41×10^{-19}
g8	100	Improved : Uniform	5.62×10^{-9}	4.33×10^{-19}
g4	100	Original : Prior8	5.97×10^{-9}	5.87×10^{-19}
g4	100	Improved : Lognormal	5.01×10^{-9}	3.95×10^{-19}
g4	100	Improved : Uniform	5.66×10^{-9}	9.42×10^{-19}
g6	150	Original : Prior5	7.23×10^{-9}	1.51×10^{-18}
g6	150	Improved : Lognormal	6.61×10^{-9}	1.33×10^{-18}
g6	150	Improved : Uniform	6.61×10^{-9}	1.62×10^{-18}
g9	150	Original : Prior8	6.04×10^{-9}	5.91×10^{-19}
g9	150	Improved : Lognormal	6.15×10^{-9}	7.22×10^{-19}
g9	150	Improved : Uniform	6.09×10^{-9}	8.05×10^{-19}
LHS	500	All 1	6.54×10^{-9}	6.95×10^{-20}
LHS	500	g9 : Improved : Uniform	6.45×10^{-9}	6.92×10^{-19}

Table 6.5: Mean and Variance of the posterior distribution for L , the variance of the CEDE

Latin Hypercube sample obtained for the classical analysis. The Bayesian methodology described in 4.3 was used to predict each of the elements of the Latin Hypercube sample. These were then compared to the true values and a maximum and an average absolute percentage error were obtained for each design. The other consistency check was based on a ‘leave-one-out’ strategy. In turn, each of the points in a design was removed and the remaining set was used to predict the excluded point. Again a maximum and an average absolute percentage error were obtained for each design. The results of these analyses are detailed in Table 6.6.

6.8. Discussion

Considering first the estimation of the expected value of the model’s output, it can be seen that the classical sample mean, 5.35×10^{-4} , is smaller than the mean of K derived from any of the Bayesian analyses. The two additional Bayesian analyses based on the Latin Hypercube samples gave different values depending on which set of smoothing parameters was applied. The use of the optimised smoothing parameters resulted in an estimate of K , 5.44×10^{-4} , that was consistent with the other Bayesian analyses. However, the estimate based on the default value (i.e. one) differed from both the Bayesian and classical estimates. This suggests that the use of the default smoothing parameters was not advisable.

The variance of the sample mean, calculated from the classical analysis, is comparable to the variance of the distribution for K calculated in the Bayesian analyses using 50 points and the original smoothing parameter values. All the other Bayesian analyses have smaller variances for the distribution of K indicating a greater confidence in the accuracy of their point estimates.

Turning to L , the variance of the model’s output, the Bayesian point estimates are of the same order of magnitude as the classical estimate. There is no comparable measure produced from the classical analysis with which to compare the variance of the estimate of L . However, these values decrease with the number of points in the design and the use of the improved smoothing parameter estimates. This indicates that the predicted accu-

			Type of check			
Design set	Size	Smoothing Parameters	LHS		Leave-one-out	
			Max error	Mean error	Max error	Mean error
g3	50	Original : Prior5	48.8	5.6	16.3	4.0
g3	50	Improved : Lognormal	32.5	3.8	10.7	2.5
g3	50	Improved : Uniform	31.9	4.0	14.9	2.9
g7	50	Original : Prior8	39.2	6.1	18.8	4.4
g7	50	Improved : Lognormal	33.1	4.4	11.2	2.6
g7	50	Improved : Uniform	32.0	5.2	13.4	2.7
g8	100	Original : Prior5	47.5	4.6	17.3	4.0
g8	100	Improved : Lognormal	39.3	3.6	12.8	2.1
g8	100	Improved : Uniform	40.5	3.5	12.9	1.9
g4	100	Original : Prior8	66.9	5.1	16.2	3.7
g4	100	Improved : Lognormal	46.0	3.5	14.1	1.9
g4	100	Improved : Uniform	42.5	3.7	7.9	1.6
g6	150	Original : Prior5	35.8	4.9	22.7	4.0
g6	150	Improved : Lognormal	33.8	3.0	12.9	1.7
g6	150	Improved : Uniform	32.1	3.0	10.8	1.5
g9	150	Original : Prior8	85.6	3.9	18.0	3.2
g9	150	Improved : Lognormal	49.9	2.5	6.1	1.5
g9	150	Improved : Uniform	40.5	2.4	6.0	1.5

Table 6.6: Consistency checks on the Bayesian analyses (values in the table are percentages).

racy of the Bayesian point estimates is increasing with increasing design size and improved parameter estimates.

The Bayesian analyses using the Latin Hypercube sample both give values of the mean and variance of L consistent with those of the standard Bayesian analyses, except for the variance of L calculated using the default values for the smoothing parameters which is rather small.

Turning finally to the table of consistency checks, Table 6.6, it can be seen that in both cases as the number of points in the design is increased or if the improved smoothing parameter estimates are used, then both the maximum error and the average error of the predictions decreases. The estimates of the maximum percentage error are considerably larger for the LHS based check than those for the 'leave-one-out' check, but there is reasonable consistency between the average percentage error values across the two checks.

Thus, the Bayesian analyses provide measures of uncertainty that are consistent with those provided by the classical methodology. Some of the above results are included in a comparison of the classical and Bayesian methodologies that has been accepted for publication [HO].

7. COMMENTS, DISCUSSION AND FUTURE WORK

This project was motivated by a desire to find a more efficient and informative methodology with which to perform parameter uncertainty analyses; in particular, on the algorithms used in the field of radiation protection. The current Monte Carlo based methodology was identified as having a number of failings. First, the range of uncertainty measures available was limited to point estimates of the mean and variance of the algorithm's output and a measure of the variance of the mean. Secondly, for computationally expensive algorithms or those with many uncertain parameters it may not be possible to obtain a sample large enough for a reliable assessment of the uncertainties.

Thus, a method was required for performing parameter uncertainty analyses that produced a wider range of uncertainty measures and that reduced the number of elements required in the sample of output.

Computationally efficient methods are used in the related field of estimating individual output values for algorithms. These are based on the use of stochastic process models to analyse a sample of the algorithm's output. These methods assume that the output of the algorithm is, in some sense, a smooth function of its uncertain inputs. This implies that knowledge about the output of the algorithm at one point also provides some information about the algorithm in the local neighbourhood of the point. The amount of extra information provided by each sample point is dependent on the degree of smoothness. Classical uncertainty analysis does not assume such smoothness, and so makes no use of this extra source of information. O'Hagan, [O'H91], showed that stochastic process models could be used, in a Bayesian context, to estimate the integrals of such

algorithms over the uncertain parameter space.

Now parameter uncertainty analysis consists mainly of estimating functions of the integral of the algorithm, e.g. the expectation. Thus, it was considered worthwhile examining these methods to determine if they had the potential to be adapted and used for the purposes of uncertainty analysis.

In the first chapter, an introduction to the field of radiation protection and an overview of the development of complex internal dosimetry algorithms were provided. The reason that it is necessary to perform uncertainty analyses on these algorithms is also discussed.

In chapter two the general concepts of uncertainty analysis were introduced and the classical methodology, including the selection of a set of design points was examined. Details were given of the use of stochastic process models, in both a Bayesian and classical framework, to predict individual values of the output of computationally expensive algorithms for specified inputs. At the end of this chapter the aims for the investigation of these methods were listed.

The first aim of the project was

‘to develop a Bayesian approach to parameter uncertainty analysis based on stochastic process models that improves on the currently available classical methods in three ways; firstly, to obtain more accurate estimates of the mean and variance of the algorithms output using fewer evaluations of the algorithm. Further, to develop extra measures of parameter uncertainty not available in the classical analysis, and to define an efficient design selection criterion’.

Classical methodology provides three main measures of parameter uncertainty. These are an estimate of the expected output, K , an estimate of the variance of this expectation, and an estimate of the variance of the output, L . All these measures are based on an analysis of a sample of the algorithm’s output. In addition, the sample mean is usually assumed to be normally distributed, according to the central limit theorem.

In chapter three, the most general form of the alternative Bayesian methodology, based on stochastic process models, was described. Here,

the Bayesian estimate of the mean, K was shown to take the form of a t distribution. The mean and variance of this distribution were considered to constitute the Bayesian measures of K comparable to the classical sample mean and variance of the sample mean.

The distribution representing the Bayesian estimate for L was also shown to be difficult to derive in a closed form. However, the mean and variance of this distribution were obtainable. Thus in contrast to the classical methodology where only a point estimate of the variance can be constructed, this methodology also provides a prediction for the accuracy of the point estimate of L .

A further, potentially useful measure, not available from the classical analysis was also identified. Part of the process of developing the measures for K and L involved defining a distribution to predict the value of an algorithm for selected individual values of the uncertain parameters. It was shown that this distribution could be used to obtain the probability that, conditional on the selected values of the uncertain parameters, the true value of the algorithm exceeds some predefined value.

Consider, for example, a radiation worker who receives an unplanned internal exposure. It would be usual to obtain an estimate of the dose received, in order to determine if the worker is likely to exceed a statutory dose limit. An algorithm would be used with its uncertain parameters set to their 'best estimate' values to calculate a measure of the dose received, usually in the form of a CEDE. This would then be compared to the dose limit. If the predicted dose were slightly under the prescribed limit, then it is plausible that the predicted dose based on a slightly different set of parameter estimates may have exceeded the limit. Using the methodology above it would be possible to obtain a probability that the true dose received by the worker exceeded the prescribed limit using the 'best estimates' of the parameters. This probability measure would take into account the uncertainty on the best estimates.

For both the Bayesian and classical methodologies the data on which all the measures of uncertainty are based come from a set of outputs from the algorithm, obtained at specific points on the space of uncertain para-

meters, referred to as the design set. The Bayesian methodology could be implemented with samples obtained for use with the classical methodology. i.e. via Monte Carlo sampling. However, one advantage of the stochastic process methodology is that it does not require the design to be selected at random. Rather, it can be selected to maximise the amount of information obtained. In chapter three, a criterion for the selection of an optimum design set, of a predetermined size was developed. At the time the design is selected the information available about the algorithm is likely to be very limited. Thus, the derived criterion is weighted to select a design that best estimates the algorithm close to the expected values of the uncertain parameters, and positions the points according to the smoothness of the algorithm.

A big incentive for choosing designs based on such a criterion is that if more information is gained from each element in the design set then fewer points will be needed in total, to obtain estimates with the same degree of predicted accuracy, as compared to the number required if the design is selected at random. Thus, even if the Bayesian analysis does not produce more accurate estimates of uncertainty for the same amount of information then, because the design can be selected more efficiently than for the classical analysis, the actual number of points required in the design set will still be reduced.

One of the major problems with the classical uncertainty analysis is that the size of the design set required to obtain a reliable result can be unrealistically large for algorithms that are computationally expensive to evaluate or which have many uncertain parameters. A large reduction in the number of evaluations required could therefore justify the use of the Bayesian methodology in itself.

In conclusion, chapter three satisfied the first aim of the project by demonstrating that in general theoretical terms it is possible to use stochastic processes, in a Bayesian framework, to obtain measures of parameter uncertainty comparable to those obtained via classical forms of analysis. It also demonstrated the potential to obtain these and other estimates using fewer evaluations of the algorithm.

The second aim of the project was

‘to objectively demonstrate any improvement in the Bayesian methodology over the classical methodology’.

To fulfil this aim, a simple computationally inexpensive algorithm with two uncertain parameters for quantifying radiation doses due to internal contamination with radioactive iodine was selected. Owing to the computational cheapness of evaluating this algorithm, independent estimates of the uncertainty generated on its output by the two uncertain parameters could be obtained with which to objectively evaluate and compare the performance of the classical and Bayesian methodologies.

To apply the stochastic process methodology to this test problem, it was necessary to rework the general theory detailed in chapter three specifically to define the uncertainty measures for an algorithm with normally distributed uncertain parameters and in which the correlation function describing the smoothness of the algorithm was of an exponential form. Chapter four detailed these definitions.

In chapter five, the ‘true’ uncertainty in the algorithm’s output induced by the two uncertain parameters was calculated along with both the classical and Bayesian uncertainty measures. A number of observations can be drawn from these analyses.

The Bayesian methodology required definition of $\mathbf{h}(\mathbf{x})$ and the smoothing parameter values whereas the classical methodology required no further assumptions. However, in this example, the Bayesian analysis generated far more accurate measures of the uncertainty on the algorithm. Had the design set that was used in the original classical analysis in 1988 been available, it might have been possible to re-run the Bayesian analysis using this design set (provided the required matrix inversions were possible). A comparison of accuracy of the classical and Bayesian based measures would then have considered solely comparing the efficiency of the methodology and would have been independent of the way in which the design was selected. However, a Bayesian analysis using a 10 point design was found to be more accurate than the classical analysis based on 1000 points. It seems unlikely that the information provided by 1000 points selected at

random, from the two dimensional parameter space, would be less than that provided by 10 points selected using the optimisation criterion. This suggests that the improved accuracy of the Bayesian analyses was due not only to the use of more informative data but also to a greater efficiency in the methodology.

This means that regardless of the method by which the design is selected it would be better to use Bayesian rather than classical uncertainty analysis methodology.

Now the problem of defining the parameters of the Bayesian analysis, i.e. the function vector $\mathbf{h}(\mathbf{x})$ and the smoothing parameter values, was minimal for this simple example. The function $\mathbf{h}(\mathbf{x})$ was taken to be a linear function in $\log w$ and $\log f$, as a default and for computational ease. In this example, it was possible to use the data from the calculation of the true values of the mean and variance of the algorithm, to obtain a picture that gives an idea of the gross relationship between w , f and the output of the algorithm, Figure 7.1.

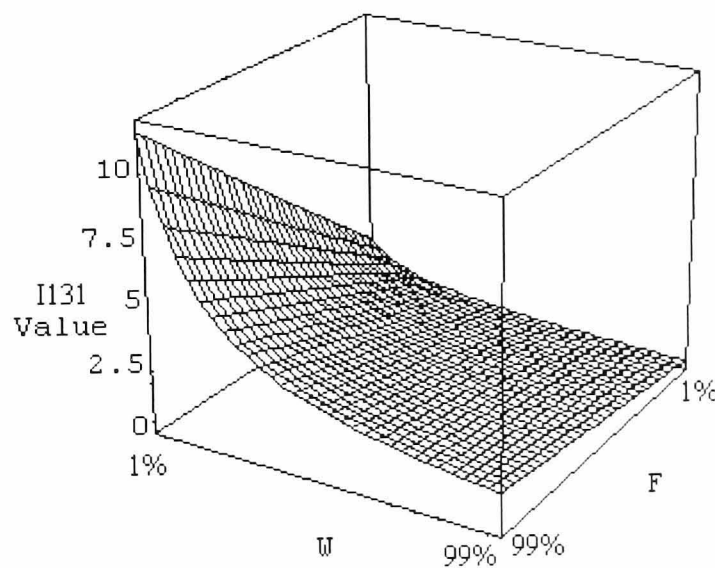


Figure 7.1: Evaluation of the ^{131}I algorithm, extending to approximately 2.5 standard deviations either side of the mean values

From this figure it can be seen that the relationship between the algorithm's output and w is approximately exponential, while for f it is

approximately linear. Unfortunately, the assumption of independence between the two parameters in defining $\mathbf{h}(\mathbf{x})$ is clearly not correct. However, the Bayesian analysis, using $\mathbf{h}(\mathbf{x})$ as defined, was able to produce far more accurate measures of the uncertainty despite the inaccuracies in the definition of the components of $\mathbf{h}(\mathbf{x})$. The ability of the Bayesian methodology to provide accurate measures despite the mis-specification of the relationship between the uncertain parameters and the algorithm's output is important. In more complex problems, it will be rare for relevant objective information to be available when the function $\mathbf{h}(\mathbf{x})$ is selected. Further, the integrals that must be evaluated to obtain the uncertainty measures will only be analytically possible for certain forms of $\mathbf{h}(\mathbf{x})$.

One of the ways in which the performance of the Bayesian analysis was measured was to examine the errors in the estimates of individual values of the algorithm's output on a regular grid over the parameter space. The maximum, minimum and mean percentage errors for the various Bayesian analyses performed are given in Table 5.5. Another way in which these errors can be usefully viewed is as a surface. The percentage errors for the design y_{15} with the default smoothing parameter values have been used to generate the surface in Figure 7.2.

This figure clearly shows that the Bayesian methodology best estimates the value of the function close to the prior mean value for each of the uncertain parameters. This was to be expected since the criterion for the selection of the best design was constructed so that the Bayesian point estimates of the algorithm would be most accurate for the expected value of each uncertain parameter.

If the same surface is generated using the 20 point design and with the improved estimates of the smoothing parameters then the Figure 7.3 is obtained (using the same plotting axes).

Comparing this figure with the previous one it can be seen that the area over which the estimates are most accurate has enlarged and also that the magnitude of the maximum and minimum errors are much smaller. Thus these figures support the conclusion that the Bayesian estimates become more accurate with both an increase in the number of points in the design

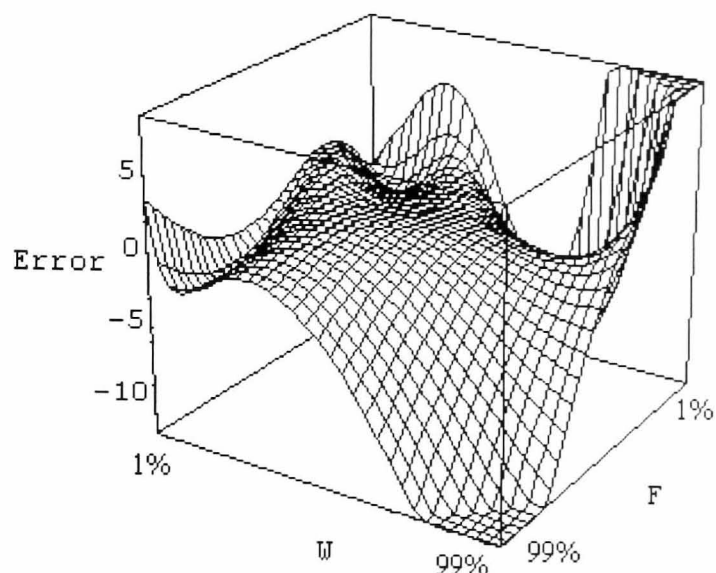


Figure 7.2: Surface representation of the percentage errors in the point estimates generated by the Bayesian analysis using design y_{15} .

and also with improved estimates for the smoothing parameters.

In fact, the magnitude of all the measures of percentage error decreased by at least an order of magnitude between the initial ten point design using the default smoothing parameter values and the twenty point design using the improved estimates. These improvements parallel the changes in the values of the variance of the expected output and the variance of the variance of the output across these designs. Hence, in ‘real-life’ examples it might be possible to get an idea of the methodologies’ accuracy at estimating the mean and variance of the algorithm’s output by determining the size of the errors in point predictions. At the end of chapter five a contour plot was provided which illustrated the use of the probability measure described above. This clearly shows the potential of this measure for determining the effect of the uncertain parameters on the output of the algorithm.

In conclusion, the comparison of the Bayesian and classical methodologies to the true uncertainties associated with the iodine algorithm have satisfied the second aim of the project. It has been objectively shown that, in a simple low-dimensional example, the Bayesian methodology out

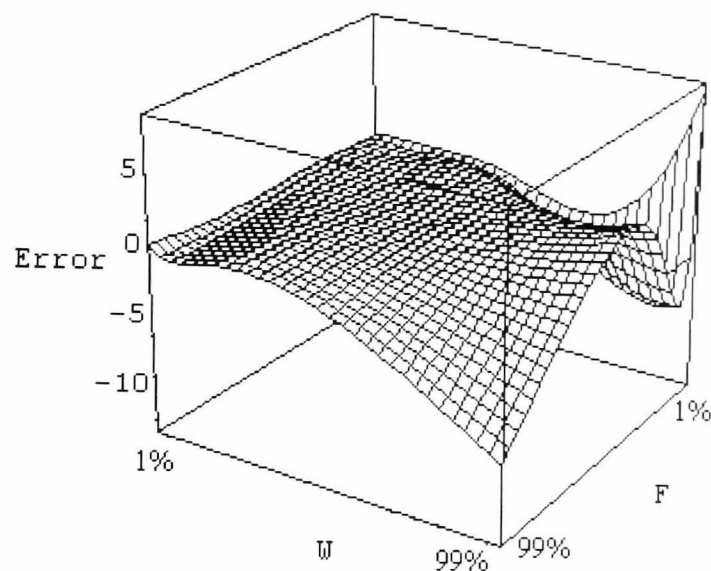


Figure 7.3: Surface representation of the percentage errors in the point estimates generated by the Bayesian analysis using design y_{20} .

performs the classical approach in its ability to produce accurate and useful measures of parameter uncertainty using fewer, but more informative, evaluations of the algorithm.

Finally, the third aim of the project was

‘to demonstrate that the Bayesian methodology is useable in a ‘real-life’ uncertainty analysis’.

The Pu239 algorithm was selected for this ‘real-life’ analysis. It was ideal for a number of reasons. Firstly, it was a large computationally expensive model with 29 parameters. Secondly, an investigation of its uncertainties was already being performed at NRPB at the time. In this analysis 14 of the 29 parameters were assumed to be uncertain. Thus, this model provided a realistic test for the Bayesian methodology.

In order to provide the closest comparison with the classical analysis the Bayesian analysis was performed assuming the same 14 parameters as being uncertain. So that the theory detailed in chapter four could be applied, all the uncertain parameters were transformed to have normal prior distributions. The function $\mathbf{h}(\mathbf{x})$ was defined as a linear function of the transformed uncertain parameters, a choice which even though incorrect

had still produced acceptable results for the iodine algorithm. Of course, for the Pu239 model it is not possible to determine if the assumed form of $h(\mathbf{x})$ was good or bad.

The construction of the optimum design sets proved more complicated, compared to the iodine example. This was due mainly to the increased number of dimensions of the problem. Choosing default smoothing parameter values for this process was accomplished, as discussed in 6.5.2, by defining independent lognormal prior distributions and then selecting either the mean or median values from them. An examination of the results revealed no consistent difference in the measures generated using either the mean or median values as the prior estimates.

A potential problem was identified when the smoothing parameters were updated using the same method as for the iodine algorithm, i.e. selecting the mode of the full posterior distribution, (4.15) assuming independent uniform prior distributions for the smoothing parameters. Some of the values obtained were either very large or very close to zero. This caused some concern, since such extreme values did not seem plausible. Therefore the optimisation process was repeated assuming the independent lognormal priors used to obtain the initial estimates, instead of the uniform priors. The use of these lognormal priors deterred the posterior estimates from taking values very large or small in comparison to the prior estimates. Examination of the results of the analyses shows little difference in the results based on the different improved smoothing parameter estimates.

The only major difference between the results of the various Bayesian analyses was in the estimates of the variance of the distributions for K and L . As expected the size of the variances decreased as the number of points increased and with the use of either type of improved smoothing parameter estimate. Unfortunately, without estimates of the ‘true’ values for K and L it is not possible to determine if 95% probability intervals generated around the Bayesian estimates of K and L would contain the true values.

The consistency checks performed on the Bayesian analyses revealed

that both increasing the number of points in the design and the use of improved smoothing parameter estimates increased the accuracy of the point predictions. As with the iodine example, the use of improved smoothing parameter estimates increased the accuracy of the estimates more than the addition of extra points, in some cases. This indicates that for future analyses it would be a more efficient use of resources to spend less time selecting the design by reducing the number of points in it and more time optimising the smoothing parameter estimates. For the iodine example an increase in the accuracy of the point predictions was associated with an increase in the accuracy of the estimate of K and L . Thus, the reductions in the variance of the K in this example could be taken to indicate an improvement in the performance of the estimation procedure.

Another way of examining the results of the Bayesian analysis was included in the software. This involved plotting the predicted uncertainty in the algorithm's output as a function of one uncertain parameter but taking the expectation of the output with respect to the uncertainty in the other parameters into account. Figure 7.4 shows this plot for the parameter 'Loss from testes' constructed using the 50 point sample and the initial smoothing parameters, Prior8. The error bounds show that the most confident prediction of the algorithms output is around the mean of the uncertain parameter, 0.00034, and that as the value of the uncertain parameter moves away from this mean value the estimate of the output becomes more uncertain.

Figure 7.5 shows the same plot but constructed using the 150 point design, selected using the Prior8 smoothing parameters. The 'Improved : Lognormal' smoothing parameters were used instead for actual construction of the plot. It can be seen that the use of the larger design set and improved smoothing parameters has resulted in reduced uncertainty in the relationship. It should be noted that it would have been equally possible to have fixed the values of the other uncertain parameters instead of averaging over them to obtain a similar plot. Both of these measures could be useful for visualising parameter uncertainty. Neither is available from a classical analysis.

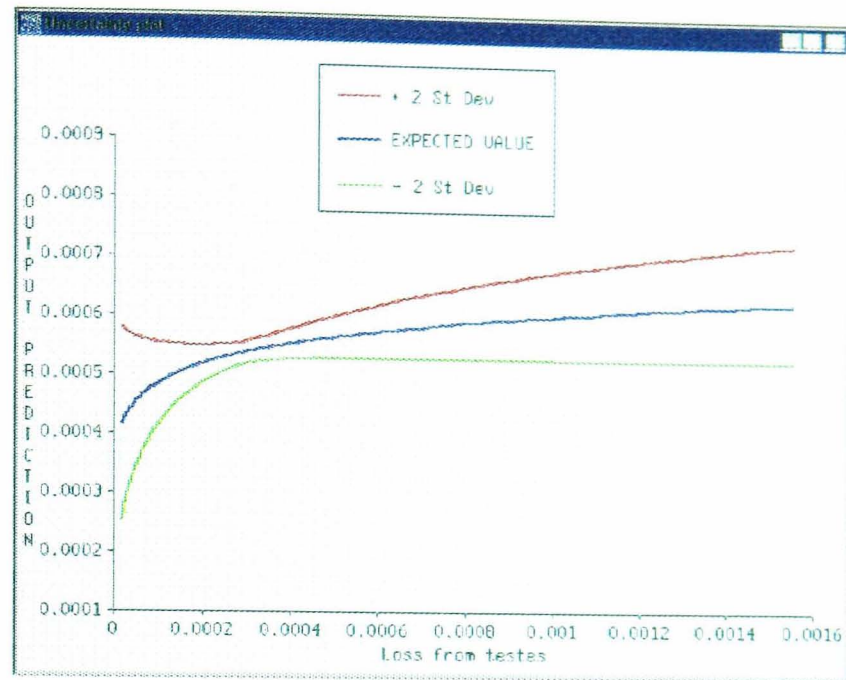


Figure 7.4: Uncertainty plot for the parameter ‘Loss from testes’ based on a 50 point design.

Thus in conclusion it can be seen that the Bayesian uncertainty analysis of this algorithm provides measures of the parameter uncertainty that are comparable to those of the classical analysis. Most importantly for the use of the Bayesian methodology, these results were achieved using design sets that contained at most one-third of the evaluations used for the classical analysis.

Overall, the Bayesian methodology has a number of advantages and disadvantages compared to the classical methodology. The main advantage is its ability to produce more accurate and informative measures of uncertainty. It can make use of any evaluations of an algorithm previously obtained whereas the classical analysis requires a set selected at random. This is a useful property since it means that no evaluations of a computationally expensive algorithms would be wasted in an uncertainty analysis. Further, if a design has already been selected at random then it has been shown that the Bayesian methodology will extract more information and produce more accurate uncertainty measures than would a classical analysis. The full potential of the Bayesian methodology is only realised, however, if a new design set is selected based on a criterion to optimise the

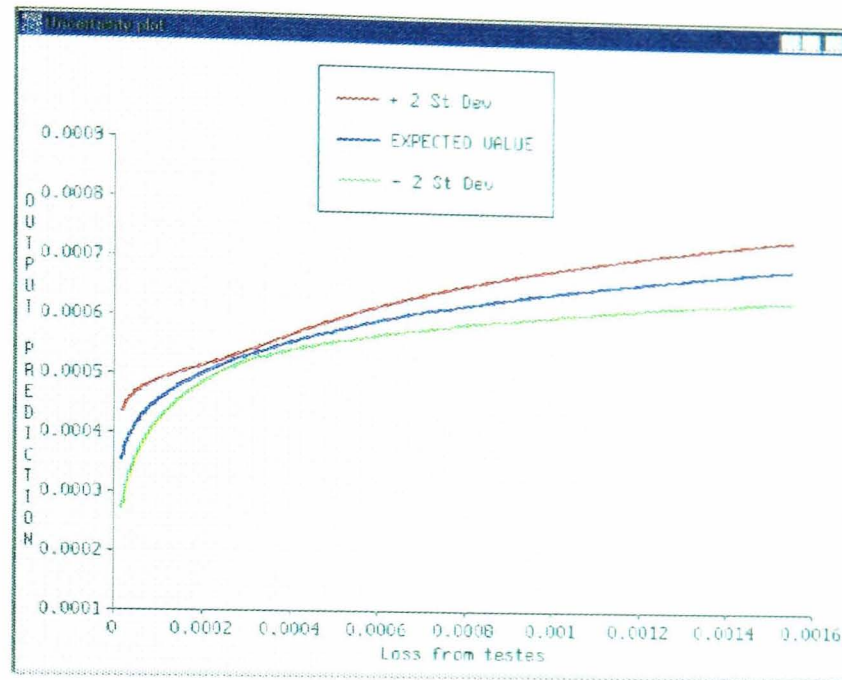


Figure 7.5: Uncertainty plot for the parameter ‘Loss from testes’ based on a 150 point design.

information provided by the evaluations.

The major disadvantage of the Bayesian methodology is its complexity. The classical methodology requires no assumptions to be made regarding the algorithm, whereas the Bayesian methodology requires the specification of the function vector $\mathbf{h}(\mathbf{x})$, the selection of a correlation function and smoothing parameters. The analyses described above, however, show that the Bayesian analysis is more powerful than the classical. Thus, the extra complication of the Bayesian calculations is likely to result in more accurate results. If some of the assumptions are avoided by, for instance, using a randomly selected design set then the Bayesian methodology is still likely to be more accurate due to its more efficient use of the data.

The other major disadvantage of the Bayesian methodology is that no software is currently available to implement it, except that written for this project. This would be a large deterrent to its use. The programs detailed in the appendix were, however, constructed to show that it is possible to write fairly user-friendly general software for the implementation of the Bayesian methodology.

In conclusion the three aims of this project have been fulfilled. The

Bayesian methodology can produce estimates of an algorithm's parameter uncertainty equivalent to those given by the classical methodology and other additional measures. It has been demonstrated objectively to outperform the classical methodology in a simple example and has been found to perform well in a 'real-life' problem. Hence, as the assessment of parameter uncertainties is becoming more important, for example in the field of radiation protection, the development of Bayesian uncertainty methodology has the potential to be very worthwhile.

7.1. Future work

There are a large number of areas of Bayesian uncertainty analysis that could benefit from further investigation. For example, it would be desirable to determine other functions that could be used to define the estimating function, $\mathbf{h}(\mathbf{x})$, and to determine what effect the use of different functions has on the results of the uncertainty analysis. Also, it would be useful to develop other measures of uncertainty. For instance, the expectation and the variance of the algorithm are estimated currently. However, it would also be useful to be able to estimate the actual distribution of the algorithm's output, not just summary measures from it.

A third area that could benefit from extra research is the selection of the design points. This is the most computationally expensive part of the analysis. Consequently, testing different selection criteria to find the best would be useful, as would be the development of efficient methods of searching for the 'best' design.

Thus, there are many areas that could be investigated to extend and improve this methodology.

8. APPENDIX : THE DESIGN SELECTION AND ANALYSIS PROGRAMS

8.1. Introduction

The analysis of the iodine model was performed using two programs written in the mathematical programming language APL. The software was written to analyse this specific problem and was sufficient to demonstrate that the basic theory of Bayesian uncertainty analysis could be put into practice. Unfortunately, these programs were not suitable for adapting to solve other uncertainty analysis problems.

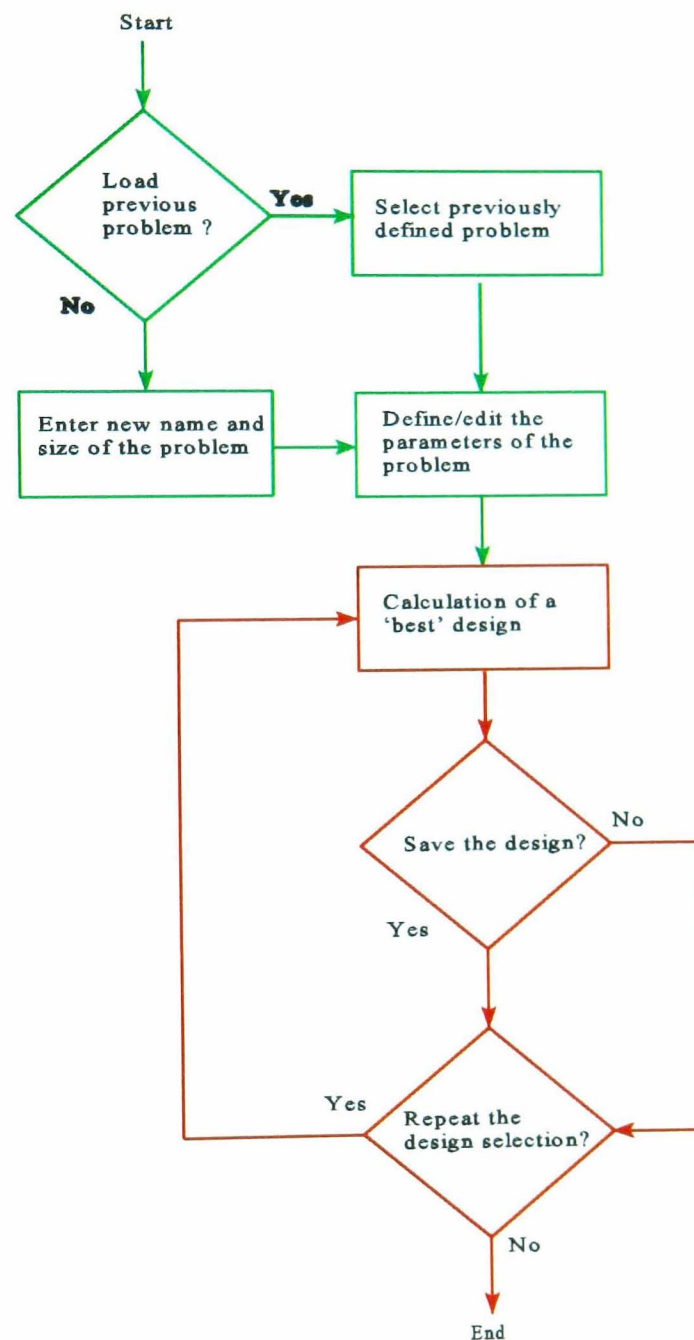
An important part of this project is to demonstrate not only the theoretical plausibility of Bayesian uncertainty analysis but also to show that the theory can be put into practice; in particular that the calculations required to obtain the measures of uncertainty can be implemented, in a reasonable time scale, using today's computer technology. With these aims in mind new software was written to perform the analysis of the plutonium model. The software is divided into two programs :

- 1) a program to select the designs,
- 2) a program to perform the analyses, including the updating of the smoothing parameters.

Both these programs were written to be able to solve a general problem (with one small exception in the analysis program), that is, where both the number of uncertain parameters and their prior distributions are specified by the user. The programs were written in the Windows version of the APL language. This enabled a user-friendly graphical interface to be incorporated.

8.2. The Design selection program

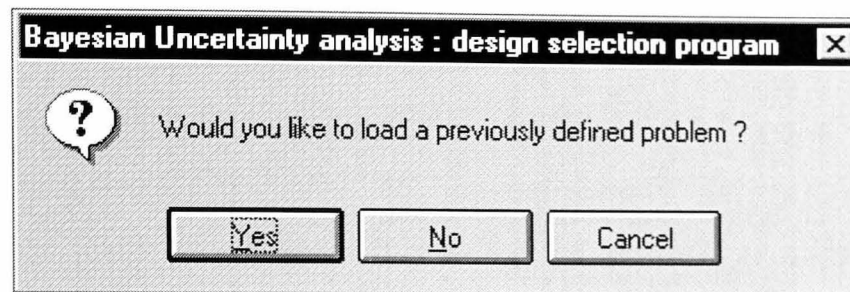
The object of this program is to enable the user to define the parameters of the uncertainty analysis problem and to select sets of 'best' design points. A flow chart of this program is detailed in Figure 8.1



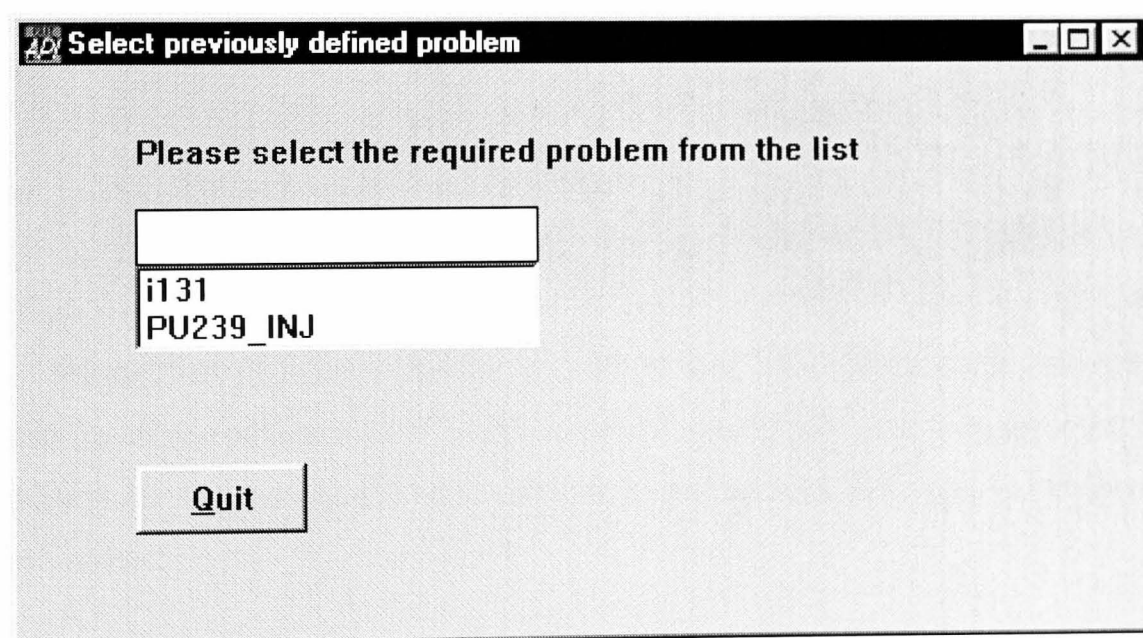
The program is comprised of two major parts; the problem specification component, where each uncertain parameter is defined along with its prior distribution (these are detailed in green on the flowchart); and the design selection part, outlined in red. The two sections of the program will now be considered in more detail.

8.2.1. Reloading an old problem or defining a new problem

The figure below details the first form of the program. It gives the user the opportunity to specify whether a previously defined problem is to be reloaded or if a new problem is to be started.



If the 'Yes' button is pressed then a new form is rendered labelled 'Select previously defined problem'. This provides the means for the user to select a previously defined problem from those listed. The figure below shows two currently defined problems.



If the 'No' button is pressed then two forms are provided for the user to enter a new name and to specify the number of uncertain parameters in the new problem.

A screenshot of a Windows-style dialog box. The title bar is black with white text that reads 'Specify a new title for the problem'. Inside the dialog, the text 'Please enter the name of the new problem' is centered. Below this text is a single-line text input field. At the bottom of the dialog, there are two buttons: 'OK' and 'Quit', both with underlined letters.

A screenshot of a Windows-style dialog box. The title bar is black with white text that reads 'New problem : Test1'. Inside the dialog, the text 'Enter the number of uncertain parameters' is centered. Below this text is a text input field containing the number '5'. At the bottom of the dialog, there are two buttons: 'OK' and 'Quit', both with underlined letters.

Regardless of which of the above options was selected the program then provides a form on to which the components of a new design can be defined or the parameters of a previously defined problem can be reviewed and edited.

8.2.2. Defining/editing the parameters of the problem

The form detailed in Figure 8.2 enables the user to specify the various components of a new problem or to edit and change those of a previously defined problem. The form is initialised with the details of the first uncertain parameter of the problem displayed. If a new problem has been specified then a default selection is displayed for each of the options, except the parameter name which is left blank.

Define the parameters of new problem PU239_INJ

Choose parameter for editing: 6

Parameter name: st1->UBC / st1->blood

Uncertainty distribution for this dimension: Lognormal

Mean: 0.4870000000

Variance: 0.0963000000

Grid range: 95

Grid spacing: 1000

Smoothing Value: . 17.81

Buttons: OK, Quit, Print, Load SM

Figure 8.2: The form provided to define or edit the parameters of a problem.

To change the details of a parameter, other than the first, the box labelled ‘Choose a parameter for editing’ (which shows the value six in Figure 8.2) should be selected and the required parameter number chosen from the list to reveal the current details of the selected parameter. Alternative values can then be specified.

The prior distribution for a new uncertain parameter is set to normal, mean equal to zero and variance set to one as a default. This default prior distribution can be changed to one of lognormal, beta or gamma. If a beta distribution is selected then two extra edit boxes are provided to specify the bounds of the distribution.

The program has an option to search for a ‘best’ design starting from one selected at random. The ‘grid range’ and ‘grid spacing’ values are used to define the initial values for each point in the random design. For each uncertain parameter the required number of points are selected from a set

of regularly spaced points. The number of points in this grid is defined by the value of 'grid spacing'; the smallest and largest values are defined as the bounds of a probability interval on the prior distribution with width defined by the 'grid range' value.

The figure entered in the smoothing value box defines the value of the smoothing parameter for the currently selected uncertain parameter. The default values for these components are 1000 for the grid spacing, 95% for the grid range, and one for the smoothing value.

At the bottom of the screen four buttons are placed: 'OK' causes the current information to be saved and the program to move on to the calculation of a best design; the 'Quit' button exits the program; the 'Print' button causes the current parameter definitions to be printed to hard copy; and the 'Load SM' button allows a previously saved set of smoothing parameter values to be loaded from an ASCII file using a standard 'Windows' dialogue box.

8.2.3. Selecting a design

The form used to define and monitor the selection of a design is illustrated in Figure 8.3.

Define the parameters of the calculation of a best design for problem PU239_INJ

Set the search start location

Set the design size

Calculation type Initial step size

Set the number of repetitions for the algorithm

Current value of the criterion Step size %

Number of iterations performed Hit rate %

Iteration on which criterion last changed

Figure 8.3: The form used to define the parameters of the search for a best design.

The elements in the top half of the form are used to define the parameters of the ‘best’ design search, while those in the bottom half are provided to enable the progress of the search to be monitored.

The search specification components

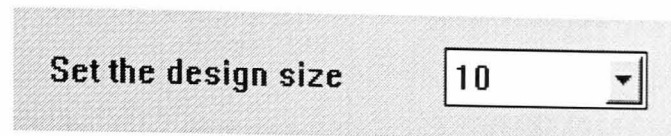
The components used to define the search are as follows :

Set the search start location

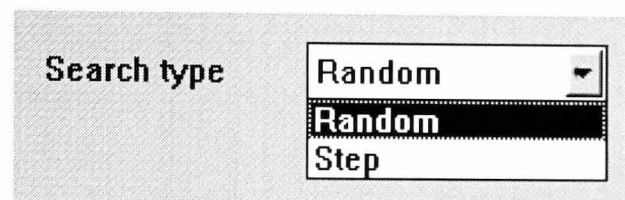
- Random
- Random
- Previous

This option defines the initial design. If ‘Previous’ is selected then a further form is displayed to enter the name of the previously saved design.

If the option ‘Random’ is selected then an initial design is selected at random using the grid range, grid spacing and smoothing value parameters entered on the ‘Define/edit the parameters of the problem’ form.

A screenshot of a user interface element. It consists of a light gray rectangular box. On the left side of the box, the text "Set the design size" is displayed in a bold, black, sans-serif font. To the right of this text is a white rectangular input field with a black border. Inside the input field, the number "10" is displayed. To the right of the input field is a small black downward-pointing arrow, indicating a dropdown menu.

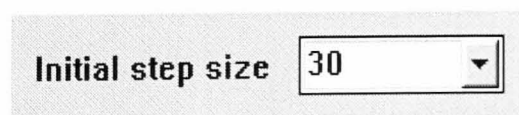
This box enables the size of the design to be defined if an initial random design has been specified or it displays the size of the design if a previously created design has been selected.

A screenshot of a user interface element. It consists of a light gray rectangular box. On the left side of the box, the text "Search type" is displayed in a bold, black, sans-serif font. To the right of this text is a white rectangular dropdown menu with a black border. The menu is currently open, showing three options: "Random" (the top option, which is highlighted with a black background and white text), "Random" (the middle option, in black text), and "Step" (the bottom option, in black text). A small black downward-pointing arrow is visible on the right side of the menu.

This box defines which of the two types of search routine will be used:

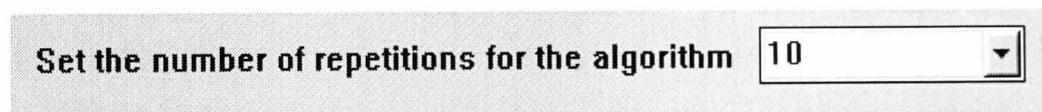
Random: This algorithm is defined as follows. The value of the criterion is calculated for the initial design. Then a new point is selected at random and the criterion is recalculated with this random point substituted in turn for each of the current points in the design. If any of the criterion values for the adjusted designs is greater than that for the original design, then the design with the largest associated criterion value is retained and the original design is discarded. If none of the adjusted designs results in an improvement in the criterion value then the original design is retained. A new random point is then selected and the process repeated.

Step: If the ‘step’ search algorithm is selected then an extra form is revealed for the user to indicate the initial step size.

A screenshot of a user interface element. It consists of a light gray rectangular box. On the left side of the box, the text "Initial step size" is displayed in a bold, black, sans-serif font. To the right of this text is a white rectangular input field with a black border. Inside the input field, the number "30" is displayed. To the right of the input field is a small black downward-pointing arrow, indicating a dropdown menu.

The ‘step’ algorithm is defined as follows. As for the random search, the value of the criterion is first calculated for the initial design. Next, the position of the first point in the design is changed. A random number of steps is selected using a predefined Poisson distribution and then assigned

randomly to be either in the positive or negative direction. The size of each step is then derived by taking the magnitude of each of the point's coordinates and calculating a percentage of these values defined by the initial 'step size' value. Each of the coordinates of the point is then moved the selected number of steps in the chosen direction along each axis. The value of the criterion is then derived again for this new design. If its value has increased then the new coordinates are kept for the point, else the original coordinates are retained. The second point in the design is then moved the selected number of steps and the value of the criterion reassessed. This process is repeated on the remaining points of the design re-using the selected number and direction of the steps. Having run through each of the points, one iteration of the algorithm has been completed. The entire process is then repeated for the whole design after first selecting a new random number of steps and step direction. If, at any time during the optimisation procedure, two iterations are completed without any changes being made to the design then the 'step size' parameter is reduced by 10% of its current value.



Set the number of repetitions for the algorithm

This form provided the means of selecting the number of iterations that the chosen algorithm will make. A value can be selected from a predefined list or a user-defined figure may be entered.

The calculation monitoring components

The number of parameters used to monitor the search for a best design depends on which algorithm has been selected to perform the calculations. The following figure illustrates those components that are common to both. Once the final parameter of the calculation has been defined the program will automatically calculate the value of the criterion for the initial design whether this is a previously defined design or a new random selection. This value is then displayed in the box labelled 'Current value of the criterion' and the number of iterations performed is displayed as '0'. A button labelled 'start' is then added for the user to initiate the rest of the calculation.

The box displaying the iteration on which the criterion was last changed will always be blank initially. It is provided to enable the user to assess the value of performing further iterations using the selected algorithm.

Current value of the criterion	<input type="text" value="0.82623454"/>
Number of iterations performed	<input type="text" value="0"/>
Iteration on which criterion last changed	<input type="text"/>

For calculations in which the step algorithm is used two extra measures are provided to enable the user to more effectively monitor the calculations.

Step size %	<input type="text" value="70"/>
Hit rate %	<input type="text" value="100"/>

The first box details the current step size and the second gives the hit rate. This represents, in percentage terms, the number of points of the design that are moved in one iteration of the algorithm.

Once the calculations are initiated by the user pressing the 'start' button, this push button is replaced by a radio button.

☐ **Stop iterations**

This enables the user to interrupt the calculation if further iterations are not likely to improve the design or if further calculations are required using a different selection algorithm.

If this button is pressed or if the full number of iterations have been performed then the user is asked to save the final design both as an internal variable and as an external, space delimited, ASCII text file or to discard it. Finally, the user is asked to choose between performing another calculation or to exit the program.

8.3. The Analysis program

The objective of this program is to perform a Bayesian uncertainty analysis. The program is designed to work in conjunction with the design selection program. A flow chart is detailed in Figure 8.4. In order to avoid the necessity of putting all the information about an uncertainty analysis problem in an ASCII file to transfer it between the two programs, the analysis program has been designed to read the relevant information directly from the workspace containing the design program.

The analysis program needs three sets of information:

- 1) the parameters of the problem, i.e. the uncertain parameters and their prior distributions. All the details of a problem will be saved under one user specified name in the design program's workspace.
- 2) the grid of coordinates representing the positions of a set of design points.
- 3) the values of the algorithm at the associated design points.

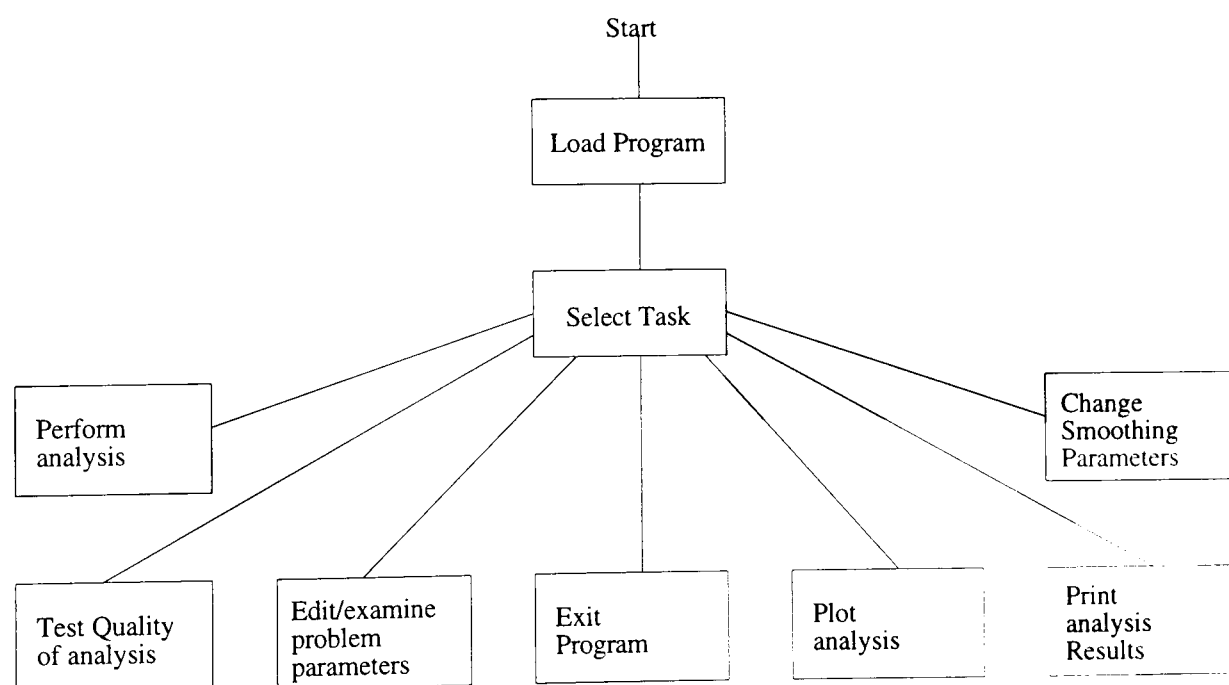
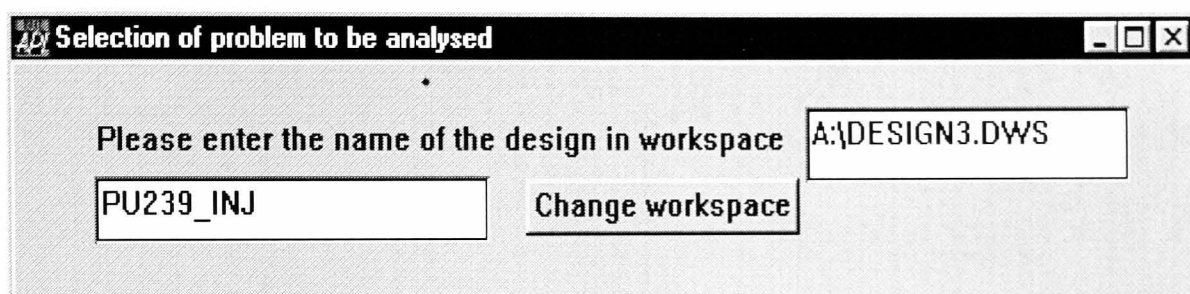


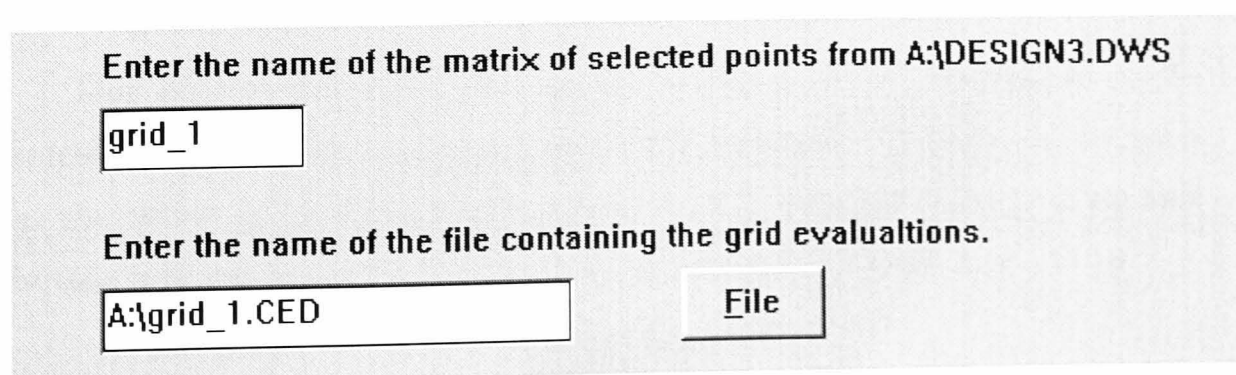
Figure 8.4: A flow chart of the analysis program.

8.3.1. Loading the details of the problem

The user enters the information required for the analysis using a single form. First the file containing the workspace for the design program, must be defined. Then the name of the variable containing the details of the selected problem must be entered, as illustrated below. The button labelled 'Change workspace' is provided to enable the user to select a workspace file, using a Windows dialogue box as opposed to entering the file name directly in the space provided.



Once these details are entered the program will perform consistency checks to determine if the user has entered a valid file name for the workspace and if the named variable exists. If an inconsistency is detected then an error message is displayed and the user is requested to re-enter the information. Once the problem's details have been successfully loaded the user is requested to enter the variable name defining a set of design points (created via the first program), and then the file containing the evaluations of the algorithm at these design points, see the illustration below. This file of outputs from the algorithm must be in ASCII format and contain only a single column of values.



Having obtained all the data required for the analysis of the problem the program moves on to the results form.

8.3.2. Analysis of the selected problem

The results form is shown in the following figure.

Analysis of problem PU239_INJ using design points grid_6

Point estimate for the expected value

Variance of the expected value

A 95 % probability range for the expected value is

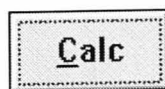
Point estimate for the variance

Variance of the variance

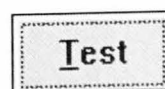
A 95 % probability range for the variance is

Calc Test Details Quit Print SM change Plot

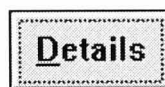
When this form is first opened no calculations have been made, so no results are presented. The row of buttons at the bottom of the form provides access to the various functions of the program.



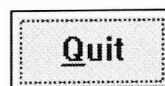
This button will cause the results of the uncertainty analysis to be calculated using the currently defined information. If information such as the values of the smoothing parameters are changed then pressing this button will cause the results to be re-calculated using the new values.



This button provides access to the analysis consistency checking facility of the program. In chapter six this checking procedure was introduced as a way of obtaining a measure of the accuracy of the Bayesian analysis in the absence of independent measures. Each point of the selected design is removed in turn (with replacement) and the remaining points are used to obtain a Bayesian point estimate of the true known value of the algorithm at the excluded point. Various summary statistics, including the maximum absolute percentage error and the average absolute percentage error are then calculated to determine the overall accuracy of the estimates. These measures were considered as substitutes for independent measures of the accuracy of the analysis. For the plutonium example, detailed in chapter six, an independent set of points, the Latin Hypercube sample, was available. An extra test procedure was implemented in the program, specifically for this problem. This used the complete design to evaluate Bayesian estimates of the output of the plutonium algorithm over the points of the Latin Hypercube sample. As in the standard test procedure summary measures of the accuracy of the estimates were calculated.



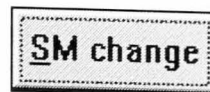
This button provides access to a screen similar to that used to define the uncertain parameters in the design program, Figure 8.2. In this program, however, the details are only provided for reference and cannot be changed.



This button will quit the program - all current results will be lost following this action.



This button enables access to the program's print facility which will provide the user with a hard copy of the current results and the details of the problem's uncertain parameters.



This button provides the user with a form, illustrated in Figure 8.5 that allows the values of the vector of smoothing parameters, δ , to be changed in various ways.

 A dialog box titled "Change the SM values" with a standard Windows-style title bar. It contains four rows of controls:

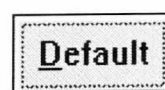
- Row 1: "Select the default SM values" label, a "Default" button, and an "Ok" button.
- Row 2: "Calculate optimum SM values" label, a "Calc" button, and an empty space.
- Row 3: "Load SM values from file" label, an empty text input field, and a "Load" button.
- Row 4: "Save current SM values to file" label, an empty text input field, and a "Save" button.

 Below these controls is a table with four columns: an index column, a "Name" column, a "Current" column, and an "Original" column. The table contains four rows of data.

	Name	Current	Original
1	Clearance from blood	46	46
2	Liver2 \rightarrow blood	30	30
3	Loss from st1	43	43
4	Loss from cort surf	15	15

Figure 8.5

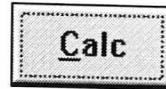
The grid object at the bottom of the form displays in four columns the number, name, current and original smoothing parameter values for each uncertain parameter in the current analysis. By placing the cursor over a particular cell in column 3 the user can change individual δ values at will.



This button will cause the current values for all the smoothing parameters in the problem to be reset to their original values, as defined in column four.

Alternatively, a previously saved set of δ parameters can be applied to the problem from an ASCII file by entering its name in the edit field next

to the button marked 'Load' and/or pressing this button (just pressing the button will cause a Windows dialogue box to specify the file to be opened). In a similar way the current set of δ parameters can be saved to a file by entering a name in the edit field next to the button marked 'Save' and/or pressing this button.



The 'Calc' button provides access to the form that controls the updating of the smoothing parameters using the design points.

8.3.3. Updating the smoothing parameters

The calculation of improved estimates of the smoothing parameter values is controlled from the form illustrated below. It is accessed by pressing the 'Calc' button on the screen illustrated in Figure 8.5.

Optimise the SM parameters

Maximum number of iterations is

The convergence tolerance is

Number of iterations performed

Optimisation criterion
☐ Uniform prior
☒ Lognormal prior

	Name	Initial	Best	Mean	Variance
1	Clearance from blood	24.78		3.21	1.25
2	Liver2 \rightarrow blood	16.28		2.79	1.25
3	Loss from st1	23.1		3.14	1.25
4	Loss from cort surf	8.499		2.14	1.25
5	Loss from trab surf	8.499		2.14	1.25


The first two columns of the grid at the bottom of this form are the same as those of the grid in Figure 8.5. The next column, labelled 'initial', contains the starting values for the search for the best smoothing values. These are set to the currently selected values when the form is opened but

can be changed at will by the user. The next column ‘Best’ will contain the results of the search when complete.

The last two columns are provided for entering the means and variances of the lognormal prior distributions that are used in the ‘lognormal’ optimisation criterion, see section 4.2. Again these means and variances can be entered either by hand or loaded from and saved to an ASCII file using the Load/Save buttons and fields above the grid.

The criterion to be used for the derivation of the improved smoothing parameter estimates is selected using the set of two radio buttons in the top right hand corner of the form. The two criteria are described in 4.2. Opposite these radio buttons are three edit fields that contain parameters of the optimisation calculations. The top value is the maximum number of iterations that the selected method will cycle through in order to obtain a best design. The second value is the convergence tolerance. The third value shows the current number of iterations performed.

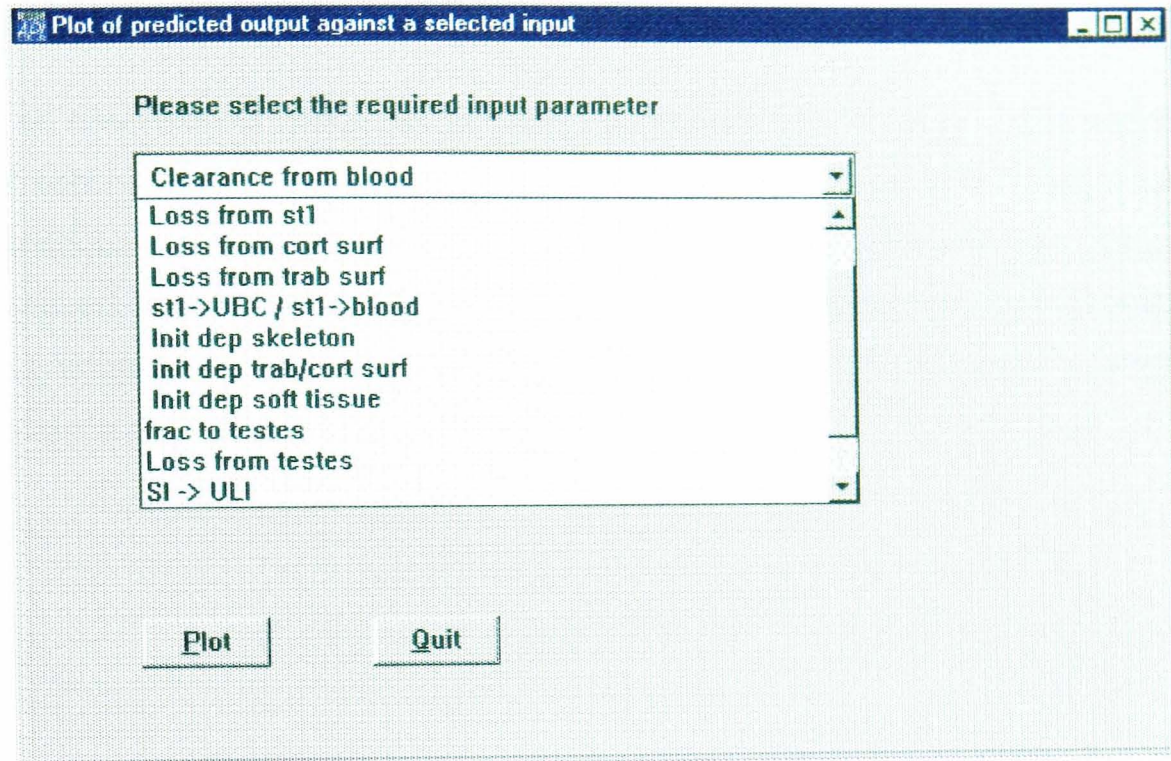
Once a selection procedure has been started the ‘Calc’ button is replaced by a radio button

 **Stop iterations**

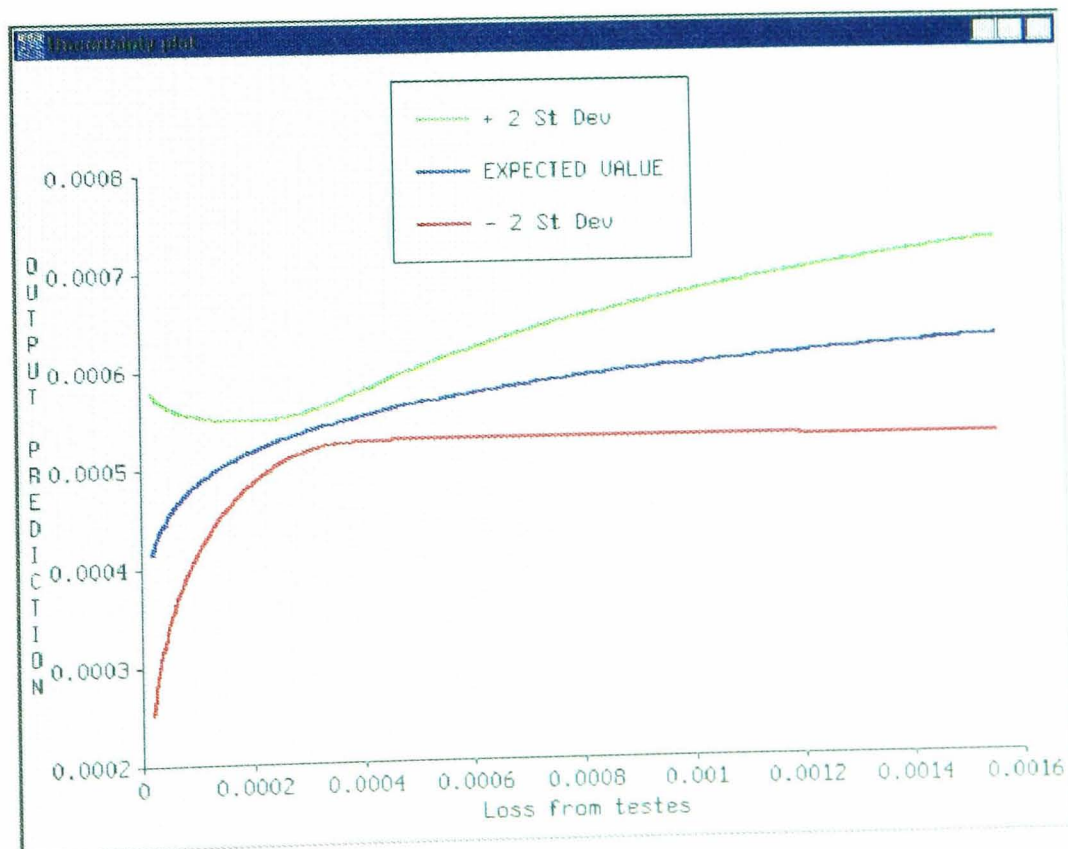
that when pressed will halt the calculations after the current iteration has been completed. The current ‘Best’ values will then be displayed in the grid. Upon exiting this form, the focus of the program returns to the form from which it was called. Before this happens the user is presented with a dialogue box to choose if the current δ values are to be replaced by the newly calculated ‘best’ values.

8.3.4. Sensitivity Plots

The sensitivity of the output of the algorithm to changes in the inputs can be assessed graphically via the ‘Plot’ button on the initial form of the program. This facility enables the user to produce a plot of the predicted output of the algorithm against one of the uncertain inputs of the algorithm but which is averaged over the others. The facility is accessed via the following window:



The user, having selected an input parameter to plot, presses the plot button. After a short wait for the calculations to be performed an annotated plot is obtained, as illustrated below.



8.4. Discussion

The descriptions above are intended to provide an overview of the scope of the software. The application of Bayesian methodology to statistical problems has long suffered from a lack of suitable software for its implementation. With these programs it is hoped that the potential for providing software in a user-friendly format for performing Bayesian uncertainty analyses has been shown.

BIBLIOGRAPHY

- [AAFH90] M Aitkin, D Anderson, B Francis, and J Hinde. *Statistical modelling in GLIM*. Oxford University Press, 1990.
- [AF88] N Adams and T P Fell. Recycling and metabolic models for internal dosimetry: with special reference to iodine. *Radiation Protection Dosimetry*, 22(3), 1988.
- [BD59] G E P Box and N Draper. A basis for the selection of a response surface design. *Journal of the American Statistical Association*, 54:622–654, 1959.
- [BSC93] K Bowman, J Sacks, and Yue-Fang Chang. Design and analysis of numerical experiments. *Journal of Atmospheric Sciences*, 50(9):1267–1278, 1993.
- [CHJH88] M J Crick, E Hofer, J A Jones, and S M Haywood. Uncertainty analysis of the foodchain and atmospheric dispersion modules of marc. Technical Report NRPB-R184, National Radiological Protection Board, 1988.
- [CMMY88] C Currin, T Mitchell, M Morris, and D Ylvisaker. A Bayesian approach to the design and analysis of computer experiments. Technical Report ORNL-6498, Oak Ridge National Laboratory, 1988.
- [CMMY91] C Currin, T Mitchell, M Morris, and D Ylvisaker. Bayesian prediction of deterministic functions with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*. 86(416):953–963. 1991.

- [Coo91] R Cooke. *Experts in Uncertainty*. Oxford University Press, 1991.
- [doc93] Estimates of late radiation risks to the uk population. Technical Report Doc. NRPB Vol 4, No 4., National Radiological Protection Board, 1993.
- [DSS81] D E Dunning, J R Schwarz, and G Schwarz. Variability of human thyroid characteristics and estimates of dose from ingested ^{131}I . *Health Physics*, 40, 1981.
- [Ham95] D Hamby. A comparison of sensitivity analysis techniques. *Health Physics*, 68(2), 1995.
- [HGMR91] J C Helton, J W Garner, R D McCurley, and D K Rudeen. Sensitivity analysis techniques and results for performance assessment at the waste isolation pilot plant. Technical Report SAND90-7103, Sandia National Laboratories Albuquerque, New Mexico, 1991.
- [HO] R G Haylock and A O'Hagan. Bayesian uncertainty analysis and radiological protection. In V Barnett, editor, *Statistics for the enviroment: 3*. In press.
- [HO96] R G Haylock and A O'Hagan. On inference for outputs of computationally expensive algorithms with uncertainty on the inputs. In J M Bernardo, J O Berger, A P Dawid, and A F M Smith, editors, *Proceedings of the 5th International Conference on Bayesian Statistics*, 1996.
- [HS93] M Handcock and M Stein. A Bayesian analysis of Kriging. *Technometrics*, 35(4):403–409, 1993.
- [ICR75] ICRP. *Report of the Task Group on Reference Man*. ICRP Publication 23. Pergamon Press, Oxford, 1975.
- [ICR89] ICRP. *Age-dependent doses to members of the public from intake of Radionuclides: Part 1*. ICRP Publication 56. Pergamon Press, Oxford, 1989.

- [ICR93] ICRP. *Age-dependent doses to members of the public from intake of Radionuclides: Part 2 Ingestion dose coefficients. ICRP Publication 67*. Pergamon Press, Oxford. 1993.
- [JMY90] M Johnson, L Moore, and D Ylvisaker. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26:131–148, 1990.
- [KP90] B W Kennedy and A W Phipps. *Calculation of doses from internal emitters using a new computer program: Quality control on the Rapid database*. NRPB-M215. 1990.
- [KW70] G S Kimeldorf and G Wahba. A correspondance between Bayesian estimation on stochastic processes and smoothing by splines. *Annals of Mathematical Statistics*, 41:495–502, 1970.
- [Lan59] W H Langham. Physiology and toxicology of plutonium-239 and its industrial medical control. *Health Physics*, 2:172–185, 1959.
- [LBHC50] W H Langham, S H Bassett, P S Harris, and R E Carter. *Distribution and excretion of plutonium administered to man*. Los Alamos Scientific Laboratory La-1151, 1950.
- [Leg85] R W Leggett. A model of the retention, translocation and excretion of systemic plutonium. *Health Physics*, 49:1115–1137, 1985.
- [Leg92] R W Leggett. A retention-excretion model for Americium in humans. *Health Physics*, 62:288–310, 1992.
- [Lin56] D V Lindley. On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*. 27:986–1005, 1956.
- [MCB79] M D McKay, W J Conover, and R J Beckman. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 21:239–245, 1979.

- [McI76] J F McInroy. *The Los Alamos Scientific Laboratory's Human Autopsy Tissue Analysis Study in The Health Effects of Plutonium and Radium*. The J W Press, Salt Lake City, 1976.
- [MMY93] M Morris, T Mitchell, and D Ylvisaker. Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction. *Technometrics*, 35(3):243–255, 1993.
- [MMY94] T Mitchell, M Morris, and D Ylvisaker. Asymptotically optimum experimental designs for prediction of deterministic functions given derivative information. *Journal of Statistical Planning and Inference*, 41:377–389, 1994.
- [MSY94] T Mitchell, J Sacks, and D Ylvisaker. Asymptotic Bayes criteria for nonparametric response surface design. *Annals of Statistics*, 22(2):634–651, 1994.
- [NM65] J Nelder and R Mead. *Computer Journal*, 7:308–313, 1965.
- [NN75] W D Norwood and C E Newton. US transuranium registry study of thirty autopsies. *Health Physics*, 28:669–679, 1975.
- [O'H78] A O'Hagan. Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society: Series B*, 40(1):1–42, 1978.
- [O'H91] A O'Hagan. Bayes-Hermite quadrature. *Journal of Statistical Planning and Inference*, 29:245–260, 1991.
- [PKKF87] D L Preston, H Kato, K J Kopecky, and S Fujita. Studies of the mortality of A-bomb survivors. 8. Cancer mortality, 1950-1982. *Radiation Research*, 111:151–178, 1987.
- [Ron95] W C Rontgen. Eine neue Art von Strahlen. Sitzungsber der Wurzberger Physik-med. *Gesellsch Jahrg.* 1895. An English version is contained in Dr W C Rontgen by Glasser O. Springfield, Illinois (1945).

- [Ros83] K L Rose. A simulation comparison and evaluation of parameter sensitivity methods applicable to large models. In W K Lauenroth, editor, *Analysis of Ecological Systems*, pages 871–877, 1983.
- [SSW85] J Sacks, S Schiller, and W Welch. Designs for computer experiments. *Technometrics*, 31(1):41–47, 1985.
- [SWMW89] J Sacks, W Welch, T Mitchell, and H Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435, 1989.
- [WBS⁺92] W Welch, R Buck, J Sacks, H Wynn, T Mitchell, and M Morris. Screening predicting and computer experiments. *Technometrics*, 34(1):15–25, 1992.

Acknowledgments

I would like to thank my supervisor Professor Tony O'Hagan for his help and advice, also the National Radiological Protection Board, especially Dr Colin Muirhead for giving me the time and financial support to undertake this research. In addition I would particularly like to thank my colleague Julia Thomas for her encouragement and understanding.