UNIVERSITY OF STRATHCLYDE
DEPARTMENT OF MANAGEMENT SCIENCE

Evaluation of Bayes Linear Modelling to Support Reliability Assessment During Procurement

By

Matthew Revie

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy

July 2008

Acknowledgements

I would like to express my gratitude to all those who have helped and supported me in completing this thesis. In particular I would like to thank Professor Tim Bedford and Professor Lesley Walls for giving me the opportunity to carry out this research and supporting me during the process. I would also like to thank Dr John Quigley, who despite not being my supervisor has been willing to offer me support throughout the entire process. I am also extremely grateful to Caroline Sisi and Alison Kerr for putting up with me and my constant demands. I would like to say a special thank you to Joseph Fragola for carrying out the much-appreciated task of inspiring me.

On a personal level, I would like to thank all my friends who have put up with me during the PhD and beyond. I would also like to thank Lauren and Craig who have offered encouragement in their own unique way. Above all, I would like to thank my parents who have always offered support and encouragement to me, no matter what I have chosen to do.

Abstract

A major task facing a number of different procuring agencies is the assessment of the developmental reliability of a product as contractors incrementally provide information. These agencies would like to use structured quantitative methodologies, as opposed to unstructured assessment currently adopted, to evaluate reliability throughout the life of a development programme as additional information becomes available. Due to resource constraints, any developed methodology must be cost and time efficient. This research attempts to develop a methodology for customers that is capable of assessing the information presented by a contractor in a reliability case to support decision making throughout the procurement process.

Bayesian Belief Networks (BBN's) have many features that make them useful for modelling complex problems as they allow for the combination of multiple sources of information, graphically represent decision problems and can formally incorporate subjective beliefs into the modelling. The Bayes linear methodology, although similar in spirit, has advantages over full probabilistic approaches, such as BBN's, when modelling complex problems. In contrast to these "traditional" Bayesian approaches, the Bayes linear methodology offers a quick, simple method to perform inference. The primitive used in a Bayes linear analysis is expectation rather than probability and as such, the elicitation burden is reduced.

Sensitivity analysis is difficult and time-consuming to carry out within a probabilistic framework. However, it is an important and necessary task to validate a model. Due to the way in which the Bayes linear inputs are elicited and the ease with which adjusting beliefs can be carried out, it is possible to carry out extensive sensitivity analysis to assess whether the decision taken is robust to small changes in the decision maker's prior beliefs.

This thesis investigates the different ways in which processes, such as the elicitation protocol, the construction of Bayes linear networks and sensitivity analysis, can be developed to build a Bayes linear model on real-life problems. The United Kingdom Ministry of Defence (MOD) is one such organisation that requires a structured model to track reliability. The usefulness of the methodology was demonstrated on two 'live' projects through the use of Action Research. The first project focused on supporting a MOD decision maker in deciding which of three contractors to choose from. The model assessed which of the systems the decision maker believed was currently the most reliable and which reliability programme was reducing her uncertainty of the reliability by the most. The second project focused on supporting a MOD decision maker in assessing the in-service reliability of two systems prior to them entering service. For both systems, test data was used along with expert opinion to assess the reliability of each system. In both projects, extensive sensitivity analysis was carried out to assess how changing the prior beliefs of the decision maker impacts on the final decision taken. The Bayes linear approach developed in this thesis provided excellent decision support to the decision makers in both projects.

Abbreviation

AGREE	Advisory Group on Reliability of Electronic Equipment
AMSAA	US Army Material Systems Analysis Activity
BLN	Bayes Linear Network
BBN	Bayesian Belief Network
DLO	Defence Logistic Organisation
DPA	Defence Procurement Agency
DS	Defence Standards
FMECA	Failure Modes and Effect Criticality Analysis
FTA	Fault Tree Analysis
HOS	Head of Specialisation
HSE	United Kingdom Health and Safety Executive
ILS	Integrated Logistic Support
IPT	Integrated Project Team
IPTL	Integrated Project Team Leader
ISRD	In Service Reliability Demonstration
ITT	Invitation to Tender
MIL-STD	US Military Standard
MOD	United Kingdom Ministry of Defence
RAC	Reliability Analysis Centre
RBD	Reliability Block Diagram
RGT	Reliability Growth Trial
R&M	Reliability and Maintainability
SC	Safety Case
SRD	System Requirements Document
TRACS	Transport Reliability Assessment and Calculation System
URD	User Requirements Document

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Reliability

The standard definition for reliability is "the ability of a system to perform required tasks under specified environmental conditions for a stated period of time" [BS 4778]. The term 'system' is used generically throughout this thesis so the same definition can be applied to platforms, sub-systems, components, structures and plants. Meeker and Escobar (2004) argue that the phrase "encountered environmental conditions" should replace "specified environmental conditions" in the above definition. Other authors define reliability in terms of other disciplines. For example, Condra (1993) believes "reliability is quality over time". This definition (and the standard definition) highlight that reliability assessment requires testing over a period of time as it may be many months before systematically poor reliability is identified in a system (Ireson et al, 1996). However, the standard definition given above will be adopted throughout this thesis.

Whilst the field of reliability theory is relatively young, with most pointing to the birth of reliability theory to have taken place during the 1950's (Bendell, 1991), the concept of producing reliable systems has been around for many years, as demonstrated by the number of systems that have been in operation for many years. Reliability is considered a key component of the performance of a system as "failure of products to function can be costly and lead to unsafe operating conditions" [Walls et al, 2006]. The desired target for reliability for different systems is dependent on the consequence of unreliability (Green and Bourne, 1972). The value placed on reliability is a function of the frustration, annoyance and financial loss incurred when a failure occurs.

The structure of the chapter will take the following form. Section 1.2 introduces the origins and initial developments of reliability modelling. Section 1.3 and Section 1.4 introduce reliability assessment and reliability management respectively. Section 1.5 discusses reliability assessment from the perspective of the United Kingdom Ministry of Defence (MOD). This section includes the history of reliability in the MOD, the reliability decisions the MOD must make and how reliability information is captured. This section also includes a summary of information gathered from the MOD Reliability and Maintainability (R&M) staff about how reliability assessment is carried out. Section 1.6 outlines the aim of the research project and Section 1.7 outlines the overall structure of the thesis.

1.2 Origins of Reliability Modelling

Prior to World War II, analysing the reliability of a system was carried out subjectively (Green and Bourne, 1972). When the aerospace industry expanded following World War I, one and two-engine aircrafts were compared with respect to reliability. "The comparisons tended to be purely qualitative at this stage and little attempt was made to express the reliability in terms, say, of the proportion of flights which were deemed successful for the different engine configurations" [Green and Bourne, 1972]. Gradually, information on aircraft flights was gathered and the results began being expressed quantitatively, such as number of fatal landings per million landings (Green and Bourne, 1972).

After World War II, reliability analysis began to be carried out quantitatively and modern reliability modelling began to evolve. The initial mathematical developments of reliability theory grew out of the new demands of modern technology; in particular, the new electronic equipment that was being extensively used (Barlow and Proschan, 1975). During this period, new electronic equipment and larger numbers of components on complex systems were introduced. This resulted in reduced reliability and final users finding systems challenging to diagnose and repair (O'Connor, 2002). During World War II, the vacuum tube was observed to be notably unreliable - requiring replacement five times more than all other other electrical equipment (Coppolla, 1984). "It is this experience with the vacuum tubes that prompted the US Department of Defense to initiate a series of studies for looking into these failures after the war; these efforts eventually consolidated and gave birth to a new discipline, Reliability Engineering" [Saleh and Marais, 2005].

In 1952 the US Department of Defence set-up the Advisory Group on Reliability of Electronic Equipment (AGREE) in order to "study the whole reliability situation and recommend measures that would increase the reliability of equipment and reduce maintenance" (Barlow and Proschan, 1975). AGREE promoted detailed test plans, recommended that formal demonstrations of reliability were carried out and quickly became a standard procedure, particular with military systems (O'Connor, 2002). The AGREE report eventually evolved into the US Military Standard (MIL-STD). In the UK, Defence Standard (DS) 00-40 - The Management of Reliability and Maintainability was first issued in 1981. Both these standards have consistently been updated and revised to take into account the evolving problems facing reliability engineers.

During the 1950's and 1960's, a number of books and standards were published that focused on developing quantitative methods to support reliability engineers. In 1956, the Reliability Analysis Centre (RAC) published TR-1100, "Reliability Stress Analysis for Electronic Equipment". This standard presented mathematical models to those wishing to estimate component failure rates and was the predecessor of the widely used Military Handbook-217 that was initially released in 1962. These initial books and standards became the initial building blocks of reliability prediction. The earliest books on reliability include Reliability Factors for Ground Electronic Equipment (Henney, 1956) , Reliability Theory and Practice (Bazovsky, 1961) and The Mathematical Theory of Reliability (Barlow and Proschan, 1965). Henney's was the first book explicitly on reliability whilst Bazovsky's book was the first on reliability engineering. Barlow and Proschan's helped to stimulate university research and was the first book to treat the subject at a graduate level. However, much of the foundations of the mathematical reliability modelling developed at this time had their origins in actuarial concepts developed initially by the insurance industry (Ebeling, 1997).

During the 1960's and 1970's, reliability engineering as a discipline became increasingly popular. At this point, reliability engineering began to diverge with engineers and mathematicians focusing on different areas of reliability. Two areas of interest to this research project are reliability growth modelling and reliability assessment tools.

Reliability growth modelling initially began to receive attention in the 1960's with the Duane model (Duane, 1964). Reliability growth testing is carried out during system development with the aim of exposing design deficiencies, identifying redesign steps and to ultimately increase reliability (Ansell et al, 1999). Three models commonly adopted by industry are the Duane model (BS 5760), the US Army Material Systems Analysis Activity (AMSAA) model (BS 5760) and the fix effectiveness model (BS 5760). These model are still routinely used (Quigley and Walls, 1999), however there are noteable theoretical and practical criticisms of them.

The Duane and AMSAA models assume the rate of occurrence of failure is infinite at time zero, and zero at time infinity. The latter issue can be addressed simply by including a constant term, but the first is more serious and because it relates to the period at the beginning of testing, it will always be a problem (Littlewood, 1984). In addition, both models fail to take into account differing test conditions, step-wise improvements in design and partial fixing of systems during testing (Ansell et al, 1999). Using these models, an instantaneous MMBF can be calculated, however this is a poor estimate of the true MMBF (Donelson, 1975), typically being optimistic during early stages of reliability growth testing (Robinson and Dietrich, 1987). The fix effectiveness model incorporates the engineer's judgement regarding the effectiveness of any change, however it fails to treat these fixes as stochastic processes (Ansell et al, 1999). For a comprehensive review of these models, see Quigley (1998).

The reliability growth models described above are those that will be encountered during Chapter 6 and not an exhaustive list of modelling in this area. Extensive research has been carried out in reliability growth modelling. Research has been carried out in areas such as multicopy testing programmes (Pulcini, 2001); modelling reliability growth with small sample sizes (Quigley and Walls, 2003) and; modelling accelerated life testing (van Dorp and Mazzuchi, 2004). In recent years, extensive research has been devoted to software reliability growth modelling (Kanoun et al, 1997, Pham and Zhang, 2003, Tian and Noore, 2005, Hu et al, 2007, Huang et al, 2007).

1.3 Tools for Assessing Reliability

A further area of development within reliability engineering focused on the tools that are available to engineers to assess the reliability of systems. Interest in assessing reliability initially focused on developing mathematical or statistical models. Books of interest included Reliability Technology (Green and Bourne, 1972), Methods for Statistical Analysis of Reliability and Life Data (Mann, 1974) and Statistical Theory of Reliability and Life Testing (Barlow and Proschan, 1975). These books primarily focused on developing and fitting distributions to failure data.

During the 1970s, reliability assessment began to shift from modelling mathematical distributions to techniques such as fault trees (Ebeling, 1997). During this period, many different techniques were developed aimed at assessing the reliability of a system. These included amongst others Failure Modes Effect and Criticality Analysis (FMECA), Fault Tree Analysis (FTA), Reliability Block Diagrams (RBD) and Parts Count Analysis. These techniques are capable of assessing the reliability of a system and are used extensively by reliability engineers during the design phase to support decisions such as design decisions or decisions regarding materials. Throughout the last thirty years, these techniques have been further developed focusing on increasing efficiency and accuracy of prediction. They have gradually evolved to the point where "methods have reached a high level of sophistication and are steadily progressing towards higher and higher degrees of precision" [Fussell, 1984].

The introduction of quantitative reliability requirements by policy makers meant that procurers needed methods that demonstrated the reliability characteristics of a system prior to the system being built (Denson, 1998). Assessing the future reliability of a system is important to both procurer and supplier for a number of reasons. Reliability prediction allows the forecasting of support costs, spares requirements, warranty costs and marketability. "Reliability prediction can be a valuable part of the study and design processes, for comparing options and for highlighting critical reliability features of design" [O'Connor, 2002]. Reliability assessment in particular supports decision making for both procurers and contractors when deciding between different engineering designs (Bendell, 1991). Modelling can also offer assistance to the analyst when deciding which reliability demonstration test to carry out and when analysing test results (Bierbaum et al, 2002).

Reliability analysis through modelling is a useful method of gathering information on a system prior to any operational testing being carried out. This modelling can focus either at component level or at system level. However, whilst modelling is often used to help the analyst predict the reliability of the system, the limitations of any modelling techniques should be understood. "Exact evaluation of system reliability is extremely difficult and sometimes impossible" [Chaudhuri et al, 2001] and prediction "can rarely be made with high accuracy or confidence" [O'Connor, 2002]. This is partially due to the fact that failure, and hence the predicted reliability of a system, is inherently linked to the demands placed on the system by the user and environment. Because of the uncertainty in how the user will interact with the system, predicting reliability is very difficult. Given this, the naive presentation of reliability predictions as fact and without necessary uncertainty assessments may have led to an undermining of the customer's confidence in the values stated by suppliers (O'Connor, 2002).

There has also been concern about the effectiveness of reliability targets / requirements. The concern has centred on the fact that complacency may set in if it is believed that reliability targets have been met (Meeker and Escobar, 2004). The problem is due to the way contractor's management seek to optimise economic performance rather than the concept of quantitative reliability targets (Meeker and Escobar, 2004). Historically, and indeed correctly, the mean time to failure (MTTF) is the primary and often only reliability parameter used by the contractor and supplier to communicate the reliability of a system (Ebeling, 1997). This is unsatisfactory however, as unless the time to failure distribution is exponential, then it is possible that two systems with the same MTTF have different reliability values at a given time. Instead, it is more useful to select the reliability parameters that are useful to the customer. It is also typical for contractors to display the reliability of a system using only one reliability characteristic (Gandy, 2007). It is unlikely that a full understanding of the reliability of a system can be gathered through a single reliability measure.

The purpose of any reliability modelling should be to identify possible failures in the design and to carry out re-design to ensure that the probability of these failures occurring in the future is reduced. It is conceivable that once a reliability prediction has been made using one of these methods, no rework is performed on the system, as the reliability prediction is greater than the reliability target. Hence the value of the modelling with respect to increasing reliability, in this case, is equal to the decrease in reliability caused by any change proposed. If no change had been proposed, then the value of modelling is zero.

1.4 Managing Reliability During a Development Project

Reliability management is the process through which reliability considerations influence design decisions throughout a developmental project. Reliability assessment tools are used to measure reliability during the design phase and as such, are useful in supporting design decisions. This could relate to decisions aimed at tracking reliability, improving reliability or at demonstrating reliability. Which of these receives the greatest priority depends on which is most important to the customer. For example, in the case of the development of an automobile, the developer/manufacturer may focus substantial effort on reliability improvement rather on than attempting to demonstrating reliability. Customers in this sector may not demand to know the reliability of the system at purchase but are interested in the reliability of the system as time progresses. Alternatively, procuring authorities such as the MOD might require that a developer demonstrates that a system meets reliability requirements prior to the system being accepted. Resources may be spent on demonstrating a target has been met in addition to improving reliability. As with the setting of requirements, the desire to know the reliability of a system prior to using it is interlinked with the consequence of failure.

An important part of managing reliability is the development of a reliability programme. "A formal reliability programme is necessary whenever the risks or costs of failure are not low" [O'Connor, 2002]. This research project focuses on the assessment of reliability programmes where one of the aims, either primary or secondary, of the programme is to demonstrate the reliability characteristics of a system to the customer. The developer will aim to demonstrate that reliability requirements have been met through a series of activities and their related outputs. Within this programme, justification should be given as to why these activities demonstrate that reliability targets have been met.

The added benefit of each activity is unique to each project and the benefit of carrying out one task might be reduced if another specific task is carried out. A reliability programme should include justification as to why certain tasks have been carried out and why other tasks have been omitted. If the aim of a reliability programme is to provide a procurer with assurance that reliability targets have been met, the level of assurance required from the procurer is an important consideration when developing the programme. This will vary from project to project and also between different decision makers on the same project. Hence the construction of a reliability programme is a task unique to each project and prescribing specific rules guiding how reliability programmes should be built is very difficult.

During the 1990's and the early part of this century, research has focused on incorporating decision making methods into the overall reliability programme attempting to manage reliability throughout the entire development project. Some of these methods have focused on assessing reliability using a combination of expert judgement, observed data and the logic structure of the system (Kerscher III et al, 2000). Other methods have focused on extending reliability growth models to incorporate cost analysis (Quigley and Walls, 1999). Other methods have looked at using value of information analysis to assess the cost-effectiveness of different forms of information (Bedford, 2004). Saleh and Marais (2006) attempt to link reliability to a financial parameter, net present value, and explore the impact of reliability on a system's revenue capability. Johnston et al (2006) investigate different techniques that could be used to develop cost effective reliability growth programmes. Chapter 2 will discuss some of these techniques in more detail.

Whilst there has been attention given to how contractors should attempt to build a reliability programme and present the evidence gathered, much of the research has focused on this task from a contractors point of view. This is strange as reliability assessment is often carried out by either a client or a procuring authority (Fenton and Neil, 2006). Little attention has been given as to how the procurer should evaluate the evidence they are presented.

For example, a procuring authority may be presented with evidence that FTA carried out regarding a system demonstrates that reliability requirements have been met, while other analysis has demonstrated that reliability requirements have not been met. Suppose furthermore that the developer of the system has a reputation for producing very reliable systems but initial testing appears to highlight that this system will not meet requirements. Under these circumstances how should the procuring authority structure this decision making problem? How much confidence should the procuring authority place in each form of evidence? How should these different forms of evidence be combined to assess the reliability of the system? Have the reliability requirements been met? How much uncertainty remains in the decision maker's assessment? These are questions that the procuring authority must answer.

In many projects the evaluation of all forms of evidence presented may be a very complex task and not one that can be carried out intuitively. In other cases, the reliability of the system may be simple to assess. Either way, this appears to be an area of study where decision support tools could be developed. As of yet, there has been a lack of academic literature concerning the development and feasibility of such methods.

The lack of attention given to assisting this decision making process is surprising yet understandable. Cost effective management of reliability should focus on growing reliability as the cost associated with improving reliability is offset by the benefit of achieving greater reliability. Hence a large amount of literature has focused on developing methods to support reliability growth. Only Fenton et al (2001) have appeared to address the issue of reliability management from a procurer's point of view. The technique adopted by Fenton et al (2001) will be discussed in detail in Chapter 2.

As there is a lack of literature addressing how procuring authorities should manage the reliability process, one must turn to the wider literature. In particular, we now focus on the methods used by contractors to support their supplier selection. There is wide range of literature on the subject with Weber et al (1991), Holt (1998) and De Boer et al (2001) offering progressive reviews of the different methods that can be used to aid purchasing decision making. There are a number of decision making methods that have been proposed to aid supplier selection. The simplest mathematical technique that has been used is the basic linear weighting model (Timmerman, 1986). Over the past fifteen years or so, such models have been extensively developed to take uncertainty into account (Soukoup, 1987). Mathematical programming has been used extensively to support decision making when choosing suppliers (Pan, 1989, Sadrian and Yoon, 1994, Rosenthal et al, 1995). Other techniques such as Multi Criteria Decision Analysis have been used (De Boer et al, 2001), Total Cost Analysis (Ellram, 1994) and neural networks (Albino and Garavelli, 1998).

The focus of this research project is to investigate different quantitative methodologies that could be used by procurers to evaluate the various forms of information presented by contractors to demonstrate reliability. The focus of this research project was not to support the contractors to develop a system that was more reliable, but instead to aid the procurers in their attempt to evaluate information presented to them. The MOD is one such procurer that is interested in developing quantitative methods to manage and assess reliability in a developmental project. The MOD have offered its services to support the development of such a methodology. They have offered 'live' reliability projects to support the development of a quantitative methodology. Because of this, the unique aspects of reliability assessment within the MOD must be addressed before research aims and goals can be identified.

The information presented in the following section is predominantly gathered from three sources; conference papers presented by MOD R&M staff, Defence Standards published by the MOD, and structured and non-structured discussions between the MOD R&M Staff and the researcher.

1.5 MOD's Perspective on Assessing Reliability

In 2006, the MOD spent approximately £16 billion procuring new equipment and providing logistic support for systems currently in-service. This money was split between two MOD departments - the Defence Procurement Agency (DPA) and the Defence Logistics Organisation (DLO). The DPA focus on procuring new equipment whilst the DLO is responsible for logistic support of in-service equipment.

Due to the size and nature of the work carried out by the DPA, processes and standards have been developed to aid the MOD support contractors in procuring equipment efficiently. During the early nineties, the MOD faced criticism regarding the bureaucratic nature of procurement and over-spending (Moore and Antill, 2001). The MOD decided to change how it procured equipment and moved towards a project management approach. The DPA adopted the Smart Acquisition process in order to procure equipment which aims "to enhance defence capability by acquiring and supporting equipment more effectively in terms of time, cost and performance" [MOD, 2002]. Part of the Smart Acquisition process is the CADMID cycle (Concept, Assessment, Demonstration, Manufacture, In-Service, Disposal) which the MOD use to identify the different stages in a procurement project. Figure 1.1 demonstrates the CADMID cycle.

During the Concept phase the Integrated Project Team (IPT) is formed. The IPT's sole task

Figure 1.1: **The CADMID Cycle**
Source: MOD, The Acquisition Handbook, 2002

is to procure a system which meets an identified capability gap. "They will drive the management of major defence equipment procurement, balance the trade-offs between performance, cost and time, within boundaries set by the approving authority" [Moore and Antill, 2001]. Within an IPT there are different roles representing the main disciplines such as cost, project management, Integrated Logistic Support (ILS), performance and R&M. During the concept phase the User Requirements Document (URD) is built which summarises the capabilities the user requires. At this point, time, cost and performance restrictions are placed on the system. Once the concept phase is complete, the project must move through Initial Gate. Initial Gate is a mini-hurdle the project must pass in which the project is approved and the time, cost and performance boundaries are noted.

During the assessment phase the IPT translates the URD into a Systems Requirements Document (SRD) which summarises the capabilities of the system. From this SRD, an Invitation to Tender (ITT) is sent out to industry. From the industry responses, the IPT must identify the most cost-effective technological solution and attempt to reduce the main risks in the project. At the end of the assessment phase, the project must go through the Main Gate. At this stage, the IPT must demonstrate that they will meet tightly defined cost, performance and time boundaries. By this stage, it is normal for a single contractor to have been identified and chosen.

During the demonstration phase, risk should be progressively reduced and contractors must demonstrate that they are able to meet requirements. During the manufacture phase, production should be undertaken and system acceptance is conducted to ensure that the system satisfies the SRD and therefore the URD. At this point, the MOD must decide whether or not the system meets performance requirements. If it does, then the MOD accepts the system into service. During the in-service phase, effective support to the user should be given to ensure that levels of performance initially agreed are met. During the disposal phase, plans for safe disposal of the system are investigated, identified and carried out.

1.5.1 MOD's Historical Perspective on Assessing Reliability

Poor reliability in military equipment reduces capabilities, places human life at risk, increases the amount of equipment required therefore increasing cost, and leads to adverse publicity for the Armed Forces and the Government. During the 1990's, there were a number of high profile projects that highlighted the DPA's reluctance to include reliability as a key performance measurement for the success of a procurement project. A House of Commons Defence Select Committee report in 2004 concluded that in the past "our Armed Forces have been let down by the organisation (DPA) tasked with equipping them" [HC 572-I, 2004]. Theoretically, the status of R&M within the MOD has equal consideration alongside time, cost and performance (Public Accounts Committee, 1989). In practice, however this is not the case.

At approximately the same time, the way in which the MOD contracted for R&M also changed. Prior to this, the MOD took a prescriptive approach to reliability. R&M standards were developed by the MOD detailing the tasks they expected a contractor to include in a R&M programme. These tasks were assumed to aid a contractor in producing (and demonstrating) reliable and maintainable equipment. After the system was designed and built, the MOD typically used trials and testing to gain statistical confidence that the contractor had delivered a system that met the R&M requirements set by the MOD.

It became apparent, for a number of reasons, that this prescriptive approach was no longer an effective way of demonstrating assurance. First, large amounts of software had to be integrated into almost every new system developed. The standard tools and techniques of reliability analysis were not easily applied to systems with extensive software (Denning, 2003). Secondly, questions were being raised by contractors and the industry as a whole as to the MOD's ability to tell contractors how to do their job (Jefferis, 2004). It was argued that contractors were better placed than the MOD to decide the tools, techniques and processes required to develop reliable equipment for each unique project. By allowing contractors to choose the correct tools and techniques for each project, more cost effective R&M Programmes could be developed. Thirdly, as the reliability of equipment being achieved by contractors was increasing, the duration of testing or trials required to demonstrate reliability requirements were being achieved was also increasing. These longer trials were costing both the MOD and contractor more money and a new approach was required.

The MOD decided that instead of placing expensive and often limited hurdles in front of the contractor at the end of a project, it would be more beneficial to both contractor and customer to build confidence that reliability requirements were being met throughout the duration of a project. The MOD and contractors believed that confidence could be built gradually during the early stages of a project by identifying and examining reliability risks and then taking steps to manage and mitigate against these risks as the project progressed. At the beginning of a project when limited information is available, there is a large amount of uncertainty regarding the future reliability of the system. As more information regarding the system is gathered, the uncertainty surrounding this reliability should decrease. Figure 1.2 demonstrates how uncertainty regarding the reliability of a system should change throughout the life of a procurement project (Bedford et al, 2006). This process of gathering information and building confidence

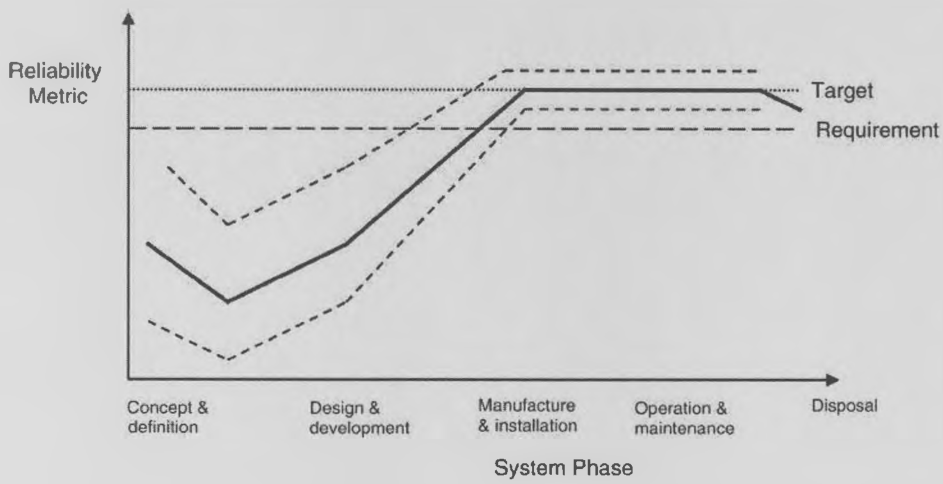

Figure 1.2: **Decreasing uncertainty in future system reliability**
Source: Bedford et al, Expert Elicitation for Reliable System Design, 2006

between contractor and customer is referred to as Progressive Assurance and is the methodology currently adopted by the MOD.

Due to this change, the role of the MOD's R&M staff has also changed. Since the introduction of progressive assurance, the MOD has adopted a role as a reliability auditor. The primary role involves evaluating the evidence presented and assessing the argument put forth by the contractor (Barnett and Denning, 2004). Any information gathered relating to R&M throughout a procurement project should be captured in the R&M Case.

1.5.2 Structure and Content of the R&M Case

The R&M Case is the body of evidence, assumptions and arguments provided by a contractor that aims to demonstrate to the MOD, or other similar procuring authorities, that a platform/system/sub-system will meet reliability requirements when placed in-service. The MOD Defence Standards state "The R&M Case is defined as a reasoned auditable argument created to support the contention that a defined system satisfies the R&M requirements" [DS 00-42, Part 3], and "The R&M Case is all the R&M evidence which will come from engineering design activities, trials and testing and in-service or field data" [DS 00-42, Part 3].

The R&M Case will contain a heterogeneous mixture of different types of evidence from different stages in the development cycle and procurement process. This means the MOD evaluator must have an understanding of the different R&M tools used by the contractor, be able to assess the relevance of each source of information presented and discard the misleading information. Ultimately, the MOD's only responsibility is to determine whether or not the R&M Case Report submitted has given them confidence that reliability requirements will be met (Barnett and Denning, 2004). While the R&M Case is described in generic terms, the details of each R&M Case is left to the contractor. This means that each R&M Case will be unique making it difficult to give specific rules for evaluating it.

Due to the size of a R&M Case, it is unlikely to be brought together within a single document (DS 00-42 Part 3). Instead a R&M Case Report is submitted by the contractor to the MOD which aims to summarise the important parts of the R&M Case; key decisions, evidence, arguments, etc. Figure 1.3 from DS 00-42 Part 3 demonstrates the concept of the R&M Case and the Case report. "All the analyses, strategies, plans, evidence, assumptions, arguments and claims that make up the R&M Case are to be found within the boundary indicated by the dotted line" [DS 00-42 Part 3]. From this summary document it should be possible to trace back all

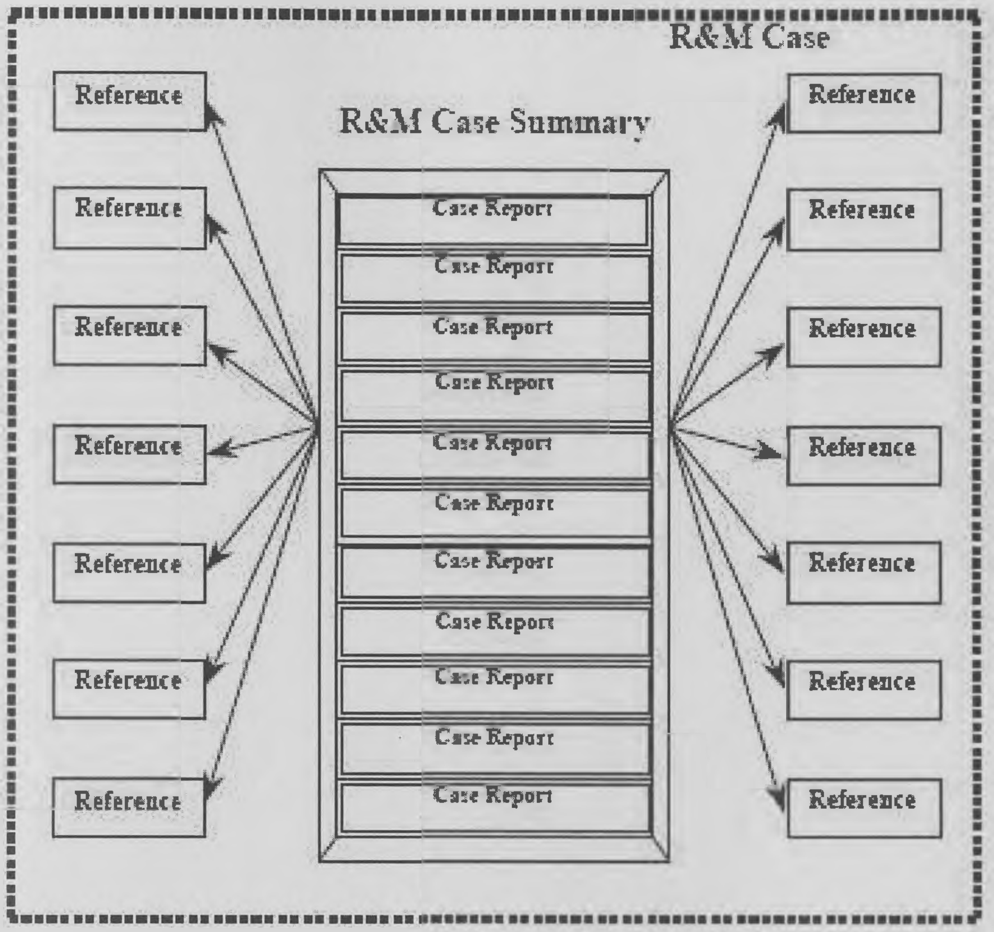

Figure 1.3: **The concept of the R&M Case**
Source: DS 00-42 Part 3, MOD, 2006

claims to the contractor's raw data.

The assessment of the R&M Case should be carried out throughout the entire procurement process, beginning when the contractors submit their initial R&M programme detailing the tasks they plan to carry out, through to the system being placed in-service and reliability data being observed. Once the system enters service, the R&M Case is not discarded. Instead, the MOD in-service manager should continue to use and update the R&M Case in order to monitor the reliability of the system.

When a R&M Case is submitted at the very early stages of a project, a crucial section of it is the R&M strategy. This strategy is supposed to demonstrate how the risks will be mitigated and how the necessary assurance will be delivered (DS 00-42 Part 3). The submitted

R&M strategy will include within it a detailed R&M Programme and R&M Plan. This should outline "a systematic programme of activities for satisfying the requirements and providing progressive R&M assurance" [DS 00-42 Part 3]. The R&M Case should contain all the forms of evidence that demonstrate that reliability requirements are being met and not necessarily only results from testing and numerical analysis. This programme could include a wide range of different sources of information; performance in previous usage, analysis of Operating Duty Cycle, design calculations, predictions and modelling, testing, simulation, expert opinion including previous recorded success of the supplier, correct implementation of best practice, etc (DS 00-42 Part 3). As the contractor carries out the planned tasks and gathers additional information regarding the system, the R&M Case and its conclusions are updated to reflect this new information. The R&M Case, due to its nature, should allow the evaluator the opportunity to evaluate reliability issues at an early stage and again periodically throughout the project. In theory, this should ensure that any problems encountered are resolved early in the project life. By resolving issues at an early stage, reliability is increased prior to the item entering service and theoretically cost is reduced.

The methodology adopted by the MOD to assess reliability is similar to the methodology used by the United Kingdom Health and Safety Executive (HSE) to assess the safety of a system (Denning, 2003). The philosophy of progressively building confidence between contractor and procurer adopted by the MOD is essentially an adaptation of the Safety Case methodology (Barnett and Denning, 2004). The Safety Case (SC) is a document required by the HSE and other regulating authorities arguing that the safety requirements of a system or installation have been met.

The SC is similar in many ways to the R&M Case. Both the SC and the R&M Case are documents that should exist throughout the life of the project and evolve as the system develops. As both documents are presented and evaluated prior to the system being extensively operated, it is unlikely that, in either case, extensive test data will be available (Kelly and McDermid, 2001). This means that the contractor has to demonstrate safety and reliability requirements have been met through means other than testing.

There are some notable differences between the SC and the R&M Case. The most notable of these differences is the legal aspect of safety. Contractors have a legal obligation to ensure that a system is safe for use. There are two main consequences of this. First, this theoretically means that procuring authorities should not have difficulty gaining access to a contractor's safety data. Second, whereas the procurer may change the reliability requirements if they are deemed unachievable, the safety requirements are fixed by law. Different methods for evaluating the SC have been proposed; Glendon and Stanton (2000) propose a form of task analysis whilst Kelly and McDermid (2001) propose using graphical tools to qualitatively map the inter-relationships between the information contained in the SC.

1.5.2.1 Reliability Decision Making within the MOD

It is the responsibility of the MOD to evaluate the R&M Case and decide whether or not the contractor has provided sufficient assurance that the system will meet reliability requirements when it enters service. On many projects, there will be two evaluators - each with different skills, knowledge base and responsibility. These evaluators are the R&M Focal Points and R&M Specialists.

The R&M representative in the IPT is the R&M Focal Point. Their task is to monitor R&M developments throughout the project, evaluate the R&M Case and report to the Integrated Project Team Leader (IPTL). It is likely that each R&M Focal Point will have detailed technical knowledge of their own project and possibly expert knowledge of the system being procured. If the R&M Focal Point require any further assistance, they may enlist the aid of a MOD R&M Specialist. The R&M Specialists within the MOD's R&M Head of Specialisation (HOS) department are likely to have have a high level of engineering knowledge; typically to degree level or equivalent and will have seen a wide range of R&M issues and problems. The breakdown of the MOD R&M community when this research started is displayed in Table 1.1.

	Land	Sea	Air
R&M Head	1	2	1
R&M Head of Specialisation member	15	11	7
R&M Focal Point	23	20	20

Table 1.1: Target population of MOD R&M Staff

Figure 1.4 displays how the internal structure of the MOD R&M department was broken down and related to the different projects. In 2006, the internal structure of R&M HOS

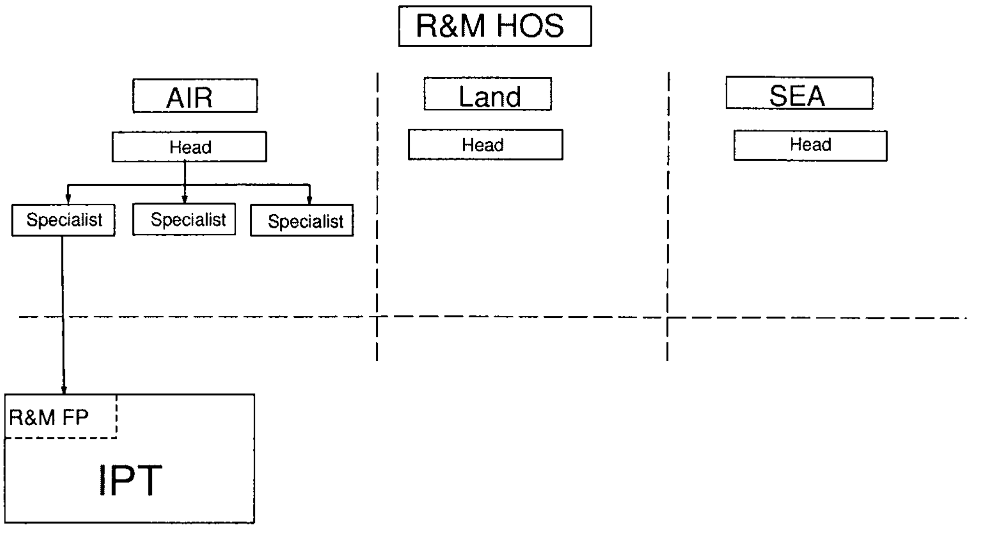

Figure 1.4: **Organisational structure of R&M in the MOD**

changed. The department is no longer broken into the three environments but instead is broken into two teams which each consist of three sections. One team consists of Land, Sea and Information Superiority. The other team consists of Air, Weapons and Support. This change of structure does not affect this research as the hierarchical R&M structure has remained: R&M HOS informing R&M Focal Point informing IPT.

The head of each environment is likely to have wide exposure to a number of R&M projects and problems. Whilst they will work on projects, most of their time is spent addressing the wider R&M issues within HOS, such as knowledge transfer, training, Defence Standards, research, etc. Within the R&M Specialist's Group, there is a wide range of experience and expertise. Some will be civil servants who have worked in the R&M HOS department for a long time, whereas some will be military personnel on secondment for three years. The main focus of the R&M Specialists is to offer support to the R&M Focal Points working on the projects.

Throughout the procurement process, there is only one MOD decision maker and that is the IPTL. Whilst the R&M Focal Point or R&M Specialist may offer support, ultimately any decision concerning the entire project rests on the IPTL. The IPTL will make decisions such as select the contractor or decide whether certain tasks should be dropped to make additional money available. However, R&M Specialists have the authority to delay a project based on reliability issues if they do not believe that a system is going to meet reliability requirements.

This research aims to support MOD decision making on reliability matters. However it is necessary to define who the "MOD decision maker" is and what decisions they are making. The methodology developed during this research project will not directly assist the actual decision maker of the project, i.e. the IPTL, but instead support the advice given to the IPTL by the R&M Specialists or R&M Focal Point. Throughout the rest of the thesis, the term "MOD decision maker" will refer to the R&M Specialist or R&M Focal Point on a project and "the decision" will refer to the advice they give to the IPTL.

It is now necessary to identify the types of decisions the "MOD decision maker" may face during a procurement project. The R&M Case exists through life from the concept phase all the way through to the disposal phase. There are two main reliability decision points during the course of a procurement project; the Main Gate and Acceptance of a system.

During the assessment phase, the IPT refine the reliability requirements and carry out simulation to determine the feasibility of the reliability requirements. At this point, an ITT is sent out and contractors are invited to submit proposals. Once these proposals have been submitted, the R&M Focal Point must evaluate the bids, assess which one is most likely to meet the reliability requirements and make a recommendation to the IPTL regarding the contractor to be chosen. In order to move through Main Gate, the IPT must demonstrate that the contractor is likely to meet reliability requirements. In some projects, several contractors may be taken through Main Gate and a decision on which contractor to choose is delayed until further evidence is gathered.

In order to achieve acceptance, the contractor must demonstrate to the MOD that reliability requirements are likely to be met when the system is placed in-service. The demonstration can be achieved through any type of evidence that the contractor chooses - however, the forms of evidence used are usually agreed upon prior to passing Main Gate. At the point of acceptance, the R&M Focal Point must assess the evidence and arguments presented by the contractor and decide whether or not he or she believes that R&M requirements are likely to be met.

These are the main decisions that the MOD R&M evaluators may face throughout procurement projects. Within these decisions are 'mini' decisions which help to support the main decision. However, it is possible that in certain projects they may have to face other decisions. For example, if the budget is reduced, the contractor may propose that one source of evidence is dropped. The MOD R&M advisor must be able to assess the effect of no longer having access to this source of evidence.

Currently there is no academic literature describing the processes adopted by procurers to support decision with respect to reliability. Interviews were carried out with members of the MOD R&M community, aiming to understand how the R&M Case is currently being evaluated and how decisions are made. Section 1.5.3 will describe the results of these interviews.

1.5.3 Interviews

1.5.3.1 Introduction

At the beginning of the research project in October 2004, interviews were carried out with MOD R&M Specialists and MOD R&M Focal Points with the primary aim of understanding how the R&M Case was currently being evaluated and how reliability was assessed by decision makers. The R&M Case is defined by the MOD as a "reasoned, auditable argument created to support the contention that a defined system will satisfy the R&M requirements" [DS 00-42 Part 3]. The focus of these interviews was understanding how this 'argument' was assessed by MOD's R&M staff throughout a procurement project. In addition to this, information on reliability decision making within the MOD and the different information sources used by the contractors to provide the MOD with assurance was also gathered.

1.5.3.2 Sampling and Analysis

At the time the interviews were carried out, the MOD R&M Department was broken into three distinct areas - Air, Sea and Land. Table 1.1 displays the target population at that time and Figure 1.4 displays the internal structure. It was necessary that an overview of the three environments and the three different levels of experience was gathered. It was apparent from an early stage that each of the three environments have unique attributes that result in different issues being raised. In order to obtain sufficient information, the R&M Head, one R&M HOS member and one R&M Focal Point for each of the three environments was interviewed.

The interviews were semi-structured and focused primarily on determining how the R&M Case was currently being evaluated by the MOD across the different environments. Each interview lasted between an hour and an hour and a half and took place at the DPA's headquarters, Abbey Wood, in Bristol. All of those interviewed were informed that their responses would be treated confidentially.

In order to bring out different themes and contrasting views from the environments and the different levels of experience, analysis matrices were used, similar to that in Table 1.2. The

Air				
	R&M Head	R&M HOS	R&M Focal Point	Analysis
Evaluating the R&M Case				
Prior knowledge				
ISRD				
Feedback to Contractors				

Table 1.2: **Analysis Matrix for Air**

analysis matrix was filled in with the corresponding responses for each R&M personnel. The analysis column summarised the views of the different personnel for that given environment. Table 1.3 gives an example of analysis matrix for one of the environments. From these analysis

Environment				
	R&M Head	R&M HOS	R&M Focal Point	Analysis
Prior Knowledge Required to Evaluate Case	Expert engineering knowledge is required to be able to work out a sampling method on the data and to identify the main R&M risks.	A mixture of project management and engineering knowledge is needed to produce or evaluate an R&M Case.	Definitely need engineering knowledge to evaluate the Case The focal points especially require a high level of engineering knowledge	From the three responses, there is agreement that engineering knowledge is required to evaluate the R&M Case.

Table 1.3: **Completed Analysis Matrix**

tables, different themes and ideas were teased out from the individual environments and the different levels. Appendix A gives an in-depth analysis of the output of the interviews. Section 1.5.3.3 will summarise the main findings of the interviews.

1.5.3.3 Summary of Interview Findings

In terms of the research project, five important themes emerged from the interviews. First, there was agreement across all three environments that the evaluation of the R&M Case was a subjective process. This subjectiveness could potentially lead to different evaluators drawing different conclusions on whether or not the R&M Case has demonstrated reliability requirements are likely to be met. At this point, there are no strict guidelines from the MOD as to how a R&M evaluator should evaluate the information presented in the R&M Case. In addition, there has been little attempt to use quantitative methods to evaluate the information provided in the R&M Case. However, most of those interviewed believed that such techniques may have a place in supporting their decision making. In addition, most of the the R&M evaluators also commented that they did not have as much time as they would have liked to work on evaluating a R&M Case.

Second, beyond the assessment phase of the CADMID cycle, the MOD has a lack of influence on reliability matters. One R&M specialist made an interesting comment about the level of influence that the IPT has at the six different stages of the procurement process, CADMID, on reliability issues. He drew Figure 1.5 depicting the level of influence.

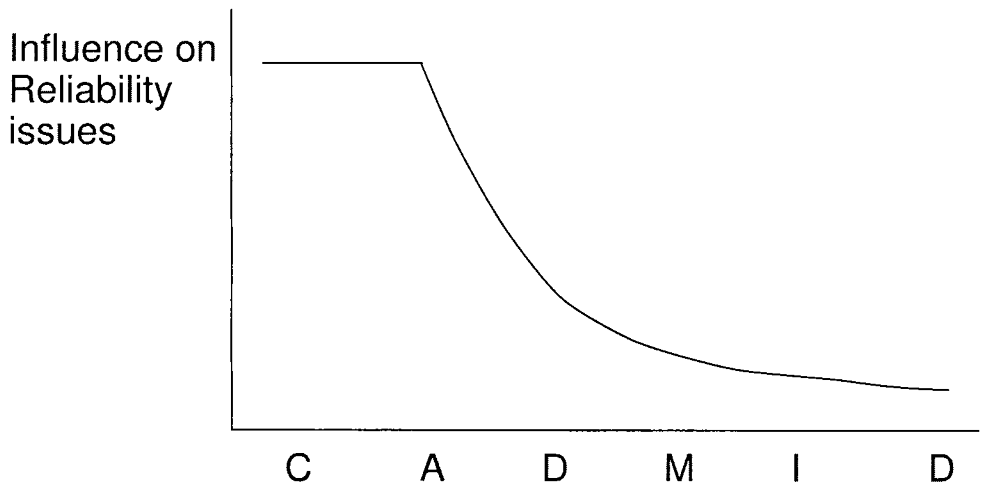

Figure 1.5: **Level of influence during a procurement project**

Figure 1.5 shows that once the contract has been signed between the MOD and contractor, the IPT has very little influence on reliability issues. Decisions that are made early on in the CADMID cycle have a huge influence on the final reliability of the system, whether this is a decision regarding contractor, solution or planned reliability tasks. Any methodology to support the MOD decision making process must take into account the fact that decisions that arise near the start of a decision making process can have a huge bearing on the success of the procurement project.

Third, the skill base of those working on evaluating R&M Cases focuses more on engineering knowledge rather than on mathematical or statistical knowledge. One R&M Specialist commented that while there is an abundance of people within the department with good engineering skills there is a distinct lack of people with mathematical or statistical skills. As far as he knew, there was only one R&M HOS member that was capable of tackling mathematical or statistical reliability problems. If R&M Specialists were to use a developed quantitative decision making tool, it would be necessary that the methodology be simple and easy to use for those without extensive mathematical backgrounds. In addition, one R&M specialist commented that another difficulty in evaluating the case is the fact that those evaluating the R&M Case may change during a procurement project. This makes it difficult for both the MOD and the contractor. For the MOD, the main difficulty arises because there is often no justification in the R&M Case as to why a certain task was chosen or dropped. Because of the subjective nature of the R&M Case, it may not be clear to a future decision maker why a certain path was chosen. For the contractor, it means that they may have to structure their arguments or evidence in a different way.

Fourth, R&M evaluators do not have as much time as desired to work on evaluating the R&M Case. If a quantitative modelling approach were to be applied within the context of the R&M Case to support decision making, it would need to be possible to set-up the model and carry out inference within a short period of time.

The final issue that makes evaluating the R&M Case difficult is the issue of burden of proof. There was almost unanimous agreement that, in theory at least, the burden of proof always lies on the contractor to demonstrate that reliability requirements have been met. However, there were a number of comments made during the interviews that appeared to contradict this. One R&M Specialist stated, ‘At the end of the day, if a piece of equipment is needed immediately, as long as the item is safe, it will be accepted warts and all" whilst another R&M Specialist stated "If a R&M Case is delivered that says requirements have been met, that is correct until such a time that the MOD prove it is not". Hence whilst there is a legal burden of proof on the contractor to demonstrate that reliability requirements have been met, due to time and cost constraints, there is a practical burden of proof on the MOD to demonstrate that reliability requirements have not been met. This means that any methodology developed must be transparent and simple to communicate the output of the model with multiple stakeholders.

1.6 Research Goals and Objectives

1.6.1 Introduction

There is a lack of academic literature addressing how procurers should evaluate the reliability evidence they are presented by contractors. The focus of this research project therefore is to investigate different quantitative methodologies that could be used by procurers to evaluate the various forms of information presented by contractors to demonstrate reliability. Much of the research carried out to date on reliability analysis and modelling has focused on designing and developing reliability. The focus of this research is not to aid a contractor develop a system that is more reliable, but instead aid procurers evaluate information presented to them.

The MOD has provided the opportunity to develop the methodology alongside 'live' projects. As such, the research goals, aims and requirements have been developed to address the decisions and take into account the limitations of the MOD. However, the problems facing the MOD are similar to those facing decision makers in other disciplines. For example, any high-technology engineering management organisation is likely to encounter situations where they are required to demonstrate through multiple sources of information that a developed product has reached a given level of performance. Hence, the wider aim of this research is not only to develop a methodology suited for the MOD, but to develop a high-level methodology that could be used by any organisation that is required to reduce a large amount of uncertainty by acquiring information.

The following is an excerpt taken from an email conversation between the University of Strathclyde and the MOD's Tim Jefferis, and summarises the direction of the research well.

> "The question exercising MOD at present is how does one assess the value of (say) an initial expert design assessment, plus some prototype testing, plus some static code analysis of an earlier version of the software, plus data on the field performance of the previous generation of the equipment, in demonstrating the

value/distribution of the final values of the R&M characteristics of the equipment."

Due to information gathered during the interviews, this has been extended to address additional decisions. The range of decisions the developed methodology would be looking to support will include the following:

- Which contractor is likely to produce the most reliable system?
- Which R&M programme is providing the MOD with the most information concerning the reliability parameter of interest?
- Which forms of information are most important in assessing the reliability of the system?
- Is the MOD satisfied that reliability requirements will be met?
- Are there sources of information that are not cost effective and hence may be removed from the R&M programme?

These are typical of the questions that a R&M Focal Point or R&M Specialist may have to answer through a procurement project. However, due to the fact that each procurement project is inherently unique, there is the possibility that some of these questions will not arise; nor should this be considered as an exhaustive list. Due to the duration of each procurement project - some projects can last in excess of ten years - it is unrealistic to expect to apply any developed methodology to an entire procurement project. First, the number of projects available would be substantially reduced if one were to restrict themselves to only entire projects as long projects and projects that are already in progress may have to be discarded. Second, if only one project is worked on, long periods of time may be wasted waiting for information to become available. Instead key time periods during a procurement project were identified during the interviews and reviewing the published MOD material. The developed methodology will be applied to different procurement projects at these key periods simultaneously to assess the effectiveness of the methodology.

Section 1.5.2.1 identified the two main decisions the MOD R&M decision makers face during a procurement project. The first was identified as prior to Main Gate where the MOD must decide which technological solution and contractor to choose. The second is prior to the system being accepted in-service, the MOD have to decide whether or not the system meets reliability requirement.

1.6.2 Research Goals

The research project can be broken down into three sections; goals, aims and research objectives. Table 1.4 summarises the goals, aims and requirements of the research.

Research Goal
Develop a quantitative methodology capable of using multiple sources of information presented in the R&M Case to support the typical decisions faced by the MOD (or other similar procuring authorities) during a procurement project.
Research Aims

1. Develop a quantitative methodology that structures the decision maker's subjective beliefs to assess the effectiveness of a contractor's reliability programme in reducing the decision maker's uncertainty surrounding a parameter of interest.

2. Develop a quantitative methodology that utilises the decision maker's subjective beliefs and observed evidence gathered to estimate the expected reliability of a system and credible uncertainty bounds around the point estimate.

Additional Requirements
In order to ensure the developed methodology is fit for purpose for the MOD, it should also meet the following criteria.

1. Mathematically justifiable and defendable
2. Traceable
3. Efficient
4. Simple to use

Table 1.4: Research Goals, Aims and Requirements

1.6.3 Research Aim 1

Research Aim 1 attempts to aid the MOD decision makers at the time where the MOD has most influence. This was identified as during the Assessment phase of the CADMID cycle prior to a contractor being chosen and prior to the IPT signing the contract between the MOD and contractor. At this point, contractors will have submitted their R&M programmes which should contain any analysis they have carried out to date and the analysis and data collection they plan to carry out in the future. There are two reasons why this time period appears to be the point where the MOD has most influence.

At this point the R&M evaluator must make a recommendation to the IPTL on whether or not they believe that a contractor shall meet reliability requirements and whether or not they believe that they will provide them with assurance that requirements have been met. If multiple contractors have submitted bids, the R&M evaluator must be able to compare the respective quality of each contractor's bid. The MOD is able to decide which contractor they wish to design and manufacture the system so it is imperative that they make the correct decision given their knowledge base and information available.

Second, the reason this is such a crucial time for the MOD is because at this point the contract has not been signed. During the interviews with MOD R&M advisors, an issue that was raised was the lack of authority that R&M advisors have once a contract has been signed. R&M Advisors believe that once a R&M Programme has been decided upon and the contract signed, there is very little they can do other than observe the contractor's work. They commented that it was unlikely a contractor would change a pre-determined list of tasks once a contract had been decided upon. Essentially the decision making powers of MOD are removed once the contract is signed and they are little more than observers. Because of this, the MOD must aim to ensure that anything they require, ranging from access to data to changing certain tasks within the R&M programme, is included clearly in the contract.

This time period is very early on in the development cycle and it is unrealistic to expect the contractor to have carried out extensive analysis. Instead, the evaluator may have to base the assessment on factors such as proposed solution, past experience of contractor or knowledge of design. During the interviews, many of the R&M specialists commented that they believed that they felt well enough informed to recommend contractor. However, they commented that they were not always capable of assessing which of the R&M programmes submitted is likely to provide them with the most confidence that the reliability of the system is what the contractor claims it to be. Decision makers are unsure how much their uncertainty is being reduced by each source of information, and in one case, a MOD R&M Specialist commented that he was unsure as to how much assurance he should be looking for.

They are unable to assess (except for extreme cases) the difference between the reduction of uncertainty from one R&M Programme and another R&M Programme. Similarly, the R&M evaluator of a problem may have an intuitive understanding of the different cost effectiveness of the different tasks the contractor could carry out, however, this belief has yet to be structured in a clear and concise manner and then related back to contractor. If this process was to be structured and captured quantitatively, it would mean the contractor would be able to produce a R&M Programme that both they and the MOD feel is cost-effective.

In order to compliment the MOD's current abilities and needs, it would be desirable to develop a methodology which is able to capture the belief structure of the decision maker, quantify the amount of initial uncertainty surrounding a reliability parameter of interest and to estimate the uncertainty being reduced by a contractors R&M programme. The developed methodology must be able to evaluate multiple R&M programmes and be able to quantify how much uncertainty will be removed by observing different forms of information. Any developed methodology should have an explanatory function (be able articulate the uncertainty remaining in the decision makers belief about relevant variables once certain information has been observed and what the expected value of the system will be) and an exploratory function (explore different scenarios to observe how much additional uncertainty is removed through certain forms of information).

1.6.4 Research Aim 2

Once a contractor has been chosen and the system is developed, different forms of information will be gathered and captured in the R&M Case. At some stage prior to the system going in-service, the MOD must accept that the system is likely to meet reliability requirements. Depending on the system being built, this acceptance is likely to occur during either the demonstration or manufacture phases of the CADMID cycle. This time period was identified as the period where the MOD has most interest in the reliability of the system. For a number of reasons, it is important for the MOD to have an accurate representation of the reliability of the system being procured prior to it entering service.

First, the final user of the system will need to have an accurate assessment of the reliability of the system in order to determine how many systems are required to meet the capabilities of a battlefield mission. Second, the DLO need to know how reliable the system is in order to be able to provide logistic support to the final user. In order to estimate spares, the DLO need an accurate assessment of the reliability of the system.

At this point, the MOD R&M evaluators must use the evidence presented in the R&M Case to determine whether or not reliability requirements have been met. Some of this evidence may have been collected very early on in the development cycle; possibly prior to design freeze, whilst some of this evidence may be observed data of the system in-service or data collected through reliability growth trials (RGT). In order to make as useful an evaluation as possible, the R&M Focal Point or Specialist must incorporate into his or her analysis an assessment of the uncertainty still remaining in their belief about the reliability of the system.

Currently, however, the MOD R&M evaluators do not use the multiple forms of information presented in the R&M Case and instead choose to use the form of information that has been gathered most recently or the information source they value greatest. This means that uncertainty is not being managed as effectively as it could be if all forms of information were being utilised.

The developed methodology should be able to use all the different forms of information presented in the R&M Case and the belief of the MOD R&M evaluator to assess the expected reliability of the system and to include with this assessment credible uncertainty bounds. This will allow the MOD to be able to estimate the reliability of the system prior to it entering service and make plans based on this value. At this stage of the procurement project, full-scale testing of any system is often expensive both in terms of cost and time. A methodology that can reduce the amount of testing that is required to be performed to reduce uncertainty to a desired level, is a benefit for the both the MOD and their contractors.

1.6.5 Research Requirements

In addition to meeting the research aims described above, in order that it the methodology be useful and practical, it must also meet these additional requirements.

Mathematically justifiable and credible

The MOD must be able to use this methodology to describe to different stakeholders their belief about the reliability of the system. This could include, amongst others; the final user, the DLO, the contractor and in some extreme cases, the government. For example, the MOD R&M evaluator may have to explain to contractors why they were not chosen for a project. The MOD R&M evaluator may have to demonstrate to a final user that given the information presented, they believed that the system would meet reliability requirements. Or the MOD R&M evaluator may have to justify to a review panel either of the above two situations.

This means that any methodology approach adopted by the MOD has to be able to withstand external scrutiny and have a solid philosophical and mathematical foundation.

Traceable

MOD policy dictates that the turnover of staff in the R&M HOS department is high and means that the R&M advisor working on a procurement project may change two or three times during the life of a project. It is important that future decision makers can understand why previous decision makers made certain decisions. Using a qualitative approach to assessment, it may not be obvious why a certain decision was taken. It should be possible, with any developed methodology, to demonstrate to future decision makers the belief structure of the of previous decision makers. This means having transparent elicitation techniques, capturing the qualitative and quantitative beliefs in a structured and easily repeatable way and adjusting these beliefs in a transparent way. By ensuring that everything is documented correctly, the MOD should be able to re-trace the steps and thought process of previous decision makers.

Efficient

Both MOD R&M Focal Points and R&M Specialists have limited time to spend on a project as their time is generally spread across multiple contracts. In particular, HOS members will have limited time as they may have no previous experience of the project they are being asked to advise on. Hence any model built must be able to be structured, populated and analysed in a short period of time. Any developed methodology that required a lot of input from the MOD yet was unable to produce results substantially better than they were currently achieving would not be used by those evaluating the R&M Case.

If a methodology can potentially be used to structure the MOD R&M evaluators beliefs, capture this in a clear and easily reproducible way, display this belief in a manner suitable for debate with a contractor, *and* can do this all in a short period of time, then the MOD R&M evaluator may look upon using the methodology as a way of saving time rather than as just an exercise to assess reliability. This issue of efficiency is very important. As systems become more and more complex, modelling techniques to assess reliability are becoming increasingly complex and time-consuming (Langseth and Portinale, 2007).

Simple to use

Across the potential MOD R&M evaluators, there is a range of differing mathematical and statistical abilities. From the MOD's perspective, there is little benefit in a methodology that requires them to consistently request external assistance to build. Therefore, any methodology developed should be capable of being utilised by those within the MOD.

Wider aspects of research

These additional requirements are specific to the MOD, however, they are unlikely to be restricted to only the MOD. As stated previously, during the development of any new product or system, there is likely to be a high amount of uncertainty which the developer wishes to reduce through time by acquiring information. In many of these cases, similar restrictions, such as resources and statistical abilities available, may be similar to those placed upon the MOD. Hence, whilst these requirements have been built primarily with the MOD in mind, it should not be assumed that these requirements narrow the focus of the research simply to the MOD.

1.7 Structure of Thesis

The thesis will follow the form of Figure 1.6

Chapter 2

Chapter 2 discusses the different statistical approaches that could be developed. The Bayesian philosophical framework is chosen as the most appropriate to develop. Two Bayesian based methodologies, Bayesian Belief Networks and Bayes linear, are investigated in further detail. From this investigation, Bayes linear is chosen as a potentially useful methodology for quantifying uncertainty in an R&M Case.

Chapter 3

Action research is the research methodology that will be adopted during this research. Chapter 3 gives an introduction regarding action research, highlights the benefits and criticisms of action research and documents how this methodology will be applied in this research project. This includes detail on how data will be gathered, the development of the model and how the modelling process will be validated.

Chapter 4

In Chapter 2, Bayes linear is identified as a potentially useful methodology for quantifying reliability uncertainty that exists during a procurement project. Chapter 4 addresses a number of practical issues regarding the application of the Bayes linear methodology. These include eliciting the necessary beliefs, building the covariance structure, constructing a Bayes linear network and carrying out inference. Two common reliability problems will be discussed to demonstrate that theoretically, the Bayes linear methodology is capable of supporting MOD decision making.

Chapter 5

In order to ensure the research goals are achieved, it is necessary to apply the Bayes linear methodology on 'real' MOD reliability problems in order to investigate the use of the methodology within the context of the research domain. Two projects were identified to apply the Bayes linear methodology to - one at the decision point where the MOD has most influence and one at the point where the MOD has most interest. Chapter 5 describes the modelling carried out in the project that addressed the problem where the MOD had most influence. The MOD were in the process of deciding which of three contractor's to choose to develop a new system. The Bayes linear modelling quantified the uncertainty of each contractor's current R&M programmes. The modelling also assessed the uncertainty remaining in each of the three contractor's R&M programmes assuming they carry out all the tasks they claim they will.

Chapter 6

The second of the two applied MOD projects focused on when the MOD had most interest in the reliability of the system - i.e. immediately prior to the system entering service. Chapter 6 describes the modelling carried out in this project. The modelling attempted to capture a number of 'softer' modelling factors that would have been difficult to capture with alternative methodologies. Extensive sensitivity analysis was carried out to ensure that the output of the model was robust to the decision maker's prior belief structure.

Chapter 7

Chapter 7 discusses validating the modelling carried out in each of the applied projects and the attempts to validate the Bayes linear methodology as an appropriate modelling process to address MOD reliability decision problems. Validation will be carried out using information gathered during interviews with MOD R&M Specialists regarding their thoughts on the usefulness of Bayes linear as a potential methodology to support reliability decision making.

Chapter 8

Chapter 8 reflects on and summarises the results of the research. Proposals for further development in this field are suggested and discussed, both in terms of developing quantitative tools for the MOD to assess reliability and the wider implication of developing the Bayes linear methodology to be applied to different problems.

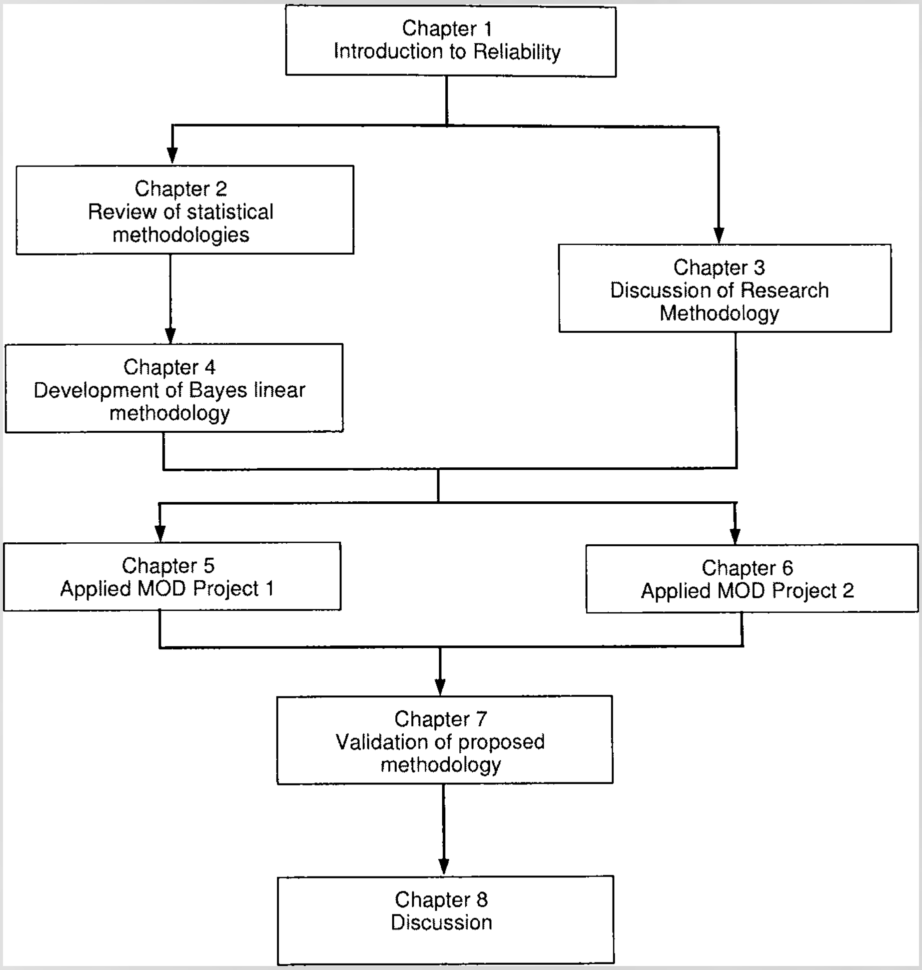

Figure 1.6: Structure of thesis

Chapter 2

Modelling Uncertainty in the R&M Case

2.1 Introduction

As discussed in Chapter 1, this research focuses on the development of tools to support decision making within a reliability context. However, the wider interest of this research project is the assessment of subjective uncertainty and modelling decision making problems where a large number of interrelated dependencies exist. The aim of this chapter is to identify a philosophical framework and methodology that can, potentially, meet the project aims and objectives. This chapter discusses the different philosophical approaches to addressing uncertainty and inference, and justifies the chosen approach.

Savage (1954) says that statistics is the study of uncertainty. It is widely accepted that probability theory offers a theoretical framework for quantifying uncertainty (Lindley, 2000). From this, however, there is much debate as to how probability should be interpreted. Despite this confusion regarding the philosophical interpretation of probability, almost all statisticians adopt Kolmogorov Axioms (Kolmogorov, 1956). The Kolmogorov Axioms state that the probability of an event is a non-negative real number, the probability of a sample space is equal to one, and the probability of two disjoint events A and B, $P(A + B)$, is $P(A) + P(B)$.

There are two main forms of inference; frequentist and Bayesian. The difference between frequentists and Bayesians is based on two key issues; the philosophical interpretation of probability and the practical issue of analysing and interpreting data. A discussion of the different philosophical viewpoints will be given in this chapter, however, the main focus is on the difference between the two approaches from a modelling perspective; particularly within reliability analysis and the MOD context.

The chapter will take the following form. Section 2.2 discusses uncertainty and the different philosophical approaches that can be taken. This section will focus on identifying an appropriate philosophical and epistemological view to proceed with. Section 2.3. will introduce the Bayesian Belief Network (BBN) methodology. Section 2.4 will introduce the Bayes linear methodology and finally Section 2.5 will conclude and summarise the chapter.

2.2 Uncertainty and Inference

2.2.1 Introduction

One of the most difficult challenges that analysts face when constructing a decision making model is how to incorporate meaningful uncertainty statements that accurately represent the decision maker's beliefs: "Decision makers are increasingly demanding some defensible representation of uncertainty surrounding the output of complicated mathematical models" [Cooke, 1991]. It is necessary that uncertainty can be communicated clearly, simply and effectively - especially in areas of decision making with multiple stakeholders. Understanding, eliciting, communicating and addressing uncertainty lie at the heart of reliability in all procurement projects. Currently the MOD communicate uncertainty qualitatively using comments such as "I'm fairly sure that the system meets requirements". However, these comments are open to misinterpretation as they are vague and ambiguous. There are a number of techniques available to communicate uncertainty in a quantitative and structured way. Whilst techniques have been developed to support decision makers, uncertainty in reliability procurement projects can never be eliminated, only managed (Aven et al, 2007). These techniques could potentially support the MOD R&M evaluators in managing reliability uncertainty more efficiently during procurement projects.

Analysts define uncertainty in different ways. Haimes and Lambert (1999) state that "Uncertainty represents partial ignorance or the lack of perfect knowledge on the part of the analysts". This statement suggests that uncertainty is a failing on the part of the analyst and that it is always possible to be certain by gathering additional information. This definition does not describe how uncertainty is reduced. A generic definition which will be adopted is given by Bedford and Cooke (2001) who state that "Uncertainty is that which disappears when we become certain ... certainty is achieved through observation, and uncertainty is that which is removed by observation".

There is disagreement about the different forms of uncertainty. It is accepted that the two most common forms of uncertainty are epistemic and aleatory [Morgan and Henrion (1990), Cooke(1991), Pate-Cornell (1996), Bedford and Cooke (2001), Aven (2003)]. Aleatory uncertainty is considered to be the "variation of quantities in a population" [Aven, 2003] compared to epistemic uncertainty which is the "lack of knowledge about the 'world' (i.e. the system performance), and observable quantities in particular" [Aven, 2003]. Aleatory uncertainty is also referred to as 'stochastic' uncertainty whilst epistemic uncertainty is uncertainty upon which it is possible to gather further information. This type of uncertainty is often referred to as 'lack-of-knowledge' uncertainty.

However, classifying uncertainties as either aleatory or epistemic can be difficult or ambiguous (O'Hagan and Oakley, 2004). For example, consider the uncertainty regarding the outcome of a toss of a coin. One may view this uncertainty as aleatoric uncertainty, however if all the physical conditions of the toss are taken into account, then the laws of physics could be applied to the decision problem and the uncertainty becomes epistemic. The context of the modelling determines whether or not an uncertainty is classified by a decision maker as epistemic or aleatoric (Bedford and Cooke, 2001). This has an impact on modelling MOD R&M problems as there are multiple stakeholders, each with their own perspective on the uncertainties contained within the problem.

Epistemic uncertainties are sometimes unreported or ignored; Pate-Cornell (1996) highlights areas of public policy where there are controversial or politically sensitive ideas as areas where epistemic uncertainty is not fully understood or appreciated. Alternatively, aleatoric uncertainty is "generally more easily acknowledged and integrated in mathematics models" [Pate-Cornell, 1996]. Also, "there appears to be wide agreement that the more epistemic uncertainty there is about a quantity the less able analysts will be to characterize the quantity with a precise probability distribution" [Ferson et al, 2004].

Inference is the process of deducing properties of the underlying distribution by analysis of data and is used to reduce uncertainty regarding a parameter of interest. The two main approaches for carrying out inference are the frequentist and Bayesian approaches. The difference between these two approaches focus on their interpretation of probability and how they analyse data. For each approach, a short description will be given describing how inference is carried out. A description of the philosophical and practical differences within the context of reliability will then be given followed by a discussion and justification of the philosophical view adopted during this research project.

2.2.2 Frequentist Inference

Moore states that frequentist inference "rests on repetition of one straightforward question, what would happen if I did this many times?" [Moore, 1997]. Frequentists believe that the probability of an event has a 'true' fixed but unknown value that can only be estimated through observations. The frequentist definition of probability is linked with the concept of relative frequency (Ushakov and Harrison, 1994) which is defined as the ratio between the number of events observed and the sample size. In a reliability context, for each system, there exists a 'true' reliability that a system has. In a frequentist framework, the system is repeatedly tested and the outcome of each test is observed. From this, the reliability of the system is estimated.

The probability of an event occuring is equal to the relative frequency observed during the trials. Assume that a one-shot device is about to be tested and that one wishes to adopt a strict frequentist framework to assess the reliability of the system. Assume that X is the probability of the one-shot device being successful. Given a sequence of tests $\{X_1, \ldots X_N\}$, for which $X_i =$ 0 if the device is unsuccessful and 1 if successful, then:

$$P(X = 1) = lim_{n \to \infty} \frac{X_1 + \ldots X_N}{n}$$

In order to be 95% confident that a target of 0.95 has been reached, 25 out of 25 successful tests would have to be observed. This approach to assessing reliability is analogous to the MOD using only ISRD data to estimate the reliability of a system. By this stage, the majority of decisions relating to the design of the system have been taken. The role of frequentist modelling in this context is to assess the reliability of the system rather than support decision-making. However, such a strict view of frequentist modelling is often not taken and other forms of information which are considered to be representative of the system, such as RBD's, can be used within a frequentist framework. For more information on the frequentist approach, see Cox and Hindley (1974).

2.2.3 Bayesian Inference

An alternative philosophical approach to assess uncertainty and carry out inference is the Bayesian methodology. All applications of Bayesian inference rest on the application of Bayes Theorem discovered by Reverend Thomas Bayes in 1763. Bayes produced a paper titled, "An essay towards solving a Problem in the Doctrine of Chances" (Bayes, 1763) in which he gave a solution to the 'inverse probability' problem.

The Bayesian framework allows the expression of uncertainty, within a structured and logical framework (Fenton et al, 2001). When modelling within a Bayesian framework, the variable of interest is not considered to be a fixed value but instead assumed to be a random variable, p. In addition, the decision maker's belief about the random variable is taken into consideration. The decision maker specifies his or her belief about the value of p through a probability distribution $f(p)$. From this, and the likelihood function $f(x|p)$, the decision maker can update their prior belief using observations x to obtain a posterior distribution $f(p|x)$ such that

$$f(p|x) \propto f(x|p)f(p)$$

This posterior distribution represents the decision maker's belief about the random variable p given the observations x.

An important distinction between frequentist and Bayesian methodologies is that frequentists believe that probability is the relative frequency of an event whereas Bayesians interpret probability as a subjective degree of belief. Subjective probability is defined as a person's degree of belief of an event occurring given the knowledge they have at a given moment in time (de Finetti, 1974). The idea of probability as a subjective belief has its foundations in Ramsey

(1926), de Finetti (1974) and Savage (1954). Those who promote subjective probability believe that a 'true' probability, as defined by frequentists, does not exist. Lindley (2000) states that "it is clear that one persons's uncertainties are ordinarily different from another's" whilst Morgan and Henrion (1990) believe that different people may "legitimately assign different probabilities to the same event" [Morgan and Henrion, 1990]. This is particularly true in this research project where a MOD R&M decision maker and a contractor's R&M engineer may have different beliefs about the uncertainty of a variable given the information that is currently accessible to them.

This idea of different decision makers having different uncertainties is at the core of Bayesian philosophy, however, it provokes contention among non-Bayesians. Non-Bayesians point to this lack of objectivity believing that it is possible to abuse the methodology. Draper (2006) distinguishes between coherency and calibration. To carry out a Bayesian analysis, a decision maker must be internally consistent, i.e. coherent. In order to be coherent, a decision maker's beliefs must adhere to the convexity rule, addition rule and multiplication rule (Lindley, 2006). De Finetti (1974) demonstrated that as long as a Dutch Book[1] is avoided, then the subjective probabilities of a decision maker are coherent (Loschi and Wechsler, 2002). Efron, a frequentist at the time, comments (sarcastically?) that it is extremely difficult to argue with a subjective Bayesian as long as their beliefs are coherent (Efron, 1986). As long as their beliefs are coherent, a subjective Bayesian's beliefs can always be justified to themselves. A decision maker is calibrated if, for example, 50% of all observed values fall within their 50% specified range. However, the output of any Bayesian modelling is ultimately unique to that decision maker. Even though the Bayesian methodology incorporates subjective belief into the analysis, if enough trials are carried out, the output of the frequentist analysis and Bayesian analysis will converge on the same value (Bedford and Cooke, 2001) [2].

Goldstein (2006) believes that there are a number of problems that can only be addressed using a subjectivist approach. However, there are Bayesians who are uncomfortable only adopting subjective probability and as such, attempts have been made to extend Bayesian analysis into the scientific community by coining the phrase "Objective Bayesians" (Berger, 2006). This is an attempt by some Bayesians to attempt to approach "scientific-objectivity". However,

[1]A Dutch Book is a gamble that results in definite loss.

[2]O'Hagan (1994) states provided the true value, θ, has non-zero prior probability, the posterior probability of θ will tend to one. This is similar to the notion of consistency in classical inference.

"there is no unanimity as to the definition of objective Bayesian analysis, nor even unanimity as to its goal" [Berger, 2006].

Kadane (2006) believes the name "objective Bayesian" does not help debate as it is thought of as being opposite to subjective. All Bayesian analysis fundamentally incorporates subjective belief at some point, hence, the term "objective Bayesian" appears at first glance to be a contradiction. Press (2003) believes that objective Bayesian inference is where very little information about an unknown parameter is known whilst Efron (1986) states that objective Bayesianism is "to produce prior distributions that capture the idea of objectivity". Berger (2006) states "Objective Bayesian analysis is a convention we should adopt in scenarios in which a subjective analysis is not tenable".

A decision-maker is said to be rational within a subjective Bayesian framework if and only if they are coherent (Williamson, 2005). Williamson (2005) says though that "Objective Bayesianism on the other hand, maintains that rationality goes beyond coherence", i.e. that further constraints must be met, such as calibration, before a coherent decision-maker can be classed as rational. A final definition is given by Kass (2006) who appears to believe that the phrase "objective Bayesian" refers to situations where a large amount of observed evidence exists which forces people to agree.

There are numerous types of Bayesians and sub-groups of Bayesians, however, for the purpose of debate it is preferable to begin by discussing subjective and objective Bayesian approaches. Volume 1, Issue 3 of the Bayesian Analysis contained two papers presented by Goldstein (2006) and Berger (2006) which debated the issue of objective versus subjective Bayesian analysis. Whilst Berger's article gives a detailed analysis of some of the benefits of using an objective approach, the approach as a whole comes under criticism from invited commentators. Berger (2006) identifies three benefits of using objective Bayesian inference. The first is that theoretically, it incorporates the best of both frequentist and Bayesian approaches. Secondly, appearing objective is necessary for the scientific community. Finally, most statistics is not carried out by statisticians (Berger, 2006).

A number of the responses to Berger's paper are strongly worded in their opposition to objective Bayesian approaches. However, the most scathing response is by Lad who states that "The marketing department has taken over from the production department. The goal is neither product quality nor service, but sales" [Lad, 2006]. Lad (2006) believes that an urge to find a 'one solution fits all' solution and a reluctance to spend the energy and effort developing and understanding a subjective framework results in people using tools that are not accurately communicating the uncertainty in a decision making problem. Kadane (2006) states that the objective approach incorporates many of the weaknesses of frequentist inference which is the driving reason why some statisticians have turned to Bayesian methods. Fienberg summarises the issue of objective Bayesian analysis well, "To me, the search for the objective prior is like the search for the Holy Grail in Christian myth. The goal is elusive, the exercise is fruitless, and the enterprise is ultimately diverting our energies from doing quality statistics" [Feinberg, 2006].

Goldstein (2006) also discusses the issue of pragmatic subjectivism which appears appealing. Pragmatic subjectivism acknowledges the difficulty of carrying out a full, subjective analysis of the necessary uncertainties but believes that under certain circumstances, in order to carry out any meaningful analysis, compromises can be made. These compromises focus on where the subjective beliefs do not affect the answer, when a preliminary or quick study of a situation is needed and when the decision maker is under-resourced (Goldstein, 2006). The first issue of where the subjective beliefs do not affect the answer is inherently interlinked with the concept of sensitivity analysis and will be discussed in detail at a later stage. The last two issues tie into the research context rather well as MOD decision makers may not wish or have the resources to carry out a detailed analysis of a problem but instead prefer to have a better understanding of a situation.

2.2.4 Comparison Between Bayesian and Frequentist Modelling in the R&M Case

The output of the frequentist approach appears at first glance naturally appealing and support for this approach remains; primarily due to its links with scientific objectivity (Efron, 1986). However, philosophical issues can be raised with applying this approach to assess reliability. An important part of the frequentist approach is the belief that trials can be perfectly replicated. This is of paramount importance when using the method and this restriction may cause issue in the field of reliability analysis. A simple example that illustrates that the output of an inference approach may not in fact equal the 'true' reliability of the system.

The example was described to the researcher at a MOD reliability conference and illustrates the difficulty of analysing data based on "replicated" trials and carrying out inference on this basis. One nation was selling a fleet of aircraft to another nation. The procuring nation analysed the reliability data and forecasted spares based on the reliability data. Once the aircrafts were purchased, the procuring nation found that their aircraft were not meeting the previously achieved reliability of the selling nation. In addition, they had exhausted their supply of air filters on the aircraft but had an abundance of extra undercarriages. Closer analysis of the data found that the selling nation deployed their aircraft at sea on aircraft carriers - hence the high reliability of air filters but the poor reliability of undercarriages. The procuring nation deployed the aircraft in desert terrain; hence their undercarriages were achieving high reliability but their air filters were continuously failing.

This simple example primarily highlights poor statistical modelling. However, it also highlights a difficulty of carrying out reliability analysis using a frequentist framework. The reliability of the two aircrafts (selling and procurer) is clearly different due to the environment that they are deployed in - however the frequentist approach adopted by the procuring nation treated them as the same. The decision maker could have used a proportional hazard model to relate the reliability of the two operating environments (Meeker and Escobar, 1998). To apply this model, the decision maker would require knowledge about the relationship of the failure rate of the system in the two environments. Whilst it is possible to apply these models at a system level, this would assume that the change in environmental effect would be uniform for all components. This would be unrealistic and so in practice these models are primarily applied to components.

Bayesian methods focus on mathematically combining the observed data with the experts prior opinion - in this case, the uncertainty about the reliability of a system. Prior information regarding the parameter of interest is used to construct a prior distribution. Bayes theorem mathematically transforms the prior distribution into a posterior distribution using the observed data. This posterior distribution represents a decision maker's belief of a random variable given the observed data and the prior distribution they specified.

Bayesian methods are very popular in the area of risk analysis (Bedford and Cooke, 2001), primarily due to the ability to incorporate prior information into the analysis, however, Bayesian methods have the opportunity to be more cost effective as less testing is often required. The Bayesian methodology offers the decision maker the potential to fix a prior distribution in order to gain a pre-determined posterior distribution. In order to address this issue, the elicitation process must be transparent, traceable and conducted in a scientific manner.

Frequentist testing on the other hand is often conservative as it does not attempt to incorporate into the analysis any form of prior knowledge. Bayesian methods have been used extensively to tackle decision making problems in many different fields, including medicine, finance, marketing, and oncology (Press, 2003). There have also been a number of books on the subject of Bayesian reliability analysis in recent years - most notably Martz and Waller (1982) and Barlow (1998) amongst others. In addition, the British Standard on reliability growth (BS 5760) has accepted the Bayesian framework as a robust methodology for carrying out analysis.

2.2.5 Statistical Philosophical Standpoint

It is necessary to explicitly identify the view of statistical foundations adopted during this research to ensure that the future direction can be understood and validated. In order to make a choice regarding which framework and method should be adopted, the nature of the problem must be understood. Chapter 1 has introduced the problem domain and described the nature of procurement within the defence industry. Based on this information gathered, a Bayesian philosophical framework to managing uncertainty and carrying out inference will be adopted. There are a number of key points that can be extracted from Chapter 1 that were influential when choosing this statistical philosophy. Mitchell (1993) identifies seven dimensions of model building. The modelling process adopted in this research must be capable of building purpose-built models and carrying out exploratory analysis on a macro level (Mitchell, 1993).

Firstly, within the R&M Case there is an abundance of information presented in different forms that contribute to the decision maker's assessment of the reliability of the system. The contractor may micro-manage reliability while the MOD assess reliability on a macro level. The Bayesian framework offers a philosophical and practical structure to carry out inference on a number of high-level inter-related sources of information that the frequentist framework does not allow.

Secondly, as reliability targets being requested by the MOD increase, demonstrating these requirements through classical testing methods becomes increasingly expensive and can only take place at the end of a development project. The frequentist modelling at this stage is predictive, whereas the progressive assurance philosophy adopted by the MOD aims to carry out exploratory analysis throughout the procurement project. The Bayesian framework is capable of carrying out this type of analysis.

Thirdly, as each reliability procurement project is unique, the modelling approach must be flexible enough to cope with the peculiarities of each project. This indicates that any modelling approach must allow for purpose-built models rather than developing a standard model. Bayesian methods offer more flexibility to those building models than the frequentist approach.

Finally, prior to a system entering service, the MOD are not actually interested in objectively assessing the 'true' reliability of the system as they understand that determining this value is unrealistic. Instead, they are interested in the decision maker's assessment of the 'true' reliability. Also, it is inevitable that the system procured will continuously change and be utilised in different environments.The MOD are interested in having an understanding of how the changes and different environments impact on reliability. In a frequentist approach, this would not be possible without testing the system extensively after the changes have been made or the system deployed in the new environment. The MOD may not be looking for an in-depth analysis of the impact, but instead an approximation as to how reliability may change.

Based on all these factors, a Bayesian philosophical approach will be adopted.

2.2.6 Bayesian Methods Applied in Decision Analysis

The Bayesian methodology has been rigourously applied in a number of projects. The aim of this section is to demonstrate the wide range of problems that Bayesian techniques have been applied to in recent years. Three main areas are covered: areas of decision making unrelated to reliability but of partial interest to this project; areas relating to reliability but not the assessment of reliability and; finally areas relating to reliability assessment.

O'Hagan et al (1992) developed a Bayesian model to estimate the assets of the UK water industry. Craig et al (1996: 2001) used a similar approach to predict future reservoir content based on the output of a computer simulation. Oakley and O'Hagan (2002) use Bayesian inference to assess the output of a computer model whilst Zhang and Mahadevan (2003) present a Bayesian approach for assessing the reliability of a model. Neil and Fenton (2005) present a Bayesian method to model operational losses. Norrington et al (2007) present a Bayesian model for assessing the reliability of search and rescue operations by the coastguard. Bayesian methods have been applied extensively in the area of value of information. Value of information is "the expected increase in the value (or decrease in the loss) associated with obtaining more information about quantities relevant to the decision process" [Dakins, 1999]. Examples of application of Bayesian Analysis in value of information problems include Azondekon and Martel (1999) and Kim et al (2003). With regards to reliability, using value of information techniques appear to be well suited to supporting those building cost effective R&M programmes.

In the reliability context, Bayesian methods have been used extensively. When modelling reliability growth, particularly with small sample sizes or when few failures are observed, frequentist methods are limited (Ansell et al, 1999). A number of authors have presented Bayesian models for assessing reliability growth. Early work (Littlewood and Verral, 1973, Mazzuchi and Soyer, 1988, Robinson and Dietrich, 1989) focused on developing Bayesian methods for modelling reliability growth. Ansell et al (1999) documents the potential shift from using a frequentist approach to a Bayesian approach. Emphasis has also been placed upon developing structured processes to elicit the necessary data to build a Bayesian reliability growth model. Walls and Quigley (2001) present a detailed elicitation process in order that the necessary data is collected in a scientific manner. Recent work has continued to develop these models and apply them in a wider range of reliability growth models, such as accelerated life testing (Van Dorp and Mazzuchi, 2004). Other application of Bayesian methods within reliability include Wooff et al (2002) who present a Bayesian methodology to support the development of software testing. This method assessed the probability of a failure remaining in the software given that a number of tests had been carried out.

With regard to reliability assessment, a number of methods have been developed to support decision making. These have focused on where extensive data regarding the structure of the system is available. Martz et al (1988) present a method for using Bayesian methods for assessing the reliability of a series binomial system using the reliability of its subsystems and components. Research at Los Alamos National Laboratory developed this method further. Kerscher III et al (2000) presented a method for assessing the reliability of a system based on using a reliability logic flow diagram of the system and updating the assessment of the sub-systems once new observations have been made. This research appears promising at assessing reliability, however, it requires that those making decisions have access to an accurate reliability logic diagram of the system. This method attempts to micro-manage reliability and requires a level of detailed knowledge about the system that may not be available to decision makers within the MOD. Kaplan et al (1990) present a methodology for progressively assessing reliability during development.

As before, there has been extensive work carried out in the area of reliability analysis using a mixture of methods. However, most of these methods focus on utilising information available to the contractor and not necessarily information that is available to the procuring authority. Only Fenton et al (2001) appear to have addressed reliability assessment in a Bayesian framework from the perspective of procuring authorities. The methodology they adopted was Bayesian Belief Networks (BBN). It seems natural to investigate this methodology and use this as a starting point for identifying a useful methodology for the MOD. The following section describes the BBN methodology.

2.3 Bayesian Belief Networks

2.3.1 Introduction

Almost all realistic decision-making problems will involve assessing a situation under uncertainty (Fenton et al, 2001). In a procurement project, there are a number of inter-related outputs, each with uncertainties attached to them which must be captured and modelled. The BBN is a methodology developed in the 1980's (Pearl, 1998) and is capable of assisting in addressing these types of problems (Fenton el al, 2001). They have been applied and used in the following fields: software quality (Littlewood et al, 2000), safety evaluation (Fenton and Neil, 2001), transport reliability prediction (Fenton et al, 2001), air safety (Roelen et al, 2003), common-cause failure modelling (Zitrou, 2007), and reliability assessment (Langseth and Portinale, 2007) amongst others. The transport reliability prediction paper (Fenton et al, 2001) is the most useful for this project as the methodology was used by the MOD's Land R&M department to assess different bids made by contractors. This particular model is referred to as the Transport Reliability Assessment and Calculation System (TRACS).

The review of the BBN literature has focused on the decision-making and engineering applications of BBN. Extensive work on the BBN methodology has also taken place in the field of computing science and artificial intelligence which for the purpose of this research is largely omitted. That research focus primarily on causal relationships and utilising objective data. This project wishes to incorporate subjective modelling into the process and perform inference on a relatively small number of variables. For more information on BBN's in the computing science field, an introduction is given by Williamson (2005).

This section will briefly discuss BBN theory, however the main focus will be on modelling within the current problem domain using a BBN approach. For a detailed description of BBN's and its theory, see Pearl (1988), Lauritzen (1996), Cowell et al (1999) and in particular, Jensen (1999).

2.3.2 BBN Theory

All definitions have been taken from Shachter and Kenley (1989). A node represents an uncertain or random variable which can be either continuous or discrete. An arc represents the causal or influential link between these uncertain variables. This influential link could represent an influence on a decision maker's belief about two uncertain quantities and not necessarily a physical causal influence. If there exists an arc from node A to node B, A is said to be a parent of B. A directed acyclic graph is a finite set of nodes and arcs. A BBN is a directed acyclic graph where information is stored in the nodes and arcs. Figure 2.1 is an example of a BBN. It is in fact a very small part of the TRACS model.

As there is an arc from 'quality of design process' to 'quality of inspection documents', 'quality of design process' is a parent node of 'quality of inspection documents'. This indicates that the value that 'quality of inspection documents' takes is conditional on the value of 'quality of design process'. Alternatively, as 'design staff quality' has no parent nodes, it is said to be unconditional and has an unconditional probability distribution.

"The main reason for building a Bayesian network is to estimate the state of certain variables given some evidence" [Jensen, 1999]. The BBN methodology is essentially a Bayesian

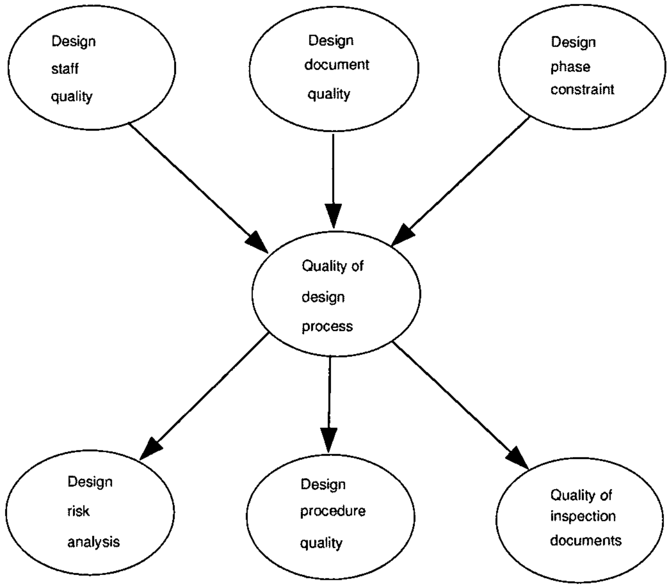

Figure 2.1: **Section of the BBN used in TRACS**

based methodology utilising graphical representations to capture the dependency and uncertainty between different variables. "The basic concept in the Bayesian treatment of certainties in causal networks is conditional probability" [Jensen, 1999]. The conditional probability of A given B is defined as the probability of event A occuring, given that event B (and no other events) has occurred. In order to populate the full BBN, it is necessary to specify the probability distribution for each node conditional on each of its parents. If a node has no parents, then it is necessary to specify its probability distribution. Whilst the distribution of a variable can, in principle, take any form, it is common that for simplicity and ease of calculations, either discrete variables or Gaussian distributions are used. This makes the updating process considerably simpler. Using all Gaussian distribution, the decision maker must specify an unconditional mean, a conditional variance, and where there is an arc, a linear coefficient representing the degree of dependency between the two variables (Shachter and Kenley, 1989). In the discrete case it is necessary to specify the conditional probability of each child node given the values of the parent nodes.

Building and utilising a BBN is broken into three different tasks. The first task is to build the model qualitatively which involves identifying all the variables that are of interest and building a graphical representation of the relationships between each of the variables. During this task, the decision maker builds a model that is subjective representation of their view of the world (Smith, 1996) and not necessarily a causal reflection of the actual objective world. Despite the extensive work carried out using BBN's, there has been a lack of description in many projects as to how this qualitative modelling was carried out (Hanea et al, 2006).

The second task involves stating all the necessary conditional and unconditional probabilities. This is sometimes a very labourous and difficult task which will be discussed in more detail in Section 2.3.4.

Once the BBN has been fully built and populated, the effect of different evidence entered into the model can be observed and propagated through the model. A posterior (or prior) distribution for any parameter of interest can be gathered. BBN's are a methodology for reasoning from uncertain evidence to uncertain conclusions. For example, if in the TRACS model, it was observed that 'design staff quality' was good, then the information can be propagated through the graph away from the observed node and the marginal distributions of the nodes can be updated.

2.3.3 Strengths of BBN

BBN's are extremely powerful computational tools and have many features that are useful when modelling complex problems.

2.3.3.1 Utilising Multiple Information Sources

BBN's offer the decision maker a methodology of combining many different sources of information in a consistent and mathematically robust way. In many reliability problems, there is often an abundance of information provided but not used efficiently by the MOD. BBN's offer a method of incorporating all the available information into the analysis.

2.3.3.2 Graphically Represent Decision Problem

BBN's take advantage of using graphical representations, most commonly in the form of an influence diagram. Those regularly using influence diagrams believe it is an effective tool for communication between decision makers and analysts (Howard, 1990). In particular, it is extremely useful with the "formulation, assessment, and evaluation of decision problems"

[Howard, 1990]. Using BBN's in this way can communicate quickly and efficiently to all those interested in a project the different influences and dependencies between different variables. BBN's are excellent for group discussion and debate regarding the structure of a model. Without this, decision makers would have to rely on mathematical equations to discuss the structure.

2.3.3.3 Traceable

One of the appeals in using BBN's, particularly in the public sector, is the transparent nature of the underlying theory and how the information is gathered. If the elicitation session is carried out in a transparent way, such that the decision maker justifies and gives evidence as to why they have chosen certain values, future decision makers can audit the process. Due to the length of some procurement projects, the reliability decision maker may change. As such, a different decision maker may decide the system should enter service as opposed to the decision maker who chooses the contractor. It would be useful to have a model which can be audited by future decision makers.

BBN's allow future decision makers the opportunity to audit, at the very least, the explicit statements of dependence and independence captured in the structure of the BBN. It would be unlikely that any auditor would be able to assess all the values stated by the decision maker, however they may be able to assess the logic which the decision maker gave to support the values stated (Smith, 1996).

2.3.3.4 Incorporating Subjective Belief into Modelling Approach

The reliability decision making problems facing the MOD are complex, have multiple variables and are generally cost-intensive. It is essential to incorporate expert judgement from decision makers and relevant experts into the process. This subjective belief can be incorporated into the modelling on two different levels. Firstly, sources of information that may not have been previously considered correlated due to a lack of causality, can now be modelled as such. An example of this in the current problem domain could be the following situation. A contractor is developing two different systems for the MOD. If the contractor meets the reliability requirement for one of the systems, the decision maker may now believe that they are more likely to meet the requirement for the second system.

Secondly, it is possible to give different weights to similar but subtly different sources of data. An example of this in the reliability field is as follows. A decision maker aims to model the reliability of a given system. The system is subjected to two different tests in different environments. Both tests were of equal length and theoretically, should be given equal weighting. In one of the tests, the system encountered three failures and in one of the tests, the system encountered 18 failures. The decision maker is aware of the unique nature of each of the environments and strongly believes that only one of the environments is representative of the actual environment the system will ever encounter. Using a structured elicitation process, the analyst can gather the conditional probability of the reliability of the system and the output of both tests.

2.3.4 Criticisms of BBN

The BBN methodology has a number of features that make it appealing to quantify and assess the information provided in a R&M Case. However, challenges in implementing the methodology exist. There are three issues in the BBN approach that make it difficult to build and populate in order to evaluate the evidence presented in a R&M Case.

2.3.4.1 Elicitation Burden

The first is the number of different conditional probability values that require elicitation (Coupe and Van der Gaag, 2002). For a problem containing many nodes, the decision maker will need to specify a large number of values. A very simple example illustrates this point. There are three manufacturers, A, B and C. Each manufacturer can interpret the R&M requirements 'correctly', 'adequately' or 'poorly'. Our third and final variable is 'failure rate of system' which can take the state 'low', 'medium' or 'high'. The BBN for this simple example is given in Figure 2.2 where M is the manufacturer, I is interpretation of requirements and FR is failure rate of the system. As the MOD R&M staff do not make decisions, the manufacturer chosen is modelled as a random variable.

The BBN represents the decision makers belief that the failure rate of the system is dependent on who the manufacturer is and how well they interpret the R&M requirements. In order to populate this simple example, the analyst would have to elicit 27 different variables from the

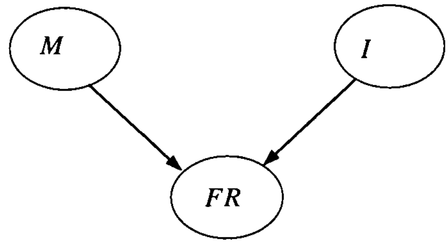

Figure 2.2: Example of simple BBN

decision maker. The values needed to be elicited is the $Pr(FR|M, I)$. For example, Pr(High|A and correctly), i.e. the probability of the failure rate being high given the manufacturer chosen was A and that A interpreted the R&M requirements correctly. The table below shows the values that need to be specified.

Manu	A			B			C		
Inter	Cor	Ade	Poor	Cor	Ade	Poor	Cor	Ade	Poor
Low									
Medium									
High									

Table 2.1: Table needed to populate BBN

This is a very simple example, yet needs considerable effort to populate. In addition, if a decision maker chooses to add another variable, the number of values needed to be elicited increases exponentially. For example, if there are two variables, each of which can take two states, then $2^2 = 4$ values need to be elicited. If there are three variables, each of which can take two states, then $2^3 = 8$ values need to be elicited and so on. As can be seen, if there is a BBN with many variables, the number of values required to be elicited soon grows to an unmanageable level.

2.3.4.2 Difficulty in Elicitation

Even if it is practical to elicit all of the necessary values from the decision maker, on occasions, questions may be raised regarding the quality of the information gathered from a decision maker (Kahneman et al, 1982). It is very difficult to translate the highly detailed but partial statements of prior knowledge from experts into the joint probability distributions required.

For simplicity, experts are usually restricted to very simple prior forms (Craig et al, 1996) in order to make the model computationally simple. If the decision maker wishes to model continuous variables, the analyst may choose to adopt a distributional form the decision maker is uncomfortable with in order to make the model computationally simpler. If the decision maker chooses a distribution that fits their beliefs but is not easily computed in a Bayesian framework, analysis will involve very complicated numerical problems of integration and maximisation of extremely high dimensional multimodal distributions (Craig et al, 1996).

In addition, the analyst may ask the decision maker to specify values which they are unable to specify, but must in order to complete the model. In BBN projects where significant time is available, these problems can be overcome with relative ease. However, in cases where decision makers are attempting to build and populate models quickly, this might be a significant issue. Goldstein (1990) believes that it is unrealistic to build a full continuous probability distribution which honestly represents a decision makers true beliefs. Instead Goldstein (1990) believes that a decision maker should state simpler values, such as expectation and variance, over a reduced number of quantities and has developed a methodology around this approach. This methodology will be discussed in detail in Section 2.4.

2.3.4.3 Setting up New Problems

BBN's have been used extensively to help support decision making problems. Examples are applications to diagnose medical problems (Heckerman et al, 1992), the Microsoft Office Help Tool (Jensen, 1998) and prognostic assessment (Coupe and Van der Gaag, 2002). However, from the review of the literature BBN's appear to have been used extensively in problems where the problem can be repeated many times and within a tight strictly-defined framework. For example, in medical diagnosis problems and the TRACS model, the questions and answers are tightly worded to limit responses. In our context, as each procurement process is unique and every R&M case different, a new BBN would have to be set up for each procurement project. The time, cost and effort required to build a full BBN suggests it is only really useful if it can be repeated many times or if the problem is of utmost importance.

Given the amount of time the MOD R&M specialists have to work on each project and their lack of detailed statistical knowledge, it is unlikely that they would be able to fully populate a

BBN for each project. Whether or not this is a fundamental limitation of the BBN methodology or a limitation of the MOD is debatable. Based on interactions with MOD R&M specialists, even with the necessary training, it is unlikely that the MOD R&M Specialists would be unable to build a BBN for a unique project. An example of this is the fact that TRACS is considered by those who use it as an extremely useful tool, yet it has not been modified within the MOD by other environments or in other projects.

2.3.5 Summary of BBN Methodology

The BBN methodology has a number of strengths that would potentially make it useful as a methodology to evaluate the information presented in the R&M Case. BBN's are capable of mathematically combining different sources of information to produce a probability distribution of a decision makers belief given some observed data. Additionally, BBN's can be used to simulate different scenarios prior to observed data being gathered, i.e. analysing the effect of different information sources. Recalling Table 1.4, the BBN methodology appears to satisfy the two research aims laid out. However, recall the additional requirements placed upon the methodology by analysing reliability from the MOD's viewpoint,

1. **Mathematically justifiable and defendable**
2. **Traceable**
3. **Efficient**
4. **Simple to use**

The BBN methodology only fulfills the first two of the additional requirements placed upon the methodology. The methodology fails in terms of time, resource efficiency and simplicity. As demonstrated above, the methodology is time consuming to build, populate and update. It is unlikely that given the number of projects that R&M evaluators have to work on simultaneously, that they would have the time to utilise the benefits of the methodology. In addition, questions would have to be raised about whether or not the MOD evaluators would be able to either build complex joint probability distributions or place faith in them if they did. Hence, whilst the BBN methodology is an extremely useful and powerful computational tool, it is unlikely that it is useful in the context of this problem. Instead, we seek a methodology with the same benefits as the BBN methodology but which can be built quickly and is simple to populate.

2.4 The Bayes Linear Methodology

2.4.1 Introduction

In practice, it is unlikely that the MOD would ever be able to apply the BBN methodology to a reliability case problem they currently face due to the elicitation burden placed upon the decision maker and the complex calculations that may need to be carried out. Hence a quicker and simpler approach that does not require the same abundance of information to build the decision maker's priori, but is able to remain justifiable and have all the benefits of the BBN methodology, is required. The benefits of using a BBN that are of particular interest are: graphical representation of dependency allowing for easier interpretation than mathematical formulas, incorporation of available expert information, transparency and the production of posterior beliefs once observations have been made.

As the method becomes increasingly accepted, it is beneficial to investigate ways in which the Bayesian methodology can be improved upon or extended further. As a Management Scientist, this equates to developing methodologies that are simpler to use, have wider practical application and are justifiable. It is believed that in terms of extending the fundamental benefits of the BBN methodology to a wider audience, the Bayes linear methodology fits this criteria. The Bayes linear methodology is an alternative to the traditional 'probabilistic' Bayesian approach. It only uses second-order information such as means, variances and covariances to make inferences, which theoretically reduces the specification burden from the expert.

The purpose of this section is to introduce the Bayes linear methodology and discuss the potential benefits and criticisms of the approach. As with the BBN section, the focus of this review will be on the modelling implications. The section will take the following form. Section 2.4.2 will briefly discuss the philosophical approach of the Bayes linear methodology and Section 2.4.3 introduces the mathematical equations used. Sections 2.4.4 and 2.4.5 discusses the benefits and limitations of the methodology. Section 2.4.6 summarises the framework and justifies applying this within a MOD context.

The majority of the literature in this section is taken from Professor Michael Goldstein's work. Goldstein has spent more than twenty years developing the Bayes linear methodology and philosophical framework culminating in the publication of *Bayes Linear Statistics: Theory and Method* (Goldstein and Wooff, 2007). As such, this section draws heavily on this publication which gives a detailed investigation of the Bayes linear framework - some of which is beyond the scope of this research project. In recent years, a number of other academics, such as Frank Coolen, Peter Craig, Michael Farrow, Anthony O'Hagan, Simon Shaw and Darren Wilkinson have supported Goldstein's theoretical developments and application of the methodology in a range of projects.

2.4.2 Bayes Linear Framework

2.4.2.1 Introduction

The Bayes linear theoretical framework is a quantitative methodology capable of expressing subjective beliefs and reviewing these beliefs once observations have been made (Goldstein, 1986). The Bayes linear methodology offers decision makers a methodology to structure, study and assess the relationships between coherent partial belief statements. These beliefs are adjusted by linear fitting as opposed to the traditional conditioning rules of Bayes theorem. It is natural to view the Bayes linear methodology as an inferential tool that is similar in philosophy to "traditional" Bayesian approaches, but has a number of distinct features which gives it certain advantages, such as simplicity and tractability, over these "traditional" approaches when modelling problems (Goldstein and Bedford, 2006). As such, more complex problems can be modeled given the same amount of time and effort. Goldstein and Wooff (2007) have suggested an alternative view of the Bayes linear methodology. They suggest that Bayes linear could be viewed as "complementary to the full Bayes analysis, offering a variety of new interpretive and diagnostic tools... where we lift the artificial constraint that we require full probabilistic prior specification before we may learn anything from the data" [Goldstein and Wooff, 2007]. If we consider the Bayes linear methodology as the foundational approach, Baye's rule actually emerges as a special case of the Bayes linear methodology where full conditioning is adopted (Goldstein and Wooff, 2007).

In a "traditional" Bayesian approach, the decision maker is forced to specify a multidimensional probability distribution over all unknown quantities, even if this is an unrealistic goal. They must specify joint probability statements about everything that could possibly occur, and in a problem of realistic complexity, this task will undoubtedly be challenging. A distribution and a likelihood function is said to be conjugate if the distributional class of the prior and posterior are the same (Bedford and Cooke, 2001). Conjugate prior distributions are used as they make it "relatively simple to carry out the process of Bayesian updating" [Bedford and Cooke, 2001]. If a conjugate prior is used for convenience, it becomes increasingly difficult to claim that there exists a connection between the model and the actual belief structure of the decision maker. However, if a conjugate prior is not used, the simple formulas for analysing belief structures are no longer available (Goldstein, 1988). In these cases, a difficult numerical integration calculation may be necessary.

Bayes linear offers a quick, simple method to perform inference in complex problems over statements of partial knowledge. Partial belief statements are used because it is believed that it is not possible to carry out a full joint prior probability specification over all possible outcomes (Goldstein and Wooff, 2007). Alternatively, the Bayes linear methodology allows the decision maker to specify coherent statements about the quantities they wish to model and revise these beliefs once observations have been made (Goldstein, 1986). Each belief statement should be considered carefully. As such, the Bayes linear methodology focuses on specifying and analysing a small number of carefully chosen collections of quantitative judgements about whichever aspects of a problem are within the ability of a decision maker to specify in a meaningful way.

2.4.2.2 Expectation as Primitive

Different authors have argued that expectation, not probability, should be used as primitive quantity when eliciting belief's De Finetti (1974) offered different operational definitions for directly eliciting expectation - most of which are based on penalties. Goldstein and Wooff (2007) revise these to give an operational definition in terms of a lottery ticket. Suppose X is a numerical random quantity. Assume that you can purchase a ticket where you receive X monetary units when the value of X is revealed. Then the expectation of X, $E(X)$, is the amount that you would be willing to pay for the ticket such that you are indifferent between the random gain X and the sure gain $E(X)$.

Uncertainty is measured in terms of variance where $var(X)$ represents the decision maker's degree of confidence in the magnitude of X. Further investigation on how this value is assessed can be found in Chapter 4. In order to assess the strength of the dependency between two related variables, the decision maker must specify $cov(X, Y)$, which represents the extent to which X may influence the decision maker's belief about Y.

Expectation is adopted as opposed to probability as it allows for simpler specification and adjustment. In problems where the decision maker is uncertain about a number of variables, it is believed that specifying the necessary joint probability distributions is a difficult and laborious task. It may be so difficult that the decision maker no longer has confidence in the values they were asked to specify, and as such, the additional benefit of the increased complexity of the model may offer no additional benefit. The Bayes linear methodology aims to overcome this limitation of a full Bayesian analysis by only requiring partial belief statements and allowing the decision maker to specify beliefs to the level of detail that they feel capable.

2.4.2.3 Temporal Sure Preference

Assume that the decision maker has specified her beliefs about a number of quantities, some of which can be observed. At a future time t, she wishes to revise her beliefs regarding these variables using a collection of observations $\mathbf{D} = [D_1, \ldots, D_k]$. In order to link the belief statements of a decision maker at different times, some form of temporal principle must be adopted. The full Bayesian temporal assumption states that a decision maker can anticipate and specify quantitative values for all possible eventualities and that such values will become the decision maker's posterior belief with no further analysis or reflection (Goldstein, 1997). It would seem unreasonable to adopt this assumption in complex decision making problems that exist over a long period of time. Instead, a weaker temporal principle is required.

Sure preference for X over Y is equivalent to $E(X) < E(Y)$ where X and Y are two small random penalties. Temporal Sure Preference (TSP) states [Goldstein, 2001]

> Suppose that we have a sure preference for X over Y at a future time t. Then we should not have a strict preference for Y over X now.

TSP can be viewed as a requirement to avoid sure loss (Goldstein and Wooff, 2007) and a method to constrain current uncertain beliefs based on knowledge of future inferences (Goldstein, 2001). However, as TSP only applies to sure preferences, it seems reasonable that the TSP is sufficiently weak to cover a large number of potential situations: "The main reason that you might have for not following the principle is that you might not now trust your future ability to make sensible judgements" [Goldstein, 1997]. In most problems, however, further observations are likely to place a decision maker in a stronger position to make judgements rather than a weaker one.

TSP appears sufficiently weak to adopt in a large number of problems but strong enough to establish the following results. Goldstein (1983) proves that if X is unbounded, $E(E_t(X)) = E(X)$, i.e. a decision maker's expected value now of their expected value of X at a future time t, should be equal to their expected value of X (Goldstein and Bedford, 2007). Using this and the TSP, Goldstein(1997) deduces

$$X = E_t(X) + R_t(X)$$

$$E_t(X) = E_D(X) + S_t(X)$$

where $R_t(X)$ and $S_t(X)$ have zero expectation and are uncorrelated with each other and all the observations of **D**. This implies that $E_D(X)$ informs the decision maker about $E_t(X)$ and $E_t(X)$ informs the decision maker about X. This is not to imply that $E_t(X) = E_D(X)$. However, some of the uncertainty of $E_t(X)$, and subsequently X, is resolved by the observations of **D**. In order to imply that $var(S_t(X)) = 0$, i.e. $E_t(X) = E_D(X)$, the Bayesian temporal principle must be adopted. As it is unlikely that the necessary assumptions can be made to set $var(S_t(X)) = 0$, a weaker temporal principle is required. As $var(S_t(X)) \neq 0$, the output of a Bayes linear model is not as precise as a full Bayesian model, however, it likely to be more realistic of the true situation. Goldstein and Wooff (2007) believe that in a number of simple situations, it is possible to consider the adjusted expectation and the posterior expectation to be the same with only a small loss of accuracy.

For more information on the foundations of the Bayes linear methodology, see Goldstein and Wooff (2007).

2.4.3 Bayes Linear Theory

A decision maker identifies an uncertain quantity of interest X or more generally, a vector of quantities $\mathbf{X}$. The decision maker specifies $E(\mathbf{X})$ and $var(\mathbf{X})$. The decision maker intends to observe some vector of quantities, $\mathbf{D}$, which they believe will improve their assessment of $\mathbf{X}$. From this, the decision maker must specify $E(\mathbf{D})$ and $var(\mathbf{D})$. In order to demonstrate the strength of the relationship between $\mathbf{X}$ and $\mathbf{D}$, the decision maker must specify $cov(\mathbf{X}, \mathbf{D})$. This matrix must be non-negative definite.

Once these values are gathered and the observation made, the decision maker can use the following formulae to adjust their prior assessments by linear fitting. The adjusted expectation of $\mathbf{X}$ given observation of a collection of quantities $\mathbf{D}$ written $E_{\mathbf{D}}(\mathbf{X})$ is defined to be the linear combination, $E_{\mathbf{D}}(\mathbf{X}) = \Sigma_{i=0}^{k} h_i D_i$, that minimises $E([\mathbf{X} - \Sigma_{i=0}^{k} h_i D_i]^2)$ over all collections $h = (h_0, h_1, \ldots, h_k)$ where $D_0 = 1$. Using the prior expectation, variance and covariance for the belief statements of $\mathbf{X}$ and $\mathbf{D}$, $E_{\mathbf{D}}(\mathbf{X})$ can be shown to satisfy

$$E_{\mathbf{D}}(\mathbf{X}) = E(\mathbf{X}) + cov(\mathbf{X}, \mathbf{D})(var(\mathbf{D})^{-\dagger})(\mathbf{D} - E(\mathbf{D}))$$

where $Var(\mathbf{D})\dagger$ is the Moore-Penrose generalised inverse.

Variance is used to quantify uncertainty. The adjusted variance of X given D is $var_{\mathbf{D}}(\mathbf{X}) = E([\mathbf{X} - E_{\mathbf{D}}(\mathbf{X})]^2)$ and is calculated as

$$var_{\mathbf{D}}(\mathbf{X}) = var(\mathbf{X}) - cov(\mathbf{X}, \mathbf{D})(var(\mathbf{D})^{-1})cov(\mathbf{X}, \mathbf{D})$$

An important part of any inference is how the output of the model is interpreted by decision makers. For a detailed description of the different interpretations of the Bayes linear output, see Goldstein and Wooff (2007). For the purposes of this project, the adjusted quantities are interpreted as follows;

- $E_{\mathbf{D}}(\mathbf{X})$ is viewed as an intuitive numerical summary of our subjective beliefs of $\mathbf{X}$ given the observations $\mathbf{D}$, i.e. an approximate estimator for $\mathbf{X}$.

- $Var_{\mathbf{D}}(\mathbf{X})$ is viewed as a strict upper bound on the expected posterior variance.

Note that the formula for adjusted variance, $Var_D(X)$, does not depend on the observed value of D. Therefore, it is possible to assess the amount of the decision maker's uncertainty is reduced prior to making any observations. If multiple observations are available, the reduction in variance can be assessed prior to making any observation.

In some special cases, this may in fact be an exact value for the expected posterior expectation and variance. One such special case is when the joint probability distribution between X and D is multivariate normal (Goldstein, 1999). Hence, it seems reasonable to assume that if the decision maker believes that the uncertain quantities are approximately joint normally distributed, then the posterior belief structure is approximately normal distributed.

2.4.3.1 Bayes Linear Networks

Graphical representations of a decision maker's collections of beliefs will be used extensively throughout the remainder of the thesis. Just as for the BBN's, graphical representations are an effective way of communicating between decision makers, analysts and those using the output of any modelling. They are useful for representing the complex structure of a real R&M problem.

The graphical representations used throughout the remainder of this research project will be referred to as Bayes linear Networks (BLN). In a BLN, a node represents an uncertain variable whilst an arc represents the potential for a source node to influence the decision maker's belief about the destination node. In the BLN, an arc does not necessarily represent a causal relationship but instead it represents the fact that the value of one variable influences a decision maker's belief about the value of another variable. Directed and undirected networks can be used within the Bayes linear framework, however, for the remainder of this project, only directed graphs are used.

Within the context of the Bayes linear framework, a BLN has a specific benefit other than simply graphically representing a situation (Goldstein and Wooff, 2007). The BLN allows the user to break down what may be a complex high-dimensional structure into a larger number of small low-dimensional components. By simplifying the overall problem into the form of a directed graph, it is only necessary to specify beliefs between neighboring nodes. These low-dimensional components provide for simpler elicitation.

There are rules that must be followed when building a BBN or BLN. An example containing three variables, A, B and C will demonstrate these rules for building BLN's. In order to build a BLN, it is necessary to first choose an ordering for the variables in the model. The order used in this example is [A, B, C]. The first step is to draw the node A. As there are no other variables in the model at this time, no arcs can be drawn. The next variable in the model is B, so another node is drawn. The decision maker believes that her belief about B is influenced by her belief about A, hence an arc is drawn from node A to node B. The next step is to introduce variable C. A node C is drawn. The decision maker believes that her belief about C is influenced by her belief about A, hence an arc is drawn from A to C.

An important part of using probabilistic modelling and building BBN's is the concept of conditional independence. Variables B and C are said to be conditionally independent given A if the conditional distribution of B given both A and C is the same as the conditional distribution of B given just A. The Bayes linear framework has a corresponding property which is important when building Bayes linear networks. A is said to be Bayes linear sufficient for B for adjusting C if

$$E_{A \cup C}(B) = E_A(B) \Rightarrow cov_A(B, C) = 0$$

The decision maker in this case believes that if she knew A, then her belief about C would not be changed by any observation of B. Hence, no arc is drawn from B to C. The graph for the above situation is represented by Figure 2.3.

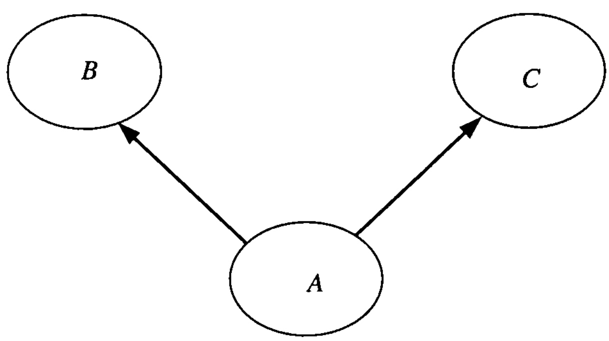

Figure 2.3: **Example of simple BBN**

2.4.3.2 Sensitivity Analysis

An important part of assessing the robustness of a model is through carrying out sensitivity analysis. Ferson et al (2004) define sensitivity analysis as "a general term for quantitative study of how the inputs to a model influence the results of the model" [Ferson et al, 2004]. The purpose of a sensitivity analysis is to "evaluate the sensitivity of model output to changes in model input" [Rabitz, 1989]. Evaluating the sensitivity of the output to changes identifies the areas of the modelling that are influential to the output and measures the robustness of the inferences of the modelling (Zitrou, 2007). Sensitivity analysis is carried out by systematically adjusting the inputs and observing the effect on the output of the model (Morgan and Henrion, 1990). Within a BBN framework, carrying out sensitivity analysis is generally difficult and time consuming (Coupe and van der Gaag, 2002).

As Bayes linear methods are computationally simpler than BBN's, carrying out extensive sensitivity analysis is simpler. Given that $\mathbf{u} = [U_1, U_2, \ldots, U_n]$ is the prior belief structure being modelled. The simplest method of carrying out sensitivity analysis is to identify a new vector $\mathbf{v}$ such that $\mathbf{v} = [V_1, V_2, \ldots, V_n]$ (Goldstein and Wooff, 2007). Goldstein and Wooff (2007) have developed a number of methods for analysing the effect of adopting the new belief structure $\mathbf{v}$ as opposed to $\mathbf{u}$. It would seem reasonable that in most decision making problems, the decision maker would not wish to carry out sensitivity analysis on all variables. Instead, it is likely that the effect of a relatively small number of variables on the output of the model will be investigated. Chapter 4, Chapter 5 and Chapter 6 will illustrate this.

For more information on sensitivity analysis, see Goldstein and Wooff (1994) and Goldstein and Wooff (2007).

2.4.3.3 Building a Non-Negative Definite Covariance Matrix

A key part of constructing a Bayes linear model is ensuring that the covariance matrix is non-negative definite. A symmetric real valued matrix with non-negative eigenvectors is always non-negative definite. However, checking a matrice's eigenvectors can be a tedious task and provide no guidance on changing the matrix if there are negative eigenvalues. It would be beneficial if another method of ensuring that the covariance matrix is non-negative definite could be adopted. We prove, by induction, that if we introduce a new variable to a Bayes linear model so that the variable is not a parent of a previous variable, the resulting variance matrix is non-negative definite. As every Bayes linear network can be built in such a way, a Bayes linear model constructed from a Bayes linear network is always coherent. In order to prove this for a $n \times n$ matrix, we first prove it for n=1 and then for n+1, assuming it holds for n.

If A is a 1×1 variance matrix with positive variance, then it is non-negative definite.

Suppose that A is an $n \times n$ variance matrix over a vector of quantities $B = [B_1, B_2, \ldots, B_n]^T$. Assume that A is non-negative definite. Hence every linear combination of B_i has non-negative variance, i.e. $var(\sum_{i=1}^{n}(\gamma_i B_i)) \geq 0$. A new variable B_{n+1} is introduced such that $B_{n+1} = \sum_{i=1}^{n} \alpha_i B_i + R$ where R is uncorrelated with everything in the model, i.e. B_{n+1} is not a parent of a previous node. Hence $cov(R, B_i) = 0 \; \forall \; i = 1, \ldots, n$. If A^+ is the new $(n+1) \times (n+1)$ variance matrix over $B^+ = [B_1, B_2, \ldots, B_n, B_{n+1}]$, then A^+ is also non-negative definite. This can be proven as follows.

$$A^+ \text{ is non-negative definite iff } var(\sum_{i=1}^{n+1} \gamma_i B_i) \geq 0 \forall \gamma_i \text{ but}$$

$$\begin{aligned}
& var(\sum_{i=1}^{n+1} \gamma_i B_i) \\
=& var(\sum_{i=1}^{n} \gamma_i B_i + \gamma_{n+1} B_{n+1}) \\
=& var(\sum_{i=1}^{n} \gamma_i B_i + \gamma_{n+1} \sum_{i=1}^{n} \alpha_i B_i + \gamma_{n+1} R) \\
=& var(\sum_{i=1}^{n} \gamma_i B_i + \gamma_{n+1} \sum_{i=1}^{n} \alpha_i B_i) + var(\gamma_{n+1} R) + cov(\sum_{i=1}^{n} \gamma_i B_i + \gamma_{n+1} \sum_{i=1}^{n} \alpha_i B_i, \gamma_{n+1} R) \\
=& var(\sum_{i=1}^{n} \gamma_i B_i + \gamma_{n+1} \sum_{i=1}^{n} \alpha_i B_i) + var(\gamma_{n+1} R) + \sum_{i=1}^{n} \gamma_i cov(B_i, R) + \gamma_{n+1} \sum_{i=1}^{n} \alpha_i B_i cov(B_i, R) \\
=& var(\sum_{i=1}^{n} \gamma_i B_i + \gamma_{n+1} \sum_{i=1}^{n} \alpha_i B_i) + var(\gamma_{n+1} R) + 0 + 0 \\
=& var(\sum_{i=1}^{n} \gamma_i B_i + \gamma_{n+1} \sum_{i=1}^{n} \alpha_i B_i) + var(\gamma_{n+1} R)
\end{aligned}$$

As A is non-negative definite, $var(\sum_{i=1}^{n} \gamma_i B_i + \gamma_{n+1} \sum_{i=1}^{n} \alpha_i B_i)$ is ≥ 0. As $var(\gamma_{n+1} R) > 0$, it can be deduced that A^+ is non-negative definite.

This result proves that if variables are added to an existing non-negative definite covariance matrix such that they are linear combinations of the previous variables included in the covariance matrix, then the subsequent covariance matrix will also be non-negative definite. From this, it is possible to deduce that a variance matrix built using a Bayes linear network will always produce a non-negative variance matrix. This result will be used extensively in Chapter 4, Chapter 5 and Chapter 6 to demonstrate that the covariance matrix is non-negative definite.

2.4.3.4 Application in Projects

The methodology has been applied in different projects in many different areas. The following are a list of projects that have applied the Bayes linear methodology. Craig et al (1996) used the Bayes linear methodology to estimate the level of hydrocarbons in a reservoir based on a computer simulation. Wooff et al (2002) reduced the number of software tests required to remove faults from software. O'Hagan et al (1992) estimated the assets of sub-parts of the UK water industry. Farrow et al (1997) used Bayes linear to support decision making within a brewery.

These papers were useful but limited with respect to their application in this project. This was for two reasons. First, these papers focused substantial time and effort on justifying that the Bayes linear was a sensible methodology to adopt. As such, little time was spent on explaining the modelling, discussing how to build the model, how to elicit the necessary values or how to carry out sensitivity analysis. Secondly, whilst these models were built to support decision making, none of them focused on the field of reliability engineering. As such, issues unique to the type of problem being addressed in this research project are not addressed in these papers.

2.4.4 Strengths of the Bayes Linear Methodology

Bayes linear has similar strengths as BBN's. The main additional strength of the Bayes linear framework is that it offers a quick, simple alternative method of performing inference in complex and unique problems. Whilst the principles are similar to a full Bayesian analysis, the specification burden placed upon the decision maker and the computational tools required to produce posterior beliefs is considerably lighter. Many of the weaknesses of the BBN methodology are addressed by this strength.

If structured in a sensible way, the number of values required from a decision maker can be vastly reduced when compared to the BBN methodology. Using mathematical equations to make deductions means the $cov(A,C)$ can be calculated if the $cov(A,B)$ and $cov(B,C)$ is known. This vastly reduces the number of values that need to be elicited, especially in large complex problems, and reduces the potential for the covariance to be non-positive definite.

As shown above, the elicitation burden placed upon the decision maker is considerably lighter in the Bayes linear framework than in the BBN methodology. Goldstein and Wooff (2007) suggest a number methods to gather the necessary quantitative specifications.

Due to the fact that it is simpler and quicker to implement models for new problems, the methodology does not require the extensive resources needed to build a BBN. It can be used to model very simple examples of only one or two variables where very little time is available, or it can be used to model very complex problems with many variables; problems that may be too complex to carry out in a BBN methodology.

2.4.5 Limitations of Bayes Linear Methodology

Whilst the Bayes linear methodology offers an alternative to BBN's and overcomes some of its weaknesses, the Bayes linear methodology also has some limitations.

One of the strengths of the Bayes linear theoretical framework is that it does not require the decision maker to specify a distributional form for the uncertain variables that they are attempting to model. As Bayes linear only requires the decision maker to specify his or her first two moments, it does not distinguish between symmetric and skewed variables. One method around this problem would be to specify and model higher moments, such as X^2.

A further limitation of the Bayes linear methodology is that the adjusted variance is always less than the original variance once an observation has been made regardless of what is observed. As can be seen from the formula for the adjusted variance above, as $var^{-1}(D) > 0$ and $cov^2(X,D) \geq 0$, then the adjusted variance must always be less than the original variance. This might not be satisfactory for all cases, however, this is also the case for the binomial/beta Bayesian models which is commonly used in Bayesian reliability analysis (Martz and Waller, 1982).

A final limitation of the Bayes linear methodology is that it does not allow for easy modelling of qunatiles. This is because no distributional form is assumed on the uncertain variables. Because of this, the information being fed back to the decision maker from the output of the model is not as detailed as the BBN methodology. The Bayes linear approach may be looked upon as a quick and simple methodology that approximates a full detailed probabilistic analysis and can potentially be used when the resources for a full analysis are unavailable.

2.4.6 Summary of Bayes Linear

Over the last twenty years, the Bayes linear theoretical and philosophical framework has been developed through papers in a number of different journals (Goldstein, 1975, Goldstein, 1983, Goldstein, 1984, Goldstein, 1986, Goldstein, 1998, Goldstein, 2001) and through a book (Goldstein and Wooff, 2007). The methodology has also been used extensively in many different areas, including sectors of public interest (O'Hagan et al, 1992).

On a theoretical level, it is a rigorous inferential methodology based on the theoretical development of de Finetti (1974). On a practical level, the Bayes linear methodology offers an alternative method to tackle challenging problems that previously would have been unfeasible or too costly using a "traditional" Bayesian approach (Goldstein and Bedford, 2006). Bayes linear methods can be used to focus on a decision maker's subjective beliefs about some uncertain quantities and provide a methodology to review these beliefs once observations have been made. Ultimately, the Bayes linear methodology should be viewed as a method to approximate the traditional 'probabilistic' Bayesian approach. However, it offers a logical and justifiable framework to handle problems in which it may only be possible to elicit partially specified beliefs.

2.5 Conclusion

Decision makers are constantly forced to make decisions where a large amount of uncertainty may exist and where observed data could be gathered to reduce this uncertainty. In the last thirty years, Bayesian methods have become increasingly popular for supporting decision making, particularly in the field of risk analysis. When building or analysing a reliability programme, decisions must be made as to what tasks should be carried out and how to interpret the observed data.

Adopting a Bayesian framework to support decision making within a reliability context seems sensible for a number of reasons. Firstly, adopting a Bayesian framework means the abundance of expert information that is collected throughout a reliability programme can be utilised. Secondly, if a strict frequentist theoretical framework is adopted, it is difficult to draw inferences from the observations collected prior to a system entering service and the system in-service. Thirdly, as reliability requirements and targets are typically becoming greater, more observations are required to demonstrate these targets within a frequentist framework to the extent that it becomes unlikely that certain targets can ever be demonstrated.

One technique that is potentially useful to support decision making is the BBN methodology. This methodology uses graphical representation to easily communicate dependencies and can produce posterior beliefs once observations have been made. However, the methodology has limitations - particularly when it comes to application. The main limitations of the BBN methodology are the elicitation burden that is placed upon the decision maker and the computational complexity. In cases where the decision maker may not have confidence in the values they have specified (they may have been forced to adopt a belief structure which does not represent their actual beliefs), the posterior beliefs generated by the model cannot be expected to give proper guidance.

An alternative technique that potentially overcomes some of the limitations of the BBN has been proposed. The Bayes linear methodology uses only first and second order information, hence reduces the specification burden placed upon the decision maker. As such, it is believed that the methodology could be applied to MOD projects to address some of the problems they currently face. However, in order to demonstrate that the methodology can be utilised by decision makers, it is necessary to develop the methodology in ways that are useful for MOD decision makers and to apply the methodology in 'live' MOD projects. Only by applying it in 'live' projects is it possible to support the claim that the Bayes linear methodology can aid MOD R&M decision making.

Chapter 3

Using Action Research to Develop the Bayes Linear Methodology

3.1 Introduction

The research field this thesis aims to make a contribution to is the assessment of subjective uncertainty in the context of reliability. However, much of the focus of this research centres on ensuring that the developed methodology can be used by those assessing reliability within the MOD. If the methodology developed cannot be practically applied or the MOD do not believe that the methodology is capable of meeting their needs, then the research goal will not have been achieved. As previously discussed, the four additional constraints placed upon the methodology by the client are

1. **Mathematically justifiable and defendable**
2. **Traceable**
3. **Efficient**
4. **Simple to use**

In order to demonstrate that the research goal has been met, these additional requirements must be achieved. To ensure that the methodology is developed to meet the needs of the MOD and that sufficient evidence can be collected to demonstrate this, the research methodology chosen must allow for interaction between the researcher and client. This chapter introduces action research, the research methodology adopted, and discusses how it was used in this project. Section 3.2 introduces action research, investigates the benefits of using action research and discusses the potential criticisms of action research. Section 3.3 discusses how action research was used in this research project. Section 3.4 will conclude the chapter.

3.2 Action Research

3.2.1 Introduction

Action research, since its inception by Lewin in the 1940's has gradually grown in popularity as a method for carrying out social-based research. Lewin first introduced the concept of action research in 1946 when he published "Action research and minority problems" (Lewin, 1946). This paper attempted to improve intergroup relations in American communities. Prior to his death in 1947, he published two further papers; "Frontiers in group dynamics, I" (1947) and "Frontiers in group dynamics, II" (1947) and a number of papers have been published posthumously since then.

Lewin was unable to systematically formulate the principles of action research due to his death however these early works introduced ideas such as 'action research', 'research in action' and 'cooperative research' that have now become common-place (Bargal, 2006). Lewin's early work was developed further by Chris Argyris (1970), Robert Rapoport (1970), Paulo Freire (1972) and Gregory Bateson (1972) amongst others. More recently, the "Handbook of action research", by Reason and Bradbury (2001) and "Fundamentals of action research" by Cooke and Cox (2005) have helped to elevate and give credible status to action research. Both these books give a comprehensive overview of action research development and the sub-methodologies that have risen out of action research in recent years.

Since it was created, the definition of action research has gradually developed and extended. Earl-Slater (2002) gives a concise summary of the evolution of the different definitions of action research. Three definitions will be used to demonstrate how action research has evolved since its inception. Lewin (1946) stated that action research was:

> "... a comparative research on the conditions and effects of various forms of

social action, and research leading to social action. Research that produces nothing but books will not suffice."

Rapoport's (1970) definition of action research was considered as the most frequently quoted by literature (Susman and Evered, 1978), and introduces the concept of collaborate research:

> "Action research aims to contribute both to the practical concerns of people in an immediate problematic situation and to the goals of social science by joint collaboration within a mutually acceptable ethical framework."

A more contemporary definition for action research is given by Reason and Bradbury (2001). They define action research as:

> "a participatory, democratic process concerned with developing practical knowledge in the pursuit of worthwhile human purposes ... It seeks to bring together action and reflection, theory and practice, in participation with others, in the pursuit of practical solutions of issues of pressing concern to people, and more generally the flourishing of individual persons and their communities

Gradually the definition of action research has developed to ensure that ethical issues were addressed and more recently, to ensure that positive social change is brought about. Recently, the focus of much effort has been to gradually develop methods to ensure that there are measures in place to assess the quality of action research projects (Editorial, 2006).

While most social science studies aim to develop casual links based on past evidence, action research focuses on using lessons from the past and the current situation to improve and develop an organisation's or individual's perception of a problematic situation (Chandler and Torbert, 2003). However, the term action research is often used loosely to generically describe action-oriented research (Coughlan and Coghlan, 2002) and due to this, it is often associated with unscientific research (Eden and Huxham, 1996). However, there is no reason why action research carried out in a controlled manner cannot contribute to knowledge in a different way to positivist approaches, but in an equally legitimate sense (Susman and Evered, 1978). Whilst the traditional positivist measurements of validation are inappropriate for action research, steps can still be taken to ensure the data is collected rigorously in order to demonstrate the validity of any conclusions.

Whilst action research is often referred to as a single method or methodology, it is more accurate to see action research as "a family of methodologies which jointly pursues action (or change) and research (or understanding) at the same time" [Earl-Slater, 2002]. In recent times, a number of different sub-groups of action research have risen including action science (Argyris et al, 1985), action inquiry (Torbert, 1976) and action learning (Revans, 1982) along with appreciative inquiry, co-operative inquiry and participatory action research (Dick, 2004).

All of these forms of action research are subtly different, however, they all aim to meet the underlying goal of action research which is to develop knowledge whilst positively changing a problematic situation. Action science is the study of improving the decision making of individuals or small groups when faced with difficult problems. Action learning aims to improve teaching methods by forcing participants to constantly review and critique their own actions and experiences. A lot of action research has focused on learning within hospitals. Participatory action research is action research where all the stakeholders participate in examining the problem, agreeing on an action aimed to improve the situation and then to review the result of any action.

The differences between forms of action research focus on the following; the degree of focus on problem-solving results versus learning and reflection, the way in which experts get involved in the process, the importance of participation and collaboration, the relative rigor of data collection and analysis procedures and the amount of time and level of system complexity (Marsick et al, 2003).

A key part of action research is that it is "a cyclical inquiry process that involves diagnosing a problem situation, planning action steps, and implementing and evaluating outcomes" [Elden and Chisolm, 1993]. Lewin himself observed that action research "proceeds in a spiral of steps, each of which is composed of a circle of planning, action and fact finding about the results of the action" [Lewin, 1946]. This idea of iteratively learning throughout the duration of the overall research project and within the individual projects identified is central to the philosophy of action research.

Action research aims to influence practice through theory. As Lewin observed, "there is nothing so practical as a good theory" [Lewin, 1951]. However, action research is no longer restricted to theory guiding practice. There is acknowledgment "that theory can and should be generated through practice" [Brydon-Miller et al, 2003] and that "theory is only useful insofar as it is put in the service of a practice focused on achieving positive social change" [Brydon-Miller et al, 2003]. Hence, action research should be seen as both theory guiding practice and practice guiding theory.

Action research theory development aims to create theory that is practical to given situations, positivist theory development aims to create universally applicable theory (Coughlan and Coghlan, 2002). Initially, action research was presented as an alternative to positivism as action research could tackle research areas that positivism was unable to (Kock et al, 1997). Researchers were disillusioned with the current techniques - "The research methods I knew well didn't fit my new situation. Either I found something else or I abandoned research altogether" [Brydon-Miller et al, 2003].

In the early development of action research, researchers were forced to take either an action research viewpoint or a positivist approach. Furthermore, International Review Boards (IRB) were created primarily from a positivist view. Therefore, many research projects that were non-positivist found it difficult to achieve status from these IRB's (DeTardo-Bora, 2004) or other journal review boards (Kock et al, 1997). Hence, for a considerable period of time, projects using action research as their research methodology were unable to gain academic recognition. However, in recent times, action research has grown in status and demonstrated that it should not be seen as the polar opposite to positivism (Kock et al, 1997).

3.2.2 Strengths of Action Research

Action research has a number of features that make it useful as a methodology in given situations. Action research is often seen as the bridge between solitary research and social practice (Kock et al, 1997). In contrast to solely using a scientific model, where the researchers objective is to study and understand the problem, an action researcher attempts to not only understand the problem but to offer potential solutions (Bargal, 2006). Action researchers share with pragmatists an underlying urge to develop theory that could be regarded as useful (Reason, 2003). They are generally less concerned with developing grand abstract theories but are more concerned with focusing on adapting techniques and current methods to address unsolved problems.

This is not to be confused with criticisms that action researchers are academic-consultants unconcerned with scientific suitability. By integrating themselves into the research problem and environment, action researchers have a deeper understanding of their research environment than their positivist counterparts. However, action researchers must take pro-active steps to ensure that, whilst embracing the research environment, they are not caught up in the action side of the research and that they continue to focus on creating scientifically defendable models or theories.

Action research is also seen as the link between similar theories in different environments, each of them adjusted for their unique setting. Eden and Huxham (2002) state, "It would be unusual for action research to deliver fundamentally new theories. Rather, the research insights are likely to link with, and so elaborate, the work of others". This idea links in with the belief that theory development is a continuous cyclical process (Eden and Huxham, 2002), which is also held by positivist Sir Karl Popper (Popper, 1963).

Another strength of action research is the rich data that can be collected through the interventions (Huxham, 2003). The data collection process becomes more that just collecting data. The process allows the researcher to further understand the context of the research and the challenges facing those who are providing the information - offering the opportunity to identify and develop the methodology or the theory. One of the underlying principles of action research is that any social science research carried out without a collaborative relationship between the researcher and the stakeholders is likely to be incomplete (Brydon et al, 2003).

With a survey or questionnaire, the researcher has no understanding of the conviction of the information attained. In action research, the researcher can witness and appreciate the difficulty that those questioned face and how what respondents say, may in fact differ from what they actually mean. The researcher and client can continuously re-diagnose the actual problem, identify possible alternatives, implement them and observe the positive outcomes of the intervention.

Action research is often performed in cycles, where the researcher develops, tests, and reflects (Kock et al, 1997). Mistakes and issues that arise can be further developed in updated versions. However, whilst many action research projects claim that this cyclical development process takes place, many do not document this process and the learning process that takes place between iterations (Kock et al, 1997). This is, however, an important feature of action research and the researcher should be clear about the "processes through which the resulting theory is developed" [Huxham, 2003]. This project aims to document the learning carried out at the different stages and how this influences future developments and actions.

Another benefit of action research is that it takes into account the expert local knowledge that stakeholders have. Whereas positivist researchers are concerned with objectivity and distance, action researchers focus on relevance, social change and validity, tested by those with the most vested interest in a project (Brydon-Miller et al, 2003). "Action and research are inherently intertwined in real life, not polar opposites of one another as they appear to be under the assumptions of empirical positivisms" [Chandler and Torbet, 2003].

Whilst in the short term action research might require more resources - time commitment from both researcher and client organisation - "it often results in a greater pay-off and impact in the longer run because it helps the organisation to:

1. Get critical input from key stakeholders, whose often-different viewpoints across complex systems are essential to an accurate read of the problem and robust design of a workable solution.

2. Increase the likelihood that studies will be implemented successfully because the implementers were involved in developing them.

3. "Fail fast" in unknown terrain by incrementing testing solutions, documentation of results, and use of lessons learned to design next steps.

4. Build individual and organisational capacity by creating habits and practices of continuous learning from experience and knowledge sharing in action." [Marsick et al, 2003]

3.2.3 Criticisms and Pitfalls of Action Research

One of the main criticisms of action research is that it often focuses too much on the action and to what extent the action can have an influence on its setting (Kock et al, 1997). This has led to the situation where action research is often likened to consultancy (McKay and Marshall, 2001) rather than scientific research. The criticism is that "if research becomes involved in practical action the ability to perform good research is lost" [Gustaven, 2003]. However, there is also criticism that if the researcher relies too much on some predetermined theory, the researcher may be unable to "see beyond their theory" [Huxham, 2003]. There is a need for a balance between the drive to develop knowledge and the drive to solve the problem.

There are criticisms that work developed in an action research framework can not be extended beyond the context that the project takes place in (Kock et al, 1997). As action research is often based within a single type of organisation or context, it is difficult to justify a generalisation of the results beyond context. This is very much a positivist criticism where the intention is to consistently develop general proofs and theories (Coughlan and Coghlan, 2002). If the results of the research, whether it be a model or findings, is intended to be extended beyond the context of the research, then using the validity criteria, as judged by positivists, the results may lie open to criticism (Huxham, 2003). However, if the results are to be confined to the research environment, defined as evaluating reliability cases within the MOD, then this issue of positivist generalisation is not an issue. The focus of this research is not to generalise the findings beyond the research environment the project is placed in.

Inter-personal issues and the biases they introduce is also an issue of concern for those who perform action research. It may not be completely clear what biases have been introduced through involvement with the research environment and therefore any findings may be difficult for other researchers to emulate in different environments (Kock et al, 1997). In order to address this, the researcher should be explicit about their role in the development of the model and their interactions with the client. This can be done through rigorous data collection methods such as recording and retaining any data collected in its natural form. This could range from questionnaires or transcripts of the clients responses or any flipchart modelling carried out. This ensures that any future reviewers can analyse the influence the researcher had on the developments.

In addition, action research requires that the stakeholders involved take an active interest in the development of the research. If the relevant stakeholders do not involve themselves in the collaborative process of development and implementation, the research is unlikely to be successful (Brydon-Miller et al, 2003). In the action research philosophy, different stages are identified (see Section 3.3) which, if carried out correctly, ensure that the client is included in the development process at a very early stage. Including the client in this early stage and allowing them to observe how the research is developing around their needs and requirements should hopefully ensure that all stakeholders are 'bought' into the process.

A further weakness of action research is the unplanned and informal structure that is sometimes encountered (Rapoport, 1970). Due to the cyclical nature of action research, it is possible that in some iterations, documentation of the steps taken are not recorded (Kock et al, 1997). This may lead to large steps in the reasoning of those carrying out action research - implying that the research was not carried out rigourously. As above, in order to ensure that the research is considered rigorous, steps must be taken to ensure that any knowledge development is captured as explicitly as possible.

Another criticism of action research is the lack of large-scale social change that comes about from the research it carries out (Brydon-Miller et al, 2003). Generally action research is carried out in a case-by-case basis. Within a very specific context the research can make a large contribution but fail to carry that contribution out of the local context and into the wider context.

To summarise, the aim of action research is to take appropriate action and create and report knowledge based on that action. Action research should be used in a situation where there is a demand for action and a demand for research. A demand for action should be justified by determining if the project is necessary or desirable for an organisation and whether or not there are any economic, political, social or technical forces pushing forth the need for modifications (Coughlan and Coghlan, 2002). With regards to research, the researcher must be asked why this project is worth studying, what contribution to knowledge it expects to make and why action research is an appropriate methodology for the project (Coughlan and Coghlan, 2002). Section 3.3 aims to cover this.

3.3 Action Research in this Research

The purpose of using action research in this research project is twofold. Firstly, in order to demonstrate that the research goals and aims have been met, it is necessary to apply the methodology on 'real' projects with 'real' decision makers. It would not be possible to justify that the methodology was simple to use and time efficient without applying it with real decision makers. Hence, the focus of the evaluation of the developed methodology will not focus heavily on the theoretical aspects of the methodology but on the practical aspects of the modelling. Secondly, by applying the methodology within the context of the MOD reliability case, rich first hand data about the issues concerning the application of Bayes linear methodology on 'real' problems can be gathered. It is hoped that this rich data will provoke further research ideas.

It is often relatively simple to determine what constitutes research, i.e. the idea that a contribution to knowledge has been achieved. A gap in knowledge is identified, sufficient research is carried out and reviewed by an academic peer. From this review and acknowledgement of achievement, knowledge is built upon.

However, it is often more difficult to define exactly what constitutes action (or intervention). Does action mean implementing a change in an organisations process or processes? Is it simply including the stakeholders of the research in the development process? Is facilitating development of an organisation action? Eden and Huxham (2002) believe that to qualify as action research, "it is not enough for the researcher simply to study the action of others". Other than this statement, the definition given to action is extremely vague and left open to the interpretation of those carrying out and reviewing action research projects. It is necessary to strictly define what action means in the context of this research project.

> *Action* is defined as any action that includes the *input* of the relevant stakeholder aimed at *developing* and *improving* the stakeholders understanding and ability to *address* a specific problem within a defined context.

This could include the following:

- Input from the client into identifying a problem of interest and the peculiarities of the problem such as attempting to provoke the client into explicitly identifying their problems, their weaknesses and their strengths.
- Identifying suitable methodologies and possible ways to develop the methodology through feedback from the client.
- Working on a specific problem faced by the MOD with the developed methodology in order to positively impact on the current problem.

- Through feedback sessions, changing how a client views a particular subject field and affect how they view similar problems in the future.

In this project, a wide range of actions will be carried out - some of which are obvious and can be pointed to and identified. It will be demonstrated that all of the above have been achieved at some point during the research project. For example, the interviews carried out with the MOD documented in Chapter 1 have aided in identifying an area of interest for research and the issues that must be taken into consideration when identifying potential methodologies. Chapters 5 and 6 will address a specific problem the MOD are facing and attempt to support their decision making. Some actions, however, may not be observable and an explanation may have to be given as to what these actions are and if they have actually taken place. Appendix B contains a time-line for the research project.

But how is action research actually carried out in practice? Stowell et al (1997) highlight that there appears to be little guidance to prospective action researchers whilst McKay and Marshall (2001) argue that there has been little attention to the reporting of the action research process. However, this is a little unfair to a number of authors who have attempted to give guidance to the *generic* action research project as well as many authors who have documented findings. Susman and Evered helped to inspire the action research cycle, Figure 3.1, as documented by Kock et al (1997). Susman and Evered (1978) (along with many other action research authors) believe action research is inherently cyclic in nature. They identify five stages in the generic action research project. Each of these five stages took place during this research project.

The first stage is diagnosing and at this stage the researcher and client work co-operatively to identify the problem facing the client organisation. In this project, the interviews carried out with the MOD R&M staff were used to define the problem and identify an improved state. The problem identified during the interview was that the MOD did not have access to the necessary techniques to fully evaluate all the information contained in the R&M Case Report. They believed that it would be beneficial if they had a technique that could assess the information in the R&M Case Report. Therefore, the improved state was identified as providing the MOD with a quantitative methodology that was capable of supporting the main decisions they faced during a procurement project. From these interviews, it was also gathered that the methodology

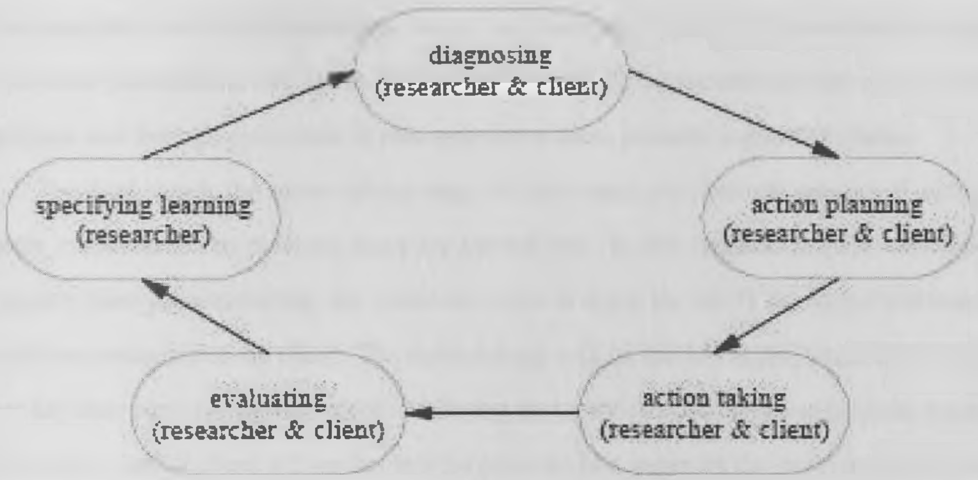

Figure 3.1: **The action research cycle by Kock et al**
Source: Kock et al, Can Action Research be Made More Rigorous in a Positivist Sense? The Contribution of an Iterative Approach, 1997

was required to be mathematically justifiable, efficient, simple to use and traceable.

The second step is action planning which involves the researcher and the client identifying alternative courses of action that are available to move from the current state to the improved state. During this phase, different methodologies were researched, a methodology was identified and discussed in detail with the client. The usefulness of the new methodology was investigated collaboratively and from this, the Bayes linear framework was identified as being potentially useful. If action research had not been used, the researcher would have been unable to gather feedback regarding the feasibility of the proposed approach. The MOD and researcher then attempted to identify potential projects where the methodology could be applied. In order to ensure that the research carried out would remain as action research and to optimise the potential of successfully completing the research, the following list of criteria were proposed.

- The project was a 'live' project with decisions still to be made.
- The decision maker on the project was enthusiastic and committed to addressing the problem using an alternative methodology.
- The project fell in the time periods identified in Chapter 1.

At this stage, the researcher presented his work to the MOD's R&M HOS department with the hope that two R&M Specialists would volunteer their time and knowledge of a project. From this presentation, two MOD R&M decision maker's volunteered to work on two different projects and through discussion, it was agreed that these projects would be suitable.

The third step is the action taking stage. At this stage, the potential courses of action that were identified in the previous stage are carried out. In this research project, once suitable projects have been identified, the researcher aims to apply the newly developed methodology with the assistance of the client. The methodology will be applied to projects at the two different key time periods that were identified during the interviews. Within these projects, the action researcher and the client will go through the previous two stages for the individual projects - i.e. diagnosing the specific problem faced on the project and what specific action to take. Chapters 5 and 6 cover the work carried out in each project.

The fourth and fifth stage of the action research cycle focus on reflecting on the research. The fourth stage is the evaluating stage. During this stage, the client and researcher study the outcome of any actions taken. This is an important stage in validating the work carried out. The effect of the modelling within each project will be assessed. This will be done by conducting interviews with the decision maker on each project prior to carrying out the modelling and gathering information on how the information in the R&M Case was assessed and the decisions that were made at that time. Once the modelling is complete, the decision maker will receive feedback regarding the output of the model. Once this is done, a further interview will be carried out to gather information on two issues; firstly, what action was taken once the modelling was complete and secondly, what benefit did the decision maker gain from carrying out the modelling. The first aspect of the interview is to identify if the modelling had any effect on the decision making in the project. If the modelling did not affect the decision made, it may be that the modelling supported their decision. This will be investigated during each interview. As above, this will be covered in Chapter 5, Chapter 6 and also addressed in Chapter 7.

The second aspect of the interview aims to measure what the decision maker gained from carrying out the modelling in a wider sense, i.e the skills they learnt, alternative methods to structure the R&M Case, etc and whether or not they would use this approach in the future. This is the final stage of specifying learning. In this stage, the researcher studies the outcome of the evaluating stage and attempts to build knowledge by understanding the situation studied. Traditionally, this is the only stage that involves the researcher working in isolation, however, recent action research projects have advocated including the client in the specifying learning stage (Elden and Chisolm, 1993). In addition to the interviews with each decision maker, a feedback session was carried out by the researcher with the client to summarise the totality of work carried out during the project. During this session, information was gathered from the participants as to their thoughts on the overall project and possible ways to improve it in the future. In addition to this feedback session, a smaller workshop session was carried out with six MOD R&M Specialists where they applied the Bayes linear methodology on a simple problem. From this, information was gathered on whether or not they believed the Bayes linear to be a useful methodology that they could use in the future. The results of this feedback session will be discussed in Chapter 7.

Due to its flexibility and dynamic nature, the course of an action research project should not be seen as linear as demonstrated in Figure 3.1. Instead, it is possible for the researcher to revisit the diagnosing stage after they have carried out the action planning stage, or any other stage. Due to this, Kock Jr et al (1997) cycle can be extended - more similar to the diagram presented by Susman and Evered (1978). This is shown in Figure 3.2.

In order to take into account that multiple projects were worked on, the cycle was extended. During the individual projects, the researcher and client had to go through the previous stages carried out, such as diagnosing the specific problem within the project, identifying the correct course of action and then taking the necessary action. As each procurement project is inherently unique, it would have been unwise to identify the 'correct' course of action prior to fully diagnosing the problem the decision maker faced. Hence, a new action research cycle has been developed to show the path the researcher travelled, as shown in Figure 3.3. The large cycle represents the overall research project that aims to ensure the research goal is met. The smaller cycle represents the individual projects carried out with the MOD. Within these smaller cycles, the individual research aims will be met as well as demonstrating that the additional requirements are achieved.

In addition to following the action research cycle, Susman and Evered (1978) highlight that the following six criteria should be met when carrying out an action research project. The

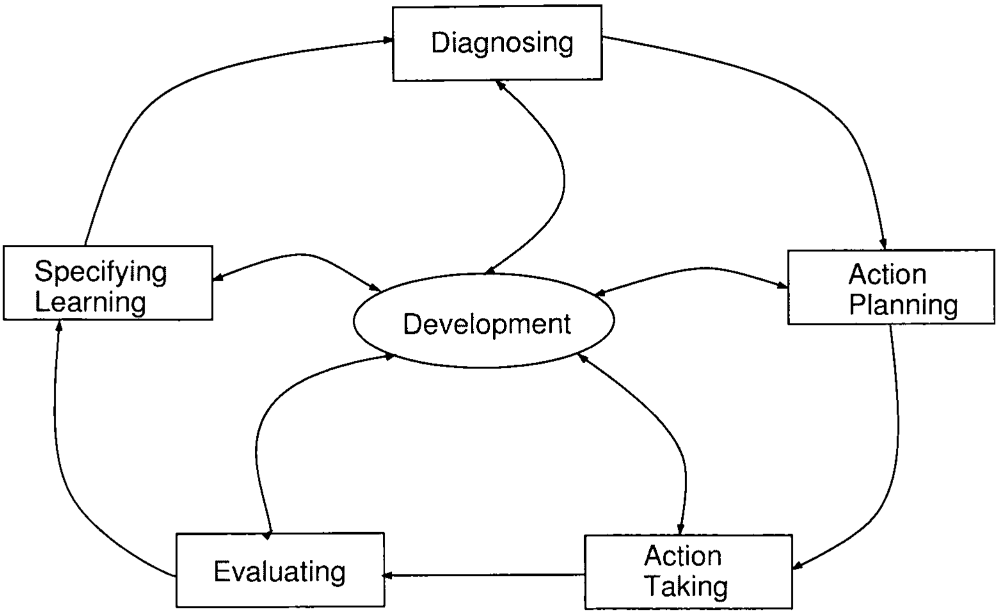

Figure 3.2: **The action research cycle by Susman and Evered**
Source: Susman and Evered, An Assessment of the Scientific Merits of Action Research, 1978

project should be future oriented, collaborative, involve system development, generate theory grounded in practice, be agnostic and situational. It is useful to identify how each of these are met in this research project.

Action research should be attempting to create a more desirable future for some organisations, such as the MOD as examined in this research. This project aims to provide the MOD R&M evaluators a methodology capable of capturing and structuring their subjective beliefs and that can carry out analysis that previously they were unable to do. This chapter has already demonstrated that a large part of this research has taken place collaboratively with the MOD. This has ensured that at all times, the MOD was consulted and their opinion valued when planning future steps.

Action research should aim to support system/organisational development. This is important in this research project as it is believed that many of the organisational developments that have taken place during the research process may not be obviously observable or measurable. For example, it is hoped that this research has provoked those within the MOD to think about reliability and uncertainty issues in a new manner. It is important these changes are identified

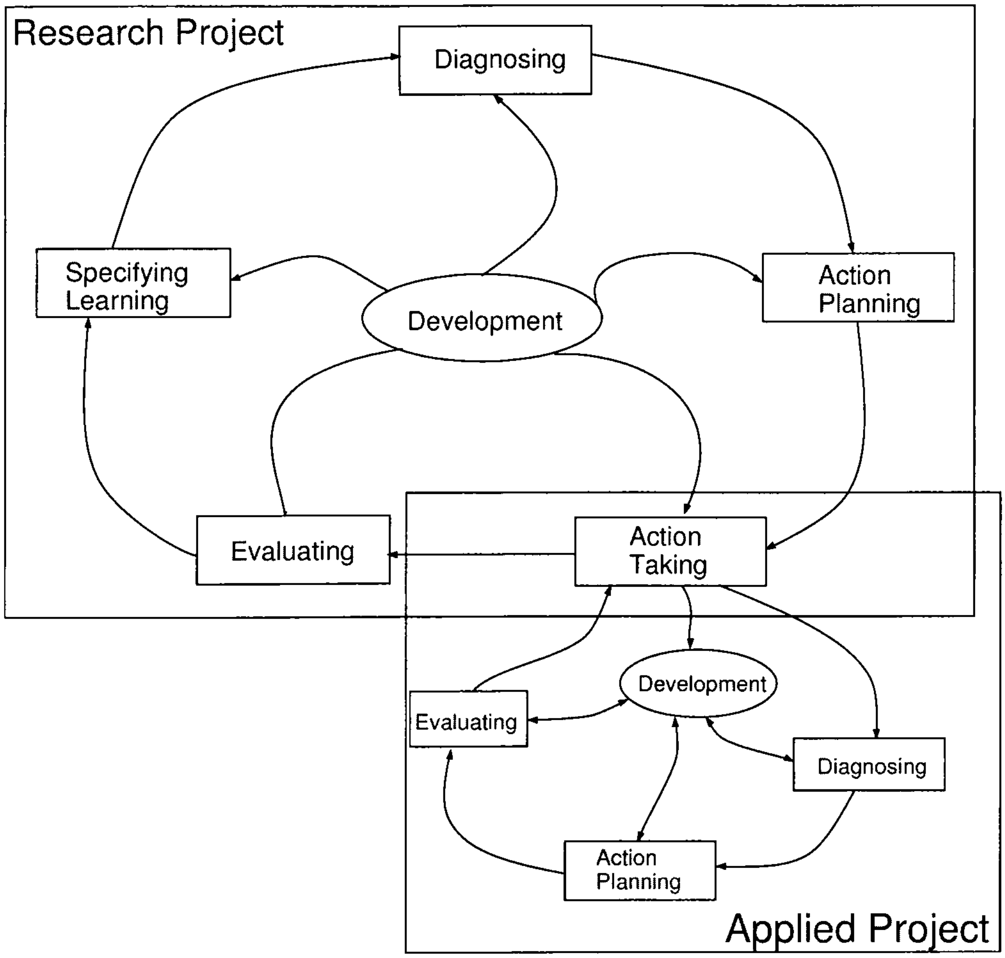

Figure 3.3: **The action research cycle in this project**

through interviews with the relevant MOD R&M participants. This will be discussed in more detail during Chapters 5 and 6.

Action research aims to generate theory that is grounded in action. The constraints placed upon the developed methodology means that the theory is not being guided by past theory but instead by the needs of those it is aiming to support. From the literature, it appears that whilst there is still a desire for approaches that can provide detailed summary of a situation, there is also a demand for methodologies similar to that developed in this research project.

As previously stated, action research is unlikely to generate new theory; instead adopting theory from one environment and adapting it for another environment. This is similar to the work carried out in this research project. The methodology chosen has already been used in decision making problems in other areas as will be demonstrated during Section 2.4.3.4. This research will focus not only on how it can be adapted to address reliability issues but also developed to address the needs of the MOD R&M evaluator.

Understanding the defence environment that this research takes place in is critical to developing any methodology. The additional constraints placed upon the project were unique to the research environment. By adopting a purely theoretical research methodology or not fully understanding the research environment, could have led to the possibility that the methodology would not have been useful to the MOD.

Different sub-methodologies of action research were identified above, such as action learning and action science. In this research project, a mixture of participatory action research and co-operative inquiry will be carried out. Both these approaches are similar with the main focus of each to perform research 'with' people rather than 'on' people in order to learn the current state of the world and how to act to make a positive change. The flow of information in this research project aimed to be two-way as much as possible. In order to do this, the action researcher aimed to ensure that the client was made aware of all developments as much as possible. This ensured that the client could offer potential feedback as often as possible.

By using participatory and co-operative inquiry, the researcher gained access to rich first hand data that may not have been possible with alternative methodologies. Furthermore, by incorporating those in the research environment into the development process, the researcher was able to gain feedback and insight, not only throughout the development process, but also at the end of the project once the model has been developed.

The relationship between the action researcher and the client is an important one that had to be managed throughout the project. The researcher did not view himself as the 'expert' entering the client organisation to solve all its problems. Instead the researcher brought with him knowledge of a particular approach which he believed, theoretically, could improve the situation of a client. The client, alternatively, brought practical knowledge, experience and a deep understanding of a situation that the researcher could not hope to develop within a short time frame. As Susman and Evered (1978) state "Neither client nor researcher has better knowledge; in a sense, they are both experts".

3.4 Conclusion

This chapter has aimed to justify the use of action research in order to gather the necessary data to demonstrate that the research questions in Chapter 1 have been met. There are a number of critical points that make action research the appropriate research methodology choice in this project.

Firstly, the research questions aim to ensure that any developed methodology meets a current knowledge gap within the MOD, i.e. how to quantify and structure a decision maker's belief and carry out inference once observations have been made. In order to demonstrate that a methodology has been developed which meets this aim and in particular, the additional restrictions the MOD face, it is necessary to develop and apply the methodology with the MOD and validate the methodology based on feedback given by the MOD. In addition, due to the subjective nature of the modelling, it is important that the projects are 'live'. It is unlikely that useful analysis could be carried out retrospectively on any MOD project due to the introduction of a number of biases and possible lack of motivation on the part of the decision maker. Instead, by working on a 'live' project, the decision maker is highly motivated to support the researcher and provide in-depth feedback.

The following chapter will discuss a number of practical issues regarding the application of the Bayes linear methodology that must be addressed before the methodology is applied on 'real' projects. These include elicitation, sensitivity analysis, building the model and carrying out inference.

Chapter 4

Applying the Bayes Linear Methodology

4.1 Introduction

The aim of this chapter is to show how the Bayes linear methodology can be applied in the reliability case context. In order to demonstrate this fully, practical issues, such as elicitation, building a Bayes linear network (BLN), carrying out inference and sensitivity analysis, have to be addressed. The chapter will take the following form. Section 4.2 will introduce how elicitation can be carried out. Section 4.3 and Section 4.4 will introduce two theoretical reliability examples and work through the detailed steps needed to build the BLN, populate and extract the maximum value from the model and Section 4.5 will conclude and summarise the chapter.

4.2 Elicitation in a Bayes Linear Framework

For a Bayes linear model with variables $X_1, \ldots, X_n$, to populate the model it is necessary to elicit $E(X_i)$ for $i = 1, \ldots, n$ and $cov(X_i, X_j)$ for $i, j = 1, \ldots, n$. This section is broken into three parts; an introduction discussing elicitation issues, a discussion of different methods to elicit means and variances, and a discussion of different methods to elicit covariances. This section is not expected to be a detailed discussion of issues surrounding the development of elicitation methods and processes, however, a short introduction is given. The aim of this section is to highlight the most appropriate methods for future users to gather the necessary information required to populate a Bayes linear model.

4.2.1 Introduction to Elicitation

For subjective Bayesians, elicitation is a key part of the overall applied process (Kadane and Wolfson, 1998). The purpose of an elicitation session is to gather quantitative data that represents the belief of the person being elicited (Garthwaite et al, 2005): "Empirical data are usually insufficient to quantify the uncertainty in the consequences of a course of action, and judgmental probabilities provide a logical means for overcoming this limit" [Merkhofer, 1987].

Whilst the work carried out in developing elicitation processes is relatively sparse compared to the numerous decision support models available to decision makers (Quigley, 1998), in recent times, more effort has been spent in documenting how elicitation has been carried out (Kadane and Wolfson, 1998, O'Hagan, 1998, Craig et al, 2001, Walls and Quigley, 2001). A number of books have been written discussing elicitation, the most notable of these are Yates (1990), Morgan and Henrion (1990), Cooke (1991), Meyer and Booker (2001), Koehler and Harvey (2004) and O'Hagan et al (2006). These books provide a comprehensive summary of the biases that may arise during an elicitation session such as availability, representativeness and anchoring (Tversky and Kahneman, 1974), the high-level steps required during an elicitation process and the different methods available to elicit probability distributions.

There is extensive research in the field of psychology evaluating the effects of different biases; ordering of questions (Payne, 1951), estimating probabilities (Hogarth, 1980) and effects of problem decomposition (Hayes-Roth, 1980) amongst others.The psychology literature suggests that it is not feasible for experts to simply state their beliefs accurately about subjective probabilities (Kahneman et al, 1982) and instead it is necessary to develop an elicitation process that support the expert/decision maker in communicating their beliefs. Different techniques to elicit prior beliefs may, however, produce different output as each "method of questioning may have some effect on the way the problem is viewed" [Winkler, 1967]. The aim when developing an elicitation procedure is to attempt to minimise these biases.

There is consensus in the literature regarding which high-level steps should be carried out during an elicitation process (O'Hagan et al, 2006). Clemen and Reilly (2001) suggest that any elicitation process should include seven stages: background, identification, recruitment of experts, motivating experts, structuring and decomposition, probability and assessment training, probability elicitation and verification, and aggregation of expert's probability distributions. This process assumes multiple experts and distinguishes expert and decision maker.

The context of the elicitation process in this research project is slightly different as the expert and decision maker is interchangeable. The expert is familiar with the problem that they are facing, they do not need to be identified or recruited and they are positively motivated to address the problem. It is likely that only one expert/decision maker will be used in the MOD projects at any one time, therefore there is no need to aggregate experts' distributions. Much of the work carried out in developing elicitation processes has focused on developing methods for multiple experts (Phillips, 1999, Clemen and Reilly, 2001, Walls and Quigley, 2001). Other authors such as Merkhofer (1987) and Garthwaite et al (2005) suggest processes which focus on one expert. Garthwaite et al (2005) suggest a four-stage process suitable for scenarios where elicitation is carried out with single experts or decision maker's. This four-stage process will be applied in the MOD projects and contains: set-up, elicitation of summaries, fit distribution and assessing adequacy of elicitation. One important change will be made; instead of 'fit distributions', this will be changed to 'calculate mean and covariances'.

In addition to identifying key *steps* in an elicitation process, it is necessary to identify the key *roles* within an elicitation process. Bedford et al (2006) identify three key roles: decision maker, expert and analyst. Alternatively, O'Hagan et al (2006) identifies four roles: decision maker, expert, statistician and facilitator. During the MOD projects, the researcher will adopt the role of the statistician and the facilitator. As the MOD participant has expert knowledge in the project and the associated reliability case, and the methodology aims to support the advice they provide to the IPTL, the MOD participant will be both the decision maker and expert. For the remainder of the report, 'she' will refer to the expert/decision maker on a project while 'he' will refer to the statistician/facilitator. It is unlikely that the MOD decision maker should adopt all four roles. If they did this, it is possible that during the development of models or during elicitation, biases will be introduced because of their close involvement in the problem. As such, it is likely that it would be beneficial to have a facilitator to support the decision maker.

4.2.2 Eliciting Means and Variances

Meyer and Booker (2001) believe that elicitation techniques should be designed to fit the way people think instead of forcing experts to answer questions in a specific format. An expert in a subject domain should not be viewed as an expert in statistics or probability theory; any elicitation method must be as simple as possible for experts to communicate their beliefs accurately. Within a Bayes linear framework, the decision maker's belief about the expectation, variance and covariance between the parameters in the model must be assessed. Eliciting means and covariances directly can often be difficult as experts do not naturally think in these terms. Kadane and Wolfson state, "experts should not be asked to estimate moments of a distributions (except possibly the first moment); they should be asked to assess quantiles or probabilities" [Kadane and Wolfson, 1998]. From these quantiles, the mean and variance can be calculated. Garthwaite and O'Hagan (2000) agree that it is necessary to use quantiles to estimate at least the variance of an expert's subjective belief.

Methods have been developed which focus on eliciting different percentiles of a variable and then calculating the mean and variance of the variable directly. Garthwaite et al (2005) provide a summary of different elicitation methods, however, all of these methods focus on fitting a distribution to the elicited quantiles. O'Hagan (1998) describes a method to elicit variance that was used to populate the Bayes linear model in O'Hagan et al (1992). O'Hagan (1998) assumes either a normal or log-normal distribution and elicits three percentiles: the median, M_X; the 17th percentile, L_X and; the 83rd percentile, U_X. No justification is given to why O'Hagan chose the 17th and 83rd percentile other than he felt that experts were overconservative when specifying 50% intervals and that larger intervals, such as a 99% range, were strongly influenced by events of small probabilities (O'Hagan, 1998). If $U_X - M_X$ is substantially greater than $M_X - L_X$, then a log-normal distribution is assumed. O'Hagan gives no guidance on the numerical value such that $U_X - M_X$ is substantially greater than $M_X - L_X$. If a normal distribution is assumed, the standard deviation is set equal to $(U_X - L_X)/2$. O'Hagan (1998) states that whilst this method is "very crude and simplistic", due to the number of quantities that were required to populate the model, a simple approach was required. This method appears reasonable if the expert's beliefs are regularly normal or log-normal. In situations where this assumption cannot be made, an alternative method is required.

Pearson and Tukey (1965) suggested a method specifically for the mean and variance which uses 3 percentiles for the mean and 5 for the variance. Keefer and Bodily (1983) further developed the method for eliciting the variance so that the analyst was only required to specify 3 percentage points: "The applicability of this approximation is significantly enhanced by eliminating the need for the 0.025 and 0.975 fractiles, especially in light of the difficulty of assessing these points in the tails" [Keefer and Bodily, 1983]. Given that the decision maker specifies her 0.05, 0.5 (the median) and 0.95 percentiles for their belief about the variable of interest, the analyst can calculate a mean and variance for X. The Pearson and Tukey formulas are as follows:

$$
\begin{aligned}
\text{mean} &= 0.63x(0.5) + 0.185[x(0.05) + x(0.95)] \\
\text{variance} &= ([x(0.95) - x(0.05)]/[3.29 - 0.1(\Delta/\sigma_0)]) \\
\text{where } \Delta &= x(0.95) + x(0.05) - 2x(0.5) \\
\text{and } \sigma_0 &= ([x(0.95) - x(0.05)]/3.25)^2
\end{aligned}
$$

Keefer and Bodily (1983) found that the 3 - point Pearson and Tukey method performed almost identically to the 5 - point Pearson and Tukey method when assessing the variance. A number of alternative methods have been proposed in the literature; the PERT method (Malcolm et al, 1959), Moder-Rogers (Moder and Rogers, 1968), Davidson-Cooper (Perry and Greig, 1975) and Swanson-Megill (Megill, 1977). Comparisons with these methods found that the 3 - point Pearson and Tukey method performed best when assessing the mean and was second only to the 5 - point Pearson and Tukey method. Keefer and Bodily (1983) and Keefer and Verdini (1993) found that when estimating the mean and variance of the Beta distribution, which is widely used in risk analysis (Johnson, 2002), the Pearson and Tukey method performs the best. The maximum error using the Pearson and Tukey method was found to be -1.7%. This was substantially lower than the maximum error found with the Moder-Rogers method, -20.7%, Davidson-Cooper method, -17.7%, and the PERT method, 5506%. Johnson (2002) extended the analysis of the accuracy of the Pearson and Tukey method to alternative distributions such as the Gamma and log-normal. Johnson found that the Pearson and Tukey Method was robust when assessing the mean with a maximum error of 0.05% for the gamma and 0.68% for the Log-normal. For the standard deviation, the method was again robust for the Gamma distribution with a maximum error of 0.7%, however, for the Log-normal distribution, the maximum error for the standard deviation was 11.6%. Martz and Waller (1982) suggest that the Beta distribution is widely used in reliability analysis. The 3-point Pearson and Tukey method performs well for the Beta distribution. Throughout the remainder of the thesis, the 3-point Pearson and Tukey method will be used to elicit the mean and variance of any variables we wish to specify. It will be referred to throughout the rest of the thesis as the Pearson and Tukey method.

4.2.3 Eliciting Covariance

4.2.3.1 Introduction to Dependency

In almost all complex but real-world decision making problems it is inevitable that there exists a degree of dependency between the variables being modeled. Where there is a lack of data on the relationships between variables, subjective expert judgement is crucial in modelling these problems and providing information on the strength of the dependencies between variables (Meyer and Booker, 2001). Whilst much attention has been given to building influence diagrams and modelling the qualitative dependencies between different variables, less attention has been given to how to assess or elicit the quantitative dependency from experts (Clemen et al, 2000). This is a crucial part of the modelling as the output of any model could be sensitive to the dependency values specified. One of the advantages of the Bayes linear methodology is the simple way dependency between variables is specified.

4.2.3.2 Proposed Methods for Assessing Covariance

Clemen et al (2000) believe elicitation methods must have strong theoretically rigorous foundations, be able to be generalised to a wide range of different problems and should have a clear intuitive interpretation. In addition to this, those using the techniques should find them "easy and credible" [Clemen et al, 2000]. In the Bayes linear framework, covariance is the measure of dependency, however, eliciting the covariance directly from an expert is likely encounter problems (Garthwaite et al, 2005). Instead, a method to calculate the covariance from other elicited values must be used.

Due to the way in which prior beliefs are specified in a Bayes linear framework, there are a limited number of ways in which the dependency value between two variables, i.e. $cov(X, Y)$, can be specified. Four different methods of determining the dependency between two variables, X and Y, in a Bayes linear framework were identified: DC (Direct Calculation), C (Direct Elicitation of Correlation), AE (Adjusted Expectation) and AU (Adjusted Uncertainty). Details of each method is given below and in each example described below, the facilitator is attempting to elicit the beliefs from an expert who is referred to as "she".

Direct Calculation (DC) For DC, two variables are modelled using the formula $Y = \alpha X + R$ where it is assumed that X and R are uncorrelated. In this case, X is viewed as an explanatory variable of Y. It is assumed that the expert has already specified her $E(X)$ and $var(X)$. The expert then specifies $E(Y_t)$, her new belief of $E(Y)$ given that $E(X)$ has increased by t. From this, α can be calculated using the formula $\alpha = \frac{(E(Y_t)-E(Y))}{t}$. She then specifies her 5, 50 and 95 percentiles for the uncertain variable R. This can be done by asking her, "given that X is known to be $\tilde{x}$ with complete certainty, what are the 5, 50 and 95 percentiles for Y?" $E(R)$ and $var(R)$ can be calculated using the Pearson and Tukey method and from this, $E(Y) = \alpha E(X) + E(R)$, $var(Y) = \alpha^2 var(X) + var(R)$ and $cov(X, Y) = \alpha var(X)$.

Direct Elicitation of Correlation (C) For C, the expert first has to specify $E(X)$, $E(Y)$, $var(X)$ and $var(Y)$. The expert is then asked to state her assessment of the correlation between variables X and Y. To do this, she must have an understanding of what correlation is. From $corr(X, Y)$, $cov(X, Y)$ can be calculated using $corr(X, Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \Rightarrow cov(X, Y) = corr(X, Y)\sigma_X \sigma_Y$.

Adjusted Expectation (AE) For AE, again the expert must specify $E(X)$, $E(Y)$, $var(X)$ and $var(Y)$. She is then told that she is to consider her belief about X given that she can observe the true value for variable Y. The expert assumes that the true value of variable Y is Y_i. She is then asked to specify her new belief about $E(X)$, $E_Y(X) = X_y$, given that she has observed Y_i. By rearranging the formula for adjusted expectation, $cov(X, Y)$ can be calculated.

$$cov(X, Y) = \frac{(E_Y(X) - E(X))var(Y)}{Y - E(Y)} = \frac{(X_y - \overline{X})\sigma_Y^2}{Y_i - \overline{Y}}$$

From $cov(X, Y)$, α can be calculated from $Y = \alpha X + R$. As $cov(X, Y) = \alpha var(X)$, $\alpha = \frac{cov(X,Y)}{var(X)}$. In order to ensure the expert is coherent, there are limitations on the values that can be specified. As $var(Y)$ and $var(X)$ have already been specified, there exists an upper and lower bound on the value for α. As $var(R) \geq 0$, $\alpha^2 var(X) \leq var(Y) \Rightarrow \alpha \leq \sqrt{\frac{var(Y)}{var(X)}}$. Assuming that $cov(X, Y) \geq 0$. Then an upper bound for the value the expert can specify, X_Y, is

$$\begin{aligned} X_Y &= E(X) + cov(X, Y)var^{-1}(Y)(Y_i - \overline{Y}) \\ &= \overline{X} + \alpha\sqrt{\frac{var(Y)}{var(X)}}var^{-1}(Y)(Y_i - \overline{Y}) \end{aligned}$$

If the expert specifies a belief greater than this value, then $cov(X, Y) > \alpha\sqrt{\frac{var(Y)}{var(X)}}$. However, as this is an upper bound, this means that one of the other values has been incorrectly specified. In this case, the facilitator must revisit and reassess with the expert the previously elicited values.

Adjusted Uncertainty (AU) For AU, again the expert must specify $E(X)$, $E(Y)$, $var(X)$ and $var(Y)$. She is then told that the value that she initially believed variable Y would take, i.e. $\overline{Y}$, has been observed. She is then asked to specify her adjusted variance for X, $\sigma_{X_i}^2$ given that she now knows the true value of Y. This can be done using the Pearson and Tukey formulas. Rearrangeing the Bayes linear formulas, $cov(X, Y)$ can be calculated using the formula:

$$cov(X, Y) = ((var(X) - var_Y(X))var(Y))^{\frac{1}{2}} = ((\sigma_X^2 - \sigma_{X_i}^2)\sigma_Y^2)^{\frac{1}{2}}$$

As with AE, there are limitations on the values that the expert can specify in order that coherency is maintained. In this case, the expert cannot specify that the adjusted variance, $var_Y(X)$ is greater than the prior variance, $var(X)$. In addition, in order that $var_Y(X) \geq 0$,

$cov(X, Y) \leq \sqrt{var(X)var(Y)}$.

4.2.3.3 Assessing the Four Methods

In order to assess the effectiveness of the four methods, elicitation sessions were carried out with volunteers. During these sessions, three dependencies were presented to each of the participants. These dependencies were: life expectancy between males and females in the same country, height and weight of male students at a university and mean miles to failure (MMTF) between a population of vehicles and the miles to failure of a single vehicle. These dependencies were chosen in order that three different scenarios were investigated. Scenario 1 assessed two dependencies with the same unit of measure; Scenario 2 assessed two dependencies with different units of measures; and Scenario 3 assessed the dependency between a population and a single observation. For each of these three dependencies, expectations and variances for each of the variables were gathered and four correlation values were elicited from each participant using each of the described methods.

There were a total of 23 participants, with a total of 69 responses for each method. All participants were postgraduate students of the University of Strathclyde Business School. None of the participants received formal training in the Bayes linear methodology. An example was carried out and a short explanation on how to interpret the values was given. There were three distinct groups; ten MSc students in Operational Research, nine PhD students from the department of Management Science and four PhD students from the Business school. The MSc students had all recently carried out a course in statistics where they had learned about correlation, means and variances. All the PhD students had attended a short course in Advanced quantitative methods where they also learned about correlation, means and variances. However within the MSc students and the PhD students in Management Science, there was a mix of statistical knowledge as some had undergraduate degrees in either Mathematics or Statistics. Future research could focus on gathering more volunteers and a comparison could have been carried out between those with detailed statistical knowledge and those who had only attended a short introductory course.

There has been criticism that much of the empirical evidence gathered to assess elicitation techniques is focused heavily on university students as the subject group (O'Hagan et al, 2006), and it is indeed dangerous to assume that the findings of this population group will necessarily translate across to another group. It is necessary to investigate the similarities and differences between the subject group, in this case the university students, and the population group, MOD R&M Specialists. The statistical knowledge within the population group, and the range of knowledge, is likely to be similar to that within the university group. For example, all the MOD R&M Specialists will have some statistical knowledge due to their engineering background. In addition, some within the group will be very proficient in statistics. Hence, the range of abilities within the MOD are very similar to that within the subject group. The main difference between the two groups is that the population group should be more positively motivated to provide accurate belief statements than the subject group. Hence, whilst we must be wary about the deductions that can be made from these types of experimental studies, it is also necessary to be pragmatic. We believe that the university students knowledge is sufficiently similar to the knowledge of the MOD R&M Specialists that recommendations for the subject group can be carried across to the population group.

4.2.3.4 Validating Elicitation Methods

Traditional methods to validate an elicitation approach are through reliability, coherence and calibration (Kadane and Wolfson, 1998). Reliability can be achieved through checking that the expert is satisfied with the overall statements given. Coherency is achieved by ensuring that the values stated conform to the laws of probability. Calibration focuses on ensuring that if an expert specifies a 95% bound for variable X, then 95% of all observed values of X are within that bound. However Kadane and Wolfson (1998) believe that calibration is not necessary as elicitation is not used to elicit 'perfect' opinion but instead it is used to elicit 'expert' opinion. This fits well with the subjective Bayesian philosophy adopted in this project; the aim is to capture the subjective belief of the expert - not necessarily to capture "reality". It is hoped that identifying a 'good' expert will lead to the expert's belief and reality coinciding. If a method captures the subjective belief of an expert, then the method is described as accurate. However, this poses a significant problem as traditional techniques can not be used to determine if the elicited prior is a "good" fit to the "true" prior (Winkler, 1967).

Other than steps such as reliability, coherence and calibration to measure the accuracy of an elicitation approach, another important consideration is practicality. "If the expert can answer the questions and feels comfortable, in the end, that to some degree her opinion had been captured, then provided that the method meets the basic mathematical criteria of coherence, and hopefully involves some reliability testing, it is a good method" [Kadane and Wolfson, 1998].

In addition, it is necessary to be flexible to the uniqueness of each project. Kadane and Wolfson (1998) state that for complicated models or application specific models, such as those that will be addressed during this research, general elicitation methods are not desirable. Instead, the elicitation method chosen should be determined by "examining the nature of the problem and whether or not the parameters have intrinsic meaning to the expert." [Kadane and Wolfson, 1998]. Therefore it is unlikely that a single method for eliciting covariance will be recommended. Instead, scenarios where each covariance method might be most appropriate will be highlighted.

Clemen et al (2000) carried out a similar experimental study. They gathered six correlation values for five different dependencies from approximately 70 respondents. From historical data, they were able to estimate the 'true' correlation for each of the dependencies. The authors were then able to compare the elicited correlations for each dependency with the 'true' value. From this they commented on the accuracy of elicited responses and methods when compared to the 'true' correlation value. One could criticise this performance measure from a subjective Bayesian perspective. A subjective Bayesian is not attempting to model the 'true' relationship between variables, but instead capture the belief of the decision maker. None of the participants were experts regarding the dependencies so it is possible that one method captured values that were close to the 'true' value but not representative of what the respondent actually believed. Hence, for this research project, a different analysis approach was used.

As the sample size was not large enough to draw any conclusive quantitative findings, analysis was three fold. The first step was to carry out some exploratory quantitative analysis. The aim of this analysis was to determine if any of the methods were producing values that were inconsistent with the participants beliefs. Since these elicitation techniques aim to gather the opinion of only the expert, any attempt to validate the techniques can only be done through other statements of belief by the same expert. In this example, it is not possible to determine which of the four techniques is capturing the 'true' belief of the respondent, however it is possible to discount some of the responses for each of the methods.

The second method of assessment was to determine which method was the most popular amongst the participants. For each dependency, each respondent was asked to comment on which method they preferred. This is because it is unlikely that any method that the participant is uncomfortable with is likely to produce accurate results. It is beneficial to gain an understanding of how those unfamiliar with covariance techniques feel about them.

Thirdly, the methods were evaluated qualitatively using observational evidence gathered during the elicitation sessions and feedback interviews carried out after the session with some of the participants. These interviews focused on where there was inconsistencies in the participants responses. Information was gathered regarding why the participant specified a given value and which of the four methods they had most confidence in.

4.2.3.5 Analysis of Covariance Elicitation Methods

Table 4.1 shows the breakdown of each method. For each of the three dependencies, all the participants agreed that there was a positive relationship between the variables. Acceptable is defined as being a correlation value of between 0 and 1 as the participants belief will fall within this range. Unacceptable is defined as producing a correlation value outside -1 and 1. Any correlation value that is believed to not capture the respondents subjective beliefs accurately were defined as dubious. This is any correlation values of 0 - each respondent agreed there was some form of dependency, 1 - no respondent believed the dependencies were fully explanatory and any negative responses - each respondent agreed there was a positive relationship. Table 4.1 summarises the elicitation session results.

Method	Acceptable	Unacceptable	Dubious	Most popular
DC	67	0	2	27
C	57	0	12	18
AE	58	0	11	17
AU	41	3	25	7

Table 4.1: Summary of Elicitation Sessions

From Table 4.1, it can be seen that the AU method is unlikely to be a useful method for gathering the covariance value from an expert. On three occasions, it produced correlation values that were outwith the acceptable range and on 25 occasions, produced values that were dubious - all of which were correlation values of 0. From the qualitative evidence gathered, some participants believed that learning about one variable may change their belief about the expectation of the other variable, but not necessarily change their uncertainty about that belief. This follows an observation made by Kahneman and Tversky (1982) that experts do not naturally think within a Bayesian theoretical framework. In this case, it is apparent that many of the respondents do not learn as quickly as the Bayes linear updating rule. In addition, out of the 69 responses, only seven responses felt that this was the easiest method to use. Therefore it is unlikely that this method would be useful in eliciting the necessary covariance values other than in unique cases. An example where this method may be potentially useful is where making an observation does not change the belief of an experts expectation of another variable but changes their belief about the uncertainty of the other variable. In this case, it may be that this method is most appropriate for eliciting the necessary values.

For the vast majority of cases, the AU method will be unsatisfactory and it is necessary to distinguish between the alternative three methods. From Table 4.1 it is noticeable that the DC method had more acceptable values than the other two methods and no unacceptable values were elicited using this method. This method was also the most popular amongst participants and also has other potential benefits. The method isolates the different uncertainties associated with a dependency relationship and attempts to force the decision maker to think about them individually. For example, for dependency three where the participant was asked to specify their belief about the relationship between the population MMTF and the single observed miles to failure, it is clear that this is a causal relationship such that the MMTF of the population influences the miles to failure of the single vehicle. As a result of this, it is possible to write the observed miles to failure, O_{MTF}, in terms of the population MMTF, P_{MMTF}, such that $O_{MTF} = \alpha P_{MMTF} + R$. If the test measuring the miles to failure was assumed to be non-biased, it would be reasonable to assume that $\alpha = 1$ and that $E(R) = 0$, as did many of the participants. However, when the $var(O_{MTF})$ was elicited from the respondent, there were a number of participants who stated that they believed that the $var(O_{MTF}) = var(P_{MMTF})$. However, when the $var(O_{MTF})$ was calculated using the DC method, $var(O_{MTF}) > var(P_{MMTF})$. The DC method forces the participant to consider the different forms of uncertainty with O_{MTF}; both epistemic and aleatoric. It is possible that when the participant is directly specifying their belief about $var(O_{MTF})$, they are underestimating their uncertainty.

Quantitatively, there is very little difference between the AE and C method as neither of them produced values that were unacceptable but over 15% of responses were deemed dubious. Both of them were relatively popular with participants with 18 preferring the C method whilst 17 preferred AE. Clemen et al (2000) found that in their experiments, the C method performed best for assessing the 'true' correlation value. This is surprising as other authors have found that directly assessing moments is a poor method for eliciting an expert's belief (Morgan and Henrion, 1990, Gohkale and Press, 1982, Kadane and Wolfson, 1998). One reason for this good result in Clemen's et al (2000) experiments may be attributed to the fact that the study population had just completed a course on correlation. It may be that a problem-domain expert may have similar experience on correlation and be keen in specifying their covariance value via a correlation value or they may have no experience in correlation and wish to not use it. In this experiment, observational evidence suggested that one reason why the C method was popular was because the participant felt that they only had to specify one value.

As these methods are attempting to model the subjective beliefs of the decision maker, the scope of the analysis that can be carried out is limited. It is impossible to determine which of the four correlation values elicited is the respondents 'true' subjective belief. However, exploratory quantitative analysis highlighted that on 73% of occasions, the value gathered using DC was greater than AE; on 63% of occasions, the value gathered using DC was greater than C and; 52% of occasions, the value gathered using AE was greater than C. However, this analysis does not assess which of the three is most accurate.

Kadane and Wolfson (1998) state, "the primary criterion in choosing an elicitation method is practicality. If the expert can answer the questions and feels comfortable, in the end, that to some degree her opinion has been captured, then provided that the method meets the basic mathematical criteria of coherence, and hopefully involves some reliability testing, it is a good method." Out of the three methods that are available, the decision as to which method to use is essentially down to the facilitator carrying out the elicitation session and the expert who is specifying their belief. If the expert is particularly comfortable and happy to adopt a specific method, then it could be argued that this method captures their belief the best.

4.2.3.6 Recommendations for Covariance Elicitation

As this experiment aimed to provide future users of the Bayes linear methodology with guidance for carrying out elicitation, it is beneficial to describe potential scenarios and offer recommendations for these examples. There are two scenarios that may typically arise during the building of a reliability case Bayes linear model. These are where it is natural to write one variable in terms of another and where it is apparent that only one of the two variables will be observed.

Assume that it is natural to write the relationship between two variables as $Y = \alpha X + R$. This may occur in situations where X is a causal factor of Y or in cases where X and Y have the same scale of measurement. In these cases, it is recommended that the DC method is used to calculate covariance. This method was the most popular amongst the participants and had the most acceptable values. In addition, the framing of the elicitation questions is in such a way that the values that are specified by the expert, in principle, are observable. Cooke (1991) believes that it is important that the values elicited from an expert can be thought of as observable values. For the DC method, the expert could be asked "given that you know with certainty that $X = x$, what is your 5, 50 and 95 percentiles for Y?". From this, $E(R)$ and $var(R)$ can be calculated using the Pearson and Tukey method and from this, $E(Y)$ and $var(Y)$ can also be calculated.

If the variables are unable to be written as explanatory variables, then it is recommended that the means and variances for both variables are elicited and the AE method is used to calculate the covariance. By setting $cov(X, Y)/var(Y) = \alpha$, the explanatory variable formula above can be written. From this, the covariance between X and Y and other variables in the model can be easily calculated. This method is recommended over C due to the amount of literature that recommends that the first order moments are not directly elicited, despite of the work carried out by Clemen et al (2000).

4.3 Using Bayes Linear to Analyse a Reliability Case

Section 4.3 and 4.4 aim to demonstrate the Bayes linear methodology addressing two hypothetical but common reliability problems that MOD R&M evaluators currently encounter. These are:

1. Comparing multiple contractor bids at the assessment phase to assess
    - the reliability of each system
    - the uncertainty reduced by each R&M programme
2. Assessing the reliability of a system prior to it entering service focusing on
    - estimating the expected reliability of the system
    - assessing the uncertainty still remaining in the expected reliability

For each example, a description of the decision maker's problem is given. For the first example a detailed description of the qualitative and quantitative modelling steps carried out is given. After this, analysis is carried out typical of that done qualitatively on a procurement project. For Example Two, part of Example One will be extended to reduce repetition. Hypothetical analysis will be carried out for hypothetical data. Analysis that can not be carried out in a qualitative way, such as sensitivity analysis, is also demonstrated.

Figure 4.1 displays the different steps that must be taken to populate a Bayes linear model and to carry out the different analysis. Each of these steps will be discussed extensively in Example One.

4.3.1 Introduction

The MOD are currently in the process of procuring an upgraded version of a previous system and two contractors have submitted proposals to build the system. Each contractor has submitted R&M programmes documenting the reliability tasks they plan to carry out throughout the duration of the procurement project to demonstrate that reliability requirements will be met. In addition, each contractor has already observed some data related to the reliability of their system.

The decision maker is interested in utilising the Bayes linear methodology to assess which contractor is likely to produce the most reliable equipment and which contractor is reducing his uncertainty by the greatest amount through the different tasks. The decision maker has two primary goals;

1. Assess the reliability of each system

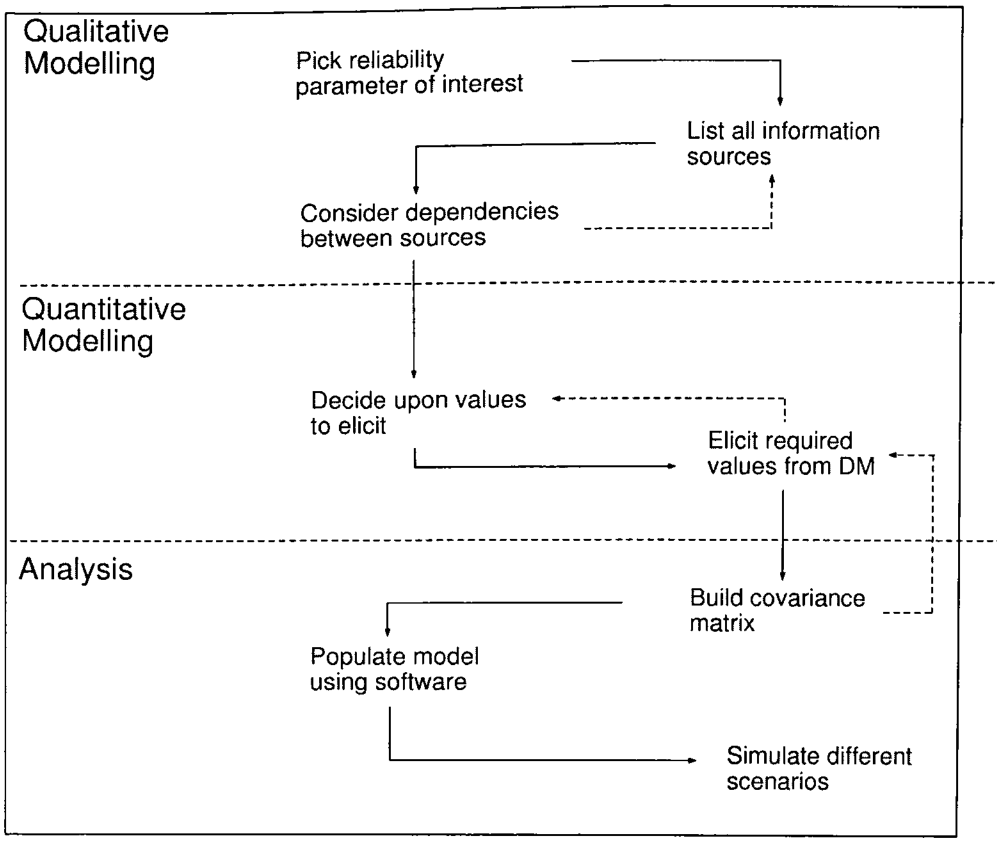

Figure 4.1: **Flowchart of steps**

2. Assess the reduction in uncertainty of the reliability of each system

Supplier A has submitted its R&M programme and plans to gather historical data regarding a previous version of the system it developed. In addition, it will carry out a RBD analysis of the system and if necessary, they plan to carry out an ISRD. Supplier A continuously makes references in its R&M Case Report to the fact it it currently on course to successfully complete a similar project for the MOD within another area of the Armed Forces.

Supplier B has also submitted its R&M programme to the MOD. It has no previous experience of designing this system but plan to carry out extensive analysis of the system continuously throughout the project. Supplier B will initially carry out a parts count analysis and initial RBD analysis. As the project progresses, they will update its RBD model as more data on the system is gathered and will carry out prototype testing. As with Supplier A, if necessary, it will carry out an ISRD.

The modelling carried out for each supplier was broken into three stages as shown in Figure 4.1: Qualitative Modelling, Quantitative Modelling and Analysis. Qualitative and quantitative modelling is carried out for both contractors however for the purposes of this report, only a detailed description of Supplier A is given. In the analysis section, comparison of both contractors is given.

4.3.2 Structuring Bayes Linear Model

The qualitative modelling is broken into three parts:

1. Identify the reliability parameter of interest
2. List all relevant information sources
3. Consider dependencies between different information sources

4.3.2.1 Identify Key Reliability Parameter

In this example, the system's reliability requirement is measured in Mean Miles Between Failure (MMBF), hence, the reliability parameter of interest to the decision maker is the MMBF of the system produced by each supplier. The MMBF of the system produced by Supplier A is denoted by $\overline{X_A}$.

4.3.2.2 List Relevant Information Sources

The next step is to list the information sources that are relevant. For Supplier A, the relevant information sources are the historical data gathered from the previous version of the system, the output of the RBD analysis and the ISRD. In addition, the decision maker comments that the reliability achievement in the other project the contractor is carrying out will influence her belief about the reliability of the system she is procuring. She would like to incorporate this into the modelling.

4.3.2.3 Consider Dependencies

The simplest scenario would be to evaluate the dependencies between the different variables qualitatively. However, by taking a few extra steps and specifying the dependencies mathematically, future elicitation specification can be made considerably simpler. Each information source is taken individually and then related to either its parent or child node (or both).

Historical Data

Begin with historical data. A sequence of historical observations $HD_1, \ldots, HD_n$ can be gathered by the decision maker, where each HD_i is the failure time of the historical system. It is assumed that this sequence $HD_1, \ldots, HD_n$ is second order exchangeable. This rather weak assumption results in the following:

$$HD_i = \overline{HD} + R_{iHD} \tag{4.1}$$

where $\overline{HD}$ is the MMBF of the historical system and each R_{iHD} is uncorrelated with everything in the model. Each R_{iHD} has prior expectation zero but will have a positive variance. This is an important part of the Bayes linear modelling. This variable represents the aleatory uncertainty of HD_i that is unexplained by $\overline{HD}$. It is assumed to be uncorrelated with everything in the model. Making this rather simple assumption has huge implications in building the model as $cov(HD_i, \overline{HD})$ can be calculated to be equal to $var(\overline{HD})$:

$$\begin{aligned}
cov(HD_i, \overline{HD}) &= cov(\overline{HD} + R_{iHD}, \overline{HD}) \\
&= cov(\overline{HD}, \overline{HD}) + cov(R_{iHD}, \overline{HD}) \\
&= var(\overline{HD}) + 0 \\
&= var(\overline{HD})
\end{aligned}$$

Instead of relating each individual observation to the historical data mean, the sample mean of the observations can be gathered, $\overline{HD_N}$, and then related to the historical data mean. This can be done as the mean is a sufficient statistic, i.e. all the necessary information is gathered through this single value. Assume that the number of historical observations, i.e $n(HD)$, is known in advance. Equation 4.1 can now be written as:

$$\overline{HD_N} = \overline{HD} + R_{nHD} \tag{4.2}$$

where R_{nHD} represents the uncertainty between the 'true' historical data mean and the observed historical data mean. As before, this variable is uncorrelated with everything else in the model. From Equation 4.2, it can be seen that in order to populate this part of the model, the decision maker does not need to specify $E(\overline{HD_N})$, $E(\overline{HD})$, $var(\overline{HD_N})$, $var(\overline{HD})$ and $cov(\overline{HD_N}, \overline{HD})$. Instead by using the following results:

$$E(\overline{HD_N}) = E(\overline{HD}) + E(R_{nHD})$$
$$var(HD_N) = var(\overline{HD}) + var(R_{nHD})$$

and

$$\begin{aligned} cov(\overline{HD_N}, \overline{HD}) &= cov(\overline{HD} + R_{nHD}, \overline{HD}) \\ &= cov(\overline{HD}, \overline{HD}) + cov(R_{nHD}, \overline{HD}) \\ &= var(\overline{HD}) \end{aligned}$$

the decision maker can simply specify $E(\overline{HD})$, $var(\overline{HD})$, $E(R_{nHD})$ and $var(R_{nHD})$ and from these, the necessary values can be calculated. If the specification questions are ordered sensibly, it is likely that $E(\overline{HD})$ and $var(\overline{HD})$ will already have been elicited. Therefore in order to add an additional variable to the model, one may only need to specify two additional values.

The next step is to relate the historical data mean to the MMBF of the system. This is done by writing the following equation:

$$\overline{HD} = \overline{X_A} + R_{HD} \tag{4.3}$$

Again, R_{HD} is uncorrelated with everything else in the model. In order to elicit R_{HD} (and subsequent R_i values), the decision maker must consider their uncertainty surrounding the value of the MMBF of the historical system, given that they were told with certainty the value of the current system, i.e. $\overline{X_A}$. By setting $var(\overline{X_A}) = 0$, then $var(\overline{HD}) = var(R_{HD})$.

ISRD

For the ISRD, as before, a sequence of observations, $ISRD_1, \ldots, ISRD_n$ are observed. The mean of this sequence of observations is $\overline{ISRD}$ and this is related to the MMBF of the system through the formula; $\overline{ISRD} = \overline{X_A} + R_{ISRD}$. R_{ISRD} is uncorrelated with everything else in the model.

RBD Modelling

The contractor also plans to carry out RBD modelling. The decision maker believes that this RBD modelling will output a point estimate of the MMBF of the system. This RBD modelling can be modelled in two different ways. Either the output of the RBD, RBD_i, could be related to the system through the formula:

$$RBD_i = \overline{X_A} + R_{RBD_n} \tag{4.4}$$

or an additional variable, $\overline{RBD}$, can be added to ensure that the different uncertainties are explicitly identified. Hence the following two equations are used.

$$\overline{RBD} = \overline{X_A} + R_{RBD} \tag{4.5}$$

$$RBD_i = \overline{RBD} + R_{RBD_I} \tag{4.6}$$

In this case, R_{RBD} represents the uncertainty difference between the decision makers assessment of the MMBF of the system and the 'true' value that the RBD would produce. R_{RBD_I} represents the uncertainty difference between the 'true' RBD value and the value observed by the contractor.

Similar System

The decision maker would also like to incorporate into the analysis the fact that the contractor is carrying out work with the MOD on another contract. The decision maker believes that it will be possible to gather some information from the MOD decision maker on the project and some in-service data of the system. The decision maker specifies $\overline{SS}$ as their assessment of the MMBF of the similar system. The decision maker specifies the following relationships;

$$\overline{SS_E} = \overline{SS} + R_{SS_E} \tag{4.7}$$

$$\overline{SS_I} = \overline{SS} + R_{SS_I} \tag{4.8}$$

where $\overline{SS_E}$ is the assessment of the system by the R&M Focal point working on the similar system and $\overline{SS_I}$ is the observed in-service MMBF of the similar system. As above, R_{SS_I} and R_{SS_E} uncorrelated with everything else in the model.

The decision maker believes that the system currently being assessed is related to the similar system in the following way;

$$\overline{SS} = \alpha_{SS}\overline{X} + R_{SS} \tag{4.9}$$

where α_{SS} is an integer and R_{SS} is uncorrelated with everything else in the model.

4.3.2.4 Building a Bayes Linear Network

Using the above information, a BLN of th R&M programme can be developed. In order to build a BLN, it is necessary to create a unique ordering. This network will be unique to the ordering that is given (Goldstein and Woolf, 2007). The ordering used in this project is [$\overline{X_A}$, $\overline{HD}$, $\overline{RBD}$, $\overline{SS}$, $\overline{HD_N}$, RBD_N, $\overline{SS_I}$, $\overline{SS_E}$]. Before building the BLN, recall the definition of Bayes linear sufficiency. A is said to be Bayes linear sufficient for B for adjusting C if

$$E_{A \cup C}(B) = E_A(B) \Rightarrow cov_A(B, C) = 0$$

The first step of building a BLN is to draw the first two nodes, $\overline{X_A}$ and $\overline{HD}$. As these two variables are correlated, an arrow is drawn from $\overline{X_A}$ to $\overline{HD}$. The next step is to draw a node for $\overline{RBD}$. As $\overline{X_A}$ is correlated with $\overline{RBD}$, an arc is drawn from $\overline{X_A}$ to $\overline{RBD}$. The decision maker specifies that if she knew $\overline{X_A}$ with certainty, learning about $\overline{HD}$ would not affect her belief about $\overline{RBD}$. Hence, $\overline{HD}$ and $\overline{RBD}$ are Bayes linear sufficient for adjusting $\overline{X_A}$, and no additional arcs are required. The next step is to introduce $\overline{SS}$. As $\overline{SS}$ is correlated with $\overline{X_A}$, an arc is drawn from $\overline{X_A}$ to $\overline{SS}$. No more additional arcs are required as $\overline{SS}$ is Bayes linear sufficient with everything else previously in the model. The next step is to introduce $\overline{HD_N}$, $\overline{ISRD}$, and RBD_i. $\overline{HD_N}$ is correlated with $\overline{HD}$ and is Bayes linear sufficient with everything else in the model for adjusting $\overline{HD}$. Therefore, only an arc is needed from $\overline{HD}$ to $\overline{HD_N}$. $\overline{ISRD}$ is Bayes linear sufficient for everything previous in the model for adjusting $\overline{X_A}$; hence only an arc from $\overline{X_A}$ to $\overline{ISRD}$ is needed. Similarly, RBD_N is Bayes linear sufficient with everything else in the model for adjusting $\overline{RBD}$. Finally, nodes for $\overline{SS_E}$ and $\overline{SS_I}$ are added. As they are both Bayes linear sufficient with everything else in the model for adjusting $\overline{SS}$, only arcs from $\overline{SS}$ to each of the new nodes are required.

The Bayes linear network for the above modelling is represented by Figure 4.2.

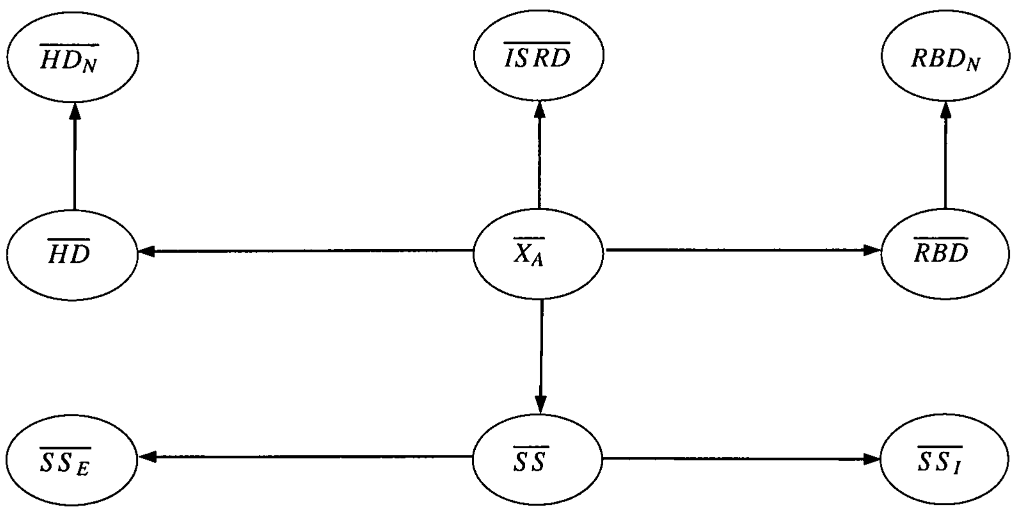

Figure 4.2: Bayes linear network for supplier A

4.3.2.5 Additional Steps

It is interesting at this point to discuss why the extra steps in the modelling are taken, i.e. why not simply relate the output of each observation directly to the new system. If a benefit of the Bayes linear methodology is to reduce the specification requirement placed upon a decision maker, then this additional step is clearly a contradiction. However, whilst it is desirable to reduce the elicitation requirements of the decision maker, it is necessary to ensure that the values elicited are done so in a sensible manner. In order to do this, it is important to ensure that the values elicited from the decision maker are observable values that the decision maker can potentially relate his beliefs to.

For example, assume that $\overline{HD}$ is removed from the modelling. Hence, $\overline{HD_N}$ can be written as:

$$\overline{HD_N} = \overline{X_A} + R_{HD_I} \tag{4.10}$$

where

$$R_{HD_I} = R_{HD} + R_{nHD} \tag{4.11}$$

The modelling above and the modelling used in the example are equivalent. Lets assume that the decision maker is now asked to specify her expectation and variance for R_{HD_I}. The decision maker must attempt to assess her uncertainty regarding the difference between the new system and the old system, as well as assess the uncertainty between the 'true' historical mean and the observed mean. This would be a difficult task for someone with little experience in statistics.

Instead of asking the decision maker to attempt to simultaneously specify her belief regarding two different uncertainties, the process is broken down into two simpler steps. In this example, R_{HD} represents the uncertainty in the difference between the historical system and the new system whilst R_{nHD} represents the uncertainty between the true historical system's MMBF and the observed mean.

4.3.3 Eliciting Beliefs to Populate a Bayes Linear Model

Referring to Figure 4.1, the main steps during the quantitative modelling part are

- Decide upon values to elicit
- Elicit required values from decision maker

This is an important part of demonstrating the usefulness of the Bayes linear methodology. Using a BBN methodology, the elicitation burden placed upon the decision maker in this example would be extensive and possibly infeasible in practice. By setting up the problem in the above manner, the number of values required in elicitation has been reduced to a minimum.

4.3.3.1 Identify Values to Elicit

The first task is to identify the necessary values to elicit in order to fully populate the model. The first step is to identify which variables need eliciting. By explicitly declaring the mathematical relationships as shown above, the minimum variables can be identified in order to fully populate the model. In this model, there are 9 different variables. Between the expectation table and the covariance matrix, 9+81=90 variables have to be directly, or indirectly, assessed. This may seem like a lot, but due to the manner in which the model has been built, only 18 values have to be elicited to fully populate the model.

The first step is to identify which variables have no parents. In this example, only $\overline{X_A}$ has no parents. Therefore the first values that must be elicited are $E(\overline{X_A})$ and $var(\overline{X_A})$, i.e. the decision makers assessment of the MMBF of the system that is procured.

The next step is to identify the children of any parents that have previously been elicited. In this example, all the children of $\overline{X_A}$ can now be elicited - $\overline{HD}$, $\overline{ISRD}$, $\overline{RBD}$ and $\overline{SS}$. In order to do this, expectations and variances of the following variables must be elicited - R_{HD}, R_{ISRD}, R_{RBD}, α_{SS}, and R_{SS}. Once this is done, the children of $\overline{HD}$, $\overline{ISRD}$, $\overline{RBD}$ and $\overline{SS}$ must be elicited. Again, in order to do this, one need only elicit the residual uncertainty between the assessment of the parent with its child. Expectations and variances of the following variables are needed to fully populate the model; R_{nHD}, R_{RBDi}, R_{SS_E} and R_{SS_I}.

4.3.3.2 Elicitation of Necessary Values

Now that the necessary values have been identified, the next step is to elicit the values from the decision maker. As identified in Section 4.2.2, the Pearson-Tukey method is used to gather the expectation and variance of each variable. From these values, direct calculation is used to calculate the entire covariance matrix. The decision maker specifies the following values;

Using the method described in Section 4.2.3.2 the decision maker specifies that she believes α_{SS} is equal to 0.2.

Variable	5th	50th	95th	Expectation	Standard Deviation
$\overline{X_A}$	500	1000	1700	1037.0	364.7
$\overline{R_{HD}}$	-500	-250	0	-250.0	152.0
$\overline{R_{ISRD}}$	-50	0	50	0.0	30.4
$\overline{R_{RBD}}$	-500	200	1000	218.5	456.0
$\overline{R_{SS}}$	800	1800	2800	1800.0	607.9
$\overline{R_{nHD}}$	-15	0	15	0.0	9.1
$\overline{R_{RBDi}}$	-20	0	20	0.0	12.2
$\overline{R_{SS_E}}$	-500	0	500	0.0	303.9
$\overline{R_{SS_I}}$	-100	0	100	0.0	60.8

Table 4.2: Values elicited from expert

4.3.4 Analysing Belief Structure

In the analysis section, the three main steps are:

- Build covariance matrix
- Populate the model
- Simulate different scenarios

4.3.4.1 Build Covariance Matrix and Populate Model

From the values elicited above, the expectation table and covariance matrix for the model can be built. In order to do this, the mathematical formulas described above and the following results are used. If $X = \alpha Y + Z$ then:

$$E(X) = \alpha E(Y) + E(Z)$$

$$var(X) = \alpha^2 var(Y) + var(Z)$$

$$cov(Y, Y) = var(Y, Y)$$

Using these formulas, it is a simple, but tedious task, to calculate the covariance matrix for Supplier A.

Assume that the qualitative and quantitative modelling for Supplier B has been carried out such that Figure 4.3 is the Bayes linear network of the R&M programme submitted by Supplier

$\overline{X_A}$	133006								
$\overline{HD}$	133006	156110							
$\overline{ISRD}$	133006	133006	133930						
$\overline{RBD}$	133006	133006	133006	340942					
$\overline{SS}$	26601	26601	26601	26601	396143				
$\overline{HD_N}$	133006	156110	133006	133006	26601	156193			
RBD_N	133006	133006	133006	340942	26601	133006	341090		
$\overline{SS_I}$	26601	26601	26601	26601	396143	26601	26601	399840	
$\overline{SS_E}$	26601	26601	26601	26601	396143	26601	26601	396143	488498

Table 4.3: Covariance matrix for Supplier A

B. The expectation table and covariance matrix for Supplier B can be found in Appendix C.

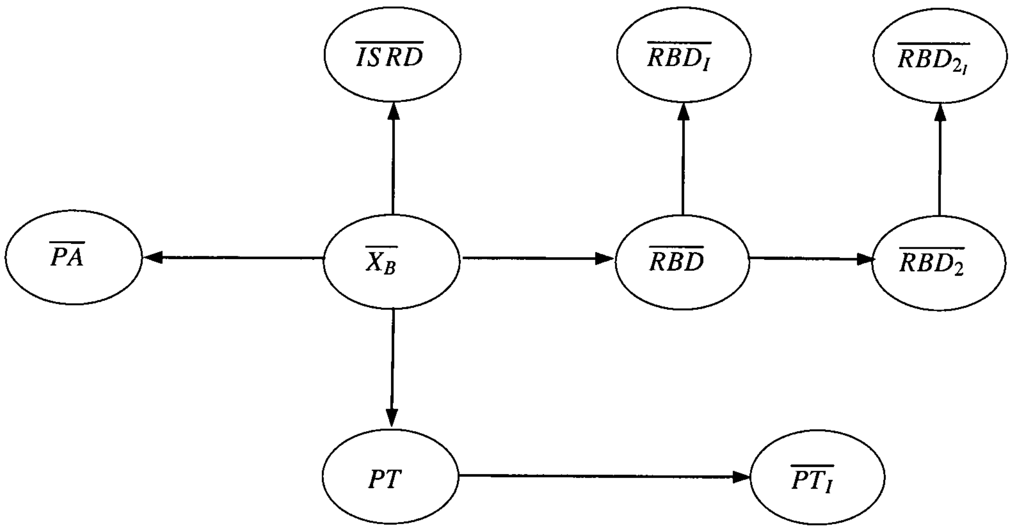

Figure 4.3: Bayes linear network for supplier B

4.3.4.2 Simulate Different Scenarios

The final step is to begin to simulate different scenarios. Recall the two decisions the model aimed to support;

- which system the decision maker believes is most reliable
- which R&M programme is reducing the decision makers uncertainty by the greatest amount

Which System is Most Reliable In order to make this example as realistic as possible, let us assume that only limited information is available for each contractor. Supplier A observed an historical data mean of 842 miles and the decision maker has spoken to the R&M Focal Point working on the similar system who has stated he expects an MMBF of 2200 miles. Supplier B has carried out some RBD modelling which gave an output of 1100 miles and a parts count analysis which gave an output of 809 miles. Using this information, the decision maker can carry out inference on X_A and X_B to assess which system the decision maker believes is most reliable at this time.

Using the following values, $D_1 = [\overline{SS_E}, \overline{HD_N}] = [2200, 842]$ and $D_2 = [RBD_{2_I}, PA] = [1100, 809]$, and the Bayes linear equations, the adjusted expectation and variance for each supplier's system can be calculated. For example:

$$
\begin{aligned}
E_{D_1}(\overline{X_A}) &= E(\overline{X_A}) + cov(\overline{X_A}, D_1)var^{-1}(D_1)(D_1 - E(D_1)) \\
&= 1037 + [26601, 133006]\begin{bmatrix} 0.00000206 & -0.000000351 \\ -0.000000351 & 0.00000646 \end{bmatrix}[2200 - 2007.4, 842 - 787]^T \\
&= 1037 + [26601, 133006][0.000378, 0.000287]^T \\
&= 1037 + 48.3 \\
&= 1085
\end{aligned}
$$

$$
\begin{aligned}
var_{D_2}(\overline{X_A}) &= var(\overline{X_A}) - cov(X_A, D_1)var^{-1}(D_1)cov(D_1, \overline{X_A}) \\
&= 133006 - 113293 \\
&= 19712
\end{aligned}
$$

Table 4.4 displays the adjusted expectation, variance and standard deviation of $\overline{X_A}$ and $\overline{X_B}$.

Variables	Adjusted Expectation	Adjusted Variance	Adjusted S.D.
$\overline{X_A}$	1085.0	19712	140.4
$\overline{X_B}$	862.6	117483	342.8

Table 4.4: Adjusted expectation and uncertainty for both suppliers

As can be seen, Supplier A appears to be producing a system that has a greater MMBF than Supplier B. In addition, the decision maker's uncertainty surrounding her assessment of Supplier A's reliability is substantially lower than that of Supplier B's. At this point, there still remains considerable uncertainty surrounding the decision maker's assessment of Supplier B's reliability.

Reducing Uncertainty In addition to this, the decision maker wishes to assess how much uncertainty remains in her assessment of both $\overline{X_A}$ and $\overline{X_B}$ if she were to observe all the information sources that they claim they will produce. She does not believe that either supplier will carry out an ISRD and even if they do, she is interested in assessing the uncertainty of the system prior to it carrying out an ISRD.

Assuming that each supplier carries out each task they stated other than the ISRD, Table 4.5 can be produced;

Variables	Adjusted Variance	Adjusted S.D.
$\overline{X_A}$	17994	131.4
$\overline{X_B}$	26332	162.3

Table 4.5: Adjusted uncertainty

Supplier A has reduced the decision maker's uncertainty by the greatest amount. Based on this information, the decision maker can now make an informed decision regarding which system is more reliable and which system will have greater uncertainty once all the information sources have been observed. As the contract may not have been signed at this stage, the decision maker also has an understanding of how much uncertainty she will be left with prior to the system entering service. Prior to the contract being signed, she may identify other forms of information that she believes the supplier should gather in order to reduce her uncertainty. In addition to this analysis, sensitivity analysis could also be carried out to assess whether or not the model is robust to changes in the belief structure specified by the decision maker. Demonstrating how sensitivity analysis can be carried out in the Bayes linear framework will be done in Section 4.4.2.

4.3.4.3 Justification For Use of Linear Models

In the above models, we are assuming that the relationship between variables are approximately linear. We are not saying that the relationships are linear; only that they are an approximation. Where we believe that the relationships are not linear, then additional variables would be added to the model, or higher moments of the current variables elicited, to address this. In order to identify whether or not assuming a linear relationship is reasonable, the facilitator can present the decision maker with a number of different scenarios of observed data and observe if the adjusted output of the model is a fair reflection of their potential future beliefs. If it is not, then the assumption of linearity may not be reasonable.

4.4 Example 2: Assessing Reliability of a System

4.4.1 Introduction

Assume that the decision maker chose supplier A to build the system. As the project progresses, the different forms of information that were described in the preliminary R&M programme are gradually observed. The decision maker uses the model developed in Section 4.3 and the data observed throughout the life of the project to assess the reliability of the system and the uncertainty surrounding this assessment prior to the system entering service.

The system is just about to be placed in-service and the decision maker would like to assess the reliability of the system and the uncertainty surrounding this assessment. In addition to the observations given above, i.e D_1, assume that the RBD modelling has been carried out and that the similar system has been placed in-service and observed data can be gathered regarding its MMBF. The qualitative and quantitative modelling carried out in Section 4.3 has been carried on and used for this section.

4.4.2 Analysing Belief Structure Given Observations

Assume that the output of the RBD modelling was 1300 miles and the observed MMBF of the similar system was 1870 miles. Using this and the previous observations, the decision makers assessment of $\overline{X_A}$ can be linearly adjusted. D_{IS} is the observations gathered prior to the system entering service, i.e. $D_{IS} = [\overline{SS_E}, \overline{HD_N}, RBD_i, \overline{SS_I}] = [2200,842,1300,1870]$. Using the

Bayes linear formulas, the adjusted expectation and adjusted variance can be calculated.

$$E_{D_{IS}}(\overline{X_A}) = E(\overline{X_A}) + cov(\overline{X_A}, D_{IS})var^{-1}(D_{IS})(D_{IS} - E(D_{IS}))$$

$$= 1037 + \begin{bmatrix} 26601 \\ 133006 \\ 133006 \\ 26601 \end{bmatrix}^T \frac{1}{1000000} \begin{bmatrix} 10.4 & -0.01 & -0.0017 & -10.3 \\ -0.015 & 9.65 & -3.7 & -0.38 \\ -0.002 & -3.7 & 4.4 & -0.042 \\ -10.32 & -0.38 & -0.042 & 12.75 \end{bmatrix} \begin{bmatrix} 2200 - 2007.4 \\ 842 - 787 \\ 1300 - 1255 \\ 1870 - 2007.4 \end{bmatrix}$$

$$= 1037 + 45.5$$

$$= 1082.5$$

$$var_{D_{IS}}(\overline{X_A}) = var(\overline{X_A}) - cov(\overline{X_A}, D_{IS})var^{-1}(D_{IS})cov(D_{IS}, \overline{X_A})$$

$$= 133006 - \begin{bmatrix} 26601 \\ 133006 \\ 133006 \\ 26601 \end{bmatrix}^T \frac{1}{1000000} \begin{bmatrix} 10.4 & -0.01 & -0.0017 & -10.3 \\ -0.015 & 9.65 & -3.7 & -0.38 \\ -0.002 & -3.7 & 4.4 & -0.042 \\ -10.32 & -0.38 & -0.042 & 12.75 \end{bmatrix} \begin{bmatrix} 26601 \\ 133006 \\ 133006 \\ 26601 \end{bmatrix}$$

$$= 133006 - 115006$$

$$= 18000$$

The adjusted expectation of $\overline{X_A}$ is now 1082 with an adjusted variance of 18000 and an adjusted standard deviation of 134.2. At this point, the decision maker can decide whether or not they believe they would like to carry out an ISRD to reduce the uncertainty further. Using the Bayes linear formula for adjusted variance, the added benefit of observing $\overline{ISRD}$ can be assessed. In this example, the adjusted variance given that the decision maker observes the ISRD in addition to what she has already observed would be 873. This is a drop in standard deviation from 134 to 29. If in this example the reliability target was 1000 hours, the decision maker may believe that with an expectation of 1082 hours and a standard deviation of 134 hours, that too much uncertainty still remains in her assessment of the X_A. The decision maker at this point must decide whether or not the added expenditure of carrying out an ISRD is worth the resulting drop in variance. Alternatively, the decision maker may attempt to identify an alternative source of information that would further reduce her uncertainty to a comfortable level.

4.4.2.1 Sensitivity Analysis

This section aims to show how sensitivity analysis is carried out in the Bayes linear framework. Due to the simplicity of the Bayes linear formulas and the belief structure, sensitivity analysis can be carried out quickly. The Bayes linear models built throughout this research project are constructed using a decision maker's degree of belief regarding a number of variables. By developing a sensible elicitation process and adopting extensively used techniques, it is hoped that the values elicited are close to the decision maker's true beliefs. However, it is inevitable that there is a degree of inaccuracy in the values elicited and sensitivity analysis will focus on assessing how changes in these elicited values affects the output of the model.

Assume that in the above example, the reliability target for the system was 1200 hours. The decision maker may wish to assess whether or not, given the covariance matrix stated, if adjusting her prior expectation of $\overline{X_A}$ by a small amount would mean the adjusted expectation of $\overline{X_A}$ is greater than 1200 hours. It is assumed that the covariance matrix remains the same. Therefore, while $E(D_{IS})$ increases as $E(\overline{X_A})$ is increased, $cov(\overline{X_A}, D_{IS})$ and $var(D_{IS})$ remain the same. Adjusting $E(\overline{X_A})$ from 1037 to 1437, a change of 400 hours, changes $E_{D_{IS}}(\overline{X_A})$ from 1082 to 1136; a change of only 54 hours. In order to reach a target of 1200 hours, the $E(\overline{X_A})$ would have to increase from 1082 to 1908 hours. Similarly, if the target for $\overline{X_A}$ was 1000 hours, the $E(\overline{X_A})$ would have to drop to 425 before $E_{D_{IS}}(\overline{X_A}) < 1000$. As can be seen, the effect of adjusting the prior expectation of $\overline{X_A}$ does not have a large effect on $E_{D_{IS}}(\overline{X_A})$.

Alternative sensitivity analysis could focus on adjusting the covariance matrix. By adjusting the covariance matrix, different weightings can be given to different observations. For example, the decision may revise their belief regarding the residual uncertainty between $\overline{X_A}$ and $\overline{HD}$. She may believe that the dependency relationship between the two variables is weaker than initially thought and wishes to increase $var(R_{HD})$ to take into account this change of belief. Doing this only changes three values in the covariance matrix; $var(\overline{HD})$, $var(\overline{HD_N})$ and $cov(\overline{HD}, \overline{HD_N})$.

Carrying out multiple simulations for different scenarios is simple but time extensive. For this example, three scenarios will be investigated: decreasing $var(R_{SS_I})$, and increasing $var(R_{HD_N})$, and doing both at the same time. These two have been specifically chosen. Only $\overline{SS_I}$ is having a negative impact on the adjusted expectation of $\overline{X_A}$. The decision maker is interested in maximsising the effect this negative observation is having on $E_{D_{IS}}(\overline{X_A})$. Conversely, $\overline{HD_N}$ is creating the largest positive impact on $E_{D_{IS}}(\overline{X_A})$. The decision maker does not wish to remove this dependency but instead she will increase $var(R_{HD_N})$ to the maximum value she believes it may take.

To simulate the first scenario, the analyst decreases $var(R_{SS_I})$ from 3697 to 1. Using the Bayes linear formula for adjusted expectation, $E_{D_{IS}}(\overline{X_A})$ remains unchanged at 1082. This is because $cov(\overline{X_A}, \overline{SS_I})$ remains at 26601. In order for this relationship to have a strong effect on $E_{D_{IS}}(\overline{X_A})$, $cov(\overline{X_A}, \overline{SS_I})$ must be increased. For the second scenario, $var(R_{SS_I})$ is set equal to its original value of 3697 and $var(R_{HD_N})$ is increased from to 23104 to 500000. Using the same observations as before, $E_{D_{IS}}(\overline{X_A})$ drops from 1082 to 1055. Simulating scenario three drops $E_{D_{IS}}(\overline{X_A})$ from 1082 to 1054.

The sensitivity analysis has demonstrated that given the observations and by making minor adjustments to the decision maker's belief structure, the adjusted expectation of $\overline{X_A}$ is unlikely to drop below 1000 hours.

4.5 Conclusion

This chapter has aimed to demonstrate that theoretically the different problems that were identified in Chapter 1 can be addressed using the Bayes linear methodology. In addition, this chapter has aimed to support the overall modelling process of the Bayes linear framework by identifying how elicitation and sensitivity analysis can be conducted.

It is recommended that Pearson and Tukey's method is used when eliciting any expectations and variances. This method has been demonstrated to be reasonably accurate for assessing the mean and variance of the normal, beta and gamma distribution. When eliciting covariance, it is recommended that one of two approaches is taken. Either the facilitator writes one of the variables in terms of the other such as $Y = \alpha X + R$ and then elicit from the expert their belief about α, $E(R)$ and $var(R)$. Alternatively, if $E(X)$, $E(Y)$, $var(X)$ and $var(Y)$ have been specified, then the expert may state their adjusted expectation of X given that they have observed Y. From this specified value, $cov(X, Y)$ can be calculated.

Two examples have been used to demonstrate that theoretically, the Bayes linear methodology can address the problems faced by the MOD. From Example 1, it can be seen that the methodology supports the investigation of different scenarios and observing the impact of different information sources. In addition, the decision maker has another criteria by which they can assess each contractor's bids. This example demonstrates that the methodology is capable of assessing the effectiveness of a contractors R&M Programme in reducing the decision makers uncertainty. This was Research Aim 1 discussed in Chapter 1. In this simple example, a MOD decision maker would be able to decide which of the R&M Programmes was more effective. In a more complex example with many more forms of information (as will be shown in Chapter 5), the methodology becomes extremely useful in determining the effectiveness of different R&M Programmes as the analysis can often not be carried out simply using an unstructured qualitative method.

As can be seen from Example 2, when multiple sources of information are contributing towards a parameter of interest, the methodology processes the information in a mathematically robust way and produces an output which is simple and concise for the decision maker to digest. In this example, the information provided was contradictory but the methodology was able to assess the effect each had on $\overline{X_A}$. Sensitivity analysis was also carried out to assess the effect of small changes to the decision maker's belief structure to the output of the model.

The examples given in this chapter are useful in order to demonstrate the application of the Bayes linear methodology but fail to take into account practical issues that arise through applying this particular methodology. These include issues such as introducing proxy variables, modelling variables where the dependency is not causal and carrying out sensitivity analysis on actual belief structures. The following two chapters will examine 'real' reliability problems with 'real' decision makers.

Chapter 5

Applied MOD Project 1: Analysing Contractor's bids

5.1 Introduction

The MOD is procuring new equipment to replace an obsolete system on a Royal Navy platform. The 'closed' architecture of the existing system software, means it is not receptive to technology refresh. Hence an IPT has been set-up with the purpose of procuring a new system with similar functionality. An ITT was issued in December 2005 with three contractors submitting bids in March 2006. The project was in the Assessment phase of the CADMID cycle in April 2006 when the modelling described in this chapter was conducted.

Each of the three contractors submitted bids to demonstrate how they planned to meet reliability requirements. For the purposes of confidentiality, the three contractors are referred to as contractor Yellow, Blue and Grey. Only the MOD decision maker knows the identity of each contractor. As part of their bids, each contractor submitted a R&M Case Report comprising a R&M programme which details the tasks planned to ensure the reliability requirements will be met. In addition, each contractor provided preliminary modelling to predict reliability for the 6 reliability requirements specified by the MOD; namely, critical, major and minor Mean Time Between Failure (MTBF) for hardware and software. A critical failure is one that causes the mission to be unsuccessful; a major failure is one that causes a drop in performance or efficiency and; a minor failure has no impact to mission efficiency but requires a maintenance action. The system is expected to enter service in late 2009 and the procurement project will cost approximately 35 million pounds.

This chapter examines how Bayes linear modelling can be developed to support decision making at the assessment phase of the CADMID cycle. In particular, the chapter explores how the Bayes linear methodology supports the decision maker in structuring her beliefs regarding competing bids and to conduct analysis to inform the evaluation of alternative bids. Section 5.2 presents the decisions to be supported by the model and explains the process normally followed by the MOD to evaluate bids. Section 5.3 states the modelling objectives. Sections 5.4 and 5.5 present the qualitative and quantitative modelling conducted respectively. Section 5.6 illustrates the analysis conducted using the quantitative data gathered from the decision-maker. Section 5.7 discusses further development of the model for the future. Section 5.8 discusses the feedback provided to the decision maker and how it informed the decision-making process. Section 5.9 summarises and concludes the chapter. Throughout the chapter, the MOD decision maker will be referred to as she and the researcher, acting as facilitator and analyst, will be referred to as he.

5.2 Decision Making and Current MOD Evaluation Process

Based on the information in the R&M Case Reports, the decision maker must make a recommendation to the IPTL regarding which contractor should be chosen in terms of reliability. The IPTL then makes a decision that takes into account all the relevant criteria. The modelling carried out is not to support the decision carried out by the IPTL but instead, the modelling is to support the recommendation given to the IPTL by the R&M Specialist. Taking into account external factors to reliability such as budget, performance and time are outwith the scope of this research. The decision maker was interested in answering the following two questions:

1. Which contractor is likely to produce a system that will meet reliability requirements?

2. Once the contractor carries out all their stated tasks in their R&M programme, will the decision maker's uncertainty regarding her assessment of the contractors reliability be reduced to an acceptable amount?

Currently the decision maker uses her subjective judgement to assess whether a contractor is likely to meet reliability requirements. Although largely unstructured, the decision-maker will seek to establish whether the contractor has made a comprehensive investigation of the R&M risks associated with the project and developed a R&M Strategy to mitigate these risks. In addition, she would expect the contractor to provide preliminary evidence to demonstrate reliability targets are likely to be met. If there is a lack of evidence in the R&M Case Report, alternative methods of evaluating the reliability of the system will have to be used. This may include investigating the technical solution, initial analysis of the contractor's design or discussions with the contractor.

Prior to this intervention, the decision maker had not formally quantified her expected MTBF for the systems proposed by each of the three contractors using the information provided in the R&M cases. Instead, the decision maker stated that she evaluated the contractor's R&M programmes by comparing the reliability tasks proposed in their R&M Case against those tasks she believed should be carried out. For those tasks common to both the decision-maker and the contractor, the MOD decision-maker asks, "Is the contractor giving me confidence that they will carry out the task well?" For those tasks the contractor is not conducting but the decision-maker deems important, the MOD decision-maker asks "How much do I believe they are losing by not carrying out these tasks?" Finally, for those tasks proposed by the contractor but that the MOD decision-maker deems unnecessary leads the MOD decision-maker to ask "Does this task offer any value?". The MOD decision-maker seeks to establish whether the contractor provides them with a 'feeling' that the planned tasks will provide the necessary assurance once implemented.

The decision maker was asked to review the R&M Programme for each of the contractor's bids but not to make any observations. The Bayes linear modelling was carried out prior to the decision maker making any observations. Based on the R&M Programmes for each contractor, the decision maker believed contractor Yellow will produce the most reliable system and they will provide sufficient assurance that they will meet reliability requirements. She believed contractor Yellow has produced a superior R&M Programme to Blue and Grey. The decision maker was unable to distinguish between Blue and Grey and subsequently, she was interested in whether or not the Bayes linear model would distinguish between contractor Blue and Grey, and if the methodology could explain any difference.

5.3 Bayes Linear Modelling

The overall goal of modelling is to assist the decision maker in answering the above two questions. To do this, the following aims must be met:

- Document the expected MTBF of each contractor at this time and the uncertainty associated with this expectation using the decision maker's subjective beliefs and the information currently submitted by the contractors
- Model and analyse each contractor's R&M programme to assess which one achieves the greatest reduction in the decision maker's uncertainty.

The decision maker believed this would help her provide detailed feedback to the IPT. In addition, she would gain an understanding of what information sources are contributing the most 'value' to reduction in the uncertainty of the reliability parameter and she would be able to explore the different amounts of uncertainty reduction associated with alternative R&M programmes.

Recall Figure 4.1 from Chapter 4 which shows that the Bayes linear modelling process has three main steps: qualitative modelling, quantitative modelling and analysis. Modelling was conducted over two-days in April 2006. During the first day, the facilitator and the decision maker spent three hours to conduct the qualitative modelling for each of the three contractors. The facilitator spent a further hour planning the elicitation session to support the quantitative modelling. The elicitation session was carried out on day 2, and lasted three hours. The facilitator carried out analysis on the model and provided feedback to the decision maker.

5.4 Structuring the Bayes Linear Model

Firstly, the reliability parameter of most interest to the decision maker was identified. There were six reliability requirements specified to the contractor - major, minor and critical MTBF of software and major, minor and critical MTBF of hardware. Due to resource constraints, it was only possible for the MOD decision maker to model one requirement. The decision maker was interested in modelling the critical MTBF of the hardware of the system as the focus of the project was to replace the system hardware whilst the software was to remain unchanged. She decided to model the critical MTBF and not the major or minor MTBF as she believed this was of most interest to the IPTL. The decision maker believed the contractor would focus on primarily meeting this target relative to the other requirements. The critical hardware MTBF requirement set by the MOD was 1800 hours.

The second step involved listing all the information sources the contractor may provide that were in any way related to the decision maker's belief of X, where X is the critical MTBF of the hardware of the new system measured in hours. All subsequent variables are also measured in hours. The decision maker believed the system has two distinct reliability values. Firstly, the reliability of the system when used as a stand-alone system. Secondly, the reliability of the system when integrated with the platform. The decision maker is interested in the latter, however, it is influenced by the value of the former. The decision maker's belief regarding X was influenced by three uncorrelated variables. Firstly, the predicted reliability of the sub-systems of the system, X_R, and the ease of integration between the system and the platform, represented by $X_{I_1} + X_{I_2}$, such that $X = X_R + X_{I_1} + X_{I_2}$.

The decision maker believes her expectation and uncertainty in X is primarily determined by the reliability of the system in a known and controlled environment, i.e. X_R. However, the decision maker believes integration issues could exist between the system and the platform that could cause the MTBF of the system to be lower. $X_{I_1} + X_{I_2}$ represents her belief about integration. She believes the reliability of the system on the platform, i.e. X will be lower than X_R if there are integration problems with the system. There are two integration variables as there exist two uncorrelated forms of information that influence her belief about X. $Var(X_{I_1})$ is set equal to $\alpha_{X_{I_1}}$ and $var(X_{I_2})$ is set equal to $\alpha_{X_{I_2}}$. A description of how $\alpha_{X_{I_1}}$ and $\alpha_{X_{I_2}}$ are calculated is given in Section 5.4.4.

Four different forms of information could be gathered: those that influence the decision maker's belief about X_R only, those that influence the decision maker's belief about X_{I_1} only, those that influence the decision maker's belief about X_{I_2} only, and those that influence the decision maker's belief about X_R, X_{I_1} and X_{I_2} i.e. X. In order to distinguish between each of the contractors, X_Y, X_{R_Y}, $X_{I_{1_Y}}$ and $X_{I_{2_Y}}$ is used to represent contractor Yellow with equivalent subscripts for Grey (G) and Blue (B). As the system is an upgrade of an existing system, the developmental part of the procurement project is short in comparison with others where new equipment is being procured. This has consequences for the resource available to conduct reliability tasks. In addition, the MOD ITT specified that the contractor must develop a FMECA and gather in-service data. Therefore the flexibility available to each contractor in building it's R&M programme was limited and as such, each contractor submitted similar R&M programmes. All the information sources and their impact on X is discussed below. There were a total of seven different information sources potentially gathered by the contractors that required modelling. The order in which the information is introduced is representative of the chronological order of observations during anR&M Programme.

5.4.1 Parts Count Analysis

Parts Count Analysis is a crude method of assessing the reliability of a system. The decision maker believes the output of a contractors parts count analysis, denoted by PA, is $PA = X_R + R_{PA}$, i.e. PA influences the decision maker's belief about X_R but not X_{I_1} or X_{I_2}. Recall from Section 4.3.2.3 that all R_i values are uncorrelated with everything in the model. R_{PA} and all subsequent R_i values defined are assumed to be uncorrelated with every other variable in the model. These R_i variables are used to explain the residual aleatory uncertainty not captured by the explanatory variable.

5.4.2 Reliability Block Diagram

Each contractor planned to build a RBD to estimate the reliability of the system. The RBD modelling will evolve as the project matures, hence the decision maker believed each contractor would submit a revised RBD at intervals throughout the project. The decision maker believed at the beginning of the procurement process, the contractor would provide a RBD in which she could not place a great deal of confidence in because it would be populated with data from information sources such as generic databases and unstructured engineering judgement, which are not considered relatively accurate. In addition, she believed the quality of the RBD will reflect the degree of system design detail at the stage in the programme at which the modelling is conducted. For example, at the beginning of the project, detailed knowledge regarding the system design may not be available. As time progresses, the contractor will be able to populate the model with data from other information sources, such as test data and historical data, and the RBD will more closely represent the detailed structure of the final system. Hence the confidence the decision maker places on the RBD analysis will grow.

The decision maker decided to assess the RBD modelling in three stages so the different confidence placed in the RBD at different phases of the procurement process could be captured. RBD_A is the output of a RBD populated with data in which the decision maker places considerable confidence in. RBD is the 'true' RBD value - i.e. the output from RBD modelling carried out perfectly and with data that had no uncertainty. The decision maker believed the critical hardware MTBF of the system in a controlled environment, i.e. X_R, influenced the output of RBD in the following way; $RBD = X_R + R_{RBD}$ and $RBD_A = RBD + R_{RBD_A}$. The decision maker believed the output of RBD and RBD_A were not influenced by PA given that she knew X_R with complete certainty. Hence X_R is Bayes linear sufficient for RBD and RBD_A for adjusting PA.

The decision maker wished to incorporate into the modelling her belief that she would receive a RBD early in the project life which would influence her belief about X_R. RBD_C is modelled as the output of the first submitted RBD and the one in which the decision maker places least confidence. RBD_B represents the intermediate RBD which will be populated with data from sources, some of which the decision maker has confidence in, and some of which she does not. When a RBD is submitted, the decision maker will decide into which of the three categories it fall, and populate the model accordingly. The following two formulas were used to model this belief; $RBD_B = RBD + R_{RBD_B}$ and $RBD_C = RBD + R_{RBD_C}$. The decision maker stated that if she knew RBD, then her belief about RBD_A would be unaffected if she learnt either RBD_B or RBD_C. Hence, RBD is Bayes linear sufficient for RBD_A, RBD_B and RBD_C for adjusting any of them.

5.4.3 Failures Modes, Effects and Criticality Analysis

The contract requires that the contractor build a functional FMECA. The decision maker believes this FMECA can be classed in one of three ways: the FMECA is considered to be well constructed indicating reliable system design; the FMECA is considered to be well constructed, but indicates an unreliable system design; or the FMECA is considered to be poorly constructed. The decision maker defined a well-constructed FMECA as one in which the system representation is sound and sufficiently detailed. She believed observing a well-carried out FMECA would reduce her prior uncertainty in X_R as it suggests reliability is a priority to the contractor. If she observed a well-carried out FMECA that indicated that the system was reliable, then she believed her expectation of X_R would rise and her variance of X_R would decrease. She believed that if she observed a well-carried out FMECA, even if it indicated the system would not be reliable, then her expectation of X_R would drop and her variance of X_R would decrease. She believed that if she observed a poorly carried out FMECA or observed no FMECA, then the mean and variance of X_R would remain the same.

In order to model this, a proxy variable FM, such that $E(FM) = 0$, $var(FM) = 1$ and $FM = \alpha_{FM}X_R + R_{FM}$, was created. The decision maker was then asked to specify her belief about the 5, 50 and 95 percentiles of X_R given that she had observed a satisfactory FMECA indicating a reliable system was being designed. From these values, $var_{FM}(X_R)$, $cov(FM, X_R)$ and α_{FM} were calculated. This was achieved using the formulas given for adjusted uncertainty in Section 4.2.3.2 For α_{FM}, the following formula was used.

$$\begin{aligned}
cov(X_R, FM) &= cov(X_R, \alpha_{FM}X_R + R_{FM}) \\
&= \alpha_{FM}cov(X_R, X_R) + cov(X_R, R_{FM}) \\
&= \alpha_{FM}var(X_R) + 0 \\
&= \alpha_{FM}var(X_R)
\end{aligned}$$

The next step involves calculating the value to be entered into the model for the two different scenarios. $E_{FM=R}(X_R)$ is the adjusted expectation of X_R given that a well-carried out FMECA indicating a reliable system is observed. $E_{FM=U}(X_R)$ is the adjusted expectation of X_R given that a well-carried out FMECA but indicating an unreliable system design is observed. The difference between $E_{FM=R}(X_R)$ (as calculated using the percentiles specified above) and $E(X_R)$ was set equal to FM_R. Using the Bayes linear adjusted expectation formula, $\widetilde{FM_R}$ was input into the model and calculated in the following way.

$$E_{FM=R}(X_R) = E(X_R) + cov(X_R, FM)var^{-1}(FM)(FM - E(FM))$$

$$E_{FM=R}(X_R) = E(X_R) + \alpha_{FM}\sigma^2_{X_R}\sigma^{-2}_{FM}(\widetilde{FM} - 0)$$

$$E_{FM=R}(X_R) - E(X_R) = \widetilde{FM}_R\alpha_{FM}\sigma^2_{X_R}\sigma^{-2}_{FM}$$

$$FM_R = \widetilde{FM}_R\alpha_{FM}\sigma^2_{X_R}\sigma^{-2}_{FM}$$

$$\widetilde{FM}_R = \frac{FM_R}{\alpha_{FM}\sigma^2_{X_R}\sigma^2_{FM}}$$

$\widetilde{FM}_R$ is entered into the model if the decision maker observes a well constructed FMECA indicating a good system design. If the decision maker observed a well-carried out FMECA that she believes indicates that the system is likely to be unreliable, then the $E_{FM=U}(X_R)$ drops. Assume $\widetilde{FM}_U$ is to be entered into the model in this scenario. The decision maker stated that the decrease in $E(X_R)$ from observing a well carried out FMECA indicating poor system design was identical to the increase in $E(X_R)$ from observing a well carried out FMECA indicating a reliable system design. From this, $\widetilde{FM}_U$ can be calculated.

If the decision maker observed a poorly carried out FMECA or no FMECA was carried out, then nothing was input into the model. From this, the variance of X_R remained the same.

5.4.4 Harbour and Sea Acceptance Test

The Harbour Acceptance Test (HAT) and Sea Acceptance Test (SAT) are tests which the system must go through before it is deemed safe to be placed on the platform and in-service. For the HAT, the system is placed on the platform and the system performs basic tasks. For the SAT, the platform operates with the system at sea for under 24 hours. Whilst these short tests are not aimed at demonstrating reliability requirements, the decision maker commented that they contributed to her overall belief about the reliability of the system. Neither test will last longer than 24 hours. Because the critical hardware MTBF is likely to be greater than 1000 hours, it is unlikely a failure will occur during either test. If a critical hardware failure does occur, the decision maker believed this indicated that there may exist integration issues between the system and the platform. Two situations could occur during the HAT or SAT that would influence the decision maker's belief about X.

- A critical hardware reliability failure could occur during either test. If this happened, the decision maker's belief about her expected value of X would decrease.
- If a critical hardware reliability failure does not occur, then she believed that her lower bound for the value of X would increase.

The decision maker stated that due to the short length of the trials, she did not believe the outcome of the two trials were correlated. She believed it was possible if the system had integration issues and failed one test, it may still pass the other test.

This scenario is modelled using proxy variables. HAT and SAT were set-up as proxy variables such that $HAT = X_{I_1}$ and $SAT = X_{I_2}$. As $E(X_{I_1}) = E(X_{I_2}) = 0$, $E(HAT) = E(SAT) = 0$. From this, $var(HAT) = var(X_{I_1})$ and $var(SAT) = var(X_{I_2})$. In order to gather the variance of each variable, it is necessary to calculate $\alpha_{X_{I_1}}$ and $\alpha_{X_{I_2}}$. In order to do this, the decision maker must state their belief about $cov(HAT, X_I)$ and $cov(SAT, X_I)$. From these values, it is possible to calculate $\alpha_{X_{I_1}}$ and $\alpha_{X_{I_2}}$ as $cov(HAT, X) = cov(HAT, X_{I_1})$ and $cov(SAT, X) = cov(SAT, X_{I_2})$.

The following discusses the modelling for HAT. Similar modelling is carried out for SAT. In the situation where the system passes the HAT, the decision maker believed her belief about $var(X)$ would fall. She specified her median and 95% quantile beliefs would remain the same, however, her belief about the 5% quantile would rise. Because of this, her belief in $E(X)$ would also rise. Using the value gathered from the reduction in variance, the $cov(X, HAT)$ can be calculated using the formula:

$$cov(X, Y) = \sqrt{(var(X) - var_{HAT}(X))}$$

From this, $\alpha_{X_{I_1}}$ can be calculated using the following result:

$$\begin{aligned} cov(X, HAT) &= cov(X_{I_1} + X_{I_2} + X_R, X_{I_1}) \\ &= cov(X_{I_1}, X_{I_1}) + cov(X_{I_2}, X_{I_1}) + cov(X_R, X_{I_1}) \\ &= var(X_{I_1}) + 0 + 0 \\ &= \alpha_{X_{I_1}} \end{aligned}$$

The next step is to calculate what value is entered into the model for the two scenarios. The difference between $E_{HAT}(X)$ and $E(X)$ when the system passes the HAT test is specified by the decision maker and set equal to P_{HAT}. Assume $\tilde{P}_{HAT}$ is input into the model:

$$E_{HAT}(X) = \overline{X} + cov(X, HAT)var^{-1}(HAT)(HAT - E(HAT))$$

$$E_{HAT}(X) - \overline{X} = \alpha_{X_{I_1}} \alpha_{X_{I_1}}^{-1} \widetilde{P_{HAT}}$$

$$\widetilde{P_{HAT}} = P_{HAT}$$

$\widetilde{P_{HAT}}$ is entered into the model if the system passes the HAT test. If the system fails the HAT test, then $E_{HAT}(X)$ drops. Assume $\widetilde{F_{HAT}}$ is to be entered into the model when the system fails. The decision maker must state F_{HAT} such that F_{HAT} is equal to the amount the decision maker believes $E(X)$ will shift by, given the system fails the test. From similar equations as above, $\widetilde{F_{HAT}}$ can be calculated. For the SAT test, similar calculations are carried out to calculate $\widetilde{P_{SAT}}$ and $\widetilde{F_{SAT}}$.

5.4.5 In Service Reliability Demonstration

An ISRD is a demonstration test carried out by the contractor over a pre-determined period of time aiming to demonstrate reliability requirements have been met. The output of the ISRD is denoted by $ISRD$. The decision maker believed the ISRD was sufficiently long that it provided her with information about the long term reliability of the system and the integration between system and platform. Hence, the following mathematical equation is used to denote the relationship between the output of the ISRD and X; $ISRD = X + R_{ISRD}$.

5.4.6 In-Service Monitoring

Each contractor has proposed an alternative to carrying out an ISRD. Instead, they have proposed that at the beginning of the in-service period, they closely monitor the reliability of the system for a pre-determined length of time. It was suggested that this length of time is similar to the length of an ISRD. However, the decision maker believes the ISM will not be monitored as closely as an ISRD. ISM is the output of the in-service monitoring period and this is modelled as $ISM = X + R_{ISM}$. The decision maker believed that the output of the ISM influenced her belief about X_R, X_{I_1} and X_{I_2}. The decision maker specified that if she knew for certain the value of X, learning about $ISRD$ would not influence her belief about ISM. Hence, $ISRD$ is Bayes linear sufficient for ISM for adjusting X.

5.4.7 Building the Bayes Linear Network

In order to build the necessary Bayes linear network, it is necessary to create a unique ordering. This network will be unique to the ordering given (Goldstein and Woolf, 2007). The ordering used in this project is chosen to minimise the number of arcs and ensure the network is simple. The ordering chosen is [X_R, X_{I_1} X_{I_2}, X, HAT, SAT $ISRD$, ISM, PA, RBD, RBD_A, RBD_B, RBD_C, FM,]. The first step is to draw the node X_R. The next step is to draw the node X_{I_1}. As X_R and X_{I_1} are uncorrelated, there are no arcs drawn at this point. The next step is to draw the node X_{I_2}. Again, as X_{I_2} is uncorrelated with everything, no arcs are drawn at this point. The next step is to draw a node for X. As X is influenced by X_R, X_{I_1} and X_{I_2}, three arcs are drawn, one from X_R to X, one from X_{I_1} to X and one from X_{I_2}.

The next two variables that need to be drawn are HAT and SAT. The decision maker believes HAT is influenced by X_{I_1}, SAT is influenced by X_{I_2} and each of them are Bayes linear sufficient with everything else in the model for adjusting X_{I_1} and X_{I_2} respectively. Hence only arc from X_{I_1} to HAT and an arc from X_{I_2} to SAT is required.

The next step is to draw the nodes $ISRD$ and ISM. As X influences the belief about both these variables, an arc is drawn from X to each of the variables. As $ISRD$ and ISM are Bayes linear sufficient for adjusting X with everything else in the model, there are no other arcs added. The next step is to draw the node PA. As X_R influences the decision maker's belief about PA, an arc is drawn from X_R to PA. No other arcs are drawn as PA is Bayes linear sufficient with everything else in the model for adjusting X_R. The next node to be drawn is RBD. As the decision maker believes that her belief of RBD is influenced by X_R, an arc is drawn from X_R to RBD. No other arcs are drawn as RBD is Bayes linear sufficient with everything else for adjusting X_R. The next nodes are RBD_A, RBD_B and RBD_C. Each of these are influenced by RBD, hence an arc is drawn from RBD to each of the three nodes. As each of them are also Bayes linear sufficient with everything else in the model for adjusting RBD, no other arcs are required.

The next variable is FM. The decision maker believed FM and X_R were correlated, hence an arc was drawn from X_R to FM. The decision maker specified that if she knew X_R, then learning about any other variables in the model would not influence her belief about FM. Therefore X_R is Bayes linear sufficient for adjusting FM with everything else in the model.

The above modelling can be represented by Figure 5.1.

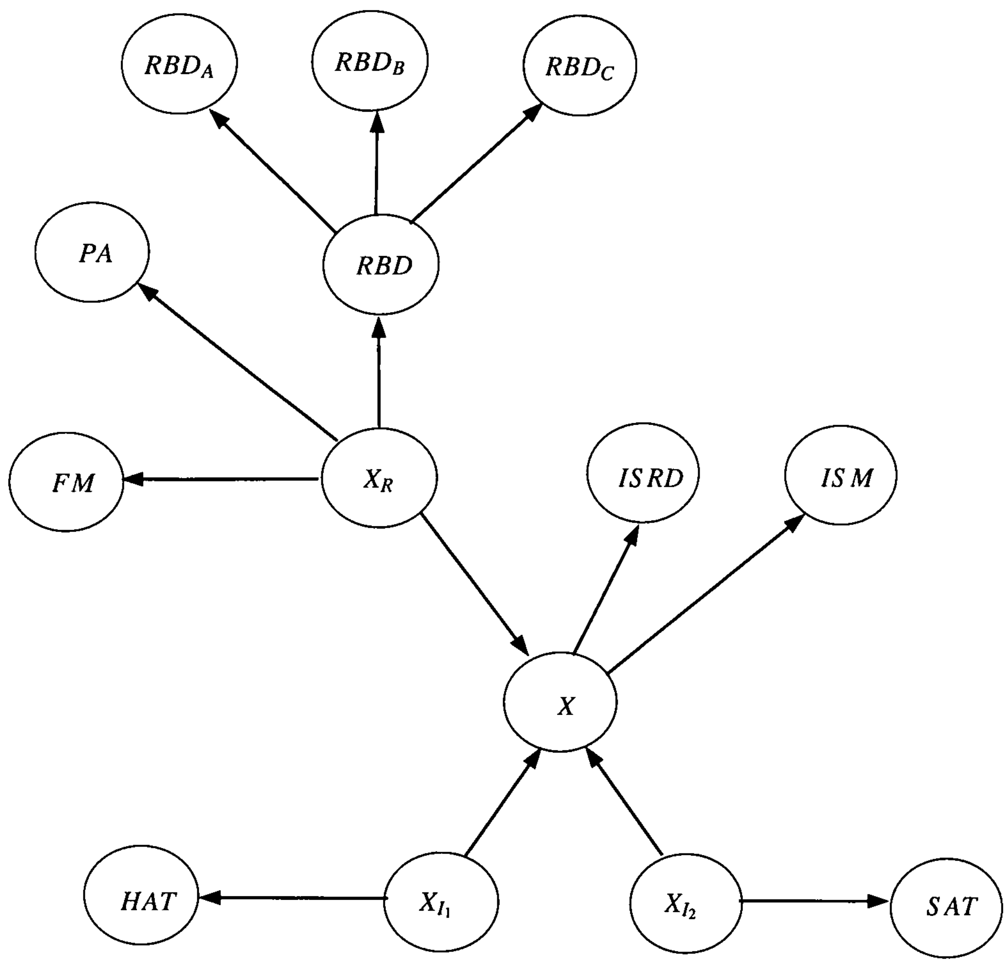

Figure 5.1: Bayes linear network of contractors R&M programme

This is the BLN for all three of the contractors combined. Contractor Yellow plans to revise their RBD once, so the RBD_B node is removed. Contractor Blue plans to carry out all the tasks so their BLN is identical to Figure 5.1. Contractor Grey plans to revise their RBD only, so the RBD_B node is removed, and they are not planning to carry out Parts Count Analysis, so PA is removed from Figure 5.1. Appendix D contains the BLN for each of the three contractors.

5.5 Eliciting Beliefs and Populating Model

This section will discuss the quantitative part of the modelling. This thesis will only illustrate how the necessary values were elicited for contractor Yellow although the same process was used for contractors Blue and Grey.

It is possible contractor Yellow will conduct out any of the following tasks: two RBD's, a FMECA, Parts Count Analysis, ISRD, ISM, HAT and SAT. In order to fully populate the model, it was necessary to gather the decision maker's belief about the expectation of 13 variables and the corresponding 13 x 13 covariance matrix. Whilst this may appear a large number of values to specify, some are already known as they are proxy variables and many of the values can be deduced from others because of the way in which the model was built. For example, if the $cov(X_Y, RBD_Y)$ and $cov(X_Y, IRSD_Y)$ is known, then it is trivial to calculate $cov(ISRD_Y, RBD_Y)$. This is achieved in the following way;

$$
\begin{aligned}
cov(ISRD_Y, RBD_Y) &= cov(X_Y + R_{ISRD_Y}, X_{R_Y} + R_{RBD_Y}) \\
&= cov(X_{R_Y} + X_{I_Y} + R_{ISRD_Y}, X_{R_Y} + R_{RBD_Y}) \\
&= cov(X_{R_Y}, X_{R_Y}) + cov(X_{R_Y}, R_{RBD_Y}) + cov(X_{I_Y}, X_{R_Y}) \\
&+ cov(X_{I_Y}, R_{RBD_Y}) + cov(R_{ISRD_Y}, X_{R_Y}) + cov(R_{ISRD_Y}, R_{RBD_Y}) \\
&= var(X_{R_Y}) + 0 + 0 + 0 + 0 + 0 \\
&= var(X_{R_Y})
\end{aligned}
$$

It is important the questions are ordered correctly. The ordering used reflects the unique ordering used above. If a variable has an explanatory variable, it is important the expectation and variance of the explanatory variable, i.e. the parent variable, has been elicited prior to the child variable. From this, the expectation and variance of the residual variable R_i, only needs to be elicited.

5.5.1 MTBF of the system

Analyst: What is your belief about the 5, 50 and 95 percentiles of the long-term critical hardware MTBF of Yellow's Sonar system?

Decision maker: She believed it was possible the system would be as low as 750 hours or as high as 3200. She believed it was equally likely to be above or below 1500 hours.

From this, the expectation and variance of X_R can be calculated using the Pearson-Tukey method; $E(X_R) = 1675.75$ and $var(X_R) = 554605.9$. As $E(X) = E(X_R) + E(X_{I_1}) + E(X_{I_2})$ and it has already been stated $E(X_{I_1}) = E(X_{I_2}) = 0$, then $E(X) = E(X_R)$. At this stage, the variance of X cannot be calculated.

However, $cov(X, X_R)$, $cov(X, X_{I_1})$ and $cov(X, X_{I_2})$ can be calculated to be equal to $var(X_R)$, $var(X_{X_1})$ and $var(X_{X_2})$ respectively. At this point, the expectation table and covariance matrix can be partially populated.

Variable	Expectation
X_R	1675.75
X_{I_1}	0
X_{I_1}	0
X	1675.75

Table 5.1: Yellow's expectations for selected variables

Variable				
X_R	554605.9			
X_{I_1}	0	$\alpha_{X_{I_1}}$		
X_{I_2}	0	0	$\alpha_{X_{I_2}}$	
X	554605.9	$\alpha_{X_{I_1}}$	$\alpha_{X_{I_2}}$	$554607.9 + \alpha_{X_{I_1}} + \alpha_{X_{I_2}}$

Table 5.2: Yellow's covariance matrix for selected variables

5.5.2 HAT

Analyst: Assume your belief of the 5, 50 and 95 percentile for X is 750, 1500 and 3200. Given that the system passes the HAT, what is your new belief about the 5, 50 and 95 percentile for X? Given that your initial expectation is 1675.75, what is your belief about the expectation given that the system fails the HAT test?

Decision maker: She believed her belief about the 50 and 95 percentile would remain the same, but her belief about the 5 percentile would be 775. If the system failed the test, her expectation of X would drop by 100 hours

Using the Pearson-Tukey method, the adjusted variance of X given that HAT is observed is 543347.816 and the adjusted expectation is 1680.375. Hence $cov(X, HAT) = 106.1$. From this $\alpha_{X_{I_1}}$ can be calculated to be equal to 106.1. If the system passes the HAT, the expectation rises by 4.625 and if it fails the HAT, the expectation drops by 100. Using the following formula, $\widetilde{P_{HAT}} = P_{HAT}$, the values input into the model can be calculated where P_{HAT} is the difference between the original expectation and the adjusted expectation. From this, if the system passes the test, the decision maker should input 4.625 into the model and if the system fails the test, the decision maker should input -100 into the model.

5.5.3 SAT

Analyst: Assume your belief of the 5, 50 and 95 percentile for X is 750, 1500 and 3200. Given that the system passes the SAT, what is your new belief about the 5, 50 and 95 percentile for X? Given that your initial expectation is 1675.75, what is your belief about the expectation given that the system fails the SAT test?

Decision maker: She believed her belief about the 50 and 95 percentile would remain the same, but her belief about the 5 percentile would be 800. If the system failed the test, her expectation of X would drop by 300 hours.

As above, $cov(X, SAT) = 149.7$ and $\alpha_{SAT} = 0.745$. If the system passes the HAT, the expectation rises by 9.25 and if it fails the HAT, the expectation drops by 300. Using the formulas above, if the system passes the test, $\widetilde{P_{SAT}} = 9.25$ and if the system fails the test, $\widetilde{F_{SAT}} = -300$

Table 5.2 can now be updated to give Table 5.3.

5.5.4 ISRD

Analyst: If you knew with complete certainty the critical hardware MTBF of Yellow's Sonar system was 1500 hours, what is your belief about the 5, 50 and 95 percentile values the output of the ISRD could take?

Variable						
X_R	554605.9					
X_{I_1}	0	106.1				
X_{I_2}	0	0	149.7			
X	554605.9	106.1	149.7	554861.7		
HAT	0	106.1	0	106.1	106.1	
SAT	0	0	149.7	149.7	0	149.7

Table 5.3: Yellow's updated covariance matrix for selected variables

Decision maker: She believed if the MTBF of the system was 1500 hours, then the output of the ISRD could be as high as 1700 hours, as low as 1300 hours and is equally likely to be above or below 1500 hours.

From this, $E(ISRD) = 1675.75$ and $var(ISRD) = 569642.6$ can be calculated. Similarly, $cov(ISRD, X) = var(X)$, $cov(ISRD, X_R) = var(X_R)$ and $cov(ISRD, X_I) = var(X_I)$. Therefore the next line in the covariance matrix is [554605.9, 106.1,149.7, 554861.7, 106.1, 149.7, 569642.6].

5.5.5 ISM

Analyst: If you knew with complete certainty the critical hardware MTBF of Yellow's Sonar system was 1500 hours, what is your belief about the 5, 50 and 95 percentile values the output of the ISM could take?

Decision maker: She believed if the MTBF of the system was 1500 hours, then the output of the ISM could be as high as 1800 hours, as low as 1200 hours and is equally likely to be above or below 1500 hours.

From this, the $E(ISM) = 1675.75$ and $var(ISM) = 587620.8$ can be calculated. As before, it is trivial to calculate $cov(ISM, X) = var(X)$, $cov(ISM, X_R) = var(X_R)$, $cov(ISM, X_I) = var(X_I)$ and $cov(ISM, ISRD) = var(X)$. The next line in the covariance matrix is [554605.9, 106.1, 149.7, 554861.7 , 106.1, 149.7, 554861.7, 587620.8].

5.5.6 Parts Count Analysis

Analyst: If you knew with complete certainty the long-term critical hardware MTBF of Yellow's Sonar system was 1500 hours, what is your 5, 50 and 95 percentiles values the output of the Parts Count Analysis could take?

Decision maker: She believed if the long-term MTBF of the system was 1500 hours, then the output of the Parts Count Analysis could be as low as 900 hours, as high as 3200 hours and is equally likely to be above or below 1700 hours.

From this, $E(PA)$ = 2005.25 and $var(PA)$ = 1043371.7. The next line in the covariance matrix is [554605.9, 0, 0, 554605.9,0,0, 554605.9, 554605.9, 1043371.1].

5.5.7 RBD

Analyst: If you knew with complete certainty the long-term critical hardware MTBF of Yellow's Sonar system was 1800 hours, what is your 5, 50 and 95 percentiles values a perfect RBD could take?

Decision maker: She believed if the long-term MTBF of the system was 1800 hours, then the output of a perfect RBD could be as low as 1500 hours, as high as 2100 hours and is equally likely to be above or below 1800 hours.

From this, $E(RBD)$ = 1675.75 and $var(RBD)$ = 587865.9. The next line in the covariance matrix is [554605.9, 0, 0, 554605.9, 0, 0, 554605.9, 554605.9, 554605.9, 587865.9]. The decision maker believed the contractor would carry out two RBD's throughout the life of the project. RBD_C is the output of an early RBD the contractor produces and RBD_A is the output of a revised RBD.

Analyst: If you knew with complete certainty the output of a perfect RBD was 1800 hours, what is your 5, 50 and 95 percentiles values RBD_A and RBD_C could take?

Decision maker: She believed the output of a RBD_A could be as low as 1650 hours, as high as 2100 hours and is equally likely to be above or below 1900 hours and the output of a RBD_C could be as low as 1600 hours, as high as 2200 hours and is equally likely to be above or below 1900 hours.

From this, the $E(RBD_A)$ = 1766.5, $var(RBD_A)$ = 606570.3, $E(RBD_C)$ = 1775.75 and $var(RBD_C)$ = 621124.2. The next two lines of the covariance matrix are [554605.9, 0, 0, 554605.9, 0, 0, 554605.9, 554605.9, 554605.9, 587865.9, 606570.3] and [554605.9, 0, 0, 554605.9, 0, 0, 554605.9, 554605.9, 554605.9, 587865.9, 587865.9, 621124.2] respectively.

5.5.8 FMECA

Analyst: You stated your belief of the 5, 50 and 95 percentiles of X_R as 750, 1500 and 3200 respectively. If you were to observe a satisfactory FMECA that indicated the new system was reliable, what are your new 5, 50 and 95 percentiles for X_R? In addition, your initial expectation of X_R was 1675.75. If you were to observe a satisfactory FMECA that indicated the new system was not reliable, what would be your new expectation for X_R?

Decision maker: The decision maker believed if she observed a well-carried out FMECA that indicated that a reliable system was being procured, her new percentiles for X_R were 900, 1850 and 2800. She believed the increase in the expectation of X_R when she observes a well-carried out FMECA that indicates a reliable system is being procured, is mirrored by a decrease in expectation if a well-carried out FMECA that indicates an unreliable system is being procured is observed.

Using the above values, the prior and adjusted variance of X_R are 554605.9 and 369536.8 respectively. Using the following formula from section 4.2.3.2 $cov(X_R, FM) = (var(X_R) - var_{FM}(X_R)var(FM))^{\frac{1}{2}}$. From this, $cov(X_R, FM) = 430.2$. α_{FM} can also be calculated. As $cov(X_R, FM) = \alpha_{FM} var(X_R)$, then $\alpha_{FM} = \frac{cov(X_R,FM)}{var(X_R)} = 0.000776$.

In addition, $\widetilde{FM}_R$ and $\widetilde{FM}_U$ can be calculated. The increase in expectation for a well-carried out reliable FMECA is 174.25. From Section 5.4.5, $\widetilde{FM}_R = \frac{174.25}{0.000776*554605.9} = 0.405047$. Similarly, $\widetilde{FM}_U = -0.405047$. The next line of the covariance table becomes [430.2, 0, 0, 430.2,0,0 430.2, 430.2, 430.2, 430.2, 430.2, 430.2, 1].

5.6 Analysing Belief Structure Given Observations

The model has been successfully populated for each of the three contractors. Appendix E contains the completed expectation table and covariance matrix for each contractor. The next step is to carry out the necessary analysis that will answer the questions the decision maker identified in Section 5.2 by meeting the aims of Section 5.3.

5.6.1 Current Assessment of Contractors Reliability

The decision maker's prior beliefs can be found in Table 5.4.

Variable	Expectation	Variance	Standard deviation
X_Y	1675.75	554861.7	744.9
X_G	1201.75	497741	683.9
X_B	1201.75	497741	683.9

Table 5.4: Decision maker's prior belief for all contractors

Contractor Yellow has carried out a Parts Count Analysis and observed an output of 1826 hours. Hence $PA_Y = 1826$ was input into Contractor Yellow's model. Contractor Grey has carried out RBD modelling. The output of the modelling was 1810 hours; hence $RBD_{C_G} = 1810$ was input into the model. Contractor Blue has carried out RBD modelling. The output of this modelling was 3993 hours; hence $RBD_{C_B} = 3993$ hours. The expectation and variance of X for each of the three contractors is adjusted based on this information.

Variable	Adjusted Expectation	Adjusted Variance	Adjusted Standard Deviation
X_Y	1580	260059	509.9
X_G	1285	159086.8	398.9
X_B	1936.2	303623	551.0

Table 5.5: Adjusted beliefs given current observed data

From Table 5.5, based on the decision maker's prior beliefs and the observations made to date, the decision maker believes the expectation of contractor Blue's system is the greatest. The standard deviation has remained high at 550.9 hours. This is a reduction in standard deviation of 19%. The expectation of contractor Grey's system is the lowest and the standard deviation has dropped to 398.9, which is a reduction of 43%. The expectation of contractor Yellow's system has dropped to 1580. The standard deviation is 509.7; a drop of 34%.

5.6.2 Uncertainty of Critical Reliability

The Bayes linear methodology is capable of assessing the uncertainty surrounding an uncertain variable given that certain information has been observed. In order to make this adjustment, the observed value does not need to be gathered. Three scenarios were of interest to the decision maker.

1. Uncertainty surrounding X given the contractor carries out the tasks stated in the R&M Programme (excluding ISM and ISRD)

2. Uncertainty surrounding X given the contractor carries out the tasks stated in the R&M Programme and ISM

3. Uncertainty surrounding X given given the contractor carries out the tasks stated in the R&M Programme and an ISRD (but no ISM)

The variance for X for each of the three above scenarios were modelled using the Bayes linear adjusted variance formula and are displayed in Table 5.6. The standard deviation of each is given in brackets.

Variable	Current	Scenario 1	Scenario 2	Scenario 3
X_Y	260378 (510.3)	39105.9 (197.8)	17548 (132.5)	10583 (102.9)
X_G	159169 (398.9)	46305 (215.2)	19363 (139.2)	11205 (105.8)
X_B	303550 (551.0)	77698 (278.7)	23288 (152.6)	12417 (111.4)

Table 5.6: Adjusted variance (standard deviation) for three different scenarios

5.6.2.1 Scenario 1

The decision maker was particularly interested in the adjusted variance of Scenario 1. This is because she believed it is possible the contractor would not gather any in-service data prior to the system being integrated with the platform. Hence, she was very interested in assessing the unresolved uncertainty once all the reliability modelling and testing has been carried out by each contractor.

Contractor Yellow reduces the decision maker's uncertainty by the greatest amount. It should also be noted that Yellow's prior uncertainty was the greatest of the three. Table 5.6 shows there is a small difference between the reduction of uncertainty of contractors Yellow and Grey. Both will reduce the decision maker's uncertainty by a substantial amount. The difference between the uncertainty of contractor Yellow after they have carried out all their tasks and the uncertainty of contractor Grey after they have carried out all their tasks is small compared to their initial uncertainty, i.e. less than 2%.

Alternatively, Table 5.6 shows contractor Yellow and Grey will reduce the uncertainty of the decision maker substantially more than contractor Blue. Once contractor Blue carries out all their reliability modelling, the variance of X_B is double the variance of X_Y and substantially greater than the variance of X_G. Now that the model has identified a difference exists between the unresolved uncertainty of the three contractors, it is necessary to carry out further investigation and attempt to identify why contractor Blue's unresolved uncertainty is so much greater than Yellow and Grey.

The Bayes linear methodology is capable of doing this. The reason why the unresolved uncertainty is so high is because of the level of confidence the decision maker placed on contractor Blue's RBD modelling. The decision maker commented that because contractor Blue uses a scaling factor within their RBD modelling which is not fully justified, she believed if X_{R_B} was equal to 1500 hours, the output from the RBD modelling may be as high as 4000 hours. As the RBD is the main source of reliability modelling being carried out by each contractor, it has a large effect on the adjusted variance of X. Table 5.7 shows the correlation between RBD_A and X for each contractor.

$corr(RBD_{A_Y}, X_Y)$	0.9562
$corr(RBD_{A_G}, X_G)$	0.9466
$corr(RBD_{A_B}, X_B)$	0.7730

Table 5.7: Correlation values for each contractor

Table 5.7 illustrates that the correlation between X_Y and RBD_{A_Y} is greatest with the correlation between X_B and RBD_{A_B} being the smallest. This lack of confidence on the main source of information means the uncertainty surrounding X_B does not drop as much as the uncertainty of X_Y when RBD_{A_B} and RBD_{A_Y} are carried out respectively. It is unlikely only Bayes linear modelling is capable of gaining this insight. Using a structured elicitation process and forcing the decision maker to quantify her beliefs about the dependencies, allows this information to be explicitly stated. Using an unstructured elicitation process, the strength of the dependencies is not, at first glance, apparent.

5.6.2.2 Scenario 2

The difference between Scenario 1 and Scenario 2 is some form of in-service data is being gathered in Scenario 2. The model allows the decision maker to assess the reduction in uncertainty of gathering the extra information prior to observing it. The decision maker can then decide whether or not the extra time and cost of carrying out the in-service monitoring is worth the extra information it provides. This scenario assumes all the reliability modelling the contractor state they will carry out has already been previously observed.

Variable	Prior Variance	Prior Standard Deviation	Adjusted Variance	Adjusted Standard Deviation
X_Y	37059	192.5	17548 (52.6%)	132.5 (31.1%)
X_G	46305	215.2	19363 (58.2%)	139.2 (35.3%)
X_B	77698	278.7	23288 (70.0%)	152.6 (45.3%)

Table 5.8: Decrease in variance and standard deviation by ISM

From Table 5.8, it can be seen that carrying out ISM has a different effect on each of the three contractors. The value in brackets represents the percentage drop in variance and standard deviation when ISM is observed compared to when just the modelling is observed. Where the prior variance is greater, the effect of ISM is greater.

5.6.2.3 Scenario 3

For Scenario 3, the decision maker observes an ISRD in addition to all the previous modelling. In this scenario, the ISM has not been observed. Table 5.9 displays the adjusted uncertainty. As before, the value in brackets represents the percentage drop in variance and standard deviation when $ISRD$ is observed compared to when just the modelling is observed.

Variable	Prior Variance	Prior Standard Deviation	Adjusted Variance	Adjusted Standard Deviation
X_Y	37059	192.5	10583 (71.4%)	102.9 (46.6%)
X_G	46305	215.2	11205 (75.8%)	105.9 (50.8%)
X_B	77698	278.7	12417 (84.0%)	111.4 (60.0%)

Table 5.9: Decrease in variance and standard deviation by ISRD

As before, the effect of the ISRD in reducing uncertainty is greatest when the prior variance is greater. Furthermore, the difference between the residual uncertainty of X_Y, X_G and X_B once $ISRD_Y$, $ISRD_G$ and $ISRD_B$ have been observed, is very small compared with the initial standard deviation. If each contractor were to carry out all the tasks they say they will carry out and the tasks the MOD wish them to carry out, then the difference between the adjusted variance for each of them is negligible when compared to the initial variance.

5.6.3 Sensitivity Analysis

Sensitivity analysis in this project will focus on modifying the covariance structure for each contractor and assessing the change needed in the decision maker's prior beliefs so that the adjusted variance of contractor Yellow's R&M Programme is greater than contractor Blue. The reason why this scenario is investigated is because the adjusted expectation of X_G is substantially lower than either Yellow or Blue and it is likely from a reliability perspective that Grey's proposal will be dismissed.

5.6.3.1 Assessing the Covariance Structure

As many of the sources of information have the same residual uncertainty, i.e. HAT, SAT, FM, ISM and $ISRD$, sensitivity analysis only needs to focus on the 'unique' observations. The decision maker believed contractor Yellow would carry out a parts count analysis and two RBD's. The decision maker believed contractor Blue would carry out a parts count analysis and three RBD's. However, the residual uncertainty associated with Blue's RBD is considerably greater than the residual uncertainty of Yellow's RBD's.

Given the reliability modelling carried out to date, the adjusted variance of X_B = 303623 and X_Y = 260059. In order for the adjusted variance of $X_B < X_Y$, it is necessary to increase the correlation between RBD_{C_B} and X_B or decrease the correlation between PA_Y and X_Y. Increasing the correlation between RBD_{C_B} and X_B was not considered an option as the decision maker firmly believed that she was being generous to contractor Blue with the values she stated. In order for the adjusted variance of $X_B < X_Y$, the residual uncertainty between PA_Y and X_Y would have to be increased from 489509 to 673509. This would be equivalent to changing the 5, 50 and 95 percentiles from -700, 200 and 1700 to -900, 200 and 1900. This information was presented to the decision maker who stated that she believed it was possible but unlikely the proposed beliefs would fit with her 'true' belief.

Once all the modelling has been carried out, the adjusted variance of X_B = 77698 and X_Y = 39101. One way for the adjusted variance of $X_Y \geq X_B$, is to increase the residual uncertainty between RBD_Y and X_{R_Y}. Currently, the residual uncertainty between RBD_Y and X_{R_Y} is equal to 33260. For the adjusted variance of $X_Y > 77698$, the residual uncertainty between RBD_Y and X_{R_Y} would have to rise from 33260 to 171260. The decision maker believed the 5, 50 and 95 percentiles of the residual uncertainty between RBD_Y and X_{R_Y} are 1500, 1800 and 2100. For the residual uncertainty to rise to 171260 and to maintain a symmetric shape, the percentiles would have to be 1100, 1800 and 2500. This information was also presented to the decision maker who stated that she believed these values definitely did not represent her prior belief.

5.6.4 Summary of Analysis

This section will summarise the analysis carried out and identify that the initial aims of the modelling were met. Recall the modelling aims from Section 5.3. Table 5.5 illustrates the current belief of the decision maker given her prior beliefs and the observations made to date. From Table 5.5, the expectation of X_B is greater than the expectation of X_Y and X_G whilst the $var(X_G)$ is the smallest, followed by $var(X_Y)$ and then $var(X_B)$. From this, the first aim above has been achieved.

The second aim was to assess how much uncertainty was reduced by the contractor's R&M Programme. The decision maker was sceptical as to whether or not any of the contractors would carry out ISM or an ISRD. Therefore she was interested in assessing how much uncertainty would remain in her assessment of X for each contractor given they carry out the tasks they state they will. From Table 5.6 the $var(X_Y)$ is lower than both $var(X_G)$ and $var(X_B)$ for this scenario. In addition, $var(X_G)$ is lower than $var(X_B)$. The modelling also identified how much uncertainty will be reduced if the contractor carries out ISM or an ISRD.

Prior to the modelling carried out, the decision maker commented that she believed the R&M Programme of Yellow would reduce her uncertainty surrounding the reliability of the system by the greatest amount. However she was unsure about the difference between Grey and Blue's R&M Programmes. At the beginning of the project, the decision maker was interested in observing how the methodology would distinguish between contractor Blue's and contractor Grey's R&M Programmes.

The modelling carried out has demonstrated that there is a substantial difference between both contractor's R&M Programme's and in addition, the modelling has highlighted where this difference occurs. This disparity existed because of information known only to the decision maker. However, due to the number of information sources and the complexity of the inter-relationships between all these information sources, the decision maker was unable to process the information intuitively. Once the modelling was presented back to the decision maker and the reason for the difference in unresolved uncertainty was explained, the decision maker commented that it was now obvious to her the difference existed.

The methodology has demonstrated that whilst it appears at first glance that contractor Blue will produce a very reliable system, it is unlikely they would be able to provide the decision maker with assurance they were going to meet requirements due to the lack of confidence she places on their RBD. Alternatively, if contractor Grey was chosen, it is unlikely reliability requirements would be met, but the decision maker may have a good understanding of how reliable the system was actually going to be. Instead, contractor Yellow appears to be producing a reliable system and in addition, providing sufficient evidence to reduce the decision maker's uncertainty to an acceptable level.

5.7 Feedback to Decision Maker

As this project was carried out using an Action Research framework, it is necessary to attempt to measure the effect of the modelling carried out. At the end of the project, an interview was carried out with the decision maker. The decision maker stated that she found the process of structuring and building the model extremely useful. When structuring the model, it forced the decision maker to explicitly identify all the information sources that influenced her belief about X. The decision maker commented that she would not have previously considered using the information gathered during the HAT or SAT to explicitly influence her belief about X. The modelling allowed her to consider the effect that these variables would have on X and hence she was able to extract all possible information out of the R&M Case.

The process of building the Bayes linear network and explicitly identifying all the possible information sources that could be observed was useful. The decision maker commented that she found it simpler to consider the inter-relationships between all the variables when she could visualise the causal relationships between them. By identifying all the information sources that could be observed, the decision maker commented that it was more likely that the information would be used in her analysis. She commented that it was possible that in some large R&M Cases, information is lost within the large amount of text. By explicitly identifying the variables that need to be observed, the decision maker commented that it was more likely that she would remember to gather the values and use them in her analysis.

The decision maker stated that she found the process of quantifying her beliefs very useful. This process forced the decision maker to consider issues such as contractor Blue's scaling factor with the RBD and also to consider the effect that the outcome of the HAT and SAT would have on the reliability of the system. In addition, being asked to quantify her beliefs about the FMECA forced the decision maker to consider the FMECA's that the contractor may carry out. Therefore the process of structuring and building the model deepened the decision maker's understanding of the situation.

All of the analysis above was provided to the decision maker to support the advice she provided to the IPTL. It is difficult to measure the added benefit the decision maker received from the analysis. This is because the decision maker had already decided how to proceed prior to the modelling. The decision maker had decided to recommend to the IPTL that contractor Yellow should be chosen. She believed that the system they designed and built would be reliable and that they would provide her with assurance that reliability requirements have been met. Alternatively, she did not believe that either contractor Grey or Blue would design a reliable system and that she did not believe that either R&M Programme would provide her with sufficient assurance.

The modelling appears to at least partially support this belief. It is possible to automatically dismiss contractor Grey as their reliability is substantially lower than both Yellow and Blue's and their adjusted uncertainty given that they carry out the modelling they claim they will, is greater than contractor Yellow.

Once contractor Grey has been dismissed, it is slightly more difficult to distinguish between Blue and Yellow. On one hand, contractor Blue appears to be designing a system with a greater reliability than Yellow, however there exists a larger amount of uncertainty regarding this assessment. Contractor Yellow is producing a R&M Programme that will provide the decision maker with a greater amount of information regarding X but may not be as reliable as contractor Blue. The decision maker must decide whether or not the risk of accepting a system that may have a high or low reliability is acceptable. Alternatively, the decision maker may decide that the risk is unacceptable and that they would prefer to take slightly lower reliability, but with more confidence that it will be achieved.

In this case, the decision maker has decided to choose contractor Yellow. It is difficult to measure whether or not the modelling strengthened this belief or was ignored. However, what can be said is that the Bayes linear modelling, whilst maybe not having a direct effect on any decisions taken, has deepened the decision maker's understanding of the situation which in turn has given her more confidence in her recommendation to the IPTL. She stated that because she has mathematical modelling that supports her recommendation, the IPTL will have more faith in her advice.

5.8 Further Development

This section aims to discuss the future developments of the modelling within this project if more time were available with the MOD decision maker. Due to work commitments on behalf of the decision maker, the amount of time the analyst could spend with the decision maker was limited. There are two developments that could have taken place in this project; the first was to assess the reliability regarding all six reliability requirements and the second was to develop the FMECA variable further.

5.8.1 Assessing all Six Reliability Requirements

First, the modelling carried out has only addressed one of the reliability requirements - the critical MTBF of the hardware. There were a total of six reliability requirements - three relating to hardware and three relating to software. It would have been possible to model all six reliability requirements. If the situation was to be modelled again and it was decided that all six reliability requirements should be modelled, then it is possible that learning about whether or not a contractor is meeting one reliability requirement, may inform the decision maker about whether or not the contractor is meeting reliability requirements for another requirement. For example, assume the decision maker did not believe the contractor was going to meet the critical reliability requirement for software. She is then provided with information such that she now believes that the critical software reliability requirement has now been met or exceeded. Without receiving any additional information, it is feasible to believe that the decision maker now believes that the major MTBF of the software is now greater than what she previously thought. This reasoning and how to model it within the Bayes linear framework will be discussed further in Chapter 7.

5.8.2 Developing the FMECA Variable

Secondly, the modelling described above for the FMECA variable seems inadequate, however for pragmatic reasons, this is the best that could be carried out given the resources available to the researcher. The decision maker commented that she was uncertain as to whether or not the FMECA would output an MTBF value or if it would only be used as a design support tool. It would have been possible to model the scenario such that there were two FMECA variables - one for an observed FMECA with no output and one for an observed FMECA with an MTBF output. Assume that FM_1 represents the observation from a FMECA with no output and FM_2 represents the observation from a FMECA with an output. Then $FM_1 = \alpha_{FM_1} X_R + R_{FM_1}$ and $FM_2 = \alpha_{FM_2} X_R + R_{FM_2}$. A Bayes linear network of the scenario can be represented by Figure 5.2.

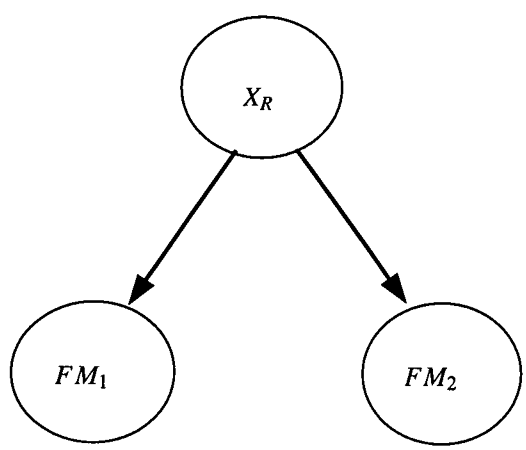

Figure 5.2: **Bayes linear network of developed FMECA modelling**

If the contractor carried out a FMECA that had no output, an observation for FM_1 could be made and FM_2 could be removed from the model. Alternatively, if the FMECA had an MTBF output, this value could be input as the observation of FM_2 and FM_1 could be removed from the model.

Another reason the FMECA modelling in this project felt inadequate was that the increase of $E(X_R)$ when a FMECA that indicated a reliable system was being designed was the same for all possible observed FMECA's. It is possible that a well-carried out FMECA may indicate that the system is substantially more reliable than the decision maker currently believed. This situation was modelled identically with the situation where a well-carried out FMECA indicated the system was marginally more reliable than the decision maker initially believed. Similarly, the decision maker commented that the FMECA may indicate that the system is unreliable, however she has confidence that the contractor has sufficient time and resources to make the necessary design changes. This means that it may be possible that over the length of the procurement project, the reliability of the system may rise again. If more time was available, it would be possible to measure the effect that the FMECA had on the reliability of the system in a more detailed manner.

Despite the fact that there were a number of improvements that could be made to the FMECA modelling, in the analysis carried out, the variable that was used was satisfactory. As no observation has yet to be made, the only effect this variable has to date is regarding the adjusted variance of X for each contractor. The decision maker was very specific that she believed that if a FMECA is "done well" it will lower the uncertainty surrounding the reliability of a system. She was very specific about the values by which this would drop. If the MOD were to continue the modelling as the project progressed, they could introduce the above modelling into the model if the contractor were to produce a FMECA with an MTBF output.

5.9 Conclusion

This chapter aimed to examine whether or not the Bayes linear methodology could be used to assess multiple contractor bids that have been submitted and support decision making. The common reliability decisions that are addressed at this stage of a procurement project were assessing the reliability of each contractor's systems and assessing how much uncertainty will remain in the assessment of the reliability once the contractor has carried out all the tasks from their R&M Programme. The Bayes linear modelling complemented the analysis already carried out by the MOD decision maker and offered additional analysis to the decision maker that has not previously been available.

The modelling was used to analyse the merit of different contractor's bids; both in terms of assessing the current reliability of the system and analysing the effectiveness of contractor's R&M Programmes in order to progressively reduce uncertainty as a procurement project proceeds. In all procurement projects, the decision maker must make an assessment of the current reliability of the system given the observations that have been made and how much uncertainty will remain once the system enters service. The Bayes linear methodology is capable of supporting both these decisions as demonstrated in the project.

Where there are multiple contractors bidding to build a system, the methodology is useful in determining which has the greatest reliability and which reduces the uncertainty of this assessment by the greatest amount. It can be used as a tool to compare multiple contractors. However, even if there is only one contractor submitting a bid, the modelling would allow the decision maker to assess at an early stage whether or not the contractor is likely to produce sufficient assurance. The decision maker could decide how much uncertainty they are willing to tolerate prior to the system entering service and use the methodology to assess whether or not the contractor has met this target.

Alternatively, this methodology could be viewed as a tool that can be used by a contractor to demonstrate the current reliability status of their system and that their reliability programme is reducing the decision maker's uncertainty surrounding the reliability of the system. If a contractor was to choose this method to demonstrate reliability requirements, they would have to be extremely transparent with how the prior beliefs were elicited.

The Bayes linear modelling was able to formally capture the subjective belief of the decision maker and report it in a concise manner. The methodology is capable of communicating to contractors and other interested parties the belief of the decision maker. It also allows a traceable path of the decision maker's reasoning to be seen by future analysts or future decision makers. This is extremely useful, particularly for the MOD where the decision maker in a project may change three or four times and future decision makers want to understand the reasoning of previous decision makers.

Chapter 6

Applied MOD Project 2: Assessing Data from a Reliability Growth Trial

6.1 Introduction

The MOD is procuring two land vehicles within a single procurement project. Both vehicles are individually designed and built, however, there are a number of sub-systems common to both vehicles. These include the hydraulic systems, power, electrical systems, suspension and other major sub-systems. In addition, the reliability programmes and analysis for both vehicles are being carried out by a single contractor. The contracted reliability requirement for both vehicles is being measured as a basic reliability and a mission reliability. Basic (and mission) reliability is defined as the probability the vehicle will complete a battlefield mission (BFM) without a basic (or mission) failure occurring. The reliability requirement for mission reliability for Vehicle A is 0.66 and for Vehicle B, 0.68. The reliability requirement for basic reliability is 0.16 for Vehicle A and 0.17 for Vehicle B. Both vehicles are purposely built to carry out tasks currently undertaken by vehicles that have been modified. Based on previous experience, these reliability requirements will be challenging to achieve, but not unrealistic. It is interesting that the MOD have made very small distinctions between the reliability of each vehicle; 0.02 for mission reliability and 0.01 for basic reliability. This suggests the MOD believe they are capable of distinguishing between the reliability of each vehicle at this level of detail, however the decision maker does not believe the end user would be able to distinguish between the a reliability of 0.66 and 0.68.

The project was in the Demonstration/Manufacture phase of the CADMID cycle when the following modelling took place. Between February 2005 and November 2005, prototypes for each vehicle were built and underwent Reliability Growth Trials (RGT). The contractor carried out 50 BFM; 26 for vehicle B and 24 for vehicle A, split into three phases. From March 2006 until September 2006, the contractor carried out a second group of RGT's on production vehicles with the intention of continuing to improve reliability and to demonstrate reliability requirements have been met. The contractor planned to carry out 30 BFM for vehicle A split into three phases with 10 in each. Similar plans were made for vehicle B. The modelling described below took place in April 2006.

Section 6.2 discusses the decisions the model was attempting to support and the current MOD process. Section 6.3 introduces the aim of the modelling carried out. Sections 6.4 and 6.5 discuss the qualitative and quantitative modelling respectively. Section 6.6 analyses the quantitative information gathered. Section 6.7 discusses the feedback given to the decision maker during the project. Section 6.8 discusses further development of the model. Section 6.9 will summarise the chapter.

6.2 Decision Making and Current MOD Modelling

Based on the evidence provided by the contractor, the decision maker must make a recommendation to the IPTL on whether or not either vehicle will meet, or has already met, their basic and mission reliability requirements. The single source of evidence presented at this time is the output of the RGT's. In order to determine whether or not either vehicle will meet their respective requirements, the MOD decision maker is interested in the current reliability of each vehicle and using her subjective engineering judgement, she will assess whether or not she believes each vehicle is likely to improve by the required amount needed to meet the requirement. To assess the current reliability of the system, the contractor uses the US Army Material Systems Analysis Activity (AMSAA) model.

The AMSAA model is a reliability growth model (Crow, 1974) that calculates an instantaneous reliability at the end of each of the three phases of the RGT. The AMSAA model is a development of the Duane model developed during an empirical study by Duane (1964). Using the AMSAA model and the data gathered during the three phases of the RGT, the contractor calculates an instantaneous Mean Missions Between Failures (MMBF) and using the formula $R_I = e^{\frac{-1}{\theta_i}}$, an instantaneous basic (or mission) reliability value is calculated, where θ_i = instantaneous MMBF. Despite the weaknesses in the AMSAA method identified in Section 1.2, the AMSAA output is the value the MOD R&M decision maker places most confidence on when assessing the current reliability of the systems.

Currently, any modelling carried out by the contractor aimed at assessing the future reliability is limited to using the AMSAA model. Using the AMSAA model and making some assumptions, the contractor can estimate the reliability of the system after the 3 phases as the duration of the RGT has been pre-determined. This method does not address softer factors that influence the decision maker's belief such as the relationship between the vehicles or the relationship between the requirements. Furthermore, whilst this analysis may be carried out by the contractor, it is not captured in the R&M Case report or communicated to the decision maker.

6.3 Bayes Linear Modelling

The purpose of the modelling was to develop a high-level methodology capable of assessing the basic and mission reliability of Vehicle A and Vehicle B using the observed data gathered during the RGT and expert assessment of the expected growth between different phases of the RGT's. The aim is to highlight to the decision maker at the earliest possible stage, whether or not the in-service basic or mission reliability of Vehicle A or Vehicle B is likely to be met. The modelling will also attempt to capture a number of softer factors that are not addressed by other statistical methods. The modelling is not aiming to analyse the data produced by the RGT's. Instead the modelling aims to assess the information that is extracted from the RGT data from other statistical techniques.

The modelling aim was subsequently broken down into 5 parts;

1. Model the basic reliability of Vehicle A
2. Model the basic reliability of Vehicle B
3. Model the mission reliability of Vehicle A

4. Model the mission reliability of Vehicle B

5. Model the relationship between all 4 of the above variables.

The modelling took place in February 2006. Between February and August 2006, Vehicle A was subjected to two phases of testing and Vehicle B was subject to three phases. All of the analysis took place in August 2006. As before, Figure 4.1 illustrates the steps needed to be taken in this project in order to carry out inference on the reliability parameters outlined above.

6.4 Structuring the Bayes Linear Model

This section will illustrate how the model was structured and built by the decision maker and the analyst for Vehicle B's basic reliability. Due to the structure of the RGT, the modelling for parts 1-4 are identical. The only difference between the four parts is the notation.

6.4.1 Structuring Parts 1-4

The first step is to identify the reliability parameter of interest. The decision maker's main interest is on the in-service basic and mission reliability of both vehicles. Hence, for this part, the parameter that inference will be carried out on is B_{IB} - Vehicle B's in-service basic reliability.

The next step is to identify all the variables that may influence B_{IB}. The decision maker stated she believed that between each phase of the RGT, modifications would be made to the system that would effect its basic reliability. As there are three phases of the RGT and then an in-service phase, four variables were identified as being equal to the basic reliability at the beginning of each phase. B_{1B} was set equal to the basic reliability of the Vehicle B prior to Phase 1 of the RGT; B_{2B} was set equal to the basic reliability of the Vehicle B prior to Phase 2 of the RGT, B_{3B} was set equal to the basic reliability of the Vehicle B prior to Phase 3 of the RGT, and B_{IB} is defined as above.

An assumption underlying the modelling carried out was at the end of each phase of the RGT, the reliability of the vehicle could be estimated using the data gathered during the RGT phase. The method for measuring reliability both the contractor and the MOD decision maker place most faith in is the reliability value calculated using the AMSAA model. Hence, the value gathered using the RGT data and AMSAA calculations is the value that will be used to assess the current reliability of each variant of each system. For example, the observed basic reliability of Vehicle B during Phase 1 of the RGT as calculated by the AMSAA model is denoted by variable B_{1Bi}. The relationship between B_{1B} and B_{1Bi} is modelled as $B_{1Bi} = B_{1B} + R_{B_{1Bi}}$ where $R_{B_{1Bi}}$ is uncorrelated with everything else in the model. This uncertain variable represents the uncertainty between the observed basic reliability and the 'true' basic reliability, i.e. the output of the AMSAA estimates the 'true' reliability; it does not equal the 'true' reliability. All R_i variables subsequently defined are uncorrelated with everything else in the model.

The decision maker believed between Phase 1 and Phase 2 of the RGT, modifications will be made to Vehicle B and from this, a new variant of the vehicle would be created. The difference in the basic reliability between the two variants is modelled as $GB_{1,2B}$ such that $B_{2B} = B_{1B} + GB_{1,2B}$, where B_{2B} is the basic reliability of the Vehicle B prior to Phase 2 of the RGT. $GB_{1,2B}$ is the uncertain variable representing the difference between the basic reliability of the Vehicle B variant that is used during Phase 1 and the basic reliability of the Vehicle B variant that is used during Phase 2 of the RGT. $GB_{1,2B}$ and B_{1B} are modelled as being uncorrelated. The decision maker stated she believed if she knew B_{1B}, then learning about B_{2B} would not influence her belief about B_{1Bi}. Hence B_{1B} is Bayes linear sufficient for B_{2B} for adjusting B_{1Bi}.

In addition, the decision maker believes she can gather expert opinion from one of the contractor's experts which would improve her initial assessment of the growth between the variant used in Phase 1 and the variant used in Phase 2. The experts assessment was modelled as $EGB_{1,2B} = GB_{1,2B} + R_{EGB_{1,2B}}$, where $R_{EGB_{1,2B}}$ is uncorrelated with everything else in the model. The decision maker does not adopt the expert's opinion. Instead she believes that her initial belief of $GB_{1,2B}$ is influenced by the observation $EGB_{1,2B}$. The decision maker stated she believed if she knew GB_{1B} with complete certainty, then learning about $EGB_{1,2B}$ would not influence her belief about B_{2B}. Hence GB_{1B} is Bayes linear sufficient for B_{2B} for adjusting $EGB_{1,2B}$.

As before with variant 1 of Vehicle B, the decision maker can make an observation of the basic reliability of Variant 2 of the vehicle during the RGT. B_{2Bi} is the output of the AMSAA model for Vehicle B's second variant during the RGT. Similar to before, this is modelled as

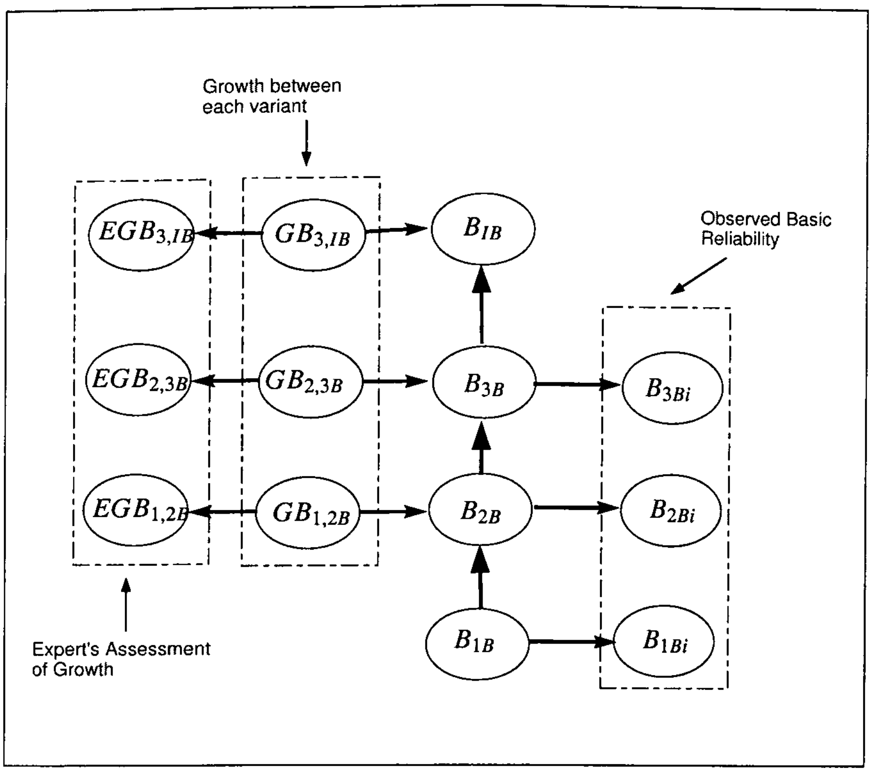

Figure 6.1: Bayes linear network for Vehicle B's basic reliability

$B_{2Bi} = B_{2B} + R_{B_{2Bi}}$. The difference in the basic reliability between the Variant 2 and Variant 3 is modelled as $GB_{2,3B}$ such that $B_{3B} = B_{2B} + GB_{2,3B}$, where B_{3B} is the basic reliability of the Vehicle B prior to Phase 3 of the RGT. As before, B_{2B} is Bayes linear sufficient for B_{3B} for adjusting B_{2Bi}. Similar to before, expert knowledge about the growth between Variant 2 and 3 can be gathered by the decision maker and is denoted by $EGB_{2,3B}$. Similarly, B_{3Bi} is the observed basic reliability of Variant 3 during the RGT. This expert assessment is an important variable as the decision maker is being asked to assess a variable she currently has very little information about but will have information about in the future.

The decision maker also believes modifications will be made between the third phase of the RGT and the in-service variant. The growth between these two variants is modelled as $GB_{3,IB}$ and B_{IB} is the basic reliability of the in-service Variant where $B_{IB} = B_{3B} + GB_{3,IB}$. Again, the decision maker believes she can gather expert assessment related to $GB_{3,IB}$ which is modelled as $EGB_{3,IB} = GB_{3,IB} + R_{EGB_{3,IB}}$.

The next step is to identify an ordering of the variables and to build the Bayes linear network. The ordering used in this project is [B_{1B}, B_{1Bi}, $GB_{1,2B}$, B_{2B}, $EGB_{1,2B}$, B_{2Bi}, $GB_{2,3B}$, B_{3B}, $EGB_{2,3B}$, B_{3Bi}, $GB_{3,IB}$, B_{IB}, $EGB_{3,IB}$]. The first step is to draw the nodes B_{1B} and B_{1Bi}. As B_{1B} and B_{1Bi} are correlated, an arrow is drawn from B_{1B} to B_{1Bi}. The next step is to draw the node $GB_{1,2B}$. $GB_{1,2B}$ is assumed to be uncorrelated with everything previously in the order so no arcs are drawn. The next variable in the ordering is B_{2B}. As B_{2B} is correlated with B_{1B}, an arrow is drawn from B_{1B} to B_{2B}. B_{2B} is correlated with B_{1Bi} but because they are Bayes linear sufficient for adjusting B_{1B}, no arrow is drawn. $GB_{1,2B}$ is correlated with B_{2B} so an arrow is drawn from $GB_{1,2B}$ to B_{2B}. The next variable is B_{2Bi}. This is correlated with B_{2B} so an arrow is drawn from B_{2B} to B_{2Bi}. $EGB_{1,2B}$ is the next node in the order. As it is correlated with $GB_{1,2B}$, an arrow is drawn from $GB_{1,2B}$ to $EGB_{1,2B}$. As $EGB_{1,2B}$ and B_{2B} are Bayes linear sufficient for adjusting $GB_{1,2B}$, no more arrows are needed.

The next node is $GB_{2,3B}$. It is uncorrelated with everything else in the model so no arrows are drawn. The next node is B_{3B}. It is correlated with both B_{2B} and $GB_{2,3B}$ so arrows are drawn from both of them to B_{3B}. $EGB_{2,3B}$ is the next node. It is correlated with $GB_{2,3B}$ so an arrow is drawn from $GB_{2,3B}$ to $EGB_{2,3B}$. As $EGB_{2,3B}$ and B_{3B} are Bayes linear sufficient for adjusting $GB_{2,3B}$, no more arrows are needed. The next node is B_{3Bi}. This is correlated with B_{3B} so an arrow is drawn from B_{3B} to B_{3Bi}.

The next node is $GB_{3,IB}$. It is uncorrelated with everything else in the model so no arrows are drawn. The next node is B_{IB}. It is correlated with both B_{3B} and $GB_{3,IB}$ so arrows are drawn from both of them to B_{IB}. The final node is $EGB_{3,IB}$. It is correlated with $GB_{3,IB}$ so an arrow is drawn from $GB_{3,IB}$ to $EGB_{3,IB}$. $EGB_{3,IB}$ is Bayes linear sufficient for adjusting $GB_{3,IB}$ with everything else in the model. Hence no more arrows are needed. As $EGB_{3,IB}$ is uncorrelated with everything else in the model, no more arrows are needed.

The above modelling can be presented by the Bayes linear network in Figure 6.1. Identical modelling for the in-service mission reliability of vehicle B, the in-service basic reliability of vehicle A and the in-service mission reliability of vehicle A was carried out using the respective notation of B_{IM}, A_{IB} and A_{IM}.

6.4.2 Structuring Part 5

An important part of demonstrating the usefulness of the Bayes linear methodology in this project was the modelling carried out for Part 5. This part was modelling the relationship between the two vehicles for the two requirements. The modelling carried out for this part was very different to the modelling carried out for Parts 1-4.

The decision maker stated she believed that because both vehicle's were being built by the same contractor and because there was an overlap between some of the mission and basic reliability failure modes, she believed learning about Vehicle A's basic reliability gave her some additional information about Vehicle A's mission reliability, and similar for Vehicle B's. For example, if the expectation of Vehicle A's basic reliability was to increase, she believed her expectation of Vehicle A's mission reliability would also increase. Similarly, she believed there was a relationship between the expectation about Vehicle A's basic reliability and the basic reliability of Vehicle B. As the vehicles shared a number of important sub-systems, she believed learning about Vehicle A's basic reliability would give her some additional information about Vehicle B's basic reliability, and similarly for mission reliability.

However, the decision maker commented as the model was being built, she belived the correlation between the basic and mission reliability would be small. She believed the contractor would focus energy on meeting the mission requirement for each vehicle. Hence a raise in the mission requirement for either vehicle, would not necessarily equate to a substantial rise in the basic reliability. The decision maker believed the contractor would focus equal time on both vehicles. Hence a rise in either reliability requirement for one vehicle, would see a rise in the same reliability requirement for the other vehicle. She believed the correlation between vehicles was stronger than the correlation between each requirement.

Both these relationships could potentially be investigated in further detail. For example, an investigation of the basic and mission failure modes could be carried out to identify those which are common to both systems and their rate of occurrence. Further investigation could focus on which of the mission failure modes are also basic failure modes and their rate of occurrence. From these investigations, an accurate representation of the relationship between the two variables could be calculated. However, the MOD decision maker is unlikely to have the time to process it. Instead, this situation is approached pragmatically and expert judgement is used to assess the relationship subjectively. Whilst this will not capture 'reality' perfectly, given the constraints placed upon the decision maker, it is the best scenario that can be modelled.

The modelling used the basic reliability of Variant 1 of Vehicle A, the basic reliability of Variant 1 of Vehicle B, the mission reliability of Variant 1 of Vehicle A and the mission reliability of Variant 1 of Vehicle B. Four new variables were introduced; R_A, R_B, R_M, R_N, each with expectation 0 and variance equal to 1. R_A represents the shared uncertainty between the basic and mission reliability of Vehicle A. Hence, A_{1B} and A_{1M} can now be written respectively as $A_{1B} = \alpha_{BRA}R_A + R_{A_{1B}}$ and $A_{1M} = \alpha_{MRA}R_A + R_{A_{1M}}$. Similarly, R_N represents the shared uncertainty between the basic reliability of Vehicle A and the basic reliability of Vehicle B. From this, A_{1B} can now be written as $A_{1B} = \alpha_{BRA}R_A + \alpha_{ARN}R_N + R_{A_{1B}}$ where R_A, R_N and $R_{A_{1B}}$ are all assumed to be uncorrelated with one another. A_{1M}, B_{1B} and B_{1M} are re-defined in a similar way such that

$$A_{1M} = \alpha_{MRA}R_A + \alpha_{ARM}R_M + R_{A_{1M}}$$

$$B_{1B} = \alpha_{BRB}R_B + \alpha_{BRN}R_N + R_{B_{1B}}$$

$$B_{1M} = \alpha_{MRB}R_B + \alpha_{BRN}R_M + R_{B_{1M}}$$

As $var(R_A)$ has already been specified, limits have to be placed on the α_i values. Assuming that $var(A_{1M}) = \sigma^2_{A_{1M}}$, then $\alpha^2_{MRA} + \alpha^2_{ARM} \leq \sigma^2_{A_{1M}}$. This means limits are placed upon the covariance's between A_{1B}, A_{1M}, B_{1B} and B_{1M}. As $corr(A_{1B}, A_{1M}) \leq 1 \Rightarrow \frac{\alpha_{MRA}\alpha_{BRA}}{\sigma_{A_{1B}}\sigma_{A_{1M}}} \leq 1$ $\Rightarrow cov(A_{1B}, A_{1M}) \leq \sigma_{A_{1B}}\sigma_{A_{1M}}$.

In order to build the Bayes linear network of this situation, an ordering of the above 8 variables is created; $[R_A, R_N, A_{1B}, R_M, A_{1M}, R_B, B_{1M}, B_{1B}]$. Nodes for R_A and R_N are drawn with no arcs between them as they are believed to be uncorrelated. A node for A_{1B} is drawn with arcs from R_A and R_N drawn to it. A node for R_M is drawn. R_M is considered to be uncorrelated with all previous variables. A_{IM} is then drawn with arcs from R_A and R_M to it. A_{1B} and A_{1M} are correlated. However R_A is Bayes linear sufficient for A_{1B} for adjusting for A_{1M}. Hence there is no arc from A_{1B} to A_{1M}. The next variables is R_B. At this point, R_B is uncorrelated with everything else in the model. The next variable B_{1M} is drawn. As B_{1M} is

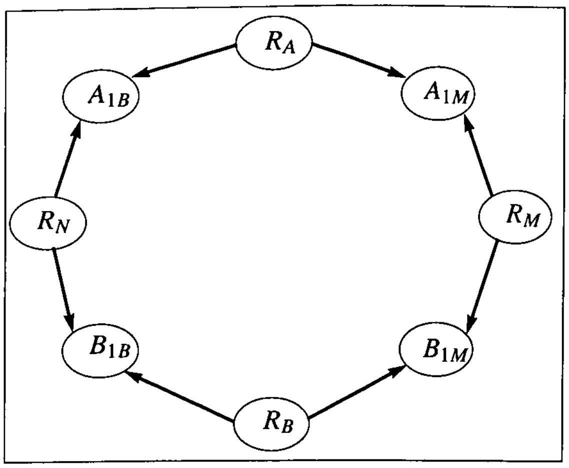

Figure 6.2: Bayes linear network for modelling Part 5

correlated with R_M and R_B, an arc is drawn from R_M and R_B to B_{1M}. As before, A_{1M} and B_{1M} are correlated, however R_M is Bayes linear sufficient for A_{1M} for adjusting for B_{1M}.. The final variable is B_{1B}. Arcs are drawn from R_N and R_B to B_{1B}. The Bayes linear network for the above modelling can be represented by Figure 6.2.

The next step is to attempt to bring together all the variables under one Bayes linear network. Due to the size of this network, this will only be demonstrated for vehicle A's basic and mission reliability. An ordering is identified - [R_A, B_{1B}, B_{1Bi}, $GB_{1,2B}$, B_{2B}, $EGB_{1,2B}$, B_{2Bi}, $GB_{2,3B}$, B_{3B}, $EGB_{2,3B}$, B_{3Bi}, $GB_{3,IB}$, B_{IB}, EGB_{1B}, B_{1M}, B_{1Mi}, $GB_{1,2M}$, B_{2M}, $EGB_{1,2M}$, B_{2Mi}, $GB_{2,3M}$, B_{3M}, $EGB_{2,3M}$, B_{3Mi}, $GB_{3,IM}$, B_{IM}, $EGB_{3,IM}$]. It is important to remember that each child node is Bayes linear sufficient with everything else in the model for adjusting its parent nodes. Using this fact and the ordering above, Figure 6.3 is the Bayes linear network for the above modelling. Extending the ordering to take into account all the variables would result in a larger Bayes linear network being built.

6.5 Eliciting Beliefs and Populating Model

6.5.1 Eliciting Parts 1-4

This section will focus on demonstrating how the necessary values were elicited from the decision maker in order to fully populate the model. As the modelling is identical for parts 1-4,

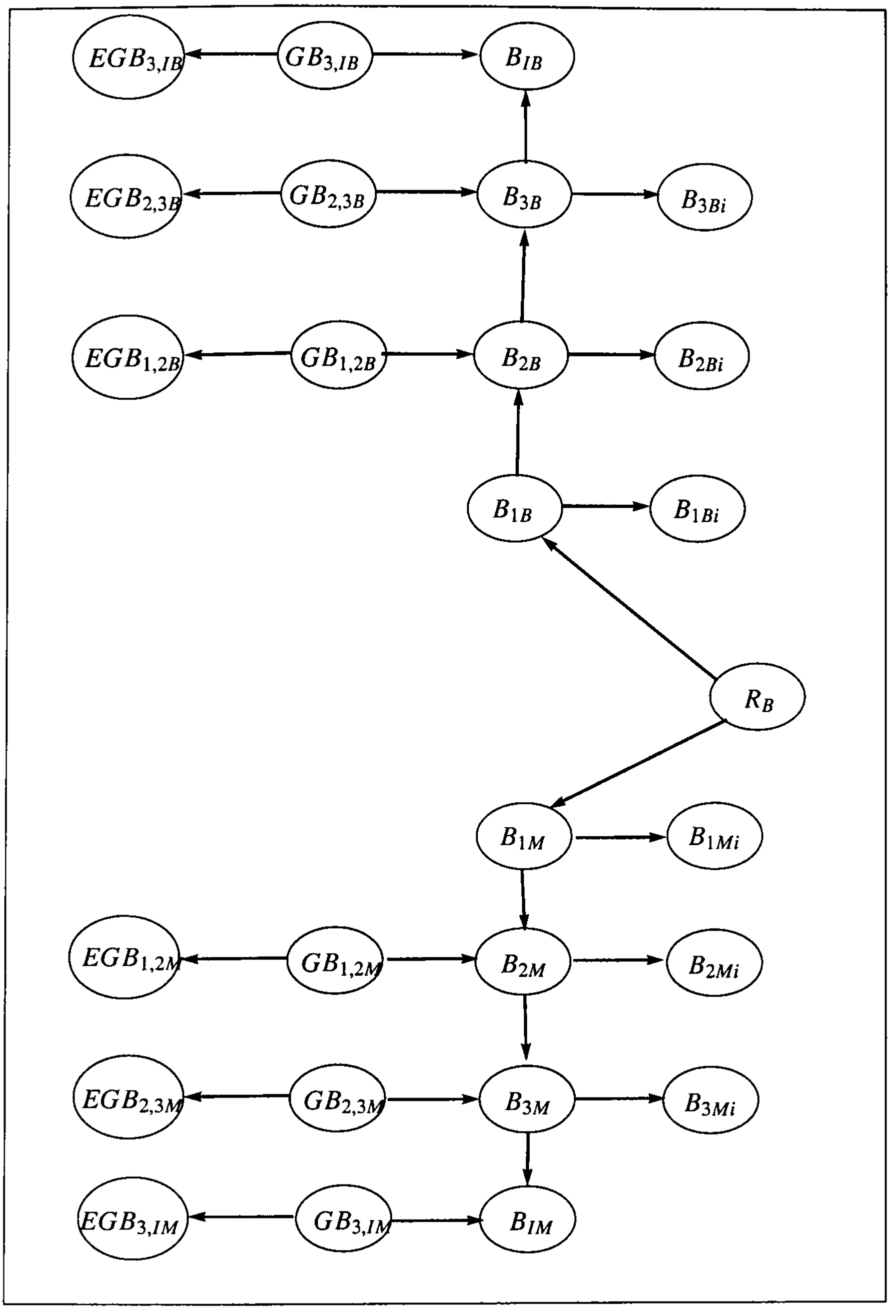

Figure 6.3: Bayes linear network for Vehicle B's reliability

this will only be demonstrated for part 1.

6.5.1.1 Basic Reliability of Vehicle B Variant 1

Analyst: What is your belief about the 5, 50 and 95 percentiles of the basic reliability of the Vehicle B's system entering Phase 1 of the RGT?

Decision maker: She believed the 5, 50 and 95 percentile values were 0.06, 0.12 and 0.16 respectively.

From this, the expectation and variance of B_{1B} can be calculated using the Pearson-Tukey formulas, giving $E(B_{1B}) = 0.01163$ and $var(B_{1B}) = 0.000343$.

6.5.1.2 Observed Basic Reliability of Vehicle B Variant 1

Analyst: If you knew for certain the basic reliability of Vehicle B Variant 1 was 0.1, what is your 5, 50 and 95 percentiles for the output of the RGT?

Decision maker: She believed it was equally likely the observed basic reliability would be above or below 0.1 and it could be as low as 0.05 and as high as 0.15. The decision maker also stated her belief about the uncertainty of the observed basic reliability of Phase 1 would be the same as the uncertainty of the observed basic reliability of Phase 2 and the uncertainty of the observed basic reliability of Phase 3.

From this, the expectation and variance of $R_{1Bi} = R_{2Bi} = R_{3Bi}$ can be calculated to be $E(R_{1Bi}) = 0$ and $var(R_{1Bi}) = 0.003695$. From this, $E(B_{1Bi}) = 0.01163$ and $var(B_{1Bi}) = 0.001266$. Similarly, it is trivial to calculate $cov(B_{1Bi}, B_{1B}) = var(B_{1B})$.

6.5.1.3 Growth Between Phases

Analyst: What is your 5, 50 and 95 percentiles of the growth between each of the three phases?

Decision maker: She believed the growth between Phase 1 and Phase 2 could be as little as 0, could be as high as 0.08 and was equally likely to be above or below 0.06. She believed the growth between Phase 2 and Phase 3 could be as little as 0, could be as high as 0.08 and was equally likely to be above or below 0.06. She believed the growth between Phase 3 and the in-service Variant of Vehicle B could be as low was -0.05, as high as 0.05 and equally likely to be above or below 0.

From this, the expectation and variance of $GB_{1,2B}$, $GB_{2,3B}$ and $GB_{3,IB}$ can be calculated. From these values, the expectation and variance of $B_{2B} = GB_{1,2B} + B_{1B}$, $B_{3B} = B_{2B} + GB_{2,3B}$ and $B_{IB} = B_{3B} + GB_{3,IB}$ can be calculated.

6.5.1.4 Experts Assessment of Growth

Analyst: Given that you know for certain the growth of vehicle B's basic reliability between phase 1 and phase 2 of is equal to 0.05, what is your 5, 50 and 95 percentiles of the value stated by the expert?

Decision maker: She believed the expert was equally likely to be above or below 0.05 and it could be as low as 0.03 and as high as 0.07. The decision maker believed her uncertainty surrounding the experts assessment of the growth is the same for all assessments.

As the decision maker believed her residual uncertainty between the true growth value and the values stated by the expert was the same, $E(R_{EGB_{1,2B}}) = E(R_{EGB_{2,3B}}) = E(R_{EGB_{3,IB}}) = 0$ and $var(R_{EGB_{1,2B}}) = var(R_{EGB_{2,3B}}) = var(R_{EGB_{3,IB}}) = 0.000148$. Using the values elicited and the values gathered above for $GB_{1,2B}$, $GB_{2,3B}$ and $GB_{3,IB}$, the expectation and variance of $EGB_{1,2B} = GB_{1,2B} + R_{EGB_{1,2B}}$, $EGB_{2,3B} = GB_{2,3B} + R_{EGB_{2,3B}}$ and $EGB_{3,IB} = GB_{3,IB} + R_{EGB_{3,IB}}$ can be calculated trivially.

From these elicited values, the necessary values to build the model for Vehicle B's basic reliability can be calculated. Similar questions were asked of the decision maker for Vehicle B's mission reliability and the basic and mission reliability of Vehicle A. The next step is to elicit the necessary values for Part 5.

6.5.2 Eliciting Part 5

As R_A, R_B, R_M and R_N are proxy variables, the expectation and variance can be arbitrarily set with expectation equal to zero and variance equal to one. The only thing that needs to be elicited from the decision maker is the covariance between, for example, Vehicle A's basic reliability and Vehicle B's basic reliability. The following can be used to elicit $cov(A_{1B}, A_{1M})$.

Analyst: Given that you know for sure basic reliability of Variant 1 of Vehicle A is 0.1, what is your new belief about the expectation of Vehicle A's Variant 1 mission reliability. Similarly, given that you know for sure that mission reliability of Variant 1 of Vehicle A is 0.7, what is *your new belief about the expectation of Vehicle A's Variant 1 basic reliability.*

As $var(A_{1B})$ and $var(A_{1M})$ have already been specified, the decision maker is restricted by the values she may specify. As we already know $cov(A_{1B}, A_{1M}) \leq \sigma_{A_{1B}}\sigma_{A_{1M}}$, then there is an upper bound of 0.0901 and 0.7433 on the values she can specify for mission and basic reliability respectively. This is calculated by using the Bayes linear formula for adjusted expectation and assuming $cov(A_{1B}, A_{1M}) = \sigma_{A_{1B}}\sigma_{A_{1M}}$.

$$\begin{aligned}
E_{A_{1M}}(A_{1B}) &= E(A_{1B}) + cov(A_{1B}, A_{1M})var^{-1}(A_{1M})(A_{1M} - E(A_{1M})) \\
&= 0.0563 + 0.00125(0.003695)^{-1}(0.1) \\
&= 0.090129
\end{aligned}$$

Any adjusted expectation value greater than this would require $cov(A_{1B}, A_{1M})$ to be incoherent as $corr(A_{1B}, A_{1M}) > 1$. If the decision maker specified beliefs outside these values, it would be necessary to revisit the values already specified.

Decision Maker: She believed her belief about the expectation of Vehicle A's mission reliability would rise to 0.61 if she observed a basic reliability of 0.1. She believed if she observed that Vehicles A's mission reliability was 0.7, Vehicle A's basic reliability would rise by 0.0663.

In order to elicit these new values, the decision maker was asked to specify their new beliefs about the 5, 50 and 95 percentiles of A_{1M} and A_{1B} respectively. Both these values are well within the limit placed upon the beliefs by her prior specified values. From this, two values for the $cov(A_{1B}, A_{1M})$ can be calculated using the following equation.

$$\begin{aligned}
cov(A_{1B}, A_{1M}) &= \frac{(E_{A_{1B}}(A_{1M}) - E(A_{1M}))var(A_{1B})}{A_{1B} - E(A_{1B})} \\
&= \frac{0.01 * 0.000343}{0.0437} \\
&= 0.0000789 \\
corr(A_{1B}, A_{1M}) &= 0.0697
\end{aligned}$$

and

$$
\begin{aligned}
cov(A_{1B}, A_{1M}) &= \frac{(E_{A_{1M}}(A_{1B}) - E(A_{1B}))var(A_{1M})}{A_{1M} - E(A_{1M})} \\
&= \frac{0.005 * 0.003695}{0.1} \\
&= 0.0001847
\end{aligned}
$$

$$
corr(A_{1B}, A_{1M}) = 0.16421
$$

Two values for the same dependency were elicited to attempt to validate the values specified by the decision maker - one was observing A_{1B} and assessing the effect on $E(A_{1M})$ and the other value was observing A_{1M} and assessing the effect on $E(A_{1B})$. As its impossible to validate which of these is the 'true' subjective belief of the decision maker, an average of the two values was taken, i.e. in this case $cov(A_{1B}, A_{1M}) = 0.000132$. It is acknowledged that due to anchoring, there is likely to be an element of bias in the second value elicited from the decision maker. However, as the decision maker was unable to carry out any calculations, this anchoring bias is likely to be small. Questions of a similar form were used to elicit $cov(A_{1B}, A_{1M})$, $cov(A_{1M}, B_{1M})$ and $cov(B_{1B}, B_{1M})$. Once these values were elicited and averaged out, the four correlation values were compared. Table 6.1 displays the four correlation values.

Variables	Correlation
A_{1B}, B_{1B}	0.6849
A_{1M}, B_{1M}	0.7
A_{1B}, A_{1M}	0.11695
B_{1B}, B_{1M}	0.11254

Table 6.1: Correlation values for Part 5

These values support the decision maker's prior belief that the correlation between the two vehicles is greater than the correlation between the requirements. In addition, it is noticeable that the correlation between the two vehicles for each requirement was similar and the correlation between each requirement for each vehicle was also similar. However, despite the similarities of the correlation values across similar dependencies, these values are very sensitive to small changes. Because very small changes can have a large effect on the correlation, it is typical to carry out sensitivity analysis to assess the effect on the output, given these small changes. This will not be done in this project. Justification for not carrying out sensitivity analysis will be given in Section 6.6.5.2

If we think of the model as being in 4 separate sections, one for Vehicle A's basic reliability, one for Vehicle A's mission reliability, etc, then due to the way in which the model was built, the covariance of A_{1B} and any other variable within that section is either equal to zero or $var(A_{1B})$. In the case of the growth variables and the expert opinion on the growth variables, the covariance is equal to zero. For all other variables, the covariance is equal to $var(A_{1B})$. Because of this, it is possible to say the covariance between a variable within Vehicle A's basic reliability section, say A_{Bi} and Vehicle A's mission reliability, say A_{Mi}, is either equal to zero or $cov(A_{1B}, A_{1M})$. It is equal to $cov(A_{1B}, A_{1M})$ if and only if $cov(A_{1B}, A_{Bi}) \neq 0$ *and* $cov(A_{1M}, A_{Mi}) \neq 0$. A simple example considering two variables within each section demonstrates this. Consider $cov(A_{3B}, A_{3Mi})$. Recall that $cov(A_{1B}, A_{3B}) = var(A_{1B}$ and $cov(A_{1M}, A_{3Mi}) = var(A_{1M})$.

$$
\begin{aligned}
cov(A_{3B}, A_{3Mi}) &= cov(A_{2B} + GB_{2,3B}, A_{3M} + R_{A_{3Mi}}) \\
&= cov(A_{1B} + GB_{1,2B} + GB_{2,3B}, A_{2M} + GM_{2,3M} + R_{A_{3Mi}}) \\
&= cov(A_{1B} + GB_{1,2B} + GB_{2,3B}, A_{1M} + GM_{1,2M} + GM_{2,3M} + R_{A_{3Mi}}) \\
&= cov(A_{1B}, A_{1M})
\end{aligned}
$$

Appendix F contains the expectation table and covariance matrix for the four reliability requirements.

6.6 Analysing Belief Structure Given Observations

Due to timing and access of data, the model was built and populated with the decision maker's subjective beliefs, however, all the necessary observations to carry out inference on the reliability parameters of interest were not gathered. The analysis section will focus on the output of the model given the values that were gathered but will also consider the added benefit of the additional information sources that were unable to be collected. The variables the analysis focused on were the in-service basic and mission reliability of each system as these were the variables of most interest to the decision maker.

6.6.1 Prior Belief of Decision Maker

Table 6.2 shows the initial belief of the decision maker. Prior to any observations being made,

Variable	Expectation	Variance	Standard Deviation (S.D.)
A_{IB}	0.222	0.003245	0.056965
A_{IM}	0.745	0.016413	0.128116
B_{IB}	0.220	0.003689	0.060741
B_{IM}	0.683	0.009270	0.117855

Table 6.2: Decision maker's prior belief for all four requirements

the decision maker believes all reliability requirements will be met, however, there is still substantial uncertainty in this assessment and it is possible neither system will meet either of the requirements.

6.6.2 Adjusted Beliefs Given Observations

The values in Table 6.3 were gathered from the contractor during the RGT's.

Variable	Observation
A_{1Bi}	0.03
A_{2Bi}	0.1
A_{3Bi}	0.03
A_{1Mi}	0.55
A_{2Mi}	0.5
A_{3Mi}	0.47
B_{1Bi}	0.04
B_{2Bi}	0.1
B_{1Mi}	0.5
B_{2Mi}	0.61

Table 6.3: Observations made during RGT's

Unfortunately, the decision maker was unable to gather the necessary values to populate the expert judgement variables on the expected growth between each system. Using the above observations, Table 6.4 displays the output of the model. In brackets is the percentage the standard deviation dropped by using the above observations.

Recall from Section 2.4.3 that the Bayes linear approximation of the adjusted expectation

Variable	Target	Adjusted Expectation	Adjusted Standard Deviation
A_{IB}	0.16	0.0973	0.0487 (14.5%)
A_{IM}	0.66	0.5132	0.0846 (34.0%)
B_{IB}	0.17	0.1908	0.0582 (4.3%)
B_{IM}	0.68	0.6302	0.1014 (14.0%)

Table 6.4: Adjusted beliefs given observations

and variance are exact in the case where all the variable are joint normally distributed. In the model, all the elicited R_i values are normally distributed. The elicited values for the growth variables and the reliability of each vehicle during Phase 1 are all skewed. We wish to measure the effect of this skewness.

Given that U_X, M_X and L_X are the 95, 50 and 5 percentiles respectively, U_X can be increased or decreased to U_{X_N} in order that $U_X - M_X = M_X - L_X$. Using U_{X_N}, a new expectation and variance for all the variables can be calculated. The new model can then be adjusted to give a new adjusted expectation and variance. Similarly, L_X can be increased or decreased in order that $M_X - L_X = U_X - M_X$. From this, a third adjusted expectation and variance can be calculated. If the three adjusted expectations and the three adjusted variances are very close for each of the reliability requirements, then it would appear reasonable to assume the decision maker's beliefs are approximately normally distributed. Table 6.5 shows the adjusted expectation (AEx) and adjusted standard deviation (ASD) for the decision maker's current beliefs, her adjusted beliefs given that U_X is increased or decreased and her adjusted beliefs given that L_X is increased or decreased.

Variable	AEx	ASD	AEx (U_X)	ASD (U_X)	AEx. (L_X)	ASD (L_X)
A_{IB}	0.0973	0.0487	0.0982	0.0481	0.0956	0.0492
A_{IM}	0.5132	0.0846	0.5147	0.0843	0.613	0.1074
B_{IB}	0.1908	0.0582	0.169	0.0687	0.1913	0.0568
B_{IM}	0.6302	0.1014	0.661	0.10369	0.6267	0.1005

Table 6.5: Adjusted normal beliefs given observations

From the table, it does not appear to be unrealistic to assume the adjusted beliefs for A_{IB} and B_{IM} are approximately normally distributed. For A_{IM} and B_{IB}, the decision maker's adjusted beliefs are substantially different to the modified adjusted beliefs to cause concern over assuming normality. However, the additional information is used where appropriate.

From Table 6.4, it appears Vehicle A is not meeting either of their reliability requirements and a large amount of uncertainty remains in the assessment. Assuming the output is normally distributed for A_{IB}, the decision maker believes there is a probability of 0.0909 that $A_{IB} \geq 0.16$. Given the information to date and the decision maker's current belief structure, the contractor is not providing sufficient evidence Vehicle A is meeting either of it's reliability requirements.

From Table 6.4, it can be seen the percentage drop in the standard deviation of both Vehicle B's reliability requirements is lower than the drop in standard deviation of Vehicle A's reliability requirements. This is because there were less observations for Vehicle B. At this time, Vehicle B, appears to be substantially more reliable than Vehicle A. One possible reason for this is the fact that less observations have been made for this vehicle. The decision maker's prior beliefs that the vehicle is reliable is having a larger effect on the adjusted expectation and variance of Vehicle B's reliability requirements than in the case of Vehicle A. At this time, the expectation of B_{IB} is greater than the reliability target of 0.17. The expectation for the mission reliability however is slightly below the target. Assuming normality again, the probability B_{IM} is greater than 0.68 is equal to 0.312.

6.6.3 Sensitivity Analysis

Due to the simplicity of the Bayes linear formulas, sensitivity analysis can be carried out quickly. Three scenarios can potentially be investigated; we could modify our prior belief about our expectations, modify our prior beliefs regarding the covariance matrix, or consider potential sources of observations that have yet to be collected. The purpose of the sensitivity analysis is to analyse whether or not different scenarios would have had an effect on whether or not we believe the systems have met reliability requirements.

6.6.3.1 Adjusting Prior Expectations

Given the current covariance matrix and observations, we assess how much the decision maker's prior belief in $E(A_{IB})$, $E(A_{IM})$, $E(B_{IB})$ and $E(B_{IM})$ would have to increase by in order that her adjusted expectation is greater than the targets. In order to do this, some assumptions are needed to be made. The Bayes linear adjusted expectation formulas is

$$E_D(X) = E(X) + cov(X, D)var^{-1}(D)(D - E(D))$$

We assume the covariance structure between the observed values and the value of interest, in this case A_{IB}, remains the same. A consequence of this is that $var^{-1}(D)$ remains the same. Because of this, this analysis has no effect on the adjusted variance of each of the four parameters of interest. The second assumption is if **D** =[A_{1Bi}, A_{2Bi}, A_{3Bi}, A_{1Mi}, A_{2Mi},A_{3Mi}, B_{1Bi}, B_{2Bi}, B_{1Mi}, B_{2Mi}], changing the decision maker's prior belief regarding the prior expectation of A_{IB} only has an effect on the prior expectation of A_{1Bi}, A_{2Bi} and A_{3Bi}. Our final assumption is that the growth between the variants remains the same.

From this information and given the observations made, in order that the adjusted expectation of A_{IB} is equal to 0.16, the prior expectation of A_{IB} must be 0.4658. The decision maker's prior belief for A_{IB} was 0.2215. Similar calculations can be carried out for $E(A_{IM})$, $E(B_{IB})$ and $E(B_{IM})$. If the prior expectation of A_{IM} was set to 0.999, the adjusted expectation of A_{IM} given the observations would be 0.5456. No matter what value was entered into the model as the prior expectation of A_{IM}, the vehicles adjusted expectation will be lower than the target and the maximum probability that the vehicle will meet reliability requirements is 0.129. If the prior expectation of B_{IB} was lowered from 0.220 to 0.1795, the adjusted expectation of B_{IB} given the observations would be equal to the target of 0.17. If the prior expectation of B_{IM} was raised from 0.683 to 0.8299, the adjusted expectation of B_{IM} given the observations would meet the target of 0.68. The decision maker stated for the three reliability requirements that are currently not meeting requirements, the proposed prior beliefs do not present her 'true' belief about the reliability of the vehicles.

6.6.3.2 Adjusting Covariance Matrix

The aim of adjusting the covariance matrix is to assess whether or not the observations are having too strong an effect on the decision maker's prior beliefs. Sensitivity analysis will be carried out on the residual uncertainty between the true reliability value and the reliability value observed during each phase of the RGT. Currently, the expectation of this value is zero and the variance is equal to 0.003695. All the observations made to date have been smaller than their expected value. As three out of the four reliability requirements are not currently being met, decreasing the residual variance, i.e. increasing the correlation between the observation and the 'true' reliability of the system, will make the adjusted expectation of the requirements decrease further. By lowering the correlation between the observations and 'true' reliability, more faith is placed in the prior belief of the decision maker and the adjusted expectation of the parameters of interest will rise.

In order to do this we build the covariance matrix as before. However, this time, the variance of all the observations, i.e. $var(\mathbf{D})$ where $\mathbf{D} = [A_{1Bi}, A_{2Bi}, A_{3Bi}, A_{1Mi}, A_{2Mi}, A_{3Mi}, B_{1Bi}, B_{2Bi}, B_{1Mi}, B_{2Mi}]$ is changed. At this time, the decision maker believes the observed basic reliability value will fall between -0.05 and 0.05 of the 'true' value. The decision maker believes the observed mission reliability value will fall between -0.1 and 0.1 of the 'true' value. Three scenarios will be investigated. The first scenario is decreasing the range to -0.0025 to 0.0025 for basic reliability and -0.05 to 0.05 for mission reliability. The second scenario is increasing the basic reliability range to -0.1 and 0.1 and the mission reliability to -0.2 and 0.2. The third scenario is increasing the basic reliability range to -0.15 and 0.15 and increasing the mission reliability range to -0.3 and 0.3. These three scenarios have the effect of changing the residual variance between the 'true' reliability and observed reliability. Tables 6.6, 6.7 and 6.8 demonstrates the adjusted expectation and the adjusted standard deviation for each of the three scenarios.

Variable	Adjusted expectation	Adjusted S.D.
A_{IB}	0.054	0.045
A_{IM}	0.483	0.074
B_{IB}	0.188	0.056
B_{IM}	0.643	0.096

Table 6.6: Adjusted beliefs given observations for Scenario 1

Variable	Adjusted expectation	Adjusted S.D.
A_{IB}	0.160	0.053
A_{IM}	0.597	0.103
B_{IB}	0.202	0.059
B_{IM}	0.635	0.109

Table 6.7: Adjusted beliefs given observations for Scenario 2

Variable	Adjusted expectation	Adjusted S.D.
A_{IB}	0.188	0.055
A_{IM}	0.655	0.114
B_{IB}	0.210	0.060
B_{IM}	0.651	0.113

Table 6.8: Adjusted beliefs given observations for Scenario 3

It can be observed from the tables above that changing the residual uncertainty of the observations does not strongly affect the adjusted beliefs of Vehicle B. The largest impact of changing the residual uncertainty of the observations occurs for Vehicle A's basic and mission reliability. For Scenario 1 where the correlation between the 'true' reliability and the observed reliability is increased, the adjusted expectation falls. In the scenario where the correlation is substantially reduced, the adjusted expectation for the basic reliability is greater that the target and the adjusted expectation for the mission reliability is only slightly below the target. However, the decision maker did not believe that it was realistic to think the residual uncertainty could potentially as high as in Scenario 3.

6.6.3.3 Potential Sources of Observations

The third form of sensitivity analysis is to consider the observations that have yet to be received. In this case, it may still be possible to gather the expert's opinion on the growth between the basic and mission reliability of each vehicle after the third phase of RGT's and prior to the vehicle entering service. We are interested in knowing what values for $EGA_{1,2B}$, $EGA_{1,2M}$, and $EGB_{1,2M}$ will result in the respective vehicle meeting its respective requirement. $EGB_{1,2B}$ will be ignored as it is already on target to meet requirements. In order for the vehicles to meet reliability requirements, the expert must state three values; $E(EGA_{3,IB}) = 0.075$, $E(EGA_{3,IM}) = 0.28$, and $E(EGB_{3,IM}) = 0.0585$. Each of these three values are outside the range of values the decision maker specified that she thought the expert may state.

6.6.4 Summary of Analysis

Based on the prior belief stated by the decision maker and the observations made, the contractor, at this stage, has not given the decision maker sufficient evidence to suggest they are meeting all four reliability requirements. The model indicates the expectation of Vehicle B's basic reliability is greater than the target. However, the expectation of both of Vehicle A's reliability requirements and Vehicle B's mission requirement is lower than the target. In the case of Vehicle A, the probability of meeting the target is small.

Extensive sensitivity analysis has been carried out to investigate whether or not the three reliability requirements currently not being met could be met given changes in the prior specification by the decision maker. Two scenarios of particular interest were investigated; first, what changes could be made to the prior expectation so that requirements are being met, and second, what changes could be made to the covariance structure in order that reliability requirements were met. For the three reliability requirements currently not met, the decision maker did not believe it was feasible to modify her prior belief structure to such an extent that the adjusted expectation of each of the three requirements was greater than the target.

6.6.5 Assessing Benefit of Additional Modelling

A lot of time and effort was placed into making the model detailed and incorporating additional information sources and expert knowledge the decision maker had regarding the two systems. Because of this additional effort, it is possible the analyst created a model that could not be replicated by the MOD decision maker without assistance. This might decrease the likelihood the method would be adopted in the near future. It is worthwhile to investigate whether or not this additional modelling effort had a large impact or whether or not a similar amount of information could have been gathered using a simple modelling approach. If the additional modelling did not have a large impact on the uncertainty of each of the four reliability requirements, then it is possible making the model more detailed had a detrimental effect on introducing the Bayes linear methodology as an alternative method of analysing information in the R&M Case.

The simplest approach would have been not to take into account the expert's assessment of the growth and to model the two vehicles and their reliability requirements as being uncorrelated. This would have reduced the amount of modelling and specification the MOD decision maker would have to have carried out, if they had been modelling this situation on their own. The following two sections will assess the reduction in uncertainty of the four reliability requirements brought on by the additional modelling.

6.6.5.1 Assessing the Effect of the Expert's Assessment of Growth

If the expert's assessment of the expected growth between each system had been gathered, then the uncertainty would have dropped further. Assume values for the following variables were gathered; $EGA_{1,2B}$, $EGA_{2,3B}$, $EGA_{1,2M}$, $EGA_{2,3M}$, $EGB_{1,2B}$, $EGB_{2,3B}$, $EGB_{1,2M}$ and $EGB_{2,3M}$, i.e. all the expert's assessment of growth other than between the current one and the in-service one. Table 6.9 shows the standard deviation if the expert's assessment had been gathered. In brackets is the percentage drop between the original standard deviation and the standard deviation if the observations had been made.

	Current S.D.	Projected S.D.
A_{IB}	0.0487	0.0486 (0.16%)
A_{IM}	0.0846	0.0833 (1.44%)
B_{IB}	0.0582	0.056 (4.35%)
B_{IM}	0.1014	0.096 (5.58%)

Table 6.9: Adjusted decision maker's beliefs given observations and early expert assessment

If the expert was able to assess his expected growth between the third variant and the in-service version of each system, the standard deviation would drop further. Table 6.10 shows the standard deviation if the decision maker was able to gather the following four variables; $EGTR_{3,IB}$, $EGTR_{3,IB}$, $EGTI_{3,IB}$ and $EGTI_{3,IM}$.

	Current S.D.	Projected S.D.
A_{IB}	0.0487	0.044 (9.39%)
A_{IM}	0.0846	0.075 (11.70%)
B_{IB}	0.0582	0.050 (17.60%)
B_{IM}	0.1014	0.082 (19.47%)

Table 6.10: Adjusted decision maker's beliefs given observations and recent expert assessment

Table 6.10 suggests that observing the expert's assessment on the growth between the current variant of both vehicles and the in-service variant, has a substantial effect on the uncertainty of each requirement. In the case of Vehicle B, this would reduce the uncertainty of B_{IB} and B_{IM} to be almost equivalent to the uncertainty of A_{IB} and A_{IM}, despite the fact an additional phase of the RGT has been carried out for Vehicle A and not Vehicle B. However, in most realistic cases, it is unlikely the expert opinion will be observed without the previous phase data being observed. Hence, we are more interested in the drop in uncertainty when the previous

RGT phase data has been observed, i.e. the drop in uncertainty for A_{IB} and A_{IM}. The drop in standard deviation for both is close to 10%. As modelling the expert's opinion is rather simple, it is likely given the effort needed to model this situation, it was worth modelling.

6.6.5.2 Assessing the Effect of Modelling Part 5

As time and effort went into modelling Part 5, both in terms of qualitative modelling (ensuring the modelling was set up in a consistent manner) and in the quantitative modelling (eliciting the necessary values from the decision maker), it is interesting to assess the benefit of this modelling and whether or not any additional value was extracted from the available data. Table 6.11 shows the uncertainty of the four reliability parameters of interest given that the modelling of Part 5 was removed and the uncertainty given that it remains. In brackets is the percentage reduced by carrying out the modelling of Part 5.

	Current Expectation	Current S.D.	Projected Expectation	Projected S.D.
A_{IB}	0.0973	0.0487	0.0978	0.0487 (0.08%)
A_{IM}	0.5132	0.0846	0.519	0.0850 (0.08%)
B_{IB}	0.1908	0.0582	0.205	0.0583 (0.26%)
B_{IM}	0.6302	0.1014	0.648	0.1019 (0.50%)

Table 6.11: Assessing effect of modelling Part 5

As can be seen, using the information for Vehicle A (and basic) reliability to assess the reliability of Vehicle B (and mission) and vice versa, reduces the uncertainty of each variable by an incredibly small amount. As the additional benefit of carrying out this analysis is very small, retrospectively, it is possible carrying out this additional modelling was not the correct course of action. Instead, keeping the model simple and more accessible to the MOD decision maker, may have resulted in the decision maker being more likely to use the methodology in the future. Because of the very small effect this additional modelling has on the adjusted variance when included, sensitivity analysis was not considered necessary.

6.7 Feedback from Decision Maker

The feedback received from the decision maker regarding the Bayes linear methodology was rather mixed. She believed that whilst the qualitative modelling was useful in "crystalising her understanding of the situation", it did not offer her much with respect to stimulating discussion between the decision maker, the IPTL and the contractor. This was largely due to the intuitive nature of the problem and the relative simplicity of the BLN. However, the decision maker believed in projects where the dependencies were more complex, the BLN would be a useful tool to allow different stakeholders to immediately view the dependencies. The decision maker was currently working on another project she believed would benefit greatly from structuring the problem using a BLN.

The decision maker believed quantifying her beliefs was useful but was nervous about sharing these beliefs with others. This was because she were wary that within the Bayes linear modelling framework, the model is built using subjective beliefs. She was worried these values could potentially be disagreed upon and a large amount of time would be wasted getting "bogged down" in fruitless discussion. When the decision maker was informed sensitivity analysis was carried out to address this issue, she seemed unimpressed and doubtful of the usefulness of it. Once the sensitivity analysis was demonstrated to them, she still felt the contractor and IPTL would focus on the values elicited and not on the overall modelling.

The decision maker believed that overall the Bayes linear modelling was useful in the project but the statistical capabilities of both the IPTL, the R&M Focal Point and the contractor would make it difficult to gain widespread acceptance of the output of any developed models. Instead, the models would be limited to being used by only MOD R&M Specialists to support their beliefs. The models would therefore not be used as the primary analysis to make decisions but rather seen as a method to support their conclusions.

The decision maker believed the modelling supported her belief of the current values of the four reliability requirements. Based on this analysis and the AMSAA model, the decision was taken to stop reliability growth through trials and to focus on the failure modes occurring. A number of failure modes and solutions were identified. Using subjective engineering belief, the decision maker evaluated these solutions and assessed how many of them would be achieved in a six month period. Using the fix effectiveness model, the requirements were demonstrated to be met. After the fixes had been made, both vehicles went through additional trials to demonstrate requirements had been met. Both basic and mission reliability targets for Vehicle B were met based on the observed data. For Vehicle A, both requirements were also met, but only by a small amount.

6.8 Further Developments

The modelling has attempted to extensively model the MOD decision makers problem. However, three possible further developments have been identified.

6.8.1 Future In-Service Data

A future development that could be carried out without further elicitation or input from the decision maker is to integrate into the model the reliability data that may be gathered in the future. The decision maker stated she believed that if the vehicles were not meeting reliability requirements, the contractor would carry out further trials with the aim of demonstrating this. Potentially, the contractor planned to carry out three phases of reliability assessment, each would contain 10 BFM's; the same amount as the RGT. The model need only be adjusted slightly to model this situation as the overall structure of the testing to assess reliability is similar to that of assessing reliability growth. The modelling for this scenario is illustrated for Vehicle A's basic reliability. Similar modelling can be carried out for the other reliability requirements.

Three new variables are introduced - A_{IB1i}, A_{IB2i}, A_{IB3i} which represent the output of the basic reliability of Vehicle A during each phase. We also use the previous variable, A_{IB}, the basic reliability of Vehicle A prior the in-service phase, to model this situation such that $A_{IB1i} = A_{IB} + R_{A_{IB1i}}$, $A_{IB2i} = A_{IB} + R_{A_{IB2i}}$ and $A_{IB3i=A_{IB}+R_{A_{IB3i}}}$. Based on the previous modelling, it is plausible to assume that A_{IB} is Bayes linear sufficient for adjusting A_{IB1i}, A_{IB2i} and A_{IB3i}. As before, we could potentially assume learning about A_{IB} informs the decision maker about B_{IB} and so on. Using the ordering of [A_{IB}, A_{IB1i}, A_{IB2i}, A_{IB3i}], Figure 6.4 is a small part of a potential Bayes linear network that would model the above scenario.

The model would require no more further specification on behalf of the decision maker. We assume the expectation and variance values for A_{IB} can be taken directly from A_{IB} in the previous modelling and the values for $R_{A_{IB1i}} = R_{A_{IB2i}} = R_{A_{IB3i}}$ could be set equal to $R_{A_{1Bi}}$ in the previous modelling.

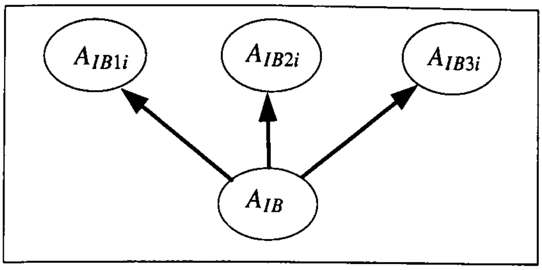

Figure 6.4: Bayes linear network for assessing reliability

6.8.2 Different Measurements of Reliability During Trial

As was seen above, the decision maker does not place a large amount of confidence in the output of the RGT phase. This is partially due to the short duration of the phases, but it can possibly also be attributed to the method used to assess the current reliability, i.e. the AMSAA model. The contractor produces three alternative assessments of the current reliability of the system. These are phase reliability, cumulative reliability and fix effectiveness reliability. A short description of these three methods for calculating reliability is given.

Phase Reliability Phase reliability is the observed reliability of the system during the particular phase. This is measured as $P_R = e^{-\frac{1}{\theta_P}}$ where $\theta_P = \frac{\lambda}{t}$ and λ is the number of failures observed during t BFM. The MOD believe this value is artificially high as it dismisses failure modes previously occurred but are assumed to have been resolved.

Cumulative Reliability The cumulative reliability value is the reliability of the system calculated across all the BFM carried out, such that $C_R = \frac{\lambda_C}{t_C}$ where λ_C is the total number of failures observed during t_C BFM. If reliability is low at the start of the testing period and there is a period of reliability growth, the actual reliability of the system at the end of the testing phase will be greater than the output of this method.

Fix Effectiveness Model The Fix Effectiveness model assumes a number of working fixes have been applied to a number of failure modes. It then assumes $x\%$ of failures have been successfully removed and calculates a new basic and mission reliability based on this. In this project, a value of 80% has been assumed and a value of over 80% has been demonstrated by

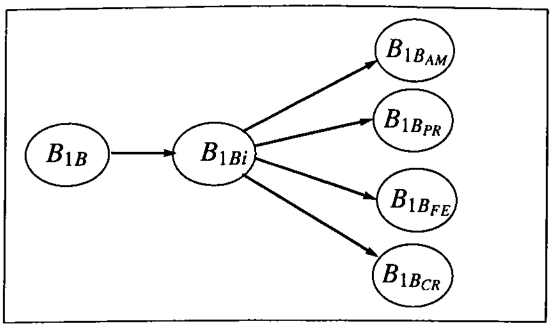

Figure 6.5: **Bayes linear network for assessing 'true' reliability**

the contractor.

In order to gain more confidence from the data, instead of using one observed value as in the previous modelling, it may be possible to combine these four forms of observations. Introduce four new variables, $B_{1B_{AM}}$, $B_{1B_{PR}}$, $B_{1B_{CR}}$ and $B_{1B_{FE}}$ respectively for the output of the AMSAA method, phase reliability, cumulative reliability and fix effectiveness reliability calculations for Vehicle A's basic reliability during Phase 1. If B_{1Bi} is redefined as the 'true' observed reliability, then each of the four variables above can be written as

$$B_{1B_{AM}} = B_{1Bi} + R_{B_{1B_{AM}}}$$

$$B_{1B_{PR}} = B_{1Bi} + R_{B_{1B_{PR}}}$$

$$B_{1B_{CR}} = B_{1Bi} + R_{B_{1B_{CR}}}$$

$$B_{1B_{FE}} = B_{1Bi} + R_{B_{1B_{FE}}}$$

One potential Bayes linear network for the above modelling is represented by Figure 6.5. Alternatively, it may be that at different stages, the decision maker uses a different method. For the phase 1 reliability, the decision maker may believe phase reliability represents the reliability of the vehicle at that stage most closely. Alternatively, at later stages, the decision maker may believe the AMSAA model output has greater accuracy.

6.8.3 Assessment of Expert

The decision maker believed the expert would be unbiased regarding the growth of the reliability between different variants. As the expert is giving a number of assessments regarding the growth, twelve in total, it would be possible to carry out learning regarding the experts assessment. For example, the expert may consistently over or under estimate the amount of growth. This learning could then be used to adjust future predictions of the expert's assessment to take into account the potential biasness. The decision maker in this project modeled that the each of the expert's assessment of the growth between variants were uncorrelated given the growth variable. However it is interesting to investigate how learning about the expert's bias could be incorporate into the model in the future. This is demonstrated solely for Vehicle A's basic reliability.

The variables $[A_{1B}, A_{1Bi}, GA_{1,2B}, A_{2B}, EGA_{1,2B}, A_{2Bi}, GA_{2,3B}, A_{3B}, EGA_{2,3B}, A_{3Bi}, GA_{3,IB}, A_{IB}, EGA_{3,IB}]$ are defined as above. A new variable Z_{A_B} is introduced such that $EGA_{1,2B} = GA_{1,2B} + Z_{A_B} + R_{EGA_{1,2B}}$, $EGA_{2,3B} = GA_{2,3B} + Z_{A_B} + R_{EGA_{2,3B}}$ and $EGA_{3,IB} = GA_{3,IB} + Z_{A_B} + R_{EGA_{3,I2B}}$ assuming that Z_{A_B} is uncorrelated with all the growth variables. Once $EGA_{1,2B}$ and A_{2Bi} have been observed, the expectation of $GA_{1,2B}$ will be adjusted which in turn will adjust Z_{A_B}. Once this adjustment is made, further adjustments are made to the expectation of $EGA_{2,3B}$ and $EGA_{3,IB}$. This variable could also potentially influence the expert assessment's of the growth for the other reliability requirements or the other vehicle.

By carrying out learning about the calibration of the expert's assessment of the growth, greater confidence can be placed upon future assessments of the growth between the variants. Where there is no observed data, as in the case of the in-service variant of the vehicle, the expert's assessment of the growth between the Phase 3 variant and the in-service variant can have a substantial effect on reducing the uncertainty of the in-service variant. Introducing the new variable and a new ordering, a new Bayes linear network, Figure 6.6, can be created for Vehicle A.

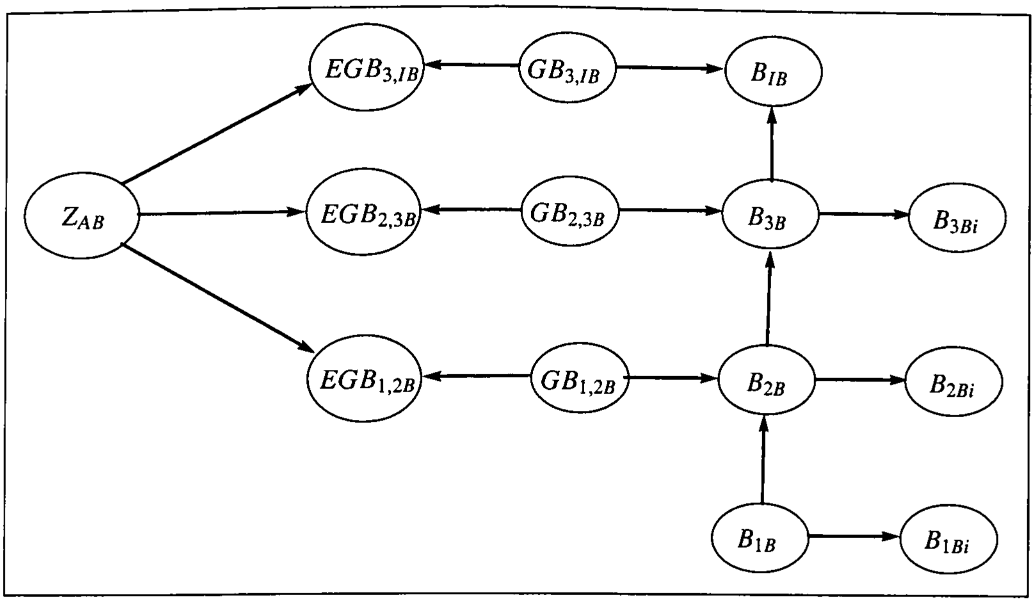

Figure 6.6: Bayes linear network for calibrating expert's assessment

6.9 Conclusion

This chapter aimed to examine how the Bayes linear methodology could be used to aid the decision maker in assessing the reliability of a system being procured. In this project, the MOD were procuring two land vehicles with similar sub-systems and the same contractor. The contractor was carrying out RGT's and was aiming to demonstrate through these RGT's that reliability requirements for basic and mission reliability of each system had already been met. The decision maker on the project, was required to make a recommendation to the IPTL on whether or not the systems currently meet reliability requirements.

As in the previous project, the modelling was broken down into three parts; qualitative, quantitative and analysis. In addition to modelling each of the system's basic and mission reliability using the observed data, the decision maker and analyst attempted to use the information gathered about one vehicle, to make inferences about the other vehicle. The reasoning was that due to the vehicles being similar and the contractor of each vehicle being the same, then if the contractor was guaranteed to meet the basic reliability requirement of one vehicle, then the likelihood of the other vehicle meeting basic reliability requirements would rise.

The decision maker initially believed the basic and mission reliability requirements of each vehicle would be met. Once the observed data from the RGT's was gathered, the output of the model suggested it was unlikely Vehicle A would meet either of the reliability requirements. The model output suggested Vehicle B would potentially meet the basic reliability target and it was possible the mission reliability target would also be met. Sensitivity analysis was carried out to explore the effect of the output of the model if the prior belief structure specified by the decision maker was adjusted. Any change to the belief structure that resulted in the three reliability requirements reaching targets was not acceptable to the decision maker. She felt these changes were greater than what she was comfortable with.

The modelling also demonstrated that building more and more complex models does not guarantee the decision maker's uncertainty is going to drop. Instead, by gathering expert's opinion the decision maker trusts, the standard deviation could be reduced by more than building complex models. In this project it was clear the 'value' of additional expert opinion was greater than the 'value' of building a more complex model.

Chapter 7

Model Validation and Verification

7.1 Introduction

The focus throughout this thesis has been to examine if the Bayes linear methodology can be developed to support decision making for the MOD when assessing reliability. In order to do this, the Bayes linear methodology has been developed through application in 'real' projects to explore the feasibility of applying the methodology in future MOD R&M problems. The models developed during these projects have been used to support decision making. Based on this experience and feedback gathered from the MOD, we consider that the methodology is capable of supporting future MOD R&M decisions.

As a result of this, it is necessary to assess the quality of the models developed and the overall modelling process. This task is often referred to as validation; however, there is no definite agreement on what constitutes a valid model (Landry and Oral, 1993) or how to actually validate a model (Dery et al, 1993). Mitchell (1993) defines validation as the process of testing that a model represents "a viable and useful alternative means to real experimentation" [Mitchell, 1993]. The use of the words "viable" and "useful" are sufficiently vague to pose the questions - "how viable" and "how useful" must the model be to be considered valid? In cases where carrying out extensive experimentation is impossible or unrealistic, then even the most basic model may be viable and useful given the alternative. In the setting of the MOD, an alternative to full experimentation has been proposed which requires the R&M Specialists to analyse the R&M Case in an unstructured subjective manner and to make decisions based on this analysis. Validation will therefore aim to examine if the Bayes linear methodology is a useful and viable alternative to the analysis currently carried out by the MOD.

If validation is taken to be a "comprehensive demonstration that a model is fully correct" [Pidd, 203], then validation is impossible (Pidd, 2003). Instead, a pragmatic approach to validation is adopted. It will be argued that a requisite model for the decision problem has been built in each of the projects. A "model is requisite if its form and content are sufficient to solve the problem" [Phillips, 1984]. Two key features of a requisite model are that it omits elements that do not significantly contribute to the output of the model and it approximates complex relationships (Phillips, 1984). In each project, the decision maker incorporated into the modelling all factors that influenced her belief. Where time was not available to gather detailed data on a dependency, structured elicitation was used to estimate the strength of the relationship. Whilst the model may not be considered 'perfect', given the resource constraints of the MOD decision maker, it is believed to be sufficient to support the decision reached. To validate the methodology, a comparison will be made with the current process. Furthermore, investigation of whether or not the additional requirements identified in Table 1.4 have been met by the Bayes linear methodology is carried out.

Validation will be carried out on two levels; validating the models and validating the modelling process. In order to do this, validation criteria for each must be identified. Section 7.2 discusses the different validation criteria used. Section 7.3 will take steps to validate the model whilst Section 7.4 examines validation of the modelling process. Section 7.5 will conclude and summarise the chapter.

7.2 Validation Criteria

In the field of statistical inference, a number of authors have proposed different methods to validate models. Aven (2003) and O'Hagan and Forster (2004) suggest that one method of validation is to explore how closely the output of the model relates to observations about the real-life system. Within a subjective Bayesian framework, this approach to validation is unsatisfactory for two reasons. Firstly, the decision maker's beliefs are modelled and do not necessarily completely represent 'reality'. Secondly, observations for the real-life system may take several years to collect while some form of validation is required in the short term.

Oral and Kettani (1993) divide validation into a number of different sub-categories such as data validation, conceptual validation and aptness validation amongst others. For each category, different criteria for validation are used. As an example, for data validation, the data gathered must be demonstrated to be sufficient, accurate, appropriate, available, reliable and cost effective (Oral and Kettani, 1993). Another example is operational validation. Operational validation investigates the usability, usefullness and cost of supporting a decision using a given model (Oral and Kettani, 1993).

It is necessary to study fields other than risk analysis to establish validation criteria. Work on validation has been carried out in the field of System Dynamics. Forrester and Senge (1980) identified 17 different validation criteria for system dynamic models which included issues such as suitability, effectiveness and consistency. Despite the age of this paper, it is considered a seminal work, as evidenced by Richardson who states that it is the "most comprehensive statement on model validation and user confidence in system dynamics models" [Richardson, 1996]. This is emphasised by the fact that between the period of 1985 and 1995, only three papers addressing validation were published in System Dynamics Review (Richardson, 1996). In simulation modelling, there appears to be a lack of agreement with many authors about the definition of terms such as validity, credibility, acceptability and confidence (Robinson and Pidd, 1998). The lack of extensive literature suggests that validation is something that is either unpopular, not carried out regularly or largely goes undocumented.

Defining validation criteria depends upon the purpose of the model (Pidd, 2003). For example, validation should be thoroughly explored when a model is used exclusively to make a decision. Where a model is used as a framework to explore different scenarios, strict validation criteria may not be necessary. Richardson (1996) believes that it is necessary for decision environments to develop their own set of validation criteria that are deemed suitable.

Dery et al (1993) offer a detailed and interesting discussion on the philosophical and historical perspectives regarding validation. Adopting a philosophical perspective, a Management Science model should satisfy four different criteria: "the model should be based on objective research, expressed mathematically, shown to be useful and should pass tests designed to show its inadequacy" [Pidd, 2003]. The historical perspective to validation suggests that the scientific community define the validation criteria and this criteria can change over time (Dery et al, 1993). Therefore, validation is carried out by the wider Management Science community.

Carrying out exhaustive validation can be extremely difficult and time consuming. Pidd (2003) identifies many difficulties when carrying out validation. One particularly relevant to the MOD is ensuring that the model represents reality, i.e. the decision maker's true beliefs, rather than representing some intended reality. Despite these difficulties, carrying out validation is an extremely useful task. It forces the analyst and the decision maker to explore scenarios they may not have previously considered and to push the model to its limits. However, to carry out feasible validation, it is necessary to be pragmatic. All models can be demonstrated to be false, yet some can be extremely useful (Box and Draper, 1987).

A mixture of the philosophical and historical validation techniques are used to validate the modelling carried out during this research. The models are shown to be based on objective research currently in the scientific literature and that the decisions the models support are robust to changes in the prior specification. In order to develop a methodology that is tractable and resource efficient, it has been necessary to reduce the elicitation specification placed upon the decision maker. This indicated that the prior information gathered from the decision maker is not as detailed as it is with other methods. Whilst the Bayes linear models developed may not be as detailed as uncertainty modelling carried out by other methodologies, such as BBN's, the models developed captured the essential features of the problem, offered the decision maker a framework to consider the problem and explored different decisions, i.e. it is a requisite model to the decision problem faced.

As the modelling process proposed is quantitative, it different to any methodology used by the MOD to date. However, it is argued that the modelling is similar to the unstructured analysis carried out by MOD R&M decision makers. In addition, the models offer the opportunity to carry out analysis that may be too difficult to carry out intuitively. The process is evaluated using interviews with MOD R&M staff to assess the usefulness of the methodology and to assess whether or not the additional requirements defined in Table 1.4 are met. Table 7.1 lists the validation criteria used in this research for the model and the modelling process.

7.3 Model Validation

In order to validate the model, three criteria were used. Phillips (1984) proposes that validating a requisite modelling is achieved by "a mixture of 'hard' and 'soft' techniques" [Phillips, 1984].

Model		Modelling Process	
Criteria	Method	Criteria	Method
Meet Project Aims	Interviews	Compare with current process	Workshop feedback session
Model is robust	Sensitivity Analysis	Additional requirements assessment	Interviews Observational evidence Workshop feedback session
Model behaves appropriately	Scenario Analysis		

Table 7.1: Validation Criteria

A mixture of interview and quantitative tools have been used to validate the model and these have been predominantly covered within Chapter 5 and Chapter 6. This section is a summary of the validation carried out for each model.

The first criteria is to assess whether or not the modelling met the aims and objectives set out by the decision maker, i.e. the modelling is capable of answering the typical questions faced by a R&M Specialist during a procurement project. In addition, information concerning what the R&M Specialist learnt from building the model was also gathered. This was achieved by interviewing the decision maker after modelling was completed, see Section 5.7 and Section 6.7. Both decision makers felt that the modelling was useful in structuring the information contained in the R&M Case. In addition, the qualitative structuring of the different information sources helped them to identify additional sources of information they could observe. In Chapter 5, this was the information contained in the HAT and SAT and in Chapter 6, this was the information provided by the contractor's expert. Both decision makers felt that the structured quantification of their belief's helped them to focus on where their main uncertainties existed. In Chapter 5, the decision maker stated that formally quantifying her beliefs helped her to identify why the two R&M Programmes that she initially believed were equally effective, were in fact, very different.

The second criteria is to assess whether or not the model behaves as the decision maker expects. This has been examined by simulating different observations and comparing the adjusted expectation with what the decision maker would expect to see. After the entire model was built, different observations were entered into the model. The decision maker was presented with the adjusted expectation and asked if she felt that this was a reasonable value given the observation.

The decision maker was given the opportunity to reflect on the results and to revisit her prior specification if she so wished. In addition, on occasions where the values elicited were counter intuitive to the facilitator, they were highlighted by the decision maker. The decision maker was then asked to explain their prior specification. One example of this occurred in Chapter 6. In this project, the correlation between the expert's assessment of growth and reliability of the vehicle, was similar to the correlation between the observed RGT data and the reliability of the vehicle. The decision maker was made aware of this and stated that she believed that this reflected her true belief. She believed that due to the short duration of the individual RGT's, she did not gain much confidence from the information. Instead, she believed that because the contractor works on the project every day, they are equally positioned to provide information about the growth than a short RGT phase.

The third validation criteria is sensitivity analysis. Mitchell (1993) uses sensitivity analysis to validate that the model is robust to changes in the input parameters and sensitivity analysis plays a major role in requisite modelling (Phillips, 1984). By demonstrating that decisions are unchanged by different input parameters, the model is considered to be robust and partially validated. In each project, sensitivity analysis has been carried out to assess how much change would be needed to the decision maker's prior beliefs before a change in the final decision would need to be made. Section 5.6.3 and Section 6.6.3 document the sensitivity analysis carried out for each project. In each, the decision maker did not believe that their prior beliefs would change by enough to impact on the final decision taken.

An important part of building Bayes linear models is deciding at which level of detail to build the model, i.e. which variables to include and which to disregard. This is a further reason why having a facilitator to support the decision maker build the model is so important. During both the applied projects in this research, the facilitator read the R&M Case Report prior to meeting the decision maker and identified all the potential sources of information that could be gathered during the life of the project. The decision maker then identified a number of variables which she believed were relevant and these were modelled. The additional sources of information identified by the facilitator were then presented to the decision maker. At this point, the decision maker decided if these additional sources of information were required. Essentially, any information source that could influence the belief of the decision maker was included in the model.

Both decision makers believed that given the amount of time they had to work on the problem, both models included everything that could be captured and quantified. The modelling also allowed the decision maker to explore the impact of different assumptions and prior beliefs on the output of the model. Phillips (1982) states, "A model is considered requisite when no new intuitions arise about the problem, at least no intuitions that would affect the decision". The decision as to whether or not a model is requisite is dependent on the current information and the problem owners (Phillips, 1984). In both projects, all forms of information that could influence the decision maker's belief were included.

7.4 Modelling Process Validation

The focus of this section is to validate the Bayes linear methodology as an alternative method to the unstructured subjective analysis that is currently carried out by the MOD to analyse the information presented in a R&M Case. Validation criteria for the modelling process will be set-up pragmatically to demonstrate that the modelling is a feasible alternative to the current methodology. Two criteria are used: does the modelling process capture everything that is currently carried out by the MOD when assessing reliability and to what extent are the additional requirements identified in Table 1.4 met. If these two criteria can be demonstrated, then it is assumed that the methodology is a feasible alternative to the current method.

After the two projects were completed, a feedback session with MOD R&M Specialists was carried out. During this feedback session, the Bayes linear methodology was described, the analysis from the two projects was illustrated and a simple theoretical reliability example was built. Six MOD R&M Specialists were then given a simple example to analyse. In this example, they had to build the influence diagram, specify the necessary values and carry out the Bayes linear adjustments. Once this example was complete, each participant was given a short questionnaire to fill in which gathered information on their thoughts regarding the Bayes linear process. This questionnaire can be found in Appendix G.

7.4.1 Comparison with Current Process

This section compares the current MOD process with the proposed Bayes linear methodology. Recall that currently, the MOD assess all evidence in a R&M Case in an unstructured subjective manner. There are a number of issues that arise because of this evaluation method. Firstly, alternative decision makers could have different opinions on whether a R&M Case is satisfactory or not and in essence, none of them are right and none of them are wrong. Secondly, as the evaluation is carried out internally by the decision maker, analysis is not transparent or documented.

Three main questions were investigated during the feedback session. Firstly, how useful did the participants find the Bayes linear process and what issues may arise with using the methodology. Secondly, how does the current process compare with the Bayes linear methodology. Finally, what sort of R&M problems could they visualise applying the methodology on.

Across the responses, there was agreement that the Bayes linear methodology was potentially useful for assessing uncertainty in a R&M Case. Participants stated that they felt that the Bayes linear methodology offered more structure to their evaluation. In addition, the qualitative aspect of building the BLN was highlighted as beneficial to capturing one's thoughts. One participant stated, "The graphical representation would provide an excellent focus for discussion". Another stated that building the BLN collaboratively would allow different parties in the procurement process to gain clarification on what is provided in the R&M Case.

However, there were comments about the simplicity of the example presented. They believed that in an example with more variables, building the BLN and specifying the necessary values would be difficult. This issue will be addressed in more detail in Section 7.4.2.3. One R&M Specialist commented that he felt that the lack of knowledge within the MOD and contractors about the concept of BBN's would also make application of the Bayes linear methodology difficult.

Some of the criticisms of the methodology, however, appeared to actually support the use of the methodology. One R&M Specialist stated, "I would be cautious about sharing the results with a contractor given that they could easily argue against any assumptions that have been made". This seems rather counter-intuitive as the aim of any methodology should be to stimulate discussion, gain consensus and carry out analysis that is meaningful to each party. If the Bayes linear methodology is able to stimulate that kind of discussion and communication between the MOD and contractor, then this is viewed as a useful feature. When disagreements do occur, it is likely to occur with the quantitative specification. As the Bayes linear methodology structures this quantitative specification transparently, the values causing the disagreement can be easily identified. Where consensus cannot be achieved, sensitivity analysis, as demonstrated in both projects, can be carried out to assess whether or not the different opinions impact on the final decision.

The R&M Specialists were asked to consider the range of R&M problems to which they could apply the Bayes linear methodology. The problems identified included supporting contractor's claims regarding the added value in including a source of information in the R&M Programme; support decision making around whether or not an ISRD is required; and justifying decisions when building a business case for a new system.

Overall, the R&M Specialists believed that the Bayes linear methodology was actually very similar to what they were already doing. However, the methodology gave them the structure to explicitly state their belief about many aspects of the R&M Case; the qualitative and quantitative relationship between different sources of information, their belief about the current reliability of the system and any assumptions they make about the information sources. One R&M Specialist stated ,"It [Bayes linear] formalises the decision making (intuitive process) allowing the documentation of decision progression and justification of action".

7.4.2 Meeting Additional Requirements

This section attempts to investigate to what extent the additional requirements identified in Table 1.4 have been achieved. These requirements were that the methodology was logically justifiable, traceable, efficient and simple to use. The methodology has been demonstrated to be mathematically justifiable in Section 2.4. This section focuses on demonstrating that on the MOD R&M projects, the methodology was shown to be traceable, efficient and simple to use. Three main sources of information were used: the interviews carried out with the decision makers on each project, feedback gathered from workshop session and observational evidence gathered during both the projects and the workshop session.

7.4.2.1 Traceable

It is desirable that the methodology is traceable and transparent. There are a number of reasons for this. It allows the contractor to observe how the MOD are analysing their R&M Case. By building the decision maker's belief structure in a transparent way, it allows for easier discussions regarding disagreement. Transparency means that future MOD R&M decision makers are able to assess the decisions made by previous decision makers.

As with the Bayes linear steps, the transparency can be broken down into three parts: qualitative, quantitative and analysis. Out of the three parts, the qualitative analysis could be seen as the least transparent of the three parts as the decision maker may not fully justify their BLN. Instead of defining all the dependencies, they may simply build the BLN and state their prior beliefs. This may lead to future problems as some of the relationships defined may not be initially obvious. In both projects carried out with the MOD, each of the relationships were explicitly defined and explained before elicitation took place. For example, in the future it may not be initially obvious why observing the HAT or SAT influence X and not X_R. In cases where the dependency may not be obvious, it is beneficial if the decision maker explicitly explains each of them to ensure that future decision makers understand why certain relationships were modelled.

For the quantitative analysis, the specification of the variables is transparent and, as illustrated in previous chapters, will be traceable if captured in a sensible manner. This suggests that both future decision makers and contractors can assess the decision maker's beliefs. This also gives the current decision maker and contractor the platform to discuss the current state of knowledge. This is obviously something that is difficult to achieve with an unstructured analysis approach. Given that the belief structure and observations can be agreed upon, there is unlikely to be any disagreement on the output of the model.

Based on this information, it is believed that the Bayes linear methodology meets the traceable requirement.

7.4.2.2 Efficient

During the interviews conducted with MOD R&M staff in October 2005 (see Section 1.5.3), it was apparent that making the methodology process resource efficient was important. MOD evaluators have limited time to work on a number of projects and so any quantitative methodology must be cost and time efficient. One of the main reasons the BBN methodology was dismissed as a potential methodology was because of the time needed to build and populate one. This was considered to be too great for MOD R&M evaluators to consider. The Bayes linear methodology was proposed as the time required to build and populate a Bayes linear model is less than for a BBN; yet still retains many of the BBN positive features.

During the interviews carried out with the decision makers on each of the two applied projects, both of them felt that building, populating and carrying out analysis with the Bayes linear model was a good use of their time. However, both of them expressed concern about their ability to repeat the modelling. Based on this information, it is believed that the Bayes linear methodology meets the efficiency requirement.

7.4.2.3 Simple to Use

This requirement was extremely important as the aim of the research is not to develop a modelling process that required external consultation every time the MOD wished to apply it. The aim is to develop a methodology that could be used by those with limited statistical knowledge. This requirement requires more subtle thinking than the previous two because the abilities of those choosing to apply the methodology must be considered. Another important consideration is motivation, however, it is assumed for the purposes of this discussion that anyone wishing to apply the methodology in the future, is sufficiently motivated.

There are four different possible scenarios that need to be explored. Firstly, the Bayes linear methodology is simple to use and can be applied to a wide range of problem by those with limited statistical knowledge. Secondly, the methodology is straightforward but those wishing to use it require basic training. Thirdly, the methodology is difficult, however a small number of tricks can be learned that means the methodology can be applied to any type of problem. Finally, the methodology is difficult and models need to be simplified in order that a wider range of models can be addressed.

It is unlikely that models of similar complexity to those built in Chapter 5 and Chapter 6 could be modelled by decision makers with similar levels of statistical knowledge to MOD R&M decision makers. A number of subtle modelling techniques had to be used during the applied projects. In Chapter 5, modelling for the HAT, SAT and the FMECA required subtle manipulation to ensure the decision maker's true beliefs were being modelled. In Chapter 6, modelling the relationship between the two vehicles and the requirements was more complex than modelling other dependencies as alternative methods of eliciting the information had to be used. Building models of similar complexity to model MOD R&M problems is unlikely to be achievable assuming no additional statistical training.

An alternative would be to give the R&M specialists some basic training in applying the Bayes linear methodology. On two occasions, basic training was given to a group of MOD R&M specialists. The first of these was the feedback session described above. On this occasion, even after basic training, the R&M Specialists found building and populating the Bayes linear model a difficult task and required substantial support. The second occasion was during a University of Strathclyde R&M course which had MOD R&M Specialists attending. As before, the Bayes linear methodology was explained, the modelling carried out in the two projects was demonstrated and a simple example was worked through. As before, the MOD R&M Specialists struggled with the modelling and it is unlikely that the models in their current forms could be addressed even with some basic training by most R&M Specialists.

The third option would be to identify a number of typical scenarios that may arise when modelling and to create a step-by-step guide to model these situations. It is likely that those wishing to use it would still require basic training or to review literature on the Bayes linear methodology. It would be possible to devise a set of 'rules' for decision makers to follow in the future. If a set of 'rules' could be developed to assist future decision makers, then it is possible that software could be developed that could support decision makers. If software could be developed, it could be set up so that once the influence diagram is built, elicitation questions are automatically generated.

The final option would be to attempt to simplify models so that difficult modelling techniques are not required. This would mean that the models may not accurately represent the decision maker's true beliefs. However, in the two applied projects with the MOD, the modelling that was considered too difficult for the decision maker to carry out alone, were the information sources that were not contributing significantly to the uncertainty. In Chapter 5, the information sources that were most difficult were HAT, SAT and FM. The correlation value for these three variables with X for contractor Yellow were 0.013, 0.016 and 0.028 respectively. This is compared to a correlation value of 0.94 between X_Y and RBD_{C_Y}; which was the weakest of the RBD variables. In Chapter 6, the difficulty in modelling occurred when modelling the relationship between the two vehicles and the two reliability requirements. As demonstrated in Section 6.6.5.2, this relationship contributed very little to the overall modelling. Hence in each of these projects, the difference in building a simple model as opposed to the complex one built would have been very little.

This poses a problem as the distinction between whether or not the methodology is simple to apply is ultimately subjective. Undoubtedly, a Bayes linear model is simpler to build than a BBN. In the context of the MOD, it is unlikely that all the R&M Specialists have sufficient statistical knowledge to adopt the methodology on R&M problems. Instead, it is possible that a handful of the R&M Specialists would be able to apply the methodology and they could be used as internal consultants on projects where a decision maker wished more support or a quantitative solution was desirable. In addition, the types of problems that were identified during the feedback session are unlikely to be as complex as the models that R&M Specialists believed they may apply the methodology on. One of the problems identified was whether or not an ISRD should be carried out. This modelling may only contain a handful of variables and is likely to be manageable for most of the R&M Specialists.

Most importantly however, the task of building, populating and analysing a Bayes linear model is broken down into the three parts of qualitative modelling, quantitative modelling and analysis. It is believed that even if a R&M Specialists was not capable of carrying out all three parts, attempting to build the Bayes linear network or to begin to structure their beliefs quantitatively, are tasks that are beneficial to carry out in most procurement projects. Even if the decision maker was not able to do an in-depth analysis as carried out in Chapter 5 and Chapter 6, the simple process of attempting to structure ones thoughts in a transparent way, can offer further insight into the problem and help to stimulate thought.

Given the type of problems to which the MOD are likely to apply the Bayes linear methodology, the statistical knowledge of some of the R&M Specialists and the breakdown of gain in the different sections of building a Bayes linear model, it is believed that the Bayes linear methodology is simple enough to be carried out fully in the rare cases this is desirable and simple enough to offer support to the decision maker in the times when an extensive analysis is not required.

7.5 Conclusion

Validation is an extremely difficult task and one that can never be carried out exhaustively. The amount of validation a model requires depends upon how much scrutiny and review the model will be placed under (Richardson, 1996). Instead of attempting to fully validate each of the models and the overall process, this chapter has attempted to address a number of the primary concerns that the literature raises with validation; both with the models that were built in Chapter 5 and Chapter 6 and with the overall Bayes linear methodology.

For the models built in Chapter 5 and 6, scenario analysis and sensitivity analysis were carried out to demonstrate that the model has behaved as the decision maker wished it to behave and that the output of the model was robust to changes in the prior beliefs specified by the decision maker. In both chapters, extensive sensitivity analysis was carried out. In both cases, the decision maker stated that it was unrealistic that her prior beliefs would change by such an extent that her decision would change. In addition, the usefulness of the modelling was assessed through interviews with the decision maker on each project. Both decision makers believed that the models provided good feedback regarding the problem and offered them a further tool to support the advice they give to the IPTL.

For the overall modelling, in order to demonstrate that it was a feasible alternative to the current analysis process, the Bayes linear model was compared to the current process to ensure it was capable of supporting the same decisions. In addition, interviews and feedback from the MOD were used to assess whether or not the four additional requirements that were identified in Table 1.4 were actually met. Based on the information gathered, it appears that the methodology provides a structured framework to the support the reliability decisions currently facing the MOD. In addition, three of the four additional requirements placed upon the methodology by the MOD were also met and one, simple to use, was partially met.

7.5.1 Removing Sources of Information

An important part of building Bayes linear models is deciding at which level of detail to build the model, i.e. which variables to include and which to disregard. This is a further reason why having a facilitator to support the decision maker build the model is so important. During both the applied projects in this research, the facilitator read the R&M Case Report prior to meeting the decision maker and identified all the potential sources of information that could be gathered during the life of the project. The decision maker then identified a number of variables which she believed were relevant and these were modelled. The additional sources of information identified by the facilitator were then presented to the decision maker. At this point, the decision maker decided if these additional sources of information were required. Essentially, any information source that could influence the belief of the decision maker was included in the model.

Chapter 8

Discussion

8.1 Introduction

The aim of this chapter is to reflect on the reported research and to discuss future work in Bayes linear; as well as reliability assessment.

Section 8.2 examines the extent to which the overall research goals and aims have been met. Section 8.3 discusses the contribution this research has made to reliability assessment. Section 8.4 examines the wider contribution of the research and Section 8.5 discusses the limitations of the current state of research and from this, areas of future research are identified.

8.2 Research Goals

Recall Table 1.4. The aim of the research has been to develop a quantitative methodology that could be used by decision makers to support decision-making at key times in a procurement project. The methodology had to be capable of evaluating the different forms of information presented in the R&M Case by contractors. In addition, the methodology had to be mathematically justifiable, traceable, efficient and simple to use.

Two candidate methodologies for this problem are BBN's and Bayes linear. BBN's have a number of benefits, however, they are time and resource intensive and the evidence of this research suggests it is beyond the statistical capabilities of the MOD to build and populate a BBN. The Bayes linear methodology shares a number of positive features with BBN's; graphical representation, transparency and the capability of combining multiple sources of information to produce a posterior belief structure.

The two methodologies use different primitives; BBN's use probability whilst Bayes linear uses expectation. Using expectation rather than probability reduces the specification required by the decision maker, hence making it tractable. This reduced specification means the output of a Bayes linear model is less rich in detail compared with a BBN model. It has been suggested that in complex problems, the extra detail necessary to populate a BBN is gathered in such a way that the decision maker may not have confidence in the output of the model.

Two theoretical R&M examples have been developed to help identify issues prior to application of the Bayes linear methodology on 'live' MOD projects. Building these models illustrated that procedures to build Bayes linear models were not documented in the literature. These included identifying elicitation guidelines, sensitivity analysis and building Bayes linear networks. This thesis builds on the current research establishing Bayes linear as an alternative inferential method by documenting these steps that often go undocumented but need to be taken to build a Bayes linear model. In order to explore the feasibility of the Bayes linear methodology as a quantitative methodology to support decision making within the context of the R&M Case, two projects, one at each of the key decision making points, were identified.

The R&M decision maker on the first project had to assess which contractor was likely to produce the most reliable system. During the assessment phase, the MOD influence the final reliability of the procured system through their choice of contractor. This project involved three contractors. A Bayes linear model has been built and the observations that each of the contractors had made were input into the model. The model has also been used to explore the uncertainty remaining in the point estimate of the reliability of each system given the contractor carries out all the reliability tasks they say they will. From this, the decision maker has been able to feedback the reliability of each contractor's system and their R&M programmes to the IPTL.

The second project involved the MOD procurement of two systems where each system was undergoing reliability growth trials. The R&M decision maker on the project had to assess if the systems were meeting basic and mission reliability requirements based on the information gathered during the trials. The model was built to assess the basic and mission reliability of each vehicle and the model addressed softer factors that other modelling was not taking into account but were influencing the decision maker's beliefs. These were the expert's assessment of the growth between phases, the dependency between the mission and basic reliability, and the dependency between the two vehicles. Analysis indicated that, at the point where the modelling was carried out, both vehicles did not meet mission reliability requirements and one vehicle was not meeting basic reliability requirements. Sensitivity analysis demonstrated that it was unlikely that changing the belief structure by a feasible amount could affect the conclusion that the three reliability requirements not being met.

Whilst the literature shows the Bayes linear methodology has been applied and documented in a number of practical projects, prior to this research, the methodology has not been applied to support decision making when evaluating reliability programmes. This research has shown that it is feasible to use thereby meeting the research goal and contributing to the existing literature.

8.3 Benefit of Adopting Bayes Linear to Assess Reliability

It is proposed that the Bayes linear methodology should be considered as an alternative to the current unstructured approach used by the MOD to assess the merit of a R&M Case. The Bayes linear methodology offers decision makers a simple method to structure and quantify their beliefs and to adjust these beliefs as new information becomes available. There are a number of features that make it useful as a method for customers to assess reliability and, potentially in the future, as a method for suppliers to present evidence.

Firstly, the proposed methodology is transparent and therefore, the contractor understands how the information they are providing is being assessed by the customer. Secondly, the Bayes linear methodology is traceable. Some developmental projects may last a long time and decision makers may change on multiple occasions. Using the Bayes linear methodology, future decision makers can understand why previous decisions were made.

Thirdly, the Bayes linear methodology improves communication between the customer and contractor on two levels. Firstly, the relationships between different forms of information are explicitly defined in the Bayes linear methodology, both qualitatively and quantitatively. From this, it is clear to the contractor which forms of information they should gather in order to give the customer confidence that reliability requirements are met. Secondly, the amount of analysis required will be dependent on the state of knowledge at the start of a procurement project and therefore it is necessary for the contractor and MOD to establish this. The Bayes linear methodology differs from other Bayesian methodologies as establishing the current state of knowledge is simpler because the quantitative specification required is less intensive.

Fourthly, reliability programmes will contain a number of dependent variables and on occasion, this can lead to complex scenarios which are difficult to monitor. Using Bayes linear networks allow the contractor and customer to quickly understand the dependencies between different information sources. The decision maker in Chapter 5 stated that she believed that by explicitly defining what observations she wished to make, when she returned to the R&M Case in the future, she would know immediately which were the important parts of the R&M Case Report to consider. This, she believed, would save her a substantial amount of time. If the Bayes linear model is not fully built, the Bayes linear network offers the decision maker a visual aid to structure the information presented in the R&M Case. This is important as the evaluators may be working on multiple projects and quickly familiarise themselves with a project they have previously worked on could potentially save a lot of time.

Fifthly, the Bayes linear model provides the decision maker with a quantitative assessment of their posterior beliefs. This should reduce the opportunity for mis-communication between different R&M decision makers and contractors who may interpret different qualitative statements in different ways. Statements such as, "I am fairly sure reliability requirements have been met" can be replaced with quantitative statements such as "the expected value of the reliability variables is greater than the reliability target".

Observational evidence suggested that there exists situations for which the MOD has the knowledge to model but are too complex to carry out in an unstructured way. An example of one of these situations was presented in Chapter 5 where the decision maker was unable to distinguish between two of the contractor's bids. As the decision maker was attempting to model the entire R&M Case for each contractor, they were missing parts of the detail. However, when the modelling was broken down into small manageable sections, the decision maker was able to identify the difference between the two contractors. This suggests that the customer has the expert knowledge to evaluate the R&M Case, however, on occasions it is necessary that a tool is required to structure the evaluation. The Bayes linear methodology is one such tool that can carry out structured analysis.

Most of these benefits could be applied to other forms of modelling; particularly BBN's. The difference between these other forms of modelling and Bayes linear is that it is unrealistic to achieve some of these benefits without substantial time and effort. In many cases, the decision maker may not have the resources or the desire to achieve such detailed results. Instead, they may simply want an approximate understanding of a situation. In these problems, building a Bayes linear model is resource efficient for decision makers to achieve all the benefits listed above. As such, the Bayes linear methodology can model more complex problems in a shorter period of time.

One criticism that could be applied to the Bayes linear methodology is that the output of any model is dependent on the subjective inputs. One could imagine that disagreement may occur between contractor and customer over the quantitative values stated if either were dissatisfied with the model output. However, this criticism highlights the principal reason why Bayes linear should be used to assess a R&M Case rather than BBN's. If a contractor and customer were to disagree with the quantitative input, sensitivity analysis can be carried out to assess if changing these input affects the output. Chapter 4, Chapter 5 and Chapter 6 have illustrated that to carry out sensitivity analysis is straightforward and can be completed in real-time. To carry out similar analysis with a BBN would be extremely time consuming and difficult. Considering the complex reliability programmes that are potentially being modelled, it is unlikely that extensive sensitivity analysis could be carried out in practice within a BBN. Therefore, it is important that the values specified in a BBN are done so accurately.

Using a Bayes linear model, a decision maker can quickly specify the necessary values and adjust the models given the observations. They can then revisit their prior beliefs to assess if changing these beliefs have an effect on the final decision. Where it does have an effect on the final decision, the decision maker can spend additional time assessing the necessary quantitative values. Given the potentially confrontational nature of the relationship between contractor and procurer, sensitivity analysis is capable of addressing contentious issues that may arise such as the state of knowledge at the start of a project. As illustrated in Chapter 6, changing these prior beliefs did not have a substantial effect on the output of the model.

8.4 Wider Implications of Modelling Using Bayes Linear

The main focus of this research was to examine how quantitative methodologies could be developed to support decision makers to assess the evidence presented in a R&M Case. As such, the research goals, aims and requirements were formulated to reflect the capabilities and resource allocation of the customer, in this case the MOD.

However, the methodology and techniques developed throughout this thesis should not be seen as restricted solely to the problems facing the MOD or to those attempting to assess reliability. As highlighted in Section 2.4, the Bayes linear methodology has been applied in a number of decision making problems. This is the first occasion that the Bayes linear methodology has been applied to assess the information in the R&M Case leading to exploration of a number of features previously missing in the literature. Due to the adversarial relationship between the MOD and contractor, sensitivity analysis was examined in detail. This research illustrates how to carry out sensitivity analysis and the extent to which it can support the decision making. Without sensitivity analysis, in Chapter 6, the contractor could contest the inputs of the model. However, it has been demonstrated that in order to change whether or not reliability requirements were met, it would be necessary to change the decision maker's belief structure by an amount that she was uncomfortable with. It is conceivable that there are other situations that the Bayes linear methodology, as presented in this thesis, could be applied. For example, an organisation procuring equipment, such as Virgin buying aircrafts from Airbus, could use the Bayes linear methodology to assess reliability. In addition, it is realistic to assume that the methodology could be applied in areas other than reliability. The obvious alternative application would be for organisations such as HSE to assess the safety of equipment.

The needing to find effective methods to elicit subjective beliefs led to investigation of new methods since the literature was lacking this respect. Much of the published application of Bayes linear to date has focused on establishing the foundations of the approach and justifying why Bayes linear should be adopted. Less effort has been given to documenting how to build a Bayes linear model and particularly, how to specify the necessary values. In Goldstein and Wooff (2007), only three out of five hundred pages are devoted to assessing the expectation and variances and only a brief discussion is given with no conclusions. Alternative methods make assumptions that would not be feasible in this research project (O'Hagan, 1998). O'Hagan

(1998) assumes that all the beliefs are either normal or log-normal. When assessing reliability, it is unlikely that a decision maker's belief would fit conveniently within either of these distributions. Therefore, a review of the potential methods has been given (see Section 4.2.2) and how these methods have been applied in practice was given in Section 5.5 and Section 6.5. Particular focus was given to developing elicitation methods to assess the covariance between two variables. This thesis provides a reference regarding the elicitation of the necessary values to future users.

8.5 Limitations and Areas for Future Research

This section identifies areas which were not addressed in the thesis but are of interest. These limitations focus on three areas: current theoretical limitations, practical limitations of the modelling carried out with the MOD with respect to reliability and practical limitations of the Bayes linear methodology in a wider context.

8.5.1 Completing the Action Research Cycle

Recall Figure 3.3 from Chapter 3. Figure 8.1 is a Figure 3.3 revised with the different stages of the research mapped on to the Action Research cycle.

The first stage, Diagnosing, consisted of interviewing MOD staff to determine their requirements, reviewing the MOD Defence Standards and reviewing the wider reliability management literature. Once this was complete, the researcher assessed the different methodologies that could be developed. Once the Bayes linear methodology was chosen, the next step was to plan the overall modelling steps needed to model a reliability case using Bayes linear. This was the Action Planning stage. Action Taking involved working on two 'live' MOD projects and Evaluating took the form of interviewing the MOD decision makers, sensitivity analysis and scenario analysis to assess to what extent the methodology met the MOD decision makers needs. Specifying Learning involved presenting to the MOD the new methodology and how it could be applied. Throughout the entire process, the researcher constantly reviewed the literature on elicitation, Bayes linear and reliability theory. The final link from Specifying Learning to Diagnosing has been removed as time and resource constraints restricted the researcher from discussing with MOD decision makers how the Bayes linear methodology could be further de-

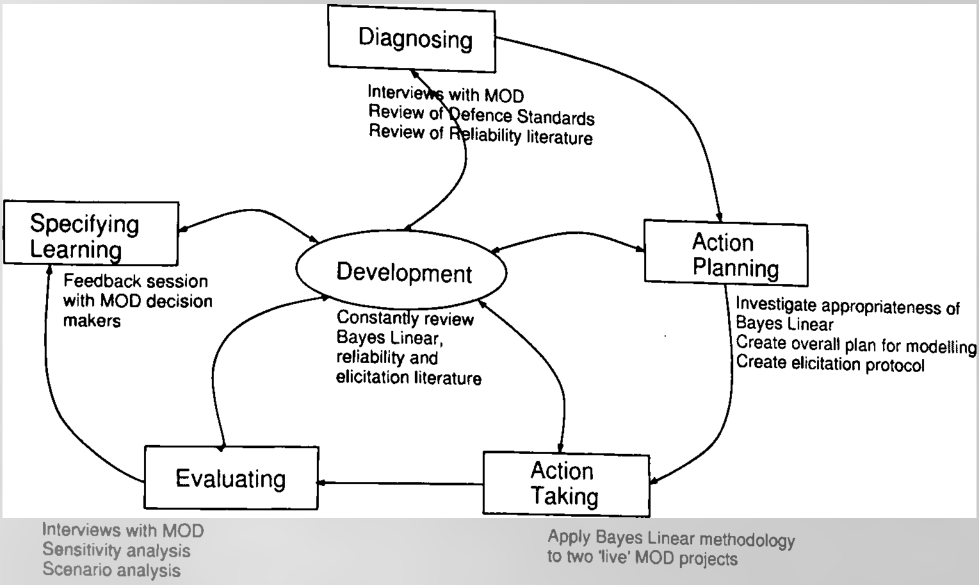

Figure 8.1: Tasks completed within the action research cycle

veloped; instead the methodology was evaluated during these interviews. If more time had been available, the cycle could have been revisited multiple times. The following paragraphs will outline the benefits of revisiting the cycle.

Re-diagnosing the problem in light of the developed methodology would complete the cycle and help to either further validate that the methodology meets the specified requirements or that further work is required. Further work could be required for two reasons. First, the MOD may decide that the developed methodology did not meet all their requirements and hence require further demonstration that the methodology meets requirements. Second, re-diagnosing the problem in light of the methodology may raise issues that were previously unseen in the first diagnosis.

Revisiting the cycle and in particular, the Action Planning phase of the cycle, would lead to revisiting and assessing the appropriateness of the elicitation methods used during the two project. In addition, further work on eliciting correlation could be carried out using the experience gathered during the two applied projects.

During this research project, the methodology was only applied to two case studies. Revisiting the cycle would mean that a number of further cases could be worked on. Increasing the number of applied cases the methodology has been successfully applied to would give the MOD added confidence that the methodology meets their requirements. This would also expose the methodology to an increased number of decision makers.

8.5.2 Increased Adjusted Variance

As identified in Section 2.4.5, within a Bayes linear framework, the adjusted variance must always decrease. It is possible that if the decision maker was to observe something they did not expect, then their adjusted variance would actually rise. It is necessary to investigate how this could be modelled using the Bayes linear framework. The simplest case to investigate is where X and D are variables such that $D = \alpha X + R$ and X and R are uncorrelated.

Using the Bayes linear formulas, that $var_D(X) = var(X) - cov(X, D)var^{-1}(X)cov(D, X)$. However one could also interpret $E_D((X - E(X))^2)$ as the variance of X once D has been observed. It can be shown that $E_D((X - E(X))^2) = E_D(X^2) - E_D(X)^2$. From this it is necessary to calculate $E_D(X^2)$. As $E_D(X^2) = E(X^2) + cov(X^2, D)var(D^{-1})(D - E(D))$, it is necessary to elicit $E(X^2)$ and $cov(X^2, D)$. Using the Pearson and Tukey method, $E(X^2)$ could be calculated. This is because the Pearson and Tukey method calculates the mean and the variance and from these two values, $E(X^2) = var(X) - E(X)^2$. Hence using the current method of elicitation, $E(X^2)$ could be gathered.

It is unrealistic to expect a decision maker to specify $cov(X^2, D)$. Instead, $cov(X^2, D)$ can be expanded to give

$$\begin{aligned}
cov(X^2, D) &= cov(X^2, \alpha X + R) \\
&= \alpha cov(X^2, X) + cov(X^2, R) \\
&= \alpha E((X^2 - E(X^2))(X - E(X))) \\
&= \alpha E(X3 - XE(X^2) - X^2 E(X) + E(X^2)E(X)) \\
&= \alpha[E(X^3) - E(X)E(X^2)]
\end{aligned}$$

In order to calculate $cov(X^2, D)$ it would be necessary to elicit $E(X^3)$. This research has not investigated the different methods that could be used to elicit higher moments such as $E(X^3)$. Papers that could be used as a sensible starting point to investigate this area further are Smith (1993) and Reilly (2002).

If it was possible to assess $E(X^3)$, then the Bayes linear model for X^2 could be built. If $E(X^2)$ was adjusted given the observation D,

$$= E_D(X^2) - E_D(X)^2$$
$$= E(X^2) + cov(X^2, D)var^{-1}(D)(D - E(D)) - [E(X) + cov(X, D)var^{-1}(D)(D - E(D))]^2$$

As the prior variance is $E(X^2) - E(X)^2$, where $cov(X^2, D)var^{-1}(D)(D-E(D)) > cov^2(X, D)var^{-2}(D)(D-E(D))^2$, the output of this method would be larger than the prior variance. In addition, if $D = E(D)$, i.e. the observed value is that which the decision maker expected to observe, then the $E_D(X^2) - E_D(X)^2$ is equal to $var(X)$.

8.5.3 Boundary of Model

During this research, the modelling carried out focused only on assessing how reliability issues supported decisions relating to the R&M Case. External factors such as budget, time and potential political issues were not captured explicitly in the modelling. The aim of the Bayes linear model in each project was to assess reliability as only the R&M decision maker was supported. It is likely that these external factors were either not impacting on the decision, or the decision maker was taking this information into account when making her final decision. Future application of the Bayes linear methodology could potentially incorporate different disciplines into the model including factors such as cost, time and performance; all of which will have an effect on the reliability of the system.

8.5.4 Non-Normal Variables

In both projects, it was assumed that all variables were approximately normal and that the relationship between these variables was approximately joint-normal. Because of this, it was deduced that the output of the model was approximately normal. This seems satisfactory when looking to approximate a higher specification probability model where the majority of variables are approximately normal. However, times may exist where the assumption of normality is dubious and an alternative is required. Skewness is measured as $s = \frac{E(X-\mu)^3}{\sigma_X^3}$. Where $|s| > a$, where a is some pre-determined value, the decision maker may believe that the assumption of normality is no longer valid.

In these situations, it is proposed that the variables are separated in such a way that the subsequent sub-variables are normal. This may require that a single variable is broken down into multiple sub-variables, i.e. $X = X_1 + X_2 + \ldots + X_n$. Adjustments are then made to these sub-variables and the adjusted expectation and variance of X is calculated using the adjusted expectation and variance of $X_1, \ldots, X_n$.

Further research is required to establish how this would be carried out in practice. More importantly, however, further investigation on establishing a suitable value for a is required. Establishing a sensible value for a would mean that additional confidence could be placed upon a model where all the variables have a skewness value less than a.

8.5.5 Elicitation from Multiple Experts

It is realistic to assume that in the future there may be an R&M Case where multiple decision makers are involved. For example, the MOD, and a number of other European defence agencies, are in the process of purchasing a number of A400 transport aircraft from Airbus. In this case, one may wish to incorporate into the analysis the expert opinion of each of the decision makers. Wisse et al (2007) propose a method for combining expert assessments of moments in a non-Bayesian way in order to construct a single prior belief structure.

8.5.6 Multivariate analysis

Chapter 6 addressed multivariate analysis to an extent as the analysis was carried out on four variables simultaneously. However, in both applied case studies, inference was only carried out on the variables that were of most interest to the decision maker. Further work in the future could focus on extending the analysis to include all the variables in the model. This would give the decision maker more detailed information about the problem they are facing.

8.5.7 Simplifying Procedures for Replication

If a decision maker chooses to build a Bayes linear model, then they must decide how they wish to assess the expectation and variance of each variable. This thesis has proposed a method to accomplish this. The decision maker must also assess the qualitative relationships between the different variables. One method of achieving this is to build a Bayes linear network as demonstrated in Section 2.4.3.1. In addition, where there exists dependency between two variables, a covariance value must be specified. A number of methods have been proposed and scenarios where each is most appropriate suggested. This research has attempted to standardise the way in which beliefs are elicited from the decision maker.

However, further research is needed to assess the accuracy of these methods and to evaluate alternative methods. In particular, further quantitative analysis of the four covariance methods is required to assess which of these methods is best at assessing the covariance under different circumstances. A number of dependencies could be identified such that the bias between the belief of a number of experts and reality is assumed to be equal to zero. Quantitative analysis could be carried out to assess which of the four methods elicits a value closest to the observed covariance value. From this it would be possible to devise a set of rules that the decision maker could follow to build and populate a Bayes linear model.

Using these rules, it would be simple to develop a Bayes linear software tool that could take limited information from the decision maker and populate the entire model. For example, assume that the Adjusted Expectation (AE) method of gathering the covariance is best under certain circumstances whilst Direct Calculation (DC) is best under other circumstances. The software would be able to ask a set of predetermined questions that could then decide which method is best to elicit the necessary covariance value.

Building software would mean checks could be introduced to ensure the decision maker is coherent in real-time, the necessary calculations could be minimised and the specification burden could be reduced. The software would expose the methodology to a larger number of decision makers and problems and would raise a number of further interesting research questions such as assessing the adequacy of the elicitation methods, evaluating the benefit of Bayes linear and investigating the potential of the methodology as a collaborative tool for customer and contractor.

Appendix A

Extended Findings of MOD Interviews

Section 1.5.3 summarised the information gathered in the MOD interviews. This appendix provides additional information regarding what was gathered to give additional context to the problem domain.

A.1 Evaluating the R&M Case

There was agreement across all three environments that evaluating the R&M Case was a subjective process. This subjectiveness could potentially lead to multiple evaluators drawing different conclusions on whether or not the R&M Case has demonstrated reliability requirements are likely to be met. There was almost universal agreement from the R&M Specialists across all environments that evaluating the R&M Case was a subjective process. One R&M Specialist stated, "It is highly possible that two people would come to two different opinions. It [the evaluation] is highly dependent on whoever is evaluating the case". At this point, there are no strict guidelines from the MOD as to how an R&M evaluator should evaluate the different information presented in the R&M Case.

As of yet, there has been little attempt to use quantitative methods to evaluate the information provided in the R&M Case. However, most of those interviewed believed that such techniques may have a place in supporting their decision making. One R&M specialist commented that he did not find this a difficult task but that the was "aware some people do find combining different forms of evidence a problem because it is not a defined process."

Some of the MOD evaluators found it difficult to communicate how they evaluate the R&M

Case. Comments such as "I get a feeling about whether or not it meets requirements" were not uncommon during interviews. What this feeling was or how they achieved this feeling could not be fully explained. It was clear that each evaluator has his or her own unique evaluation process. In addition, most of the the R&M evaluators also commented that they did not have as much time as they would have liked to work on evaluating an R&M Case.

One issue that makes evaluating the R&M Case difficult is the issue of burden of proof. There was almost unanimous agreement that, in theory at least, the burden of proof always lies on the contractor to demonstrate that reliability requirements have been met. However, there were a number of comments made during the interviews that appeared to contradict this. One R&M Specialist stated, 'At the end of the day, if a piece of equipment is needed immediately, as long as the item is safe, it will be accepted warts and all" whilst another R&M Specialist stated "If an R&M Case is delivered that says requirements have been met, that is correct until such a time that the MOD prove it is not". Both these statements appear to contradict the notion that the burden of proof truly lies on the contractor. Because of this, it means the MOD must have evidence a system is not meeting requirements rather than simply stating that they do not believe there is sufficient evidence in the R&M Case.

A.2 Prior Knowledge Required to Evaluate R&M Case

There was agreement across all three environments that in order to evaluate the R&M Case, the evaluator required some engineering knowledge of the given system. One R&M Specialist stated, "You do need some expert engineering knowledge but not a huge amount of deep detailed R&M knowledge is required to produce or evaluate an R&M Case." One Air R&M Specialist commented, "A mixture of project management and engineering knowledge is needed to produce a good R&M Case. These factors are equally as important as having an understanding of R&M issues."

One R&M Specialist raised an interesting issue regarding knowledge contained in R&M HOS. He commented that there is an abundance of people within the department with good engineering skills however there is a distinct lack of people with mathematical or statistical skills. As far as he knew, there was only one R&M HOS member that was capable of tackling mathematical or statistical reliability problems.

A.3 Decision Making

One R&M specialist made an interesting comment about the level of influence that the IPT has at the six different stages of the procurement process, CADMID, on reliability issues. He stated that once the contract has been signed between the MOD and contractor, the IPT has very little influence on reliability issues. This means that the decisions they make at the assessment phase are very important. The same R&M HOS member stated that from the design stage onwards there is not much point having an R&M point of contact on the project as they have very little influence and are performing an observing role rather than an active advising role.

In addition, one R&M specialist commented that another difficulty in evaluating the case is the fact that those evaluating the R&M Case may change during a procurement project. This makes it difficult for both the MOD and the contractor. For the MOD, the main difficulty arises because there is often no justification in the R&M Case as to why a certain task was chosen or dropped. Because of the subjective nature of the R&M Case, it may not be clear to a future decision maker why a certain path was chosen. For the contractor, it means that they may have to structure their arguments or evidence in a different way. For example, one MOD decision maker may place a lot of emphasis on modelling whilst another MOD decision maker may place of a lot of emphasis on softer issues, such as meeting certain build standards.

A.4 ISRD

In the past, an In-Service Reliability Demonstration was the principal method through which the MOD would determine whether or not reliability requirements had been met. The aim of an In-Service Reliability Demonstration (ISRD) is to demonstrate whether or not reliability requirements have been met by the system under in-service conditions. Due to increased reliability, longer testing times and more expensive systems, the ISRD, as a method of determining whether or not reliability requirement have been met, is no longer financially viable.

In terms of a decision-making context within the MOD, the ISRD is almost a pointless exercise. Once a system has been designed, developed and manufactured to the extent where it is ready to undergo an ISRD, only if there is a major failure of the system will the results of an ISRD make an impact on the decision of the IPTL on whether or not to accept or reject a system. There are two predominant reasons for this both due to the nature of the MOD and the military sector.

First, due to the nature of the industry, the MOD are forced into using a small number of specialised contractors. They have to maintain a good relationship with them and have an obligation not to place them under unnecessary financial pressure. If the MOD were to force contractors to carry out additional ISRD's when reliability requirements, this would place contractor's under severe financial pressure and may make them reconsider when submitting bids in the future. Second, re-designing a system and carrying out further ISRD's have a time implication. For some systems, they may be replacing a system currently in-service that is being retired on a given date, or filling a capability gap currently not being filled. On these occasions, it is more important to provide the Armed Forces with a system of reduced reliability performance rather than providing the Armed Forces with nothing at all. The strategy adopted by the MOD and contractors when reliability requirements are not being met is to provide the Armed Forces with the current system and to make a focused effort to increase reliability throughout the in-service life of the system.

Within the MOD, there are mixed feelings about the ISRD across the different environments and different personnel. Whilst most R&M personnel agree that there is no longer a need for a statistically justified ISRD on many systems, all of them state there is a need for some observed data to be collected before a decision can be made on whether or not the system will meet reliability requirements. One R&M Specialist stated that he believed that the contractors seem to have a similar point of view. He claimed it appears contractors are unwilling to have an R&M Case rejected without some observed data being collected.

Whilst the ISRD is no longer viable as a method for determining if reliability requirements have been met, it should not be seen as a redundant source of information. Along with other forms of information, a truncated ISRD should be seen as another form of information in helping the contractor provide assurance.

Appendix B

Time-line of Research

Interview MOD R&M Specialists	October 2004
Covariance Elicitation Sessions	October 2005
Identify projects to apply Bayes linear	November 2005
Applied Project 1: Elicitation	March 2006
Applied Project 2: Elicitation	March 2006
Applied Project 1: Feedback	April 2006
Applied Project 2: Feedback	September 2006
MOD Feedback Session	October 2006

Table B.1: Time-line of Research

Appendix C

Expectation Table and Covariance Matrix for Supplier B

Variables	Expectation
$\overline{X_B}$	1000
$\overline{ISRD}$	1000
$\overline{PA}$	1000
$\overline{PT}$	1300
$\overline{PT_i}$	1300
$\overline{RBD}$	1218.5
$\overline{RBD_I}$	1218.5
$\overline{RBD_2}$	1437
$\overline{RBD_{2_I}}$	1437

Table C.1: Expectation table for Supplier B

$\overline{X_B}$	236509								
$\overline{ISRD}$	236509	237433							
$\overline{PA}$	236509	236509	768655						
$\overline{PT}$	236509	236509	236509	269768					
$\overline{PT_i}$	236509	236509	236509	269768	273463				
$\overline{RBD}$	236509	236509	236509	236509	236509	444378			
$\overline{RBD_I}$	236509	236509	236509	236509	236509	444378	444526		
$\overline{RBD_2}$	236509	236509	236509	236509	236509	444378	444378	652248	
$\overline{RBD_{2_I}}$	236509	236509	236509	236509	236509	444378	444378	655248	652396

Table C.2: Covariance matrix for supplier B

Appendix D

Bayes Linear Networks for Each Contractor

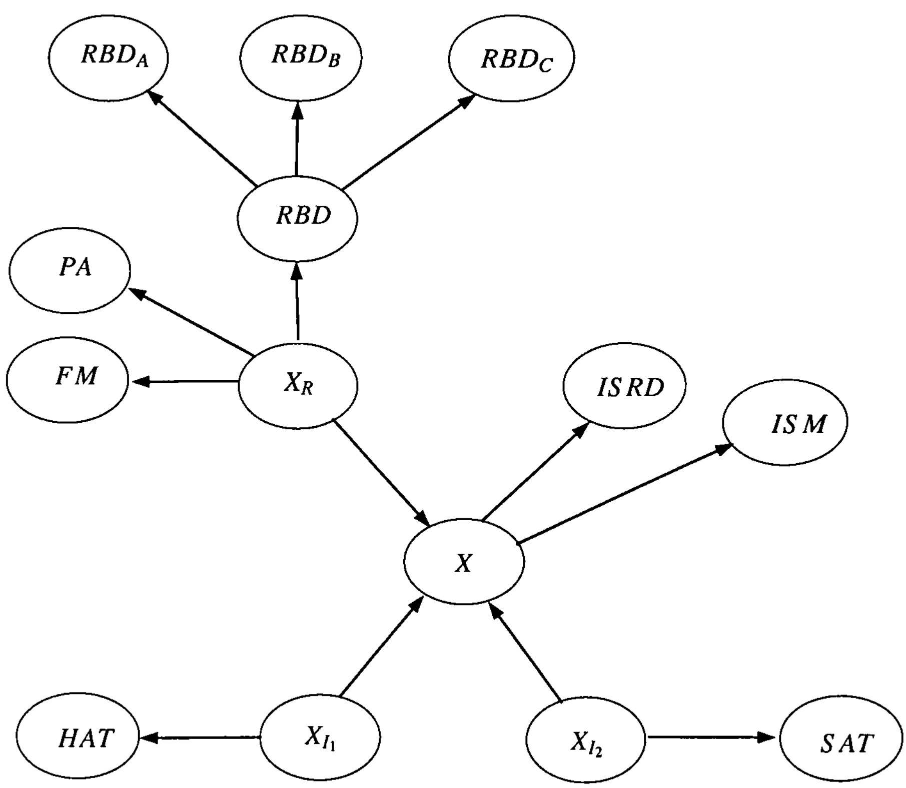

Figure D.1: Bayes linear network of contractor Blue's R&M programme

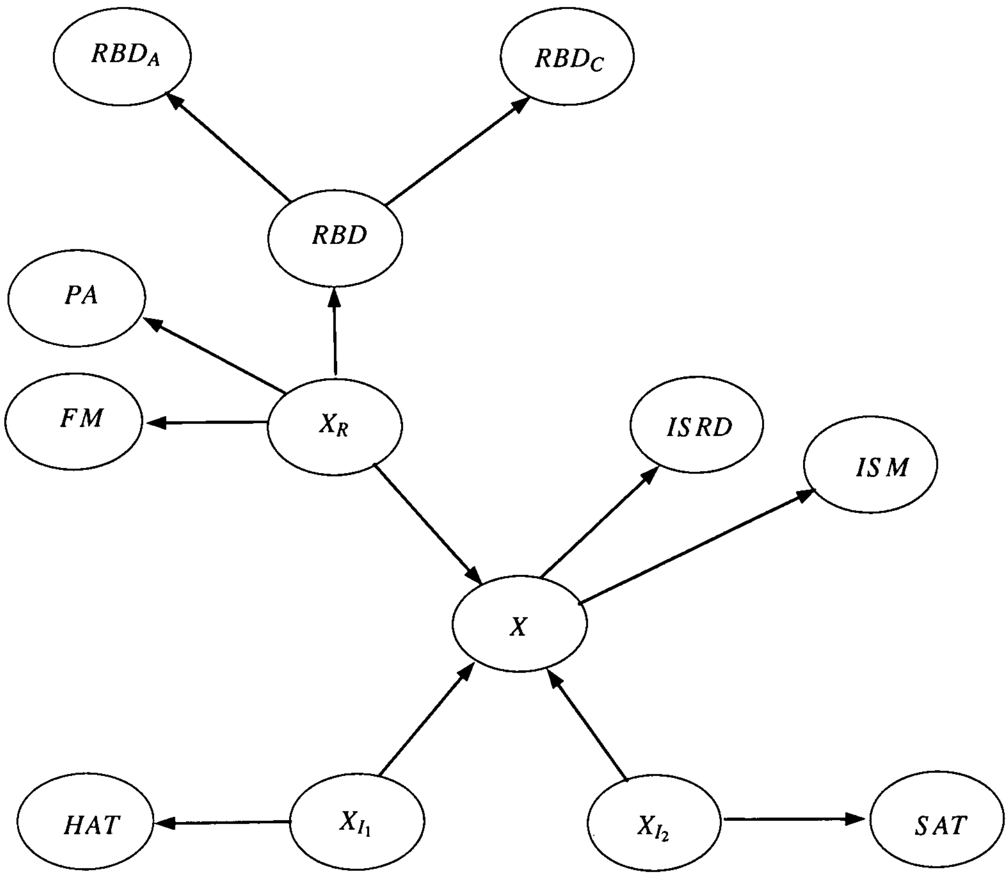

Figure D.2: Bayes linear network of contractor Yellow's R&M programme

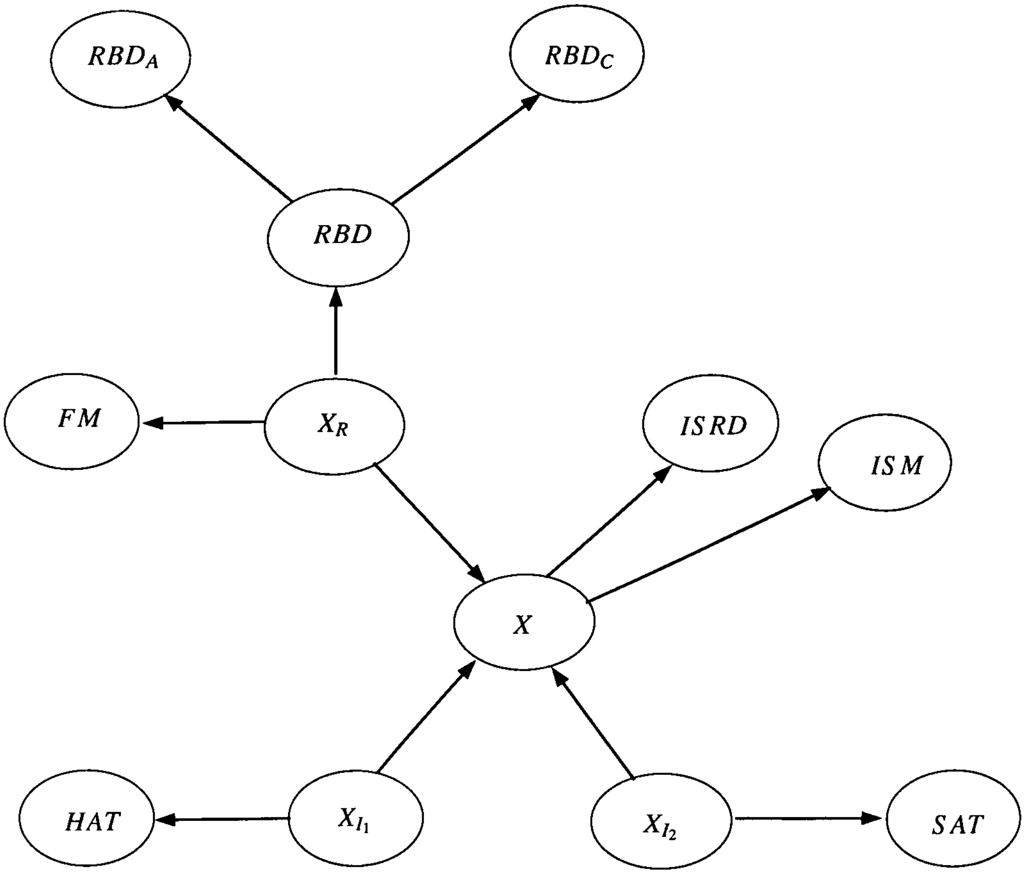

Figure D.3: **Bayes linear network of contractor Grey's R&M programme**

Appendix E

Expectation Table and Covariance Matrix for Each Contractor

Variable	Expectation
X_R	1675.75
X_{I_1}	0
X_{I_1}	0
X	1675.75
HAT	0
SAT	0
$ISRD$	1675.75
ISM	1675.75
PA	2005.25
RBD	1675.75
RBD_A	1766.5
RBD_C	1775.75

Table E.1: Yellow's expectations for all variables

Variable	Expectation
X_R	1201.75
X_{I_1}	0
X_{I_1}	0
X	1201.75
HAT	0
SAT	0
$ISRD$	1201.75
ISM	1201.75
PA	1786.75
RBD	1201.75
RBD_A	1401.75
RBD_B	1775.75
RBD_C	1901.75

Table E.2: Blue's expectations for all variables

Variable	Expectation
X_R	1201.75
X_{I_1}	0
X_{I_1}	0
X	1201.75
HAT	0
SAT	0
$ISRD$	1201.75
ISM	1201.75
RBD	1201.75
RBD_A	1401.75
RBD_C	1683.25

Table E.3: Grey's expectations for all variables

Variable													
X_R	554605.9												
X_{I_1}	0	106.1											
X_{I_2}	0	0	149.7										
X	554605.9	106.1	149.7	554861.7									
HAT	0	106.1	0	106.1	106.1								
SAT	0	0	149.7	149.7	0	149.7							
$ISRD$	554605.9	106.1	149.7	554861.7	106.1	149.7	569642.6						
ISM	554605.9	106.1	149.7	554861.7	106.1	149.7	554861.7	587620.8					
PA	554605.9	0	0	554605.9	0	0	554605.9	554605.9	1043371.1				
RBD	554605.9	0	0	554605.9	0	0	554605.9	554605.9	554605.9	587865.9			
RBD_A	554065.9	0	0	554065.9	0	0	554605.9	554605.9	554605.9	587865.9	606570.3		
RBD_C	554605.9	0	0	554605.9	0	0	5546505.9	5546059	587865.9	587865.9	587865.9	621124.2	
FM	430.2	0	0	430.2	0	0	430.2	430.2	430.2	430.2	430.2	430.2	1

Table E.4: Yellow's covariance matrix for all variables

Variable														
X_R	467739													
X_{I_1}	0	106.1												
X_{I_2}	0	0	149.7											
X	467739	106.1	149.7	467792										
HAT	0	106.1	0	106.1	106.1									
SAT	0	0	149.7	149.7	0	149.7								
$ISRD$	467739	106.1	149.7	467792	106.1	149.7	482773							
ISM	467739	106.1	149.7	467792	106.1	149.7	467992	501251						
PA	467739	0	0	467739	0	0	467739	467739	624592					
RBD	467739	0	0	467739	0	0	467739	467739	467739	500998				
RBD_A	467739	0	0	467739	0	0	467739	467739	467739	500998	560125			
RBD_B	467739	0	0	467739	0	0	467739	467739	467739	500998	500998	870568		
RBD_C	467739	0	0	467739	0	0	467739	467739	467739	500998	500998	500998	1332476	
FM	366	0	0	366	0	0	366	366	366	366	366	366	366	1

Table E.5: Blue's covariance matrix for all variables

Variable												
X_R	467739											
X_{I_1}	0	106.1										
X_{I_2}	0	0	149.7									
X	467739	106.1	149.7	467992								
HAT	0	106.1	0	106.1	106.1							
SAT	0	0	149.7	149.7	0	149.7						
$ISRD$	467739	106.1	149.7	467992	106.1	149.7	482772					
ISM	467739	106.1	149.7	467992	106.1	149.7	467992	501250				
RBD	497739	0	0	467739	0	0	467739	467739	554605.9			
RBD_A	497739	0	0	467739	0	0	467739	467739	500998	587865.9		
RBD_C	497739	0	0	467739	0	0	467739	467739	500998	500998	524242	
FM	366	0	0	366	0	0	366	366	366	366	366	1

Table E.6: Grey's covariance matrix for all variables

Appendix F

Expectation Tables and Covariance Matrix for Reliability Growth Trials

Variable	Expectation
A_{1B}	0.0563
A_{1Bi}	0.0563
A_{2B}	0.13335
A_{2Bi}	0.13335
$GA_{1,2B}$	0.07705
$EGA_{1,2B}$	0.07705
A_{3B}	0.2204
A_{3Bi}	0.2204
$GA_{2,3B}$	0.08705
$EGA_{2,3B}$	0.08705
A_{IB}	0.2204
$GA_{3,IB}$	0
$EGA_{3,IB}$	0

Table F.1: Expectation for Vehicle A's Basic Reliability

Variable													
A_{1B}	0.000343												
A_{1Bi}	0.000343	0.000343											
A_{2B}	0.000343	0.000343	0.00063										
A_{2Bi}	0.000343	0.000343	0.00063	0.001554									
$GA_{1,2B}$	0	0	0.000288	0.000288	0.000288								
$EGA_{1,2B}$	0	0	0.000288	0.000288	0.000288	0.000435							
A_{3B}	0.000343	0.000343	0.00063	0.00063	0.000288	0.000288	0.001842						
A_{3Bi}	0.000343	0.000343	0.00063	0.00063	0.000288	0.00028	0.001842	0.002766					
$GA_{2,3B}$	0	0	0	0	0	0	0.000288	0.000288	0.000288				
$EGA_{2,3B}$	0	0	0	0	0	0	0.000288	0.000288	0.000288	0.000435			
A_{IB}	0.000343	0.000343	0.00063	0.00063	0.000288	0.000288	0.001842	0.001842	0.000288	0.0002888	0.00369		
$GA_{3,IB}$	0	0	0	0	0	0	0	0	0	0	0.000924	0.000924	
$EGA_{3,IB}$	0	0	0	0	0	0	0	0	0	0	0.000924	0.000924	0.001072

Table F.2: Covariance Matrix for Vehicle A's Basic Reliability

Variable	Expectation
A_{1M}	0.6
A_{1Mi}	0.6
A_{2M}	0.62815
A_{2Mi}	0.62815
$GA_{1,2M}$	0.02815
$EGA_{1,2M}$	0.02815
A_{3M}	0.6852
A_{3Mi}	0.6852
$GA_{2,3M}$	0.05705
$EGA_{2,3M}$	0.05705
A_{IM}	0.6852
$GA_{3,IM}$	0
$EG_{3,IM}$	0

Table F.3: Expectation for Vehicle A's Mission Reliability

Variable													
A_{1M}	0.003695												
A_{1Mi}	0.003695	0.007391											
A_{2M}	0.003695	0.003695	0.005287										
A_{2Mi}	0.003695	0.003695	0.005287	0.008983									
$GA_{1,2M}$	0	0	0.001592	0.001592	0.001592								
$EGA_{1,2M}$	0	0	0.001592	0.001592	0.001592	0.005287							
A_{3M}	0.003695	0.003695	0.005287	0.005287	0.001592	0.001592	0.009271						
A_{3Mi}	0.003695	0.003695	0.005287	0.005287	0.001592	0.001592	0.009271	0.012966					
$GA_{2,3M}$	0	0	0	0	0	0	0.000288	0.000288	0.000288				
$EGA_{2,3M}$	0	0	0	0	0	0	0.000288	0.000288	0.000288	0.000435			
A_{IM}	0.003695	0.003695	0.005287	0.005287	0.001592	0.001592	0.009271	0.009271	0.000288	0.0002888	0.01389		
$GA_{3,IM}$	0	0	0	0	0	0	0	0	0	0	0.000924	0.000924	
$EGA_{3,IM}$	0	0	0	0	0	0	0	0	0	0	0.000924	0.000924	0.001072

Table F.4: Covariance Matrix for Vehicle A's Mission Reliability

Variable	
B_{1B}	0.1163
B_{1Bi}	0.1163
B_{2B}	0.1689
B_{2Bi}	0.1689
$GB_{1,2B}$	0.0526
$EGB_{1,2B}$	0.0526
B_{3B}	0.2215
B_{3Bi}	0.2215
$GB_{2,3B}$	0.0526
$EGB_{2,3B}$	0.0526
B_{IB}	0.2215
$GB_{3,IB}$	0
$EGB_{3,IB}$	0

Table F.5: Expectation for Vehicle B's Basic Reliability

Variable													
B_{1B}	0.000343												
B_{1Bi}	0.000343	0.000343											
B_{2B}	0.000343	0.000343	0.000408										
B_{2Bi}	0.000343	0.000343	0.000408	0.001332									
$GB_{1,2B}$	0	0	0.000065	0.000065	0.000065								
$EGB_{1,2B}$	0	0	0.000065	0.000065	0.000065	0.000213							
B_{3B}	0.000343	0.000343	0.000408	0.000408	0.000065	0.000065	0.001397						
B_{3Bi}	0.000343	0.000343	0.000408	0.000408	0.000065	0.000065	0.001397	0.002321					
$GB_{2,3B}$	0	0	0	0	0	0	0.000065	0.000065	0.000065				
$EGB_{2,3B}$	0	0	0	0	0	0	0.000065	0.000065	0.000065	0.000213			
B_{IB}	0.000343	0.000343	0.000408	0.000408	0.000065	0.000065	0.001397	0.001397	0.000065	0.000065	0.003245		
$GB_{3,IB}$	0	0	0	0	0	0	0	0	0	0	0.000924	0.000924	
$EGB_{3,IB}$	0	0	0	0	0	0	0	0	0	0	0.000924	0.000924	0.001072

Table F.6: Covariance Matrix for Vehicle B's Basic Reliability

Variable	Expectation
B_{1M}	0.65
B_{1Mi}	0.65
B_{2M}	0.69075
B_{2Mi}	0.69075
$GB_{1,2M}$	0.04075
$EGB_{1,2M}$	0.04075
B_{3M}	0.74445
B_{3Mi}	0.74445
$GB_{2,3M}$	0.0537
$EGB_{2,3M}$	0.0537
B_{IM}	0.74445
$GB_{3,IM}$	0
$EGB_{3,IM}$	0

Table F.7: Expectation for Vehicle B's Mission Reliability

Variable													
A_{1M}	0.003695												
A_{1Mi}	0.003695	0.007391											
A_{2M}	0.003695	0.003695	0.004403										
A_{2Mi}	0.003695	0.003695	0.004403	0.008099									
$GA_{1,2M}$	0	0	0.000708	0.000708	0.000708								
$EGA_{1,2M}$	0	0	0.000708	0.000708	0.000708	0.000856							
A_{3M}	0.003695	0.003695	0.004403	0.004403	0.000708	0.000708	0.008247						
A_{3Mi}	0.003695	0.003695	0.004403	0.004403	0.000708	0.000708	0.008247	0.011942					
$GA_{2,3M}$	0	0	0	0	0	0	0.000148	0.000148	0.000148				
$EGA_{2,3M}$	0	0	0	0	0	0	0.000148	0.000148	0.000148	0.000296			
A_{IM}	0.003695	0.003695	0.004403	0.004403	0.000708	0.000708	0.008247	0.008247	0.000148	0.000148	0.01389		
$GA_{3,IM}$	0	0	0	0	0	0	0	0	0	0	0.000924	0.000924	
$EGA_{3,IM}$	0	0	0	0	0	0	0	0	0	0	0.000924	0.000924	0.001072

Table F.8: Covariance Matrix for Vehicle B's Mission Reliability

Vehicle	Vehicle	Covariance
A_B	A_M	0.000078
A_B	B_B	0.000235
A_M	B_M	0.0025868
B_B	B_M	0.000070

Table F.9: Covariance Values Between Different Vehicles and Reliability Measurements

Variable	Expectation
R_A	0
R_N	0
R_M	0
R_B	0

Table F.10: Expectation for Proxy Variables

Appendix G

Questionnaire for Feedback Session

Bayes linear Evaluation Questionnaire

I would be grateful if you could take a few minutes to jot down your thoughts on some issues regarding the Bayes linear methodology. Please do not simply answer yes or no. Any information gathered will be used to improve the current process. Your responses to this questionnaire may be used anonymously and are for research evaluation purposes only.

Name and Role:

1. How easy/difficult did you find the overall process?

    - Building the network
    - Specifying your beliefs
    - Using the software
    - Interpreting the output

2. Would you feel comfortable using the Bayes linear methodology in the future?

3. How does the Bayes linear process compare to your current process? Do you think it could support your current process?

4. Was the output of the model useful in supporting your decision making?

5. Does building the model support your decision making? Could it aid discussion with the contractor?

6. What sort of problems could you visualise using the Bayes linear process for in the future, if any? If not any, why not?

Bibliography

[1] Albino, V. and A. C. Garavelli (1998). "A neural network application to subcontractor rating in construction firms." International Journal of Project Management 16(1): 9-14.

[2] Ansell, J., J. Quigley and L. Walls (1999). "Achieving growth in reliability." Annals of Operations Research 91: 11-24.

[3] Argyris, C. (1970). Intervention theory and method. Reading, MA, Addison-Wesley.

[4] Argyris, C., R. Putnam and D. M. Smith (1985). Action science. San Francisco, Jossey-Bass.

[5] Aven, T. (2003). Foundations of Risk Analysis. Chichester, John Wiley & Sons Ltd.

[6] Aven, T., J. E. Vinnem and H. S. Wiencke (2007). "A decision framework for risk management, with application to the offshore oil and gas industry." Reliability Engineering and System Safety 92: 433-448.

[7] Azondekon, S. H. and J.-M. Martel (1999). "'Value' of additional information in multicriterion analysis under uncertainty." European Journal of Operational Research 117: 45-62.

[8] Bargal, D. (2006). "Personal and intellectual influences leading to Lewins paradigm of action research." Action Research 4(4): 367-388.

[9] Barlow, R., E. (1998). Engineering Reliability. Philadelphia, SIAM.

[10] Barlow, R., E. and F. Proschan (1965). The Mathematical Theory of Reliability. New York, John Wiley & Sons.

[11] Barlow, R., E. and F. Proschan (1975). Statistical Theory of Reliability and Life Testing. Holt, Winston, Rinehart.

[12] Barnett, N. and R. Denning (2004). "The R&M Case - Practical Experience". Assembling evidence of reliability: The reliability case as a decision support tool, Glasgow.

[13] Bateson, G. (1972). Steps to an Ecology of Mind. New York, Ballantine.

[14] Bayes, T. (1763). "An Essay Towards Solving a Problem in the Doctrine of Chances." Philosophy Transactions of the Royal Society 61(53): 370-418.

[15] Bazovsky, I. (1961). Reliability Theory and Practice. Engelwood Cliffs, N.J., Prentice-Hall.

[16] Bedford, T. (2004). "The "Value of Information" in a Reliability Programme". Proceedings of the 27th ESREA Seminar, Glasgow.

[17] Bedford, T. and R. Cooke (2001). Probabilistic Risk Analysis: Foundations and Methods, Cambridge University Press.

[18] Bedford, T., J. Quigley and L. Walls (2006). "Expert Elicitation for Reliable System Design." Statistical Science 21(4): 428-450.

[19] Bendell, A. (1991). Reliability. Reliability Data Banks. A. G. Cannon and A. Bendell. London, Elsevier Science Publishers.

[20] Berger, J. (2006). "The Case for Objective Bayesian Analysis." Bayesian Analysis 1(3): 385-402.

[21] Bierbaum, R. L., T. D. Brown and T. J. Kerschen (2002). "Model-based reliability analysis." IEEE Transactions on Reliability 51(2): 133-140.

[22] Box, G. and N. Draper (1987). Empirical Model Building and Response Surfaces John Wiley & Sons.

[23] Brydon-Miller, M., D. Greenwood and P. Macguire (2003). "Why action research?" Action Research 1(1): 9-28.

[24] BS4778 Glossary of Terms Used in Quality Assurance Inclduing Reliability and Maintainability Terms. B. S. Insitution. London.

[25] BS5760 Reliability of Constructed or Manufactured Products, Systems, Equipment and Components, Introductory Guide to Reliability. B. S. Institution, London.

[26] Chandler, D. and B. Torbert (2003). "Transforming inquiry and action." Action Research 1(2): 133-152.

[27] Chaudhuri, G., K. Hu and N. Afshar (2001). "A new approach to system reliability." IEEE Transactions on Reliability 50(1): 75-84.

[28] Clemen, R. T., G. W. Fischer and R. L. Winkler (2000). "Assessing Dependence: Some Experimental Results." Management Science 46(8): 1100-1115.

[29] Clemen, R. T. and T. Reilly (2001). Making Hard Decisions with Decision Tools. Pacific Grove, CA, Duxbury Press.

[30] Condra, L. W. (1993). Reliability Improvement with Design of Experiments. New York, Marcel Dekker.

[31] Cooke, B. and J. W. Cox (2005). Fundamentals of action research. London, Sage.

[32] Cooke, R. M. (1991). Experts in uncertainty. Opinion and subjective probability in science. New York, Oxford University Press.

[33] Coppola, A. (1984). "Reliability engineering of electronic equipment: a historical perspective." IEEE Transactions on Reliability 33(1): 29-35.

[34] Coughlan, P. and D. Coghlan (2002). "Action research for operations management " International Journal of Operations and Production Management 22(2): 220-240.

[35] Coupe, V. M. H. and L. C. van der Gaag (2002). "Properties of Sensitivity Analysis of Bayesian Belief Networks " Annals of Mathematics and Artificial Intelligence 36(4): 323-356.

[36] Cowell, R. G., A. P. Dawid, S. L. Lauritzen and D. J. Spiegelhalter (1999). Probabilistic Networks and Expert Systems. New York, Springer.

[37] Cox, D. and D. Hindley (1974). Theoretical Statistics, Chapman and Hall.

[38] Craig, P. S., M. Goldstein, J. C. Rougier and A. H. Seheult (2001). "Bayesian Forecasting for Complex Systems Using Computer Simulations." Journal of American Statistical Association 96(454): 717-729.

[39] Craig, P. S., M. Goldstein, A. H. Seheult and J. A. Smith (1996). Bayes Linear Strategies for Matching Hydrocarbon Resevoir History. Bayesian Statistics 5. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Oxford University Press: 69-95.

[40] Crow, L. H. (1974). "Reliability analysis for complex repairable systems." Reliability and Biometry: 379-410.

[41] Dakins, M. E. (1999). "The Value of the Value of Information." Human and Ecological Risk Assessment 5(2): 281-289.

[42] De Boer, L., E. Labro and P. Morlachhi (2001). "A review of methods supporting supplier selection." European Journal of Purchasing & Supply Management 7: 75-89.

[43] De Finetti, B. (1974). Theory of probability: a critical introductory treatment John Wiley & Sons Ltd.

[44] Denning, R. (2003). Three years experience of the R&M Case. Safety and Reliability, Swets and Zeitlinger.

[45] Denson, W. (1998). "The History of Reliability Prediction." IEEE Transactions on Reliability 47(3): 321-328.

[46] Dery R., M. Landry and C. Banville (1993). "Revisiting the issue of model validation in OR: An epistemological view" European Journal of Operational Research, 66 (2) 168-183

[47] DeTardo-Bora, K. (2004). "Action Research in a world of positivist-oriented review boards." Action Research 2(3): 237-253.

[48] Dick, B. (2004). "Action research literature: Themes and Trends." Action Research 2(4): 425-444.

[49] Donelson, J. (1975). Duane's Reliability Growth Model as a Nonhomogenous Poisson Process, Institute for Defense Analyses, Arlington, Virgiana, Paper P-1162.

[50] Draper, D. (2006). "Coherence and calibration." Bayesian Analysis 1(3): 423-428.

[51] Duane, J. T. (1964). "Learning curve approach to reliability modelling." IEEE Transactions on Aerospace 2(2): 563-565.

[52] Earl-Slater, A. (2002). "Critical appraisal of clinical trials - The superiority of action research?" British Journal of Clinical Governance 7(2): 132-135.

[53] Ebeling, C., E. (1997). An Introduction to Reliability and Maintainability Engineering. New York, McGraw-Hilll.

[54] Eden, C. and C. Huxham (1996). Researching organization using action research. Handbook of Organization Studies. S. Clegg, C. Hardy, W. Word and T. Lawrence. Beverly Hills, Sage.

[55] Eden, C. and C. Huxham (2002). Action Research. Essential Skills for Management Research. D. Partington. London, Sage.

[56] Efron, B. (1986). "Why Isn't Everyone a Bayesian?" The American Statistician 40(1): 1-11.

[57] Elden, M. and R. F. Chisolm (1993). "Emergent variteties of action research: introduction to the special issue." Human Relations 46(2): 121-142.

[58] Ellram, L. M. (1994). "A taxonomy of total cost of ownership models." Journal of Business Logistics 15 (1): 171-192.

[59] Farrow, M., M. Goldstein and T. Spiropoulos (1997). Developing a Bayes linear decision support system for a brewery. The practice of Bayesian analysis. S. French and J. Q. Smith, Edward Arnold.

[60] Fenton, N. and M. Neil (2001). "Making decisions: using Bayesian nets and MCDA." Knowledge-Based Systems 14: 307-325.

[61] Fenton, N. and M. Neil (2006). "Comment: Expert Elicitation for Reliable System Design." Statistical Science 21(4): 451-453.

[62] Fenton, N., M. Neil, S. Forey and H. R. (2001). "Using Bayesian belief networks to predict the reliability of military vehicles." Computing & Control Engineering Journal 12(1): 11-20.

[63] Ferson, S., C. A. Joslyn, J. C. Helton, W. L. Oberkampf and K. Sentz (2004). "Summary from the epistemic uncertainty workshop: consensus and diversity." Reliability Engineering and System Safety 85: 355-369.

[64] Fienberg, S. E. (2006). "Does it Make Sense to be an Objectve Bayesian?" Bayesian Analysis 1(3): 429-432.

[65] Freire, P. (1972). Pedagogy of the Oppressed. Harmondsworth, Penguin.

[66] Fussell, J. B. (1984). "Nuclear Power System Reliability: A Historical Perspective." IEEE Transactions on Reliability 33(1): 41-46.

[67] Gandy, A., P. Jager, B. Bertsche and U. Jensen (2007). "Decision support in early development phases - A case study from machine engineering." Reliability Engineering and System Safety 92: 921-929.

[68] Garthwaite, P. H., J. B. Kadane and A. O'Hagan (2005). "Statistical Methods for Eliciting Probability Distributions." Journal of American Statistical Association 100(470): 680-700.

[69] Garthwaite, P. H. and A. O'Hagan (2000). "Quantifying expert opinion in the UK water industry: an experimental study." The Statistician 49(4): 455-477.

[70] Glendon, A. I. and N. A. Stanton (2000). "Perspectives on safety cultures." Safety Science 34: 193-214.

[71] Gokhale, D. V. and S. J. Press (1982). "Assessment of a Prior Distribution for the Correlation Coefficient in a Bivariate Normal Distribution." Journal of the Royal Statistcal Society A 145: 237-249.

[72] Goldstein, M. (1975). "A note on some bayesian nonparametric estimates." The Annals of Statistics 3(3): 736-740.

[73] Goldstein, M. (1983). "The Prevision of a Prevision." Journal of American Statistical Association 78(384): 817-819.

[74] Goldstein, M. (1984). "Turning probabilities into expectations." The Annals of Statistics 12(4): 1551-1557.

[75] Goldstein, M. (1986). "Exchangeable Belief Structures." Journal of the American Statistical Association 81(396): 971-976.

[76] Goldstein, M. (1990). "Influence and Belief Adjustment. Influence Diagrams, Belief Nets and Decision Analysis. R. M. Oliver and J. Q. Smith. Chichester, Wiley: 143-174.

[77] Goldstein, M. (1997). Prior inferences for posterior judgements. Structures and norms in Science, M.C.D. C. e. al, Kluwer: 55-71.

[78] Goldstein, M. (1998). "Adjusting Belief Structures." Journal of the Royal Statistcal Society B 50(1): 133-154.

[79] Goldstein, M. (2001). "Avoiding foregone conclusions: geometric and foundational analysis of paradoxes of finite additivity." Journal of Statistical Planning and Inference 94: 73-87.

[80] Goldstein, M. (2006). "Subjective Bayesian Analysis: Principles and Practice." Bayesian Analysis 1(3): 403-420.

[81] Goldstein, M. and T. Bedford (2006). "The Bayes linear approach to inference and decision-making for a reliability programme." Reliability Engineering and System Safety 92(10): 1344-1352.

[82] Goldstein, M. and D. Wooff (1994). "Robustness measures for Bayes linear analyses." Journal of Statistical Planning and Inference 40(2-3): 261-277.

[83] Goldstein, M. and D. Wooff (2007). Bayes Linear Statistics: Theory and Methods. Chichester, UK, John Wiley & Sons.

[84] Green, A. E. and A. J. Bourne (1972). Reliability Technology. New York, Wiley-Interscience.

[85] Gustaven, B. (2003). "New forms of knowledge production and the role of action research." Action Research 1(2): 153-164.

[86] Haimes, Y. Y. and J. L. Lambert (1999). "When and How Can You Specify a Probability Distribution When You Don't Know Much? ." Risk Analysis 19(1): 43-68.

[87] Hanea, D., R. Cooke and B. Ale (2006). The Methodology to Build the Network used in a Bayesian Belief Net Approach. International Conference on Probabilistic Safety Assessment and Management, New Orleans, LA.

[88] Hayes-Roth, B. (1980). Estimation of time requirements during planning: interactions between motivation and cognition. Santa Monica, California, Rand Corporation Report.

[89] Heckerman, D. E., E. J. Horvitz and B. N. Nathwani (1992). "Toward normative expert systems. Part 1: The Pathfinder projcet." Methods of Information in Medicine 31: 90-105.

[90] Henney, K. (1956). Reliability factors for ground electronic equipment.

[91] Hogarth, R. M. (1980). Judgement and Choice: The Psychology of Decisions. Chicago, Illinois, Wiley-Interscience.

[92] Holt, G. D. (1998). "Which contractor selection methodology?" International Journal of Project Management 16(3): 153-164.

[93] House of Commons Defence Committee, (2003-04). Defence Procurement Report 572-I.

[94] Howard, R. A. (1990). From Influence to Relevance to Knowledge. Influence Diagrams, Belief Nets and Decision Analysis. R. M. Oliver and J. Q. Smith. Chichester Wiley: 3-23.

[95] Hu, Q. P., M. Xie, S. H. Ng and G. Levitin (2007). "Robust recurrent neural network modeling for software fault detection and correction prediction." Reliability Engineering and System Safety 92: 332-340.

[96] Huang, C.-Y., S.-Y. Kuo and M. Lyu, R. (2007). "An Assessment of Testing-Effort Dependent Software Reliability Growth Models." IEEE Transactions on Reliability 56(2): 198-211.

[97] Huxham, C. (2003). "Action research as a methodology for theory development." Policy and Politics 2: 239-248.

[98] Ireson, W. G., C. F. Coombs Jr and R. Y. Moss (1996). Handbook of Reliability Engineering and Management. New York, McGraw-Hill.

[99] Jefferis, T. (2004). The Drive for Progressive Assurance within MOD. Assembling evidence of reliability: The reliability case as a decision support tool, Glasgow.

[100] Jensen, F. V. (1999). An introduction to Bayesian Belief Networks. London, UCL Press Limited.

[101] Johnson, D. (2002). "Triangular approximations for continuous random variables in risk analysis." Journal of the Operational Research Society 53: 457-467.

[102] Johnston, W., J. Quigley and L. Walls (2006). "Optimal allocation of reliability tasks to mitigate faults during system development." IMA Journal of Management Mathematics 17(2): 159-169.

[103] Kadane, J. B. (2006). "Is "Objective Bayesian Analysis" objective, Bayesian, or wise?" Bayesian Analysis 1(3): 433-436.

[104] Kadane, J. B. and L. J. Wolfson (1998). "Experiences in elicitation." The Statistician 47(1): 3-19.

[105] Kahneman, D., P. Slovic and A. Tversky (1982). Judgment under Uncertainty: Heuristics and Biases Cambridge, Cambridge University Press.

[106] Kanoun, K., M. Kaaniche and J.-P. Laprie (1997). "Qualitative and quantitative reliability assessment." IEEE Software 14(2): 77-87.

[107] Kaplan, S., G. D. M. Cunha, A. A. Dykes and D. Shaver (1990). A bayesian methodology for assessing reliability during product development. Annual Reliability and Maintainability Symposium, Los Angeles, CA, USA.

[108] Kass, R. E. (2006). "Kinds of Bayesians." Bayesian Analysis 1(3): 437-440.

[109] Keefer, D. L. and S. Bodily, E. (1983). "Three-point approximatinos for continous random variables." Management Science 29(5): 595-610.

[110] Keefer, D. L. and W. A. Verdini (1993). "Better estimation of PERT activity time parameters." Management Science 39: 1086-1091.

[111] Kelly, T. P. and J. A. McDermid (2001). "A systematic approach to safety case maintenance." Reliability Engineering and System Safety 71: 271-284.

[112] Kerscher III, W. J., J. M. Booker, M. A. Meyer and T. R. Bennett (2000). Chareterizing Reliability During a Product Development. ESREL, Edinburgh, Scotland.

[113] Kim, J. B., B. F. Hobbs and J. F. Koonce (2003). "Multicriterion Bayesian Analysis of Lower Trophic Level Uncertainties and Value of Research in Lake Erie." Human and Ecological Risk Assessment 9(4): 1023-1057.

[114] Kock, N., R. J. McQueen and J. L. Scott (1997). "Can Action Research be Made More Rigorous in a Positivist Sense? The Contribution of an Iterative Approach." Journal of Systems and Information Technology 1(1): 1-24.

[115] Koehler, D. J. and N. Harvey (2004). Blackwell Handbook of Judgement and Decision Making. Oxford, UK, Blackwell Publishers.

[116] Kolmogorov, A. N. (1956). Kolmogorov: Foundations of the Theory of Probability. New York, Chelsea Publishing Company.

[117] Lad, F. (2006). "Objective Bayesian Statistics ... Do you buy it? Should we sell it?" Bayesian Analysis 1(3): 441-444.

[118] Landry, M. and M. Oral (1993). "In search of a valid view of model validation for operations research." European Journal of Operational Research 66: 161-167.

[119] Langseth, H. and L. Portinale (2007). "Bayesian Networks in Reliability." Reliability Engineering and System Safety 92: 92-108.

[120] Lauritzen, S. L. (1996). Graphical Models. Oxford, Oxford Science Publications.

[121] Lewin, K. (1946). "Action research and minority problems." Journal of Social Issues 2: 34-46.

[123] Lewin, K. (1947). "Frontiers in group dynamics I." Human Relations 1(1): 5-41.

[123] Lewin, K. (1947). "Frontiers in group dynamics II." Human Relations 1(2): 143-153.

[124] Lewin, K. (1951). Field Theory in Social Science. New York, Harper.

[125] Lindley, D. V. (2000). "The philosophy of statistics." The Statistician 49(3): 293-337.

[126] Lindley, D. V. (2006). Understanding Uncertainty. New Jersey, John Wiley & Sons.

[127] Littlewood, B. (1984). "Rationale for a Modified Duane Model." IEEE Transactions on Reliability 33(2): 157-159.

[128] Littlewood, B. and J. L. Verrall (1973). "A Bayesian Reliability Growth Model for Computer Software." Applied Statistician 22: 332-346.

[129] Loschi, R. and S. Wechsler (2002). "Coherence Bayes theorem and posterior distributions." Brazilian Journal of Probability and Statistics 16: 169-185.

[130] Malcolm, D. G., J. H. Roseboom, C. E. Clark and W. Fazar (1959). "Application of a technique for research and development program evaluation." Operations Research 5: 646-669.

[131] Mann, N. R. S. (1974). Methods for statistical analysis of reliability and life data. London, Wiley.

[132] Marsick, V. J. and M. A. Gephart (2003). "Action Research: Building the Capacity for Learning and Change." Human Resource Planning 26(2): 14-18.

[133] Martz, H. F. and R. A. Waller (1982). Bayesian Reliability Analysis. New York, Wiley.

[134] Martz, H. F., R. A. Waller and E. T. Fickas (1988). "Bayesian Reliability Analysis of Series Systems of Binomial Subsystems and Components." Technometrics 30(2): 143-154.

[135] Mazzuchi, T. and R. Soyer (1988). "A Bayes Empirical-Bayes Model for Software Reliability." IEEE Transactions on Reliability 37: 248-254.

[136] McKay, J. and M. Oeter (2001). "The dual imperatives of action research." Information Technology and People 14(1): 46-55.

[137] Meeker, W. Q. and L. Escobar, A. (1998). Statistical Methods for Reliability Data. New York, John Wiley & Sons.

[138] Meeker, W. Q. and L. Escobar, A. (2004). "Reliability: The Other Dimension of Quality." Quality Technology and Quantitative Management 1(1): 1-25.

[139] Megill, R. E. (1977). An Introduction to Risk Analysis. Tulsa, UK, Petroleum Publishing Company.

[140] Merkhoffer, M. W. (1987). "Quantifying Judgemental uncertainty: Methodology, experiences and insights." IEEE Transactions on Systems, Man and Cybernetics 17(5): 741-782.

[141] Meyer, M. A. and J. M. Booker (2001). Eliciting and Analyzing Expert Judgement. London, Academic Press Limited.

[142] Mitchell, G. (1993). The Pracitce of Operational Research. Chichester, John Wiley & Sons.

[143] MOD (2002). The Acquisition Handbook.

[144] MOD (2006). Defence Standard 00-42 Issue 2 - Reliability and Maintainability (R&M) Assurance Guidance Part 3, .

[145] Moder, J. J. and E. G. Rodgers (1968). "Judgement estimates of the moments of PERT type distributions." Management Science 15: B76-B83.

[146] Moore, D. (1997). "Bayes for beginners? Some reasons to hesitate." The American Statistician (51): 254-261.

[147] Moore, D. M. and P. D. Antill (2001). "Integrated Project Teams: the way forward for UK defence procurement." European Journal of Purchasing & Supply Management 7: 179-185.

[148] Morgan, M. G. and M. Henrion (1990). Uncertainty. Cambridge, Cambridge University Press.

[149] Neil, M., N. Fenton and M. Tailor (2005). "Using Bayesian Networks to Model Expected and Unexpected Operational Losses." Risk Analysis 25(4): 963-972.

[150] Norrington, L., J. Quigley, A. Russell and R. Van der Meer (2007). "Modelling the Reliability of Search and Rescue Operations with Bayesian Belief Networks." Reliability Engineering and System Safety.

[151] Oakley, J. E. and A. O'Hagan (2002). "Bayesian inference for the uncertainty distribution of computer model outputs." Biometrika 89(4): 769-784.

[152] O'Connor, P. D. T. (2002). Practical Reliability Engineering - Third Edition Revised. New York, John Wiley & Sons.

[153] O'Hagan, A. (1998). "Eliciting expert beliefs in substantial practical applications." The Statistician 47(1): 21-35.

[154] O'Hagan, A., C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley and T. Rakow (2006). Uncertain Judgements: Eliciting Experts' Probabilities. Chichester, John Wiley and Sons.

[155] O'Hagan, A. and J. Forster (2004). Kendall s Advanced Theory of Statistics, Volume 2B: Bayesian Inference. London, Arnold.

[156] O'Hagan, A., E. B. Glennie and R. E. Beardsall (1992). "Subjective Modelling and Bayes Linear Estimation in the UK Water Industry." Applied Statistician 41(3): 563-577.

[157] O'Hagan, A. and J. E. Oakley (2004). "Probability is perfect, but we can't elicit it perfectly." Reliability Engineering and System Safety 85: 239-248.

[158] Oral, M. and O. Kettanii (1993). "The facets of the modeling and validation process in operations research." European Journal of Operational Research 66: 216-234.

[159] Pan, A. C. (1989). "Allocation of order quantities among suppliers." International Journal of Purchasing and Materials Management 25(2): 36-39.

[160] Pate-Cornell, M. (1996). "Uncertainties in risk analysis: Six levels of treatment." Reliability Engineering and System Safety (54): 95-111.

[161] Payne, S. (1951). The art of asking questions. Princeton, New Jersey, Princeton University Press.

[162] Pearl, J. (1998). Probabilistic reasoning in intelligent systems: networks of plausible inference. San Fransico, CA, Morgan Kaufmann.

[163] Pearson, E. S. and J. W. Tukey (1965). "Approximate means and standard deviations based on distances between percentage points of frequency curves." Biometrika 52(3 and 4): 533-546.

[164] Perry, C. and I. D. Greig (1975). "Estimating the mean and variance of subjective distributions in PERT and decision analysis." Management Science 21: 1447-1480.

[165] Pham, H. and X. Zhang (2003). "NHPP software reliability and cost models with testing coverage." European Journal of Operational Research 145(2): 443-454.

[166] Phillips, L. D. (1982). "Requisite Decision Modelling: A Case Study " The Journal of the Operational Society 33(4): 303-311.

[167] Phillips, L. D. (1984). "A theory of requisite decision models." Acta Psychologica 56: 29-48.

[168] Phillips, L. D. (1999). Group elicitation of probability distributions: Are many heads better than one? Decision Science and Technology: Reflections on the Contributions of Ward Edwards. J. Shanteau, B. A. Mellers and D. A. Schum. Norwell, MA, Kluwer Academic Publishers: 313-330.

[169] Pidd, M. (2003). Tools for Thinking: Modelling in Management Science. Chichester, John Wiley & Sons.

[170] Popper, K. (1963). Conjectures and Refutations: The Growth of Scientific Knowledge London, Routledge.

[171] Press, S. J. (2003). Subjective and Objective Bayesian Statistics. Hoboken, NJ, John Wiley & Sons.

[172] Public Accounts Committee, (1989). Reliability and Maintainability of defence equipment and departmental response in treasury minutes.

[173] Pulcini, G. (2001). "An exponential reliability-growth model in multicopy testingprogram." IEEE Transactions on Reliability 50(4): 365-373.

[174] Quigley, J. (1998). Managing Information from a Reliability Growth Development Programme. Department of Management Science. Glasgow, University of Strathclyde. PhD.

[175] Quigley, J. and L. Walls (1999). "Measuring the effectiveness of reliability growth testing." Quality and Reliability Engineering International 15: 87-93.

[176] Quigley, J. and L. Walls (2003). "Confidence intervals for reliability-growth models with small sample-sizes." IEEE Transactions on Reliability 52(2): 254-262.

[177] Quigley, J. and L. Walls (2007). "Trading reliability targets within a supply chain using Shapley's value." Reliability Engineering and System Safety 2007(92): 1448-1457.

[178] Rabitz, H. (1989). "System analysis at molecular scale." Science 246: 221-226.

[179] Ramsey, F. P. (1926). Truth and probability. Foundations of Mathematics and Other Logical Essays. R. B. Brathwaite, Kegan Paul: 156-198.

[180] Rapoport, R. N. (1970). "Three dilemmas of action research." Human Relations 23: 499-513.

[181] Reason, P. (2003). "Pragmatist philosophy and action research." Action Research 1(1): 103-123.

[182] Reason, P. and H. Bradbury (2001). Handbook of action research: Participative inquiry into practice. London, Sage.

[183] Revans, R. W. (1982). The Origins and Growth of Action Learning. Bickley, Kent, Chartwell-Bratt.

[184] Richardson, G. P. (1996). "Problems for the future of system dynamics." System Dynamics Review 12(2): 141-157.

[185] Robinson, D. G. and D. Dietrich (1987). "A New Nonparametric Growth Model." IEEE Transactions on Reliability 36(4): 411-418.

[186] Robinson, S. and M. Pidd (1998). "Provider and customer expectations of successful simulation projects." Journal of the Operational Research Society 49(3): 200-209.

[187] Rosenthal, E. C., J. L. Zydiak and S. S. Chaudhry (1995). "Vendor selection with bundling." Decision Sciences 26(1): 35-48.

[188] Sadrian, A. A. and Y. S. Yoon (1994). "A procurement decision support system in business volume discount environments." Operations Research 42(1): 14-23.

[189] Saleh, J. H. and K. Marais (2005). "Highlights from the early (and pre-) history of reliability engineering." Reliability Engineering and System Safety 91: 249-256.

[190] Savage, L. J. (1954). The Foundations of Statistics. New York, NY, Dover Publications.

[191] Shachter, R. D. and R. Kenley (1989). "Gaussian influence diagrams." Management Science 35(5): 527-550.

[192] Smith, J. W. (1996). Plausible Bayesian Games. Bayesian Statistics 5. J. M. Bernardo, B. A.P. and S. A.F.M. Oxford, Oxford University Press: 387-405.

[193] Soukoup, W. R. (1987). "Supplier selection strategies." International Journal of Purchasing and Materials Management 23(3): 7-12.

[194] Stowell, F., D. West and M. Stansfield (1997). Action research as a framework for IS research. Inofrmation systems: An Emerging Dicispline. J. Mingers and F. Stowell. London, McGraw-Hill.

[195] Susman, G. I. and R. D. Evered (1978). "An Assessment of the Scientific Merits of Action Research." Administrative Science Quarterly 3582-603.

[196] Tian, L. and A. Noore (2005). "Evolutionary neural netwrok modeling for software cumulative failure time prediction." Reliability Engineering and System Safety 87: 45-51.

[197] Timmerman, E. (1986). "An approach to vendor performance evaluation." International Journal of Purchasing and Materials Management 22(1): 27-32.

[198] Torbert, W. R. (1976). Creating a community of inquiry: conflict, collaboration, transformation. New York, John Wiley.

[199] Tversky, A. and D. Kahneman (1974). "Judgement under uncertainty: Heuristics and biases." Science 185: 1124-1131.

[200] Ushakov, I. and R. A. Harrison (1994). Handbook of Reliabiltiy Engineering. New York, John Wiley & Sons.

[201] Van Dorp, J. R. and T. Mazzuchi (2004). "A general Bayes inference model for accelerated life testing." Journal of Statistical Planning and Inference 119: 55-74.

[202] Walls, L. and J. Quigley (2001). "Building prior distributions to support Bayesian reliability growth modelling using expert judgement." Reliability Engineering and System Safety 74: 117-128.

[203] Walls, L., J. Quigley and J. Marshall (2006). "Modeling to Support Reliability Enhancement during Product Development with Applications in the UK Aerospace Industry." IEEE Transactions on Engineering Management 53(2): 263-274.

[204] Weber, C. A., J. R. Curent and A. Desai (1991). "Vender selection criteria and methods." European Journal of Operational Research 50: 2-18.

[205] Williamson, J. (2005). Bayesian Nets and Causality - Philosophical and Computational Foundations. Oxford, Oxford University Press.

[206] Winkler, R. L. (1967). "The Assessment of Prior Distributions in Bayesian Analysis." Journal of American Statistical Association 62(319): 776-800.

[207] Wooff, D., M. Goldstein and F. Coolen (2002). "Bayesian Graphical Models for Software Testing." IEEE Transactions on Software Engineering 28(5): 510-525.

[208] Yates, J. F. (1990). Judgement and Decision Making. New Jersey, Prentice-Hall.

[209] Zhang, R. and S. Mahadevan (2003). "Bayesian methodology for reliability model acceptance." Reliability Engineering and System Safety (80): 95-103.

[210] Zitrou, A. (2007). Exploring a Bayesian Approach for Structural Modelling of Common Cause Failures. Department of Management Science. Glasgow, University of Strathclyde.