STATE AND PARAMETER ESTIMATION TECHNIQUES FOR STOCHASTIC SYSTEMS

Matthew J. CARR

Centre for Operational Research and Applied Statistics
(CORAS)
University of Salford, Salford, UK

Submitted in Partial Fulfilment of the Requirements of
the Degree of Doctor of Philosophy, June 2006

CONTENTS

Acknowledgements

I would like to thank my supervisor Professor Tony Christer who sadly passed away earlier this year, for all the opportunities, support and guidance that he gave to me. He is sorely missed, by myself, and everyone who knew and worked with him at the University.

I would also like to thank my co-supervisor Dr Wenbin Wang for the support and direction he has given to me during my PhD and beyond.

I also wish to thank Miss Susan Sharples for her continuous support throughout all my time at Salford.

To my parents, thank you for all your love and support.

To my Sudeh, thank you for everything!

ABSTRACT

This thesis documents research undertaken on state and parameter estimation techniques for stochastic systems in a maintenance context. Two individual problem scenarios are considered. For the first scenario, we are concerned with complex systems and the research involves an investigation into the ability to identify and quantify the occurrence of fault injection during routine preventive maintenance procedures. This is achieved using an appropriate delay time modelling specification and maximum-likelihood parameter estimation techniques. The delay time model of the failure process is parameterised using objective information on the failure times and the number of faults removed from the system during preventive maintenance. We apply the proposed modelling and estimation process to simulated data sets in an attempt to recapture specified parameters and the benefits of improving maintenance processes are demonstrated for the particular example. We then extend the modelling of the system in a predictive manner and combine it with a stochastic filtering approach to establish an adaptive decision model. The decision model can be used to schedule the subsequent maintenance intervention during the course of an operational cycle and can potentially provide an improvement on fixed interval maintenance policies.

The second problem scenario considered is that of an individual component subject to condition monitoring such as, vibration analysis or oil-based contamination. The research involves an investigation into techniques that utilise condition information that we assume is related stochastically to the underlying state of the component, taken here to be the residual life. The techniques that we consider are the proportional hazards model and a probabilistic stochastic filtering approach. We investigate the residual life prediction capabilities of the two techniques and

construct relevant replacement decision models. The research is then extended to consider multiple indicators of condition obtained simultaneously at monitoring points. We conclude with a brief investigation into the use of stochastic filtering techniques in specific scenarios involving limited computational power and variable underlying relationships between the monitored information and the residual life.

Chapter 1. Introduction

The general objective of the research documented in this thesis is to provide a contribution to the goal of optimising the performance of operational systems that are stochastic by nature and subject to some form of degradation over time. This categorisation incorporates almost any operational system from a complex industrial production line with many sub-systems to a simple photocopier or printer. The objective is achieved through the efficient scheduling of activities that are often overlooked as a viable means of boosting operational availability and performance, such as the use of planned preventive maintenance and effectively timed component replacements. These are activities that are often carried out in an opportunistic fashion when systems either fail (or are not currently operational for some other reason) or are conducted according to an inappropriate model producing decisions that are not cost effective or result in excessive downtime.

The particular focus of this research is on the techniques that assist in the characterisation of stochastic systems including both complex systems and individual replaceable components. Accurate representation of said systems is achieved through the use of an appropriate model specification and parameterisation. It is the parameter and state estimation techniques that are the primary topics of interest here. Using the constructed models, maintenance and replacement decision modelling can be optimised to reduce costs, identify areas of the current operational procedure that are lacking and limit the downtime of the system, thus increasing availability and operational efficiency.

Two types of scenario are addressed in this thesis with the first being covered by chapters 3 and 4 and the second by chapters 5 to 9. The first scenario involves the modelling of complex systems that incorporate the potential for human error at

maintenance interventions. We primarily consider the type of human error that manifests itself in the form of artificial fault injection during the course of planned inspection and preventive maintenance procedures although the model constructed is not limited to specifically human error based injections. The focus of the research is on the ability to characterise from relevant failure data, using an appropriate model specification and parameter estimation techniques, the fault arrival and failure processes of the system with emphasis being placed on the estimation of the level of fault injection that may be taking place. The subsequent modelling of complex systems concerns the on-line estimation and characterisation of the underlying fault arrival process using stochastic filtering and a hidden Markov model formulation. Modelling the system in a dynamic manner allows for the construction of adaptive decision models rather than fixed interval maintenance policies.

The second scenario concerns the condition-based maintenance (CBM) of an individual component or a piece/part of machinery with a single dominant failure mode. The research involves comparing CBM models where indicatory condition monitoring (CM) information, such as vibration levels or metal concentrations in oil samples, may be used to estimate the probability of component failure within a specified time frame and schedule maintenance or replacements accordingly. We assume that the CM parameters are stochastically related to the actual condition or residual life of the component. The two techniques that we compare are the proportional hazards model and a probabilistic stochastic filtering approach. The comparisons are conducted using industrial CM data, the first using the overall vibration level as a single CM input and the second considering multiple oil-based CM parameters. Issues regarding the handling of multiple information parameters

are also addressed and finally, some theoretical developments on the use of filtering theory in the context of CBM are introduced in the final chapter of the thesis.

The outline of the thesis is now discussed;

Chapter 2 documents the necessary modelling and theoretical background and presents the key introductory points for the techniques that are applied in subsequent chapters.

Chapter 3 covers research undertaken regarding the provision for human error in maintenance models of complex systems using delay time modelling; Initially a discussion of the relevant delay time modelling and general modelling background and literature is presented. Then the work undertaken using maximum likelihood estimation to capture the necessary parameters is documented. A number of case studies are presented to demonstrate the application of the proposed techniques and the ability to compare differing model forms and combinations of the different aspects of human error during inspection and maintenance procedures is also addressed.

Chapter 4 is a continuation of the fault injection work contained in the previous chapter. Initially we present an alternative description and solution methodology for the problems addressed in chapter 3 by combining delay time modelling and a hidden Markov model (HMM). The ability to construct adaptive decision models that respond to the failure history of the system is of particular interest in this chapter.

Chapter 5 contains a discussion of relevant CM techniques and some associated literature. We then document the necessary CBM literature that is available for modelling replacement decisions that are associated with the monitoring of

individual components and discuss the differences when modelling information that is directly or indirectly related to the underlying state of the component.

Chapter 6 presents two models for condition-based maintenance applications that are compared for industrial case studies in subsequent chapters. The techniques described are the proportional hazards model and a probabilistic stochastic filtering approach. We consider the potential for utilising multiple information sources and the need for data reduction techniques, such as principal components analysis (PCA), is addressed. The chapter concludes with a discussion of additional means of tackling some of the problems raised. These techniques include dynamic principle components analysis (DPCA) and independent components analysis (ICA). Techniques for comparing the models and establishing optimal replacement decisions are also introduced in this chapter.

Chapter 7 is a case comparison of proportional hazards modelling and probabilistic stochastic filtering when applied to vibration based CM information.

Chapter 8 is a comparison of the PHM and the filtering approach for condition-based maintenance applications using oil-based CM parameters. In this chapter, we also discuss the use of incomplete condition monitoring information and the impact on both parameter estimation and the ability to compare the two techniques.

Chapter 9 presents some further uses of filtering theory in the context of condition based maintenance applications. Attention is reserved for non-linear problems and approximate means of tackling the state estimation problem. The techniques described are applicable in situations where limited computational power is available, a large number of components are under scrutiny or the underlying dynamics of the systems degradation and the relationship with the monitored condition information are not known to a satisfactory degree of precision.

The thesis concludes with a list of the associated references that are cited within the body of the text.

Chapter 2. Modelling background

2.1 Introduction

In this chapter, the necessary modelling background and preliminaries for the research documented in subsequent chapters is presented. The techniques used for estimating the parameters of the various stochastic models developed in the thesis are introduced in section 2.2, section 2.3 introduces the hazard and reliability functions for systems that will fail at some unknown time and sections 2.4 and 2.5 introduce some of the techniques that are available for estimating the underlying state of discrete time and continuous time stochastic systems, respectively. The references and background that are relevant to the research on complex systems are given in the introduction to chapter 3 and those pertaining to the research on the monitoring of individual components are given in chapter 5. However, there are some general references that have been particularly useful in the development of the research contained here and these are now introduced. For information on system identification and related topics, see Ljung (1999), for state space models and the Kalman filter, the primary references have been Harvey (1989) and Jazwinski (1970), and for further non-linear stochastic filtering information, see Krishnan (1984), Jazwinski (1970) and Kallianpur (1980). Other general references have been Bernardo & Smith (2000) for background on Bayesian inference and analysis and Aoki (1967) and Liptser (1997) for information on control theory for the stochastic systems considered here.

2.2 Parameter estimation

To estimate the parameters of the stochastic models discussed in this thesis, we seek an estimator that exhibits some nice statistical properties. A good estimator utilises

all the relevant and required information from the data under investigation. The first statistical property of an estimator that we consider is the level of bias, which is the difference between the expected value of a parameter given by an estimator and the true underlying value of that parameter. Amongst the class of unbiased estimators for a given problem, an efficient estimator is the one with a minimal variance and as such, a minimal mean-square error (MSE). An additional property for consideration is that of consistency. A consistent estimator is one that converges probabilistically to the true value of a parameter with an increasing sample size. In fact it is often necessary to consider the asymptotic (large sample) properties of estimators when selecting an approach for practical scenarios.

2.2.1 Maximum likelihood estimation

The models used in the research presented in this thesis are characterised by parameters that are estimated from data using an appropriate modelling approach and maximum likelihood estimation. The maximum likelihood estimates of the parameter set are the values that maximise the likelihood function but they may not necessarily be unbiased estimates. The bias of ML estimators may be quite large and the estimator may not be unique or even exist for particular cases. Under the regularity conditions that the first and second derivatives of the log-likelihood function must be defined and the Fisher information matrix must not be zero, the MLE can be considered to be asymptotically optimal, see Kendall & Stuart (1979). For instance, the estimate is asymptotically unbiased in that the bias tends to zero as the number of samples gets large. This property is a result of the fact that the distribution of the estimate tends to a Gaussian distribution as the sample size increases. In addition, the MLE is asymptotically efficient achieving the Cramer-

Rao lower bound which is an asymptotic lower bound on the variance of any unbiased estimator.

If x is a continuous random variable with probability density function $f(x;\underline{\theta})$, where $\underline{\theta} = \theta_1, \theta_2, ..., \theta_k$ is the set of k parameters under scrutiny, the likelihood of observing the information set $\underline{x} = x_1, x_2, ..., x_r$ is given by the product

$$\mathcal{L}(\underline{\theta} \mid \underline{x}) = \prod_{i=1}^{r} f(x_i;\underline{\theta}) \qquad [2.1]$$

Maximisation is often eased by taking logarithms of the likelihood function. The optimal parameter estimates are equivalent under either function. The log-likelihood function is

$$l(\underline{\theta} \mid \underline{x}) = \sum_{i=1}^{r} \log\left(f(x_i;\underline{\theta})\right) \qquad [2.2]$$

The maximum likelihood parameter estimates are then obtained as the simultaneous solutions of the k equations

$$\frac{\partial l(\underline{\theta} \mid \underline{x})}{\partial \theta_j} = 0 \qquad [2.3]$$

for $j = 1, 2, \ldots, k$. The covariance matrix for the estimated parameters is established as

$$\underline{\Sigma} = \begin{bmatrix} -\dfrac{\partial^2 l(\underline{\theta} \mid \underline{x})}{\partial \theta_1^{\,2}} & \cdots & -\dfrac{\partial^2 l(\underline{\theta} \mid \underline{x})}{\partial \theta_1 \partial \theta_k} \\ \vdots & & \vdots \\ -\dfrac{\partial^2 l(\underline{\theta} \mid \underline{x})}{\partial \theta_k \partial \theta_1} & \cdots & -\dfrac{\partial^2 l(\underline{\theta} \mid \underline{x})}{\partial \theta_k^{\,2}} \end{bmatrix}^{-1} \qquad [2.4]$$

When using the standard maximum likelihood approach to parameter estimation, the observations are assumed to be independent from one another and identically distributed. As such, the likelihood of observing the given data is simply the product

of the individual probability associated with each observation. In chapters 6 - 8, the observed information is assumed to be conditional on previous observations and as a result, the standard approach is to establish the likelihood of observing the r pieces of information as the product of conditional probabilities, see Harvey (1989). The functional form of the likelihood function becomes

$$\mathcal{L} = f(x_1) \times f(x_2 \mid x_1) \times f(x_3 \mid x_2, x_1) \times \ldots \times f(x_r \mid x_{r-1}, x_{r-2}, \ldots, x_2, x_1) \qquad [2.5]$$

where, $f(a \mid b)$ is the probability of observing event a given that event b has already been observed.

With the likelihood function in hand, an optimisation algorithm is still required for maximisation of the expression with respect to the unknown parameters. In general, a local optimisation method is designed to generate a sequence of points that will converge to a local minimum. The algorithm is stopped or the sequence terminated once a convergence criterion or criteria are met. Often a criterion is that the norm of the gradient is small because theoretically at a local minimum the norm of the gradient is zero. One approach that is available is the Broyden-Fletcher-Golfarb-Shanno quasi-Newton (BFGS) algorithm. The BFGS algorithm is based upon the second order Taylor polynomial of the objective function and the Newton-Raphson method and has a convergence rate that is much faster than other optimisation algorithms such as conjugate gradient methods. This is due to the fact that the search directions for the BFGS are often more accurate however, more computational power is required for each iteration. The approximate second order representation for the log-likelihood function is

$$l(\underline{\theta} \mid \underline{x}) \approx l(\underline{\alpha} \mid \underline{x}) + \delta\underline{\alpha}^T \nabla l(\underline{\alpha} \mid \underline{x}) + \frac{1}{2}\delta\underline{\alpha}^T H(\underline{\theta} \mid \underline{x})\delta\underline{\alpha} \qquad [2.6]$$

where, both $\nabla l(\underline{\theta} \mid \underline{x}) = (\partial l(\underline{\theta} \mid \underline{x})/\partial\theta_1, \ldots, \partial l(\underline{\theta} \mid \underline{x})/\partial\theta_k)$ and the Hessian matrix

$$H(\underline{\theta}\,|\,\underline{x}) = \begin{bmatrix} \dfrac{\partial^2 l(\underline{\theta}\,|\,\underline{x})}{\partial\theta_1^{\,2}} & \cdots & \dfrac{\partial^2 l(\underline{\theta}\,|\,\underline{x})}{\partial\theta_1\partial\theta_k} \\ \vdots & & \vdots \\ \dfrac{\partial^2 l(\underline{\theta}\,|\,\underline{x})}{\partial\theta_k\partial\theta_1} & \cdots & \dfrac{\partial^2 l(\underline{\theta}\,|\,\underline{x})}{\partial\theta_k^{\,2}} \end{bmatrix} \qquad [2.7]$$

are evaluated at $\underline{\theta} = \underline{\alpha}$ and $\delta\underline{\alpha} = (\delta\alpha_1,...,\delta\alpha_k)^T$ is $\underline{\theta} - \underline{\alpha}$. The well-known Newton-Raphson method for solving $\nabla l(\underline{\theta}\,|\,\underline{x}) = \underline{0}$ is given by the algorithm

$$\underline{\theta}^{(c+1)} = \underline{\theta}^{(c)} - [H(\underline{\theta}^{(c)}|\,\underline{x})]^{-1}\nabla l(\underline{\theta}^{(c)}|\,\underline{x}) \qquad [2.8]$$

where, c is the index. $\underline{A}^{(c)}$ is defined as an approximation to the inverse Hessian matrix $[H(\underline{\theta}^{(c)}|\,\underline{x})]^{-1}$. The Hessian matrix is the square matrix of second partial derivatives and is often used in optimisation algorithms. This is due to the fact that if the Hessian is negative definite at a critical point (when the gradient of a scalar function is zero), then we have a local maximum. The inverse of the Hessian matrix also gives the variance-covariance matrix for the estimated parameter values.

The BFGS algorithm is

$$\underline{A}^{(c+1)} = \underline{A}^{(c)} + \left(1 + \frac{(\underline{b}^T \Delta\underline{g}^{(c)})}{(\underline{a}^T \Delta\underline{g}^{(c)})}\right)\frac{\underline{a}\underline{a}^T}{(\underline{a}^T \Delta\underline{g}^{(c)})} - \frac{\underline{a}\underline{b}^T + \underline{b}\underline{a}^T}{(\underline{a}^T \Delta\underline{g}^{(c)})} \qquad [2.9]$$

where, $\underline{a} = \underline{\theta}^{(c+1)} - \underline{\theta}^{(c)}$, $\underline{b} = \underline{A}^{(c)}\Delta\underline{g}^{(c)}$ and

$$\Delta\underline{g}^{(c)} = \nabla l(\underline{\theta}^{(c+1)}|\,\underline{x}) - \nabla l(\underline{\theta}^{(c)}|\,\underline{x}) \qquad [2.10]$$

When the algorithm converges, the optimal parameter estimates are obtained. The BFGS algorithm is presented here and used in the proceeding research in preference to other candidate solutions to the optimisation problem, such as the Davidon-Fletcher-Powell (DFP) method, due to the fact that it is widely regarded as being the most robust search algorithm.

2.2.2 Alternative methods

In this section, we will discuss some (but by no means all) of the alternatives to maximum likelihood estimation. An alternative and frequently used means of estimating the parameters of a stochastic model or probability density is the least squares approach which involves the minimisation of the squared errors between the observed information and the model or density expectation. However, unlike maximum likelihood estimation, probabilistic statements can not be made about the estimated values using the least squares approach. A method similar to the least squares approach involves the minimisation of the chi-square function using the expectation from the model or density, see Kendall & Stuart (1979). Another technique is the Rao-Blackwell theorem, see Rao (1965), that utilises sufficient statistics for a given data set and modifies existing estimators to find an improved estimator. A sufficient statistic is an observable random variable constructed from a set of data that provides enough information to construct the conditional probability distribution for the data set and is not a function of the population parameters. Applications of the Rao-Blackwell theorem often use a maximum likelihood estimator as a starting point. Alternatively, if the original estimator is unbiased and complete then, according to the Lehmann-Scheffé theorem, the Rao-Blackwell technique provides a means of finding the minimum unbiased estimator.

Another parameter estimator is given by the generalised method of moments which, as the name implies, is a generalisation of the method of establishing the moments of a probability distribution. In principle, it is similar to the minimum chi-square estimator. Minimum variance unbiased estimators also exist however, although theoretically sound, the restrictions placed on the bias can easily produce unrealsitic and misleading parameter estimates. Maximum a-posteriori (MAP) parameter

estimates are obtainable with the availablility of prior information, see Sorenson (1980). The MAP estimates are achieved by maximising the product of the likelihood function and an a-priori probability distribution for the parameter. Another technique that utilises prior distributions for the parameters is the expectation-maximisation (EM) algorithm which is a recursive procedure that defines some of the unknown information as latent variables and can also utlise the likelihood function in aquiring optimal estimates. In terms of constructing probability distributions, some techniques are available which do not require parameterisation. The Kaplan-Meier approach is a nonparametric technique for survival function estimation based on data only, see Kalbfeisch & Prentice (1973).

2.3 Hazard and reliability

The hazard and reliability functions are often utilised in applications involving the analysis of the expected life of a system. For a given system or individual component, we define $f(t)$ as a continuous failure time distribution on $t \geq 0$, and the reliability function $R(t)$, also known as the survival function, is the probability that the system survives beyond time t. We have the relationship

$$R(t) = 1 - F(t) \qquad [2.11]$$

where, $F(t)$ is the cumulative failure time density given by $F(t) = \int_0^t f(s)ds$. The hazard $h(t)$ is often referred to as the instantaneous failure rate and is given by

$$h(t) = f(t) / R(t) \qquad [2.12]$$

We also have

$$F(t) = 1 - \exp\{-H(t)\} \qquad [2.13]$$

where, $H(t)$ is the cumulative hazard at time t given by $H(t) = \int_0^t h(s)ds$. Using equation [2.11], equation [2.13] can be manipulated to give

$$h(t) = -d \ln R(t) / dt \qquad [2.14]$$

Chapters 6 - 8 utilise a proportional hazards model (PHM) for determination of the life expectancy of individual components. In a general sense, to establish the hazard, the PHM weights the impact of the unit's age and the input from monitored information that is assumed to be in some way related to the degradation of the component. The hazard for the PHM is given by

$$h(t, y) = h_0(t)\lambda(\gamma y) \qquad [2.15]$$

where, $h_0(t)$ is a baseline hazard function that is dependent on the age of the system only, y represents the information available about the system at time t and $\lambda(\gamma y)$ is a function of y with a co-efficient γ. The function λ can be extended to incorporate the information from multiple sources (denoted by the vector $\underline{y}$) as $\lambda(\underline{\gamma}'\underline{y})$. In chapters 6 - 8, the PHM is investigated as a methodology for condition based maintenance applications and its derivation is contrasted with the technique of stochastic filtering that is used in this context as a probabilistic approach to the problem that arises from the methodology outlined in the next section.

2.4 State estimation for discrete time stochastic systems

2.4.1 The conditional estimation problem

Much of the research documented in this thesis centres around the estimation of the state or condition of a complex system or individual component using both the age and any available monitored information when the state is not directly observable. It may not be the state of the system that is measured directly but rather, some signal or output that is assumed to be stochastically correlated with the system state and as such, the underlying condition of the system is inferred from this information. When modelling the underlying state of a complex system (see chapter 4), our monitored

information consists of failure times and the number of defects removed during the course of planned preventive maintenance interventions. The modelling objective is the characterisation of the underlying dynamics of the system regarding the fault arrival process and the potential for human error related fault injections, with a view to improving a given process and establishing a fixed decision model. As such, the underlying (unobservable) state is the number of faults that have arisen in the system by that point in time. However, in the context of condition based maintenance (see chapters 5 – 9), the state of an individual unit is not easily definable and we are required to consider variables that are related to or are functions of the severity of a fault at a given moment, or functions of the general operational capability, such as, the remaining useful life of the component before the defect leads to failure. Monitoring techniques such as vibration monitoring and oil analysis provide the indicator information that is used to estimate the state.

Techniques such as statistical process control (SPC) are limited when considering data that is non-stationary and evolving stochastically. One of the techniques that is suitable for data of this type is the stochastic filtering approach to state estimation that utilises any knowledge of the system and the characteristics of the particular indicatory information that is in use. The filtering approach assumes a statistical description for the system and the observation noise (measurement errors) and incorporates any uncertainties in the dynamics of the system. The recursive filtering process is relatively straightforward for systems that are characterised by linear relationships and perturbed by Gaussian white noise, (Harvey, 1989). The methodology can be generalised into a general Bayesian filter using a probabilistic approach and relaxing the linear assumptions to reveal an optimal filter designed to handle non-linear state space models, see Meinhold & Singpurwalla (1983, 1986)

and Jazwinski (1970). As such, it can be demonstrated that the linear filter is merely a special case of the general non-linear filter. However, by relaxing the assumptions of normally distributed noise and linear system equations, the computational complexity is greatly increased and approximate or numerical solutions are required, as is illustrated in chapters 4, 7 and 8. An understanding of multivariate random variables, Gaussian distributions, white noise processes, conditional probabilities and Bayes' theory are of particular relevance to the proceeding research. In this section, we focus on the discrete time state estimation problem where, information is received and knowledge of the state is updated at discrete time points during the life of the system or component. The estimation problem is addressed using the least squares method and a probabilistic filtering approach for a general state and condition input where, the state and observation processes are described by vector processes.

An important element of the discrete probabilistic filtering approach is the Markov property of independent increments that implies that current states are independent of their history and as such, future states may be inferred solely from knowledge of the current state vector. For discrete time systems, the conditional probability density function

$$p(\underline{x}_{i+1} \mid \underline{x}_0, \underline{x}_1, \ldots, \underline{x}_i) = p(\underline{x}_{i+1} \mid \underline{x}_i) \qquad [2.16]$$

describes the transition for a state vector $\underline{x}$ from one stage to the next where, in some applications, the stages may correspond to monitoring intervals. Using the transition densities, the state at the next discrete time point $\underline{x}_{i+1}$ can be modelled as the conditional mean

$$\mathbf{E}[\underline{x}_{i+1} \mid \underline{x}_i] = \int_{-\infty}^{\infty} \underline{x}_{i+1}\, p(\underline{x}_{i+1} \mid \underline{x}_i) d\underline{x}_{i+1} \qquad [2.17]$$

Many stochastic systems can be modelled effectively using a first-order Markovian discrete time representation of the form

$$\underline{x}_{i+1} = f(\underline{x}_i, t_{i+1}) + \underline{v}_i \qquad [2.18]$$

where, t_i is the age of the system at the ith check-point and $\underline{v}_i$ is a random disturbance that is often assumed to be normally distributed and independent in time. The expression describing the evolution of the stochastic process, equation [2.18], is used in control theory with the inclusion of an additional control input $\underline{u}_i$ as $f(\underline{x}_i, t_{i+1}, \underline{u}_i)$. See Aoki (1967) for details on control theory for discrete time stochastic systems.

The observed information vector $\underline{y}_i$ is assumed to be a function of the state of the system and a level of measurement noise $\underline{e}_i$ is incorporated in the observation expression as

$$\underline{y}_i = h(\underline{x}_i, t_i) + \underline{e}_i \qquad [2.19]$$

When considering systems where the observed data is not directly related to the state or condition that is the objective measure of the estimation (filtering) or prediction process, the observations and the state are assumed to be correlated stochastically and in many estimation problems (see the hidden Markov modelling undertaken in chapter 4), it is a necessary requirement that the subsequent state $\underline{x}_{i+1}$ may be determined uniquely from the current state $\underline{x}_i$ as

$$p(\underline{x}_{i+1} \mid \underline{x}_i, \underline{y}_i) = p(\underline{x}_{i+1} \mid \underline{x}_i) \qquad [2.20]$$

where, the state and measurement noise are assumed to be mutually independent. However, it is often possible to transform the situation when the assumption does not hold by augmenting the state vector to create a model in the standard state-space

form. See Harvey (1989) for details. The conditional mean can be used to estimate the condition $\underline{x}$ when an observation $\underline{y}$ is obtained as

$$\mathbf{E}[\underline{x}_i \mid \underline{y}_i] = \int_{-\infty}^{\infty} \underline{x}_i p(\underline{x}_i \mid \underline{y}_i) d\underline{x}_i = \int_{-\infty}^{\infty} \frac{\underline{x}_i p(\underline{x}_i, \underline{y}_i)}{p(\underline{y}_i)} d\underline{x}_i = \frac{\int \underline{x}_i p(\underline{x}_i, \underline{y}_i) d\underline{x}_i}{\int p(\underline{x}_i, \underline{y}_i) d\underline{x}_i} \quad [2.21]$$

assuming that the various probabilistic relationships are defined. The conditional mean gives the mean square optimal estimate, however, it is complicated to update for dynamic systems, see Jazwinski (1970). A notable exception is that of linear systems perturbed by Gaussian noise where, the approach using the properties of conditional mean estimation is known as the Kalman filter. Estimation and the resulting analysis is far less complex when considering linear estimates, the Linear Least Mean Square (LLMS) estimation process produces results identical to the conditional mean approach for Gaussian distributed random variables. The LLMS approach involves the estimation of a and b in the expression $\hat{x}_i = a + by_i$ to find the parameter values that minimise the error covariance. However, the LLMS filter only gives optimal estimates in the special linear case described.

2.4.2 The least squares approach

We now consider a general least squares approach to the state estimation (not parameter estimation) problem for the system and observation process described by equations [2.18] and [2.19]. With the probabilistic stochastic filtering approach described in the next section, the errors in the system and observation expressions are defined as random inputs with known properties. However, when applying the least squares approach, they are defined as merely errors of an unknown quantity. Assuming that an estimate of the initial state $\hat{\underline{x}}_0$ is available, upon the availability of the ith observation at time t_i, the least squares approach minimises

$$J_i = \frac{1}{2}(\underline{x}_0 - \hat{\underline{x}}_0)^T \underline{P}_0^{-1}(\underline{x}_0 - \hat{\underline{x}}_0) + \frac{1}{2}\sum_{k=1}^{i}(\underline{y}_k - h(\underline{x}_k, t_k))^T \underline{Q}_k^{-1}(\underline{y}_k - h(\underline{x}_k, t_k)) + \frac{1}{2}\sum_{k=1}^{i}\underline{v}_k{}^T \underline{R}_k^{-1}\underline{v}_k \quad [2.22]$$

with respect to $\{\underline{x}_0,...,\underline{x}_i\ ;\ \underline{v}_1,...,\underline{v}_i\}$ and subject to the constraints

$$\underline{x}_{k+1} = f(\underline{x}_k, t_{k+1}) + \underline{v}_k \quad [2.23]$$

for $k = 0, 1, \ldots, i\text{-}1$. In this context, the contributions $\underline{P}_0^{-1}$, $\underline{Q}_k^{-1}$ and $\underline{R}_k^{-1}$ are simply defined as weighting matrices. The series of estimated state vectors $\{\hat{\underline{x}}_0, \hat{\underline{x}}_1, ..., \hat{\underline{x}}_i\}$ that minimise equation [2.22] are the smoothed solution and $\hat{\underline{x}}_i$ is the filtering solution at time t_i. The major drawback of the least squares approach is that upon observing each new vector of stochastically related information, $\underline{y}_{i+1}$, an entirely new problem must be solved. An alternative is the recursive least squares (RLS) algorithm involving the minimisation of J_{i+1} in terms of $\underline{y}_{i+1}$ and the estimate of the state from the previous recursion $\hat{\underline{x}}_i$. Analogous to the least squares approach for estimating parameters, probabilistic statements can not be made about an estimated state using least squares state estimation.

2.4.3 Probabilistic stochastic filtering

When conditioned on stochastically related observations, an optimal estimate of the state is obtained using the following general framework for the majority of systems. Consider a non-linear stochastic system with state and measurement equations;

$$\underline{x}_{i+1} = f(\underline{x}_i, t_{i+1}, \underline{v}_i) \quad [2.24]$$

$$\underline{y}_i = h(\underline{x}_i, t_i) + \underline{e}_i \quad [2.25]$$

where, the state and measurement error $\underline{v}_i$ and $\underline{e}_i$ are vectors with elements that are assumed to follow 0-mean white noise processes. If $\underline{Y}_i$ denotes all the available measurements at the ith discrete time point, then the optimal mean square estimate is

$\hat{\underline{x}}_i = \mathbf{E}[\underline{x}_i \mid \underline{Y}_i]$ which must be computed recursively. Firstly, using Bayes' rule, the functional form of the conditional probability density function $p(\underline{x}_i \mid \underline{Y}_i)$ must be determined as

$$p(\underline{x}_i, \underline{y}_i \mid \underline{Y}_{i-1}) = p(\underline{x}_i \mid \underline{y}_i, \underline{Y}_{i-1}) p(\underline{y}_i \mid \underline{Y}_{i-1}) = p(\underline{y}_i \mid \underline{x}_i, \underline{Y}_{i-1}) p(\underline{x}_i \mid \underline{Y}_{i-1}) \quad [2.26]$$

For the non-linear system defined in equations [2.24] and [2.25], the measurement contains white noise and we may therefore assume that the estimated condition contains all the necessary information regarding the measurements and as such, we have

$$p(\underline{y}_i \mid \underline{x}_i, \underline{Y}_{i-1}) = p(\underline{y}_i \mid \underline{x}_i) \quad [2.27]$$

Using equations [2.26] and [2.27], the probability density function of a particular state given the monitored condition history to date can be established as

$$p(\underline{x}_i \mid \underline{Y}_i) = \frac{p(\underline{y}_i \mid \underline{x}_i)}{p(\underline{y}_i \mid \underline{Y}_{i-1})} p(\underline{x}_i \mid \underline{Y}_{i-1}) \quad [2.28]$$

Using the transition density, the probability of a one-step predicted state is given by

$$p(\underline{x}_i) = \int p(\underline{x}_i \mid \underline{x}_{i-1}) \, p(\underline{x}_{i-1}) \, d\underline{x}_{i-1} \quad [2.29]$$

where, all the densities can be conditioned on the observed data and due to the Markovian nature of the process, the conditional predictive density becomes

$$p(\underline{x}_i \mid \underline{Y}_{i-1}) = \int p(\underline{x}_i \mid \underline{x}_{i-1}, \underline{Y}_{i-1}) p(\underline{x}_{i-1} \mid \underline{Y}_{i-1}) d\underline{x}_{i-1} = \int p(\underline{x}_i \mid \underline{x}_{i-1}) p(\underline{x}_{i-1} \mid \underline{Y}_{i-1}) d\underline{x}_{i-1} \quad [2.30]$$

Combining equations [2.28] and [2.30] produces

$$p(\underline{x}_i \mid \underline{Y}_i) = \frac{p(\underline{y}_i \mid \underline{x}_i)}{p(\underline{y}_i \mid \underline{Y}_{i-1})} \int p(\underline{x}_i \mid \underline{x}_{i-1}) p(\underline{x}_{i-1} \mid \underline{Y}_{i-1}) d\underline{x}_{i-1} \quad [2.31]$$

and this can be written as

$$p(\underline{x}_i \mid \underline{Y}_i) = \frac{\int p(\underline{y}_i \mid \underline{x}_i) p(\underline{x}_i \mid \underline{x}_{i-1}) p(\underline{x}_{i-1} \mid \underline{Y}_{i-1}) d\underline{x}_{i-1}}{\iint p(\underline{y}_i \mid \underline{x}_i) p(\underline{x}_i \mid \underline{x}_{i-1}) p(\underline{x}_{i-1} \mid \underline{Y}_{i-1}) d\underline{x}_{i-1} d\underline{x}_i} \qquad [2.32]$$

Equation [2.32] is the formulation used in chapter 4 and is adapted for a model with a discrete state space. In chapters 6 - 8, the particular definition of the state as the time remaining before failure provides a deterministic relationship between subsequent underlying states and enables the updating equation

$$p(\underline{x}_i \mid \underline{x}_{i-1}) p(\underline{x}_{i-1} \mid \underline{Y}_{i-1}) = p(\underline{x}_i \mid \underline{Y}_{i-1}) \qquad [2.33]$$

to be established.

For non-linear, non-Gaussian systems the conditional probability density given by equation [2.32] is intractable and sub-optimal policies or approximations are required to obtain an estimate of the state. For linear systems, the aforementioned Kalman filter provides a convenient means of updating the conditional density using the properties of the Gaussian distribution. In highly non-linear problems, it may be desirable that the system be modelled continuously between observations. Defining $\hat{x}_{t_i+s}$ as the estimate of the underlying state at time $(t_i + s)$, the system expression that describes the evolution of the state is utilised as $\hat{x}_{t_i+s} = f(\hat{x}_i, t_i, (t_i + s))$ for $0 < s < (t_{i+1} - t_i)$.

2.4.4 The Kalman filter

There are a number of different ways of representing and deriving the Kalman filter and the approach described here is based on the propagation of the conditional distribution. The Kalman filter may be applied to any linear system model in the state space form, see Harvey (1989) and Aoki (1967) for extensive information on the Kalman filter in the probabilistic framework and the augmentation of state expressions to produce a model in the state space form.

Although the various definitions of state in this thesis are univariate, here we present the matrix form of the linear system expressions and the Kalman filtering equations in order to provide the appropriate background for the development of approximate Kalman filters applied to non-linear systems in chapter 9. Consider the following observation expression in the state space form for a system observed at equidistant time points;

$$\underline{y}_i = \underline{H}_i \underline{x}_i + \underline{e}_i \qquad [2.34]$$

where, the vector $\underline{y}$ has N elements and $\underline{H}$ is a time dependent matrix of dimension $(N \times m)$. The underlying state is then defined as a first-order m-vector Markov process

$$\underline{x}_{i+1} = \underline{F}_i \underline{x}_i + \underline{G}_i \underline{u}_i + \underline{v}_i \qquad [2.35]$$

where, $\underline{F}$ is an $(m \times m)$ time dependent matrix. The contributions $\underline{v}_i$ and $\underline{e}_i$ are N-vector white noise processes that are assumed to be mutually independent and the $(a \times 1)$ vector $\underline{u}$ is a control input such as, the effect on the system state of a maintenance intervention with a time dependent matrix $\underline{G}$ of order $(m \times a)$. It is assumed that the control input is to some extent within the control of the user and that the effect is either known or determined uniquely from knowledge of the observations. However, one cannot always be certain of the effects that may arise as a result of actions taken and attempting to quantify the impact of maintenance procedures can prove difficult. The control is usually derived subject to some criterion function and is often used in control theory to obtain some type of balance when deviations in the estimated value of the state are occurring, see Aoki (1967). The units of the control function are likely to be shared by the state that is the focus of the particular filtering application. For instance, if the state of the system is taken to be the residual time of a defect in the system before it results in failure, the

corrective action taken at maintenance interventions may result in a decrease in the systems virtual age and as such, an increase in the residual life would be expected. The notation t_i is defined as the time of the ith iteration of the filtering process that, in the context of a maintenance process, could represent an intervention or monitoring point. Assuming that $\underline{Y}_j$ represents all the available information at t_j including any control actions taken and that the objective is to ascertain the state at t_i, the problem is one of prediction when $t_j < t_i$, filtering or estimation when $t_j = t_i$ and smoothing or hindsight when $t_j > t_i$. The covariance of the estimation error is given by

$$\underline{P}_{i|j} = E\left[\underline{x}_i - \underline{\hat{x}}_{i|j}\right]\left[\underline{x}_i - \underline{\hat{x}}_{i|j}\right]^T \qquad [2.36]$$

where, $\underline{\hat{x}}_{i|j}$ represents the optimal estimate. As noted previously, the LLMS estimate and the conditional mean propagation approach derive identical results for this scenario under different assumptions. This is due to the fact that, the LLMS estimate is reliant on the assumption of white noise disturbances and coincides with the conditional mean for Gaussian distributed data. Here, we consider the system described by equations [2.34] and [2.35] and assume that both the measurement and system disturbances are 0-mean Gaussian white noise processes and mutually independent as

$$\mathbf{E}\begin{pmatrix}\underline{v}_i\\ \underline{e}_i\end{pmatrix}\left(\underline{v}_i^T, \underline{e}_i^T\right) = \begin{pmatrix}\underline{R}, & \underline{0}\\ \underline{0}, & \underline{Q}\end{pmatrix} \qquad [2.37]$$

For most of the estimation approaches used in this research, some prior knowledge of the initial state or condition of the system is required and it is often a requisite that this initial state has no relationship with either of the noise sequences included in the model. With this particular case, we assume the initial state to be normally distributed as $\underline{x}_0 \sim N(\underline{\hat{x}}_0, \underline{P}_0)$. The conditional distributions utilised for state

estimation and prediction purposes, given the available information, are also taken to be Gaussian;

$$p(\underline{x}_i \mid \underline{Y}_i) \sim N(\hat{\underline{x}}_i, \underline{P}_i) \qquad [2.38]$$

$$p(\underline{x}_{i+1} \mid \underline{Y}_i) \sim N(\hat{\underline{x}}_{i+1|i}, \underline{P}_{i+1|i}) \qquad [2.39]$$

and the parameters are obtained using the Kalman filtering process. The expressions are presented in two stages for prediction and updating purposes. The prediction equations that are used between observations are

$$\hat{\underline{x}}_{i+1|i} = \underline{F}_i \hat{\underline{x}}_i + \underline{G}_i \underline{u}_i \qquad [2.40]$$

$$\underline{P}_{i+1|i} = \underline{F}_i \underline{P}_i \underline{F}_i^T + \underline{R} \qquad [2.41]$$

The updating equations upon observing the next piece of information are

$$\hat{\underline{x}}_i = \hat{\underline{x}}_{i|i-1} + k_i^f [\underline{y}_i - \underline{H}_i \hat{\underline{x}}_{i|i-1}] \qquad [2.42]$$

$$\underline{P}_i = \underline{P}_{i|i-1} - k_i^f \underline{H}_i \underline{P}_{i|i-1} \qquad [2.43]$$

where, the Kalman gain function is

$$k_i^f = \underline{P}_{i|i-1} \underline{H}_i^T [\underline{H}_i \underline{P}_{i|i-1} \underline{H}_i^T + \underline{Q}]^{-1} \qquad [2.44]$$

The system is initialised using the mean and variance of the prior distribution for the initial state. Parameter estimation for the Kalman filter is undertaken using the conditional probability of observing each piece of information to formulate the joint density. The parameter values are then obtained using the maximum likelihood approach discussed earlier in this chapter. See Harvey (1989) and a furnace erosion prediction case study in Christer et al (1997) for more details of the parameter estimation process for the discrete Kalman filter.

As noted previously, when considering non-linear systems (as all the scenarios modelled in subsequent chapters are), sub-optimal schemes and approximations are

required for estimation and prediction of the underlying state. Applying an extended Kalman filter (EKF) to non-linear systems essentially involves applying the standard Kalman filter to linearised versions of the non-linear systems (see chapter 9 for details). There are many variations on the EKF in the literature. For instance, the 'iterated' EKF is designed to consider situations where it is not obvious what the relevant linearisation point is when computing the Kalman gain function. The process involves iterating over the measurement equation and the iteration means that the linearisation point is changed. In some situations, this modification can result in improved performance of the filter, (Jazwinski, 1970).

Another variation on the basic EKF is a 2^{nd}-order EKF where, the recursive procedure deals with the 2^{nd}-order terms in the Taylor series expansion of the state expression. See chapter 9 for details. The resultant equations contain quadratic terms that are replaced by their expected values. Gaussian sum estimators are frequently described in the literature available on approximate non-linear state estimation techniques. The complexity associated with the recursive computation of the conditional density functions is simplified through approximation. The method involves approximating an arbitrary density function by a weighted sum of Gaussian distributions. The process of propagating the conditional means and covariance matrices involves applying a number of EKF's simultaneously and weighting the respective output. To ensure that the covariance matrices are small, it is a general requirement that a large number of filters are used in achieving the best possible approximation. The primary approach pursued in the research presented here is the probabilistic approach whereby, the characteristics of the various relationships and the associated error processes are described by probability distributions. For the examples considered in chapters 7 and 8, a closed form solution to the filtering

problem is attainable however, an analytical means of solving the expression is not and approximate or numerical techniques are required.

2.5 State estimation for continuous time stochastic systems

2.5.1 Stochastic calculus and Brownian motion

In this section, we are concerned with modelling the state of a system or a component as a continuous-time stochastic process $\{X_t\}_{0\le t<\infty}$. A realisation of the process X is called a sample path and is usually continuous however, jump discontinuities are viable on the condition that the functions involved are right-continuous. A necessary component of much of the continuous time non-linear filtering theory is Brownian motion. A Brownian motion is a process that has independent and Gaussian distributed increments with a mean of zero and an incremental variance that is proportional to the size of the increment. This property is attributable to the fact that each increment is comprised of many smaller, independent sub-intervals. Brownian motion is utilised in Ito processes (or diffusions) which are the primary tools in stochastic calculus. An Ito process is described by a stochastic differential equation

$$dX_t = U_t dt + V_t dW_t \qquad [2.45]$$

where, $\{W_t\}$ is a Brownian motion. For a small time increment Δt, we have

$$X_{t+\Delta t} - X_t \sim N(U_t \Delta t\ , V_t^2 \Delta t) \qquad [2.46]$$

where, the Gaussian distribution is conditioned on the processes W, U, V and X over the interval $[0,t]$. The process U represents the rate of change in X and V represents the level of random diffusing. Both U and V are assumed to be non-anticipating or adapted processes. The adaptive property means that they are not reliant on the outcome of the Brownian motion after time t although, they may be dependent on it

until that point. With V being an adapted process, the stochastic integral of X can be given with respect to a Brownian motion W as

$$X_t = \int_0^t V_s \, dW_s = \lim_n \sum_{i=1}^n V_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}) \qquad [2.47]$$

The limit is very complex in most situations and indeed often does not exist. However, the limit always exists in quadratic mean or if the adapted process has bounded variation. The Ito process given by equation [2.45] can be expressed as a sum of stochastic integrals as

$$X_t = X_0 + \int_0^t U_s \, ds + \int_0^t V_s \, dW_s \qquad [2.48]$$

For $X_t = f(t, W_t)$, partial differentials are taken to ascertain the change in the rate of the process as

$$dX_t = f_t(t,W_t)dt + f_w(t,W_t)dW_t + \frac{1}{2} f_{ww}(t,W_t)dt \qquad [2.49]$$

and this is known as Ito's lemma. Equation [2.49] can be written in the form of a stochastic integral as

$$X_t = X_0 + \int_0^t f_w(s,W_s)dW_s + \int_0^t \left(f_t(s,W_s) + \frac{1}{2} f_{ww}(s,W_s) \right) ds \qquad [2.50]$$

where, the final term is included because even squared increments of a Brownian motion can have an impact on the overall process. Another integral form that is available for stochastic processes is the Stratonovich integral which can be expressed in terms of the Ito integral. Although, it is a useful approach for problems involving stochastic differential equations, it lacks some of the necessary properties that are required for stochastic filtering, see Krishnan (1984) for details.

The continuous observation process $\{Y\}$ is stochastically related to the state as

$$Y_t = h(t, X_t) \qquad [2.51]$$

The change in the observation process is described by

$$dY_t = h_t(t, X_t)dt + h_x(t, X_t)dX_t + \frac{1}{2}h_{xx}(t, X_t)(dX_t)^2 \qquad [2.52]$$

where, $(dX_t)^2 = V_t^2 dt$ is the quadratic variation of X. In keeping with the description of the observation process, a stochastic differential equation for a general Ito process X_t is given by

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t \qquad [2.53]$$

for functions f and g that are dependent, in this context, only on the current value of the process X. Meaning that the process is of a Markovian nature where, f and g are analogous to the transition probabilities of a Markov chain. A further important element of the stochastic calculus discussed here are martingale processes. A process X_t is a martingale if for $t < u$, $\mathbf{E}[X_u \mid X_s, 0 \le s < t] = X_t$ and as such, the expectation is $\mathbf{E}[X_t] = \mathbf{E}[X_0]$ for all t. An Ito process is a martingale if it satisfies the property $dX_t = V_t dW_t$. Importantly, Brownian motion is itself a martingale and can be expressed as

$$X_t = \int_0^t V_s dW_s \approx \sum_{i=1}^{n} V_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) \qquad [2.54]$$

where, the expectation is $\mathbf{E}[X_t] = 0$ and $\mathbf{E}[X_t^2] = \int_0^t \mathbf{E}[V_s^2]ds$.

2.5.2 Continuous time filtering preliminaries

Liptser & Shiryaev (1989) show that, if the state is governed by a stochastic differential equation and the process describing the evolution of the state over time is of the diffusion type, the diffusion process has an equivalent Ito process

representation. An Ito process can be represented as a diffusion process relative to the innovations process V. In the filtering representations discussed here, the innovations process is a Brownian motion process that represents the new information that is available and consists of the differences between what is expected to be observed and what is actually observed. The innovations process is derived with respect to the σ-field generated by the observation process $\mathcal{F}_t$ and is crucial when deriving the non-linear filtering representations. The innovations process is an $\mathcal{F}_t$-martingale, this property is implied by the fact that the innovations process is assumed to be of the Brownian motion type. Doob's decomposition theorem also illustrates this property and Kallianpur (1980) gives a more in depth discussion of the innovations process.

We are concerned with the estimation of the state $\{X_t, t \in T\}$ of a system with respect to a σ-field. As with the problem scenario for the discrete time case in section 2.4, the state is not observable directly and must be ascertained via an observation process that is assumed to be correlated with the state. In contrast to the discrete time case, we are interested in updating our knowledge of the state continuously as observations arrive continuously. The state estimates are derived using the observation process and an optimal criterion such as the MSE function. As with the discrete systems, the conditional expectation of the state provides an optimal estimate for most criterion functions however, the expectation will in general be a non-linear function of the observations. The issue is further complicated by the fact that both the state and observation process are assumed to be governed by stochastic differential equations and the resultant expression for the conditional expectation of the state will also be a stochastic differential equation. For non-linear cases, the conditional mean is dependent on higher moments of the conditional distribution and

as such, approximations or sub-optimal policies are required. The linear scenarios require only the second moments of the conditional density and therefore the evaluation of the conditional mean is a tractable problem. Analogous to the treatment of discrete time systems, the estimation procedure for linear systems can be derived within the framework of a general non-linear model, see Krishnan (1984). Using the properties of martingales, a closed-form representation for the general non-linear stochastic system is given by Liptser, Krishnan (1984) and Kallianpur (1980) and some approximation is required to obtain the estimate. Under general conditions, any martingale can be given as a stochastic integral. Krishnan (Theorem 8.4.1) shows that a square integrable martingale can be given as a stochastic integral with respect to the innovations process and it is this fact that makes the closed-form representation possible.

2.5.3 Non-linear filtering for continuous time systems

We are considering the complete probability space $(\Omega, \mathcal{F}, P)$ and are attempting to compute the least squares estimate of the state of the system $\{X_t, t \in T\}$ given the availability of the current value and history of the observation process $\{Y_s, s \leq t, t \in T\}$ and as noted previously, this involves finding the conditional expectation of the state given the σ-field generated by the observation process $\{\mathcal{F}_t, t \in T\}$. Further, we require that the conditional expectation be updated recursively and continuously. The state of the system adheres to the process $\{X_t, \beta_t, t \in T\}$ and is taken to be an Ito process defined on the complete probability space as

$$X_t = X_0 + \int_0^t f_\tau \, d\tau + W_t \qquad [2.55]$$

The observation process $\{Y_t, \beta_t, t \in T\}$ is also an Ito process and is defined on the complete probability space as

$$Y_t = X_0 + \int_0^t h_\tau \, d\tau + V_t \qquad [2.56]$$

where, unique solutions exist for both the state and observation at time t and X_0 is an arbitrary initial condition that is assumed to be independent of all processes involved in the system equations, $\{\beta_t, t \in T\}$ is the filtration σ-field defined by $\beta_t \supset \sigma\{X_0, X_s, W_s, Y_s, V_s, s \le t, t \in T\}$ and $\mathcal{F}_t = \sigma\{Y_s, s \le t, t \in T\}$ is the filtration σ-field generated by the observation process, where $\mathcal{F}_t \subset \beta_t$. The state and the observation process are defined as semi-martingales on the σ-field β_t, and f_t and h_t may be functions of X_t and are β_t-measurable. Finally, W_t is a general right-continuous martingale process and $\{V_t, \beta_t, t \in T\}$ is a Brownian motion process with parameter σ_v. The innovations process v_t has the same statistics as V_t and is given by

$$dv_t = dY_t - \hat{h}_t dt \qquad [2.57]$$

where, the σ-field generated by the innovations process is equal to the σ-field generated by the observation process. The same representation for the non-linear filtering theorem can be derived without this assumption but the derivation is much more complex.

The conditional estimate for the state of the system with respect to the σ-field at time t is given by

$$d\hat{X}_t = \hat{f}_t dt + \frac{1}{\sigma_v^2}\left[\mathbf{E}^{\mathfrak{I}_t} \frac{d}{dt}\langle W, V \rangle_t + \mathbf{E}^{\mathfrak{I}_t}(X_{t\_} h_t) - \mathbf{E}^{\mathfrak{I}_t} X_{t\_} \mathbf{E}^{\mathfrak{I}_t} h_t \right] dv \qquad [2.58]$$

or as a sum of integrals as

$$\hat{X}_t = \hat{X}_0 + \int_0^t \hat{f}_\tau d\tau + \frac{1}{\sigma_v^2}\int_0^t \left[\mathbf{E}^{\Im_\tau}\left(X_{\tau-} h_\tau - \hat{X}_{\tau-}\hat{h}_\tau + \frac{d}{d\tau}\langle W,V\rangle_\tau \right)\right] dv_\tau \qquad [2.59]$$

where, $\langle W,V\rangle_t$ is the quadratic covariance between W and V. However, as noted previously, this expression is generally intractable except in the linear Gaussian case. See Krishnan (1984) and Kallianpur (1980) for proofs of the filtering theorem. A useful modification is to consider the error quantities $\tilde{X}_\tau = \hat{X}_{\tau-} - X_{\tau-}$ and $\tilde{h}_\tau = \hat{h}_{\tau-} - h_{\tau-}$ that enable the representation

$$d\hat{X}_t = \hat{f}_t dt + \frac{1}{\sigma_v^2}\left[\mathbf{E}^{\Im_t}\left(\hat{X}_t\hat{h}_t + \frac{d}{dt}\langle W,V\rangle_t \right)\right] dv_t \qquad [2.60]$$

If W and V are independent martingales then $\langle W,V\rangle_t = 0$ and the filtering expression becomes

$$d\hat{X}_t = \hat{f}_t dt + \frac{1}{\sigma_v^2}\mathbf{E}^{\Im_t}(\hat{X}_t\hat{h}_t)\, dv_t \qquad [2.61]$$

which is a useful representation for many cases. Krishnan (1984), Kallianpur (1980) and Liptser & Shiryaev (1989) present very similar treatments of the filtering problem from a martingale perspective. Engineering applications including Koch (1986) tend to utilise a Poisson counting process with independent positive increments as

$$N_t = N_0 + \lambda t + m_t \qquad [2.62]$$

where, N_t is the number of events and m_t is a martingale.

2.5.4 An alternative approach

An alternative approach to the continuous conditional estimation problem involves partitioning the range [0, t] into sub-intervals $\Delta_0 = 0$, Δ_1, Δ_2, ..., $\Delta_n = t$. Then, defining $\alpha = \max_i(\Delta_{i+1} - \Delta_i)$, the objective is to establish

$$p(X_t \mid \underline{Y}_t) = \lim_{\substack{\alpha \to 0 \\ n \to \infty}} p(X_t \mid \underline{y}_{\Delta_0}, \underline{y}_{\Delta_1}, \underline{y}_{\Delta_2}, ..., \underline{y}_{\Delta_n}) \qquad [2.63]$$

for the continuous information path $\underline{Y}_t = \{\underline{y}_s \; ; 0 \le s \le t\}$. Redefining the observation expression in a manner consistent with the probabilistic framework developed in section 2.4.3, we have

$$d\underline{y}_t = h(X_t, t) + d\underline{\eta}_t \qquad [2.64]$$

where, $\mathbf{E}(d\underline{\eta}_t d\underline{\eta}_t^{\,T}) = Q(t)dt$ for $Q(t) > 0$, and $p(X_0)$ is assumed known and independent of $\underline{\eta}_t$. To establish a conditional density for the underlying state at time t, the approximation

$$p(X_t \mid \underline{Y}_t) = \frac{E\left\{\exp\left[-\frac{1}{2}\int_0^t h(X_s,s)^T Q^{-1}(s) h(X_s,s) ds + \int_0^t h(X_s,s)^T Q^{-1}(s) d\underline{y}_s\right] \mid X_t\right\} p(X_t)}{E\left\{\exp\left[-\frac{1}{2}\int_0^t h(X_s,s)^T Q^{-1}(s) h(X_s,s) ds + \int_0^t h(X_s,s)^T Q^{-1}(s) d\underline{y}_s\right]\right\}} \qquad [2.65]$$

can be used where, the expectation is taken over $\{X_s \; ; 0 \le s \le t\}$.

Chapter 3. Recognising and measuring the potential for human error at maintenance interventions using delay time modelling

Within the context of Delay Time modelling, we investigate an approach whereby the injection of defects at maintenance maybe ascertained from basic inspection and failure data. The model is developed in the context of a competing risks scenario and the objective of the research is not to assist management in optimising existing situations that incorporate substandard procedures but rather, the intention is to highlight the existence of such inefficiencies and to demonstrate the benefits that may be achieved through improved practice. Although the model is constructed with the provision for human error based fault injection in mind, the structure of the model does not limit its application to situations that incorporate specifically human error. The form of the model is appropriate for scenarios incorporating any kind of fault injection. A number of cases are investigated using simulated data to test a methodology for establishing the existence and indeed the level of potential human error injected faults. This entails the selection of an appropriate form for the model and accurate estimation of the necessary parameters.

Initially, a review of the relevant background and supporting literature is presented. The basic delay time model and the associated parameter estimation approach are then introduced with a view to extending the modelling and estimation for the cases incorporating human error. The impact of potential fault injection on the resulting downtime modelling phase is discussed and the methodology adopted for simulating the process and obtaining the necessary failure and preventive maintenance (PM) repair information is introduced. Some modelling options are proposed and implemented on the simulated data sets and the ability to differentiate between a process with and without human error at PM using the suggested techniques is

assessed. The model selection approach is then extended and tested for comparison of models incorporating fault injection and models incorporating fallible detection of existing defects at PM. The chapter concludes with a discussion of general modelling recommendations, potential extensions to the work and the limitations of the modelling approach encountered.

3.1 Literature review

3.1.1 Modelling literature

For the research documented in this chapter, delay time modelling is used to represent the failure and inspection process for complex dynamic systems. The delay-time concept was first introduced in an appendix to Christer (1973) and then in a cost based decision model utilising subjective estimation in Christer (1982). The first formal presentation of the delay-time model for systems in a steady state of operation was given in Christer & Waller (1984a). Initially, the model was developed for systems with a homogenous fault arrival rate and then extended for the non-homogenous case. The paper also addressed the issue of non-perfect detection of existing faults at inspection. Industrial case studies include the application of delay time modelling to a canning line, Christer & Waller (1984b), and to the maintenance of coal mining equipment, Chillcott & Christer (1991).

In this chapter, objective parameter estimation techniques are applied to failure times and data collected at inspections in order to characterise the proposed delay time models. Baker & Wang (1991, 1993) introduced the objective approach for estimating the parameters of a delay time model when applied to a single-component system and Christer et al (1995) applied delay-time modelling to the maintenance of a copper production plant with objective estimation being used to ascertain the parameters for a complex system representation. In situations where incomplete data

sets only permit the accurate estimation of some of the model parameters using objective methods, subjective input is required to establish the model of the system. Christer & Waller (1984a) discussed the subjective estimation of delay time distributions using expert opinion obtained from experienced engineers. Further work on the subjective engineering-based estimation of delay time distributions can be found in Wang (1997) and Christer & Redmond (1992) presented objective parameter updating techniques for subjectively estimated delay time models. The maximum likelihood estimation of optimal inspection intervals is addressed in Baker et al (1997) and Christer et al (1998, 2000) considered parameter estimation problems with either limited or deficient data sets. Christer & Wang (1992, 1995) developed a delay-time model to represent the condition monitoring of a plant for the single component case and subsequently, a multi-component system. Reviews of the developments in delay time modelling can be found in Baker & Christer (1994) and Christer (1999).

With regard to the type of physical process and preventive action under investigation, Ascher & Feingold (1984) present an extensive account of maintenance techniques for repairable systems. When modelling the impact of potential fault injection in this chapter, the delay time model is presented in the context of the competing risks model. See Bedford & Cooke (2003) and Crowder (2001) for information on competing risks and the associated problems of parameter identifiability. Counting processes are used to model the system failures as a stochastic process and a thorough treatment of the Poisson process, and variations of, can be found in Ross (1983). Barlow & Hunter (1960) give the initial presentation of the non-homogenous Poisson process (NHPP) and Barlow & Proschan (1965) showed that for complex

systems incorporating negligible failure repair times, the type of failure process that we consider in section 3.2 follows a NHPP in the limiting steady-state case.

3.1.2 Literature on fallible maintenance

Many studies have highlighted the presence of human error related defect injections at maintenance interventions. Steedman & Whittaker (1973) estimated that in a particular ICI plant, up to 30% of system failures were directly attributable to defects injected at some point during the course of the previous PM. An ASRS air transport report, Patankar & Taylor (2003), claimed that up to 40% of defects that are present in an aircraft at any given time are due to the unintentional release of further errors during substandard inspection and repair procedures. It is feasible that the inspection or repair process for existing faults could result in the accidental injection of further and potentially more severe defects. For instance, Jia et al (2002) encountered a case where one specific error-prone maintenance procedure consistently produced defects that subsequently resulted in a system failure. In cases such as this, it may be beneficial to reduce the level of maintenance or indeed forgo it altogether and rely solely on breakdown maintenance. Alternatively, the modelling and identification of a defective process can reveal areas for potential improvement. In a limited PM data and 'selective repair' case, Christer et al (1998), the modelling process revealed defects that could potentially be removed and the resulting improvement in maintenance procedures produced downtime savings of approximately 15%. This illustrates the potential benefits of modelling human error, and although the focus of that particular study was on the failure to identify and remove existing defects, the quantities and levels of reduction in downtime are of interest.

Similar scenarios and alternative modelling solutions for problems that include fallible maintenance can be found in the following references; Kaio & Osaki (1989)

present a comparison of inspection policies for a single component that may be detected as failed only by potentially fallible inspections, Jack (1991) investigates a scenario with non-perfect inspection repairs and a finite planning horizon, Makis & Jardine (1992) present a replacement model which again facilitates the existence of human error in the form of imperfect defect repair, Parmigiani (1996) looks at the scheduling of fallible inspections where, the inspections take the form of time consuming tests with two possible test types considered namely, fallible and error-free, and Dagg & Newby (1998) consider a scheduling problem with imperfect inspection and repair for three system states; good, faulty and failed and a Markov structure is utilised in the computation of average costs that are subsequently used to determine the optimal number of inspections before overhauling. There are numerous cases in the maintenance literature involving poor data detection capabilities, McKone & Weiss (1998) and Baker & Wang (1992) are just a couple of examples. However, although the existence of fault injection based human error in maintenance procedures is recognised as being common-place, few modelling-based studies have considered the actual creation of defects as a direct consequence of the inspection process or indeed the potential for poor quality maintenance resulting in the insertion of further defects into the system.

3.2 Modelling and analysis

This research is concerned with industrial plants that are subject to regular periodic inspections. The aim is to incorporate the potential for fault injection during routine planned preventive maintenance (PM) into the modelling process and to determine whether we can recognise and quantify the level of injection when it is indeed present. Also, by optimising resulting downtime or cost models with and without human error, we can then quantify any potential benefits, such as a reduction in

downtime that may be achieved through improved maintenance practice. Carr & Christer (2003) discussed how the expected number of defect injections could subjectively be estimated under certain convenient assumptions and illustrated how an existing maintenance process could be optimised using downtime or cost control functions. The downtime functions utilise the interval between inspections as the key decision variable. The potential benefits of improving maintenance procedures can then be demonstrated by comparing the optimal maintenance policies obtained using delay time models with and without the facilitation for defect injection at PM. In this chapter, we investigate the ability to characterise, through appropriate model specification and parameter estimation, the fault arrival process, the fault injection process and the resulting failure process from objective failure data. We require the approximate failure times where for instance, they may be recorded as occurring 'n' hours/days/weeks etc after the last inspection. The primary objective is the identification of human-error based fault injection when it is not necessarily known to be taking place.

3.2.1 The basic delay time model

The basic fault, failure and inspection process for defects arising naturally during standard operation of the system is illustrated in figure 3.1 and provides the framework upon which this research is developed.

Figure 3.1 - Illustrating the failure process for defects arising during the course of operation

Figure 3.1 illustrates the alternative outcomes for a defect, consisting of a failure later in the same cycle or detection and removal at the subsequent PM. The process can be represented by the basic delay time model, see Christer (1999), and is subject to the following assumptions;

- Defects are assumed to arise during standard operation according to a homogenous Poisson process (HPP) with a constant arrival rate k. As such, the expected number of faults arising over a regular inspection interval $((i-1)T, iT)$ is kT.
- The delay time h of a standard fault is assumed to be independent of its time of origin u and is governed by a probability density function $f(h)$.
- Defects detected at inspection are repaired within the PM interval d.
- Upon failure of the system, repairs are initiated immediately and only the defect that has resulted in the particular failure is attended to. All other defects that are present in the system remain untouched.
- The plant or equipment being modelled is in a steady state, i.e. has been in a similar operational state for a substantial amount of time. This is reflected in the constant fault arrival rate k.

Once an appropriate form has been selected for the delay time distribution, $f(h)$, we are required to estimate the parameters of the delay time model. There are both subjective approaches, see Christer & Waller (1984b) and Wang (1997), and objective approaches, see Baker & Wang (1992) and Christer et al (1995), that are available for parameter estimation purposes. $\mathbf{E}[N_f((i\text{-}1)T, iT)]$ is defined as the expected number of failures over the interval between inspections $((i-1)T, iT)$ and is given by

$$\mathbf{E}[N_f((i-1)T,iT)] = \mathbf{E}[N_f(0,T)] = \mathbf{E}[N_f(T)] = k\int_0^T F(T-u)du \qquad [3.1]$$

The steady-state assumption implies a consistency in operational conditions. As a result, we use the notation $\mathbf{E}[N_f(T)]$ in preference to $\mathbf{E}[N_f((i-1)T,iT)]$ as all the intervals can be treated equally from a statistical perspective. Another factor in establishing this property is the fact that, from a fault detection perspective, we initially assume a perfect inspection process to be in place and as a result, any failures that arise in a given interval are assumed to be attributable to defects that originated within the same interval. Therefore, we can classify each operational interval $((i-1)T, iT)$ as independent. For fault arrivals that adhere to a homogenous Poisson process (HPP), the resulting failure process follows a non-homogenous Poisson process (NHPP);

$$N_f((i-1)T,iT) \sim \text{Poisson}\{\mathbf{E}[N_f(T)]\} \qquad [3.2]$$

Ross (1983) states that the following properties characterise a non-homogenous Poisson process (NHPP);

1. $N_f(0) = 0$,
2. $N_f(t)$ has independent increments for $t \geq 0$,
3. $P(N_f(t,t+dt) = 1) = r(t)dt + o(dt)$ where, $r(t)$ is a time-dependent rate function,
4. $P(N_f(t,t+dt) > 1) = o(dt)$.

Properties 1, 2 and 4 are easily verified in practical situations. Addressing property 3 for the basic delay time model, we have

$$r(t) = \frac{d}{dt}\mathbf{E}[N_f(t)] = k\int_0^t f(t-u)du + kF(0) = k\int_0^t f(t-u)du \qquad [3.3]$$

The Poisson process properties are very useful as they enable the construction of the probabilities associated with observing a specific number of failures in a given interval. These probabilities are then utilised to formulate the maximum likelihood expression that is used to estimate the required parameters. A similar result applies to the number of faults detected and removed at the ith inspection (PM); $N_p(iT) \sim \text{Poisson}\{\mathbf{E}[N_p(T)]\}$. Under steady state conditions, $\mathbf{E}[N_p(T)]$ represents the expected number of faults found and removed at any given inspection;

$$\mathbf{E}[N_p(iT)] = \mathbf{E}[N_p(T)] = kT - k\int_0^T F(T-u)du \qquad [3.4]$$

As the observations are independent, the likelihood of observing the given data set is just the product of the Poisson probability of observing each cycle of data where, m_i is the number of failures observed in the ith interval and j_i is the number of faults removed at the ith PM. The reasons for selecting the maximum likelihood approach for estimating the parameters (regarding the asymptotic properties of the estimator) are discussed in chapter 2. The likelihood function for L intervals of data is

$$\mathcal{L} = \prod_{i=1}^{L}\left\{P(N_f((i-1)T, iT) = m_i)\,P(N_p(iT) = j_i)\right\}$$

$$= \prod_{i=1}^{L}\left\{\left(\frac{e^{-E[N_f(T)]}E[N_f(T)]^{m_i}}{m_i!}\right)\left(\frac{e^{-E[N_p(T)]}E[N_p(T)]^{j_i}}{j_i!}\right)\right\} \qquad [3.5]$$

An alternative likelihood function can be derived that utilises the exact failure times, see Christer (1999), however in practical situations, data is rarely recorded in such a precise fashion. As such, we opted to pursue the likelihood formulation described here. The likelihood function is maximised with respect to the parameters to obtain the estimated values. The optimisation process can be simplified by taking natural logarithms of the likelihood function as

$$l = \left(\sum_{i=1}^{L} m_i\right)\log(\mathbf{E}[N_f(T)]) + \left(\sum_{i=1}^{L} j_i\right)\log(\mathbf{E}[N_p(T)]) - L(\mathbf{E}[N_f(T)] + \mathbf{E}[N_p(T)])$$

$$\ldots + \sum_{i=1}^{L}(\log(m_i!) + \log(j_i!)) \qquad [3.6]$$

Omitting the constant terms that are not a function of any of the parameters under investigation and inserting the expressions for $\mathbf{E}[N_f(T)]$ and $\mathbf{E}[N_p(T)]$ gives

$$l = \left(\sum_{i=1}^{L} m_i\right)\log\left(k\int_0^T F(T-u)du\right) + \left(\sum_{i=1}^{L} j_i\right)\log\left(kT - k\int_0^T F(T-u)du\right) - LkT \qquad [3.7]$$

We can obtain an estimate for k by taking the partial differential

$$\frac{\partial l}{\partial k} = \frac{\sum_{i=1}^{L} m_i \int_0^T F(T-u)du}{k\int_0^T F(T-u)du} + \frac{\sum_{i=1}^{L} j_i\left(T - \int_0^T F(T-u)du\right)}{k\left(T - \int_0^T F(T-u)du\right)} - LT \qquad [3.8]$$

Setting $\partial l / \partial k = 0$ and re-arranging for k, we obtain

$$\hat{k} = \frac{\sum_{i=1}^{L}(m_i + j_i)}{LT} \qquad [3.9]$$

which we could have arrived at by induction. However, the derivation serves to illustrate the parameter estimation process that is built upon in subsequent sections. Each event, whether a failure or defect removed at inspection, represents the outcome of one defect and as we are only interested in the average rate of arrival, it is obvious that this is given by

$\hat{k}$ = (Total number of fault arrivals) / (Total time over all cycles) [3.10]

which is a standard result in statistics when dealing with Poisson processes, see Ross (1983). The estimate of k can then be inserted into the likelihood function thus

easing the estimation process for the parameters of $f(h)$. However, with the introduction of faults injected at maintenance interventions, the derivation of k and the other parameters becomes more complex. Maximum likelihood estimation is not the only approach available for parameter estimation problems of this nature. As discussed in chapter 2, a number of alternative methods are available for parameter estimation purposes such as the least squares approach.

Using the established delay time model with the estimated parameters, a downtime control function can be constructed and then optimised for the decision variable T (the regular interval between inspections). An appropriate control function for the basic inspection model is the expected downtime per unit time and is given, under steady-state conditions, as;

$$D(T) = \frac{Expected\ Downtime}{Cycle\ Length} = (\mathbf{E}[N_f(T)]d_f + d)/(T + d) \qquad [3.11]$$

where, d_f is the average duration of a failure repair.

3.2.2 Modelling the injection of faults at PM

Extending the basic process, our objective is to model the potential for defect injection at maintenance interventions. Figure 3.2 illustrates the potential failure process for defects injected at inspection. The model is constructed in the context of a competing risks scenario (see Bedford & Cooke (2003)) which implies that if a system or piece of equipment can fail, it can usually fail in a number of ways. A problem that is typically associated with the modelling of competing risks is one of 'identifiability' and can result in an inability to identify the marginal distributions associated with the times to failures that are attributable to different sources of risk. The problem often manifests itself during the parameter estimation process as observed later in this chapter.

As with standard defects, the injected faults may result in failures within the same operational cycle or they may be captured at the next inspection. The outcome is dependent on the duration of their respective delay-time.

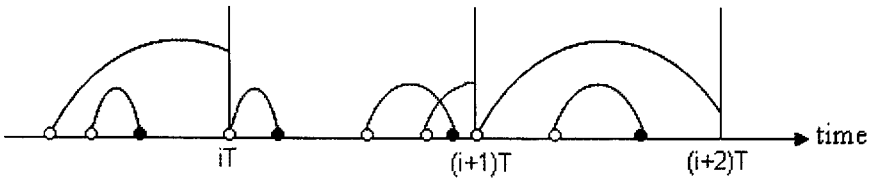

Figure 3.2 - Illustrating the failure process for standard faults and faults injected during PM

In addition to the considerations associated with the arrival and subsequent outcome of standard defects and the basic delay time model inspection process, we make the following additional assumptions regarding the injection of defects at PM;

- Inspections are not benign, in that defects may be injected as a direct result.
- Detection at inspection is perfect.
- Injected defects begin deteriorating at the start of the next operational cycle.
- The average number of defect injections at a maintenance intervention is v.

  The delay time of an injected defect has pdf $g(h)$. Note that, in principal we allow the delay-time distribution for injected defects to differ from that of standard defects. This implies that, although the defects are assumed to be of the same type, their severity may differ due to the fact that their creation is likely to be more abrupt and attributable to some form of system incursion. However, estimation becomes much more complicated when the distributions are allowed to differ.

As with the basic inspection process, a downtime control function can be constructed. The structure of the function is unchanged and the expected number of failures occurring over an interval $(0, T)$ is now discussed. In the context of competing risks, we have

$$\mathbf{E}[N_f(T)] = k\int_0^T F(T-u)du + \sum_{i=1}^{n} v_i G_i(T) \qquad [3.12]$$

for n potential sources of fault injection where v_i is the expected number from source i and $F(\cdot)$ and $G_i(\cdot)$ are cumulative distribution functions. Competing risks models are often constructed for scenarios where a number of different faults from a number of potential sources can cause a failure but, we only observe the first failure and the system/component is then renewed/replaced with all other existing faults also being attended to. In addition, the cause of the particular failure is often identifiable. For our particular scenario, we are only considering a single source of fault injection, the source of individual failures is not identifiable from an engineering perspective and we potentially have multiple failures in the interval between inspections with an unknown origin for each failure from which the parameters of our model must be estimated. The expected number of failures over $(0, T)$ is

$$\mathbf{E}[N_f(T)] = k\int_0^T F(T-u)du + vG(T) \qquad [3.13]$$

and the non-homogenous failure rate function is established as

$$r(t) = \frac{d}{dt}\mathbf{E}[N_f(t)] = k\int_0^t f(t-u)du + kF(0) + vg(t) = k\int_0^t f(t-u)du + vg(t) \qquad [3.14]$$

Therefore, as with the basic case without the provision for fault injection, the number of failures arising over an interval between inspections, $((i-1)T, iT)$, is Poisson distributed with the mean being the expected number of failures, $\mathbf{E}[N_f(T)]$, as

$$N_f((i-1)T, iT) \sim \text{Poisson}\{\mathbf{E}[N_f(T)]\} \qquad [3.15]$$

The number of defects detected and repaired at an inspection is also Poisson distributed with the mean being the expected number;

$$N_p(iT) \sim \text{Poisson}\{\mathbf{E}[N_p(T)]\} \qquad [3.16]$$

where, $N_p(iT)$ represents the number of existing faults that are detected and removed at the ith PM.

The potential benefits of this research can be seen in figure 3.3. Accurate estimation of the model parameters, including v, would enable the construction of the downtime control function (see equation [3.11]) and determination of the optimal inspection interval T^*, with and without the injection of defects at inspection for comparison. The parameters used for the demonstration in figure 3.3 are representative of values that have been observed in practical scenarios. We assume the average downtime for a failure to be $d_f = 0.5$ hours and for an inspection, we have $d = 0.35$ hours. The delay-time distribution is taken as negative exponential with $\lambda = 0.05$ for both faults created during production and those injected during the course of an inspection,

$$f(h) = g(h) = 0.05e^{-0.05h} \quad ; \; h \geq 0$$

The average fault arrival rate during standard operation is $k = 0.1$ per hour. The average number of defect injections v is taken to be a binomial random variable with 5 being the maximum number of potential fault injections. Each potential fault injection is independent and has an associated probability of $p = 0$, 0.3 and 0.7. Expected values are then obtained for the mean number of fault injections of $v = 0$, 1.5 and 3.5 respectively. The downtime expression is given by equation [3.11].

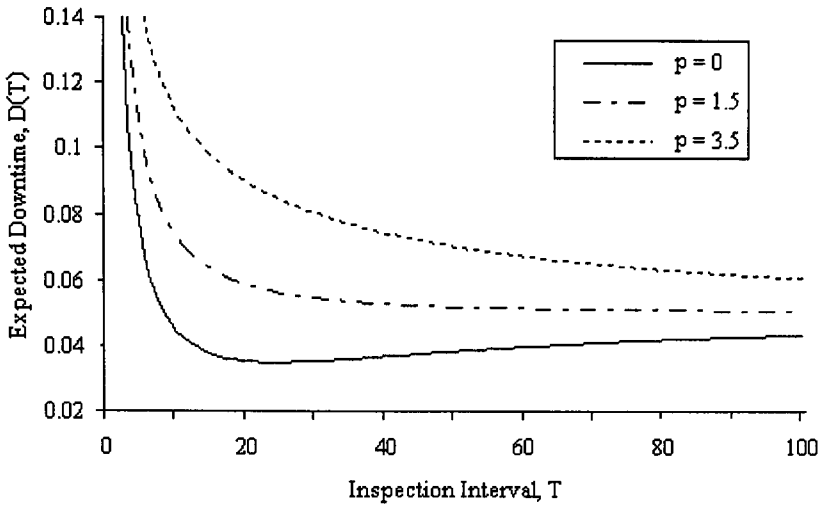

Figure 3.3 - The expected downtime under a prospective inspection interval where the maximum number of fault injections is 5.

As can be seen in figure 3.3, with probability 0.3 of injecting a defect at each of five statistically identical opportunities, the minimum expected downtime increases by over 70% when compared with the perfect PM case, where $p = 0$, which corresponds to the basic delay time model. As more faults are injected during inspections, the optimal region becomes flatter, and any change in the inspection interval within this region has little effect on the resulting downtime. However, with the introduction of an excessive number of faults, the behaviour of the downtime function alters to the extent that, an optimal solution is no longer available as the curve continuously decreases well beyond any reasonable range for setting a regular PM interval. In this situation, the recommended procedure would be to invest in improving the actual inspection process or to forgo inspections altogether in favour of a breakdown or 'contingency-repair' policy with an associated downtime $D(T) = kd_f$, which is the asymptotic expected downtime per unit time for all the models as $T \rightarrow \infty$. However, it is the parameters of $\mathbf{E}[N_f(T)]$ and their estimation from failure data that are the focus of this investigation. In a practical scenario, the construction of the downtime

(or cost) control function would form the final stage of the analysis and would be used to demonstrate the benefits of improving maintenance practice by preventing further human-error based defect injections at PM. As mentioned previously, we assume the plant to be in a steady state of operation and therefore, we assume the data to be reflective of this condition. The type of data we are concerned with primarily consists of; the number of recorded failures and their associated times within each operational cycle and the number of defects detected and subsequently repaired at each inspection. In a practical scenario, we would also have information on the duration of the interval between inspections T (assumed throughout this section to be a regular interval) and the number of cycles of data, represented by L.

3.2.3 Simulating a process with potential fault injection at PM

In our investigations into the ability to accurately estimate the necessary model parameters, we have used simulated sets of data. Simulating the data has a number of advantages; firstly, we can simulate many data sets and therefore try numerous runs and combinations of the model parameters to investigate the behaviour of the specified models structure, and secondly, given that we have specified the form of the delay time model and the parameters that are used to simulate the data, we already have knowledge of the actual underlying parameter values that we are hoping to recapture. The steps for simulating the arrival and subsequent outcome of standard defects are given as follows;

1. Generate the number of defects arising in an operational cycle as a Poisson random variable with the expected number being kT.

Each defect is now treated individually;

2. Assign a time of origin u uniformly over the cycle $((i\text{-}1)T, iT)$.

3. Assign a delay time h from the delay time distribution $f(h)$ using inversion techniques on the cumulative density function.
4. Extrapolate the outcome of the defect according to the time of origin and the associated delay time until failure. The outcome will be either a repair at the next inspection or a failure within the same cycle.

An analogous approach is utilised to simulate the failures and inspection repairs that arise as a consequence of fault injection at inspection. The number of injections is generated as a binomial random variable and the delay times are acquired from the delay-time distribution $g(h)$. The time of origin for each injected defect is the start of the subsequent operational cycle. As we have prior knowledge of the inspection interval, the parameters and the number of cycles of data specified for simulation, the following expression can be used to validate the simulated data set;

$$kT + v \approx \frac{M + J}{L} \qquad [3.17]$$

where, M is the total number of failures and J is the total number of repairs at PM. Also, by inserting the parameter values used for simulating the data into the expressions $\mathbf{E}[N_f(T)]$ and $\mathbf{E}[N_p(T)]$, the following approximate equalities are useful for validation of the simulation process,

$$\mathbf{E}[N_f(T)] \approx M/L \quad \text{and} \quad \mathbf{E}[N_p(T)] \approx J/L \qquad [3.18]$$

where, $\mathbf{E}[N_p(T)]$ represents the expected number of defects that are found and removed at any given inspection as

$$\mathbf{E}[N_p(T)] = kT - k\int_0^T F(T-u)du + v\big(1 - G(T)\big) \qquad [3.19]$$

In summary, the simulated data sets include the approximate failure times within specific increments of each interval and the number of faults removed at inspection.

3.2.4 Model specification and parameter estimation

It is a necessity when commencing modelling that we already have possession of the data. The modelling process begins after an initial analysis of the type of data we have at our disposal. We initially apply an entirely objective approach to the parameter estimation process using maximum likelihood estimation (MLE), see Baker & Wang (1992). The key steps to consider when attempting to estimate or recapture the required parameters from maintenance data using MLE are;

1. Collect the data into appropriate groups/intervals.
2. Determine the functional form for the expected number of failures that occur during the course of an operational cycle $\mathbf{E}[N_f(T)]$ and the expected number of defects detected and repaired at an inspection $\mathbf{E}[N_p(T)]$.
3. Select forms for the delay time distributions $f(h)$ and $g(h)$.
4. Determine the functional form of the likelihood function as a product of the probabilities associated with observing each piece of available information i.e. the number of failures that we observe in specified groups or intervals and the number of faults removed at inspections.
5. Insert the available objective data into the likelihood expression.
6. Maximise the specified likelihood function with respect to the parameters under investigation.
7. Use selection criterion such as the AIC (Akaike Information Criterion) to choose between potential model forms when the number of parameters under investigation differs.

The AIC provides a means of comparing the maximum likelihood values obtained for different candidate distribution forms and is derived on the assumption that the actual underlying fault arrival process and resulting failure process can be described

by a given delay time model if its parameters are suitably adjusted, see Akaike (1974). The AIC was developed around the concept of entropy and is an estimator based on the maximised log-likelihood function and corrected for asymptotic bias. The AIC provides an estimate of the expected, relative Kullback-Leibler information. The Kullback-Leibler information is a quantification of the meaning of information that is related to the concept of sufficient statistics. The AIC for a given likelihood function applied to a particular data set is

$$\text{AIC} = -2\log(\mathcal{L}) + 2q \qquad [3.20]$$

where, $\mathcal{L}$ is the maximum likelihood value for the formulation and q is the number of parameters under investigation. Naturally, we seek to minimise the AIC and as can be seen from the $2q$ component of the function, a penalty is applied for excessive parameterisation. A model with a larger number of parameters may be more tailored and hence provide a better fit to the data used to establish it, but is likely to be less flexible when applied to new data and the AIC takes this into consideration i.e. discourages over-fitting.

However, grouping the data for parameter estimation purposes is more complicated than the basic case. Essentially, the problem is that a range of values for k and v will satisfy the equality given by expression [3.17] and grouping the data according to the number of failures in each interval will not suffice. We must examine the behaviour of the failure process within the intervals to obtain the parameter estimates. Each interval $((i-1)T, iT)$ is partitioned into z non-overlapping, equidistant increments of length Δ as

$$(i\text{-}1)T + z\Delta = iT \qquad [3.21]$$

and we apply the following index for $j = 1, 2, \ldots, z$;

$$I_j^i = [(i-1)T + (j-1)\Delta,\ (i-1)T + j\Delta] \qquad [3.22]$$

If we are considering the limiting case (steady-state conditions), in terms of the associated probability, we expect the corresponding increment j from each interval i to exhibit similar behaviour as regards the number of observed failures. The definitions given in [3.21] and [3.22] reduce to the following for $j = 1, 2, \ldots, z$;

$$I_j = [(j-1)\Delta, j\Delta] \quad \text{and} \quad z\Delta = T \qquad [3.23]$$

Similarly, the number of failures in each increment of every interval is Poisson distributed where, under steady-state conditions the mean is simply the expected number of failures for that increment regardless of the particular cycle;

$$N_f(I_j^i) \sim \text{Poisson}\{\mathbf{E}[N_f(I_j)]\} \qquad [3.24]$$

This follows from the fact that a Poisson process has the property of independent Poisson distributed increments and the expected number of failures over an interval is the sum of the number expected over all non-overlapping sub-intervals;

$$\mathbf{E}[N_f(T)] = \sum_{j=1}^{z} \mathbf{E}[N_f(I_j)] \qquad [3.25]$$

The likelihood function can then be established as the product of the individual probabilities associated with observing the number of failures in each increment of all the intervals and the number of repairs undertaken at each inspection as

$$\mathcal{L} = \prod_{i=1}^{L} \left(P(N_p(iT) = j_i) \left(\prod_{j=1}^{z} P(N_f(I_j^i) = m_{ij}) \right) \right)$$

Inserting the probabilities, the expression becomes

$$\mathcal{L} = \prod_{i=1}^{L} \left(\frac{E[N_p(T)]^{j_i} e^{-E[N_p(T)]}}{j_i!} \left(\prod_{j=1}^{z} \frac{E[N_f(I_j)]^{m_{ij}} e^{-E[N_f(I_j)]}}{m_{ij}!} \right) \right) \qquad [3.26]$$

where, $\mathbf{E}[N_p(T)]$ is given by equation [3.19] and the expected number of failures over an increment I_j is derived as

$$\mathbf{E}[N_f(I_j)] = \mathbf{E}[N_f((j-1)\Delta, j\Delta)] = \mathbf{E}[N_f(0, j\Delta)] - \mathbf{E}[N_f(0, (j-1)\Delta)]$$

$\mathbf{E}[N_f(0,T)]$ is given by equation [3.13] and it follows that

$$\mathbf{E}[N_f(I_j)] = k\left(\int_0^{j\Delta} F(j\Delta - u)\,du - \int_0^{(j-1)\Delta} F((j-1)\Delta - u)\,du\right) + \nu(G(j\Delta) - G((j-1)\Delta)) \qquad [3.27]$$

Again, taking logarithms of the likelihood expression reduces the complexity from an optimisation perspective;

$$l = \sum_{i=1}^{L}\left(j_i \log(\mathbf{E}[N_p(T)]) - \mathbf{E}[N_p(T)] - \log(j_i!) + \sum_{j=1}^{z} m_{ij} \log(\mathbf{E}[N_f(I_j)])\ldots\right.$$

$$\left.\ldots - \mathbf{E}[N_f(I_j)] - \log(m_{ij}!)\right) \qquad [3.28]$$

3.2.5 Assessing the fit of the model to the data

A simulation test can be used to check the validity of the general form of the proposed likelihood function and to examine whether or not the maximum likelihood estimates are subject to bias. A number of sets of data are simulated under the specified process with known parameter values and the ML estimates are obtained. The bias is the difference between the mean parameter estimates and the true values used to simulate the data set and the standard error of the mean (SEM) is the standard deviation of the mean parameter estimates obtained from the various simulations. If the parameters can be recovered whereby the bias is less than the SEM; the likelihood formulation can be deemed appropriate and subsequent parameter estimation for the actual data set can commence. If the results are successful, the process also lends weight to the validity of any optimisation algorithm that may also

be under scrutiny. To analyse and give statistical verification of the fit of a model to data we use a chi-squared (χ^2) test.

Although we are using simulated data for the analyses conducted in this chapter, the testing process is described and applied in order to demonstrate the methodology for practical situations. In order for the test to be conducted, the data has to be arranged into a reasonable number of groups, each containing at least 5 events. Appropriate grouping of the available case data is an essential part of the parameter estimation process. Similarly, it would be impractical to consider every failure time or the number of failures in each interval. Instead we group the data into a smaller number of classes for the analysis of the model fit. As we are considering the steady-state case, the failure data can be conveniently grouped according to the increments with all cycles expected to exhibit the same pattern of behaviour within the interval. However, with cases that are not assumed to be steady state, the expected number of defects and the parameter estimation process is influenced by the number of PM's the system has previously been subjected to. As a result, the issue of grouping the data is more complicated as the same increments from different cycles cannot be treated equally. The steady-state assumption may not hold in cases where there are non-perfect inspections taking place and the data has been collected from a system in a new or post-overhaul/restoration state or in cases that incorporate an age based fault arrival rate and again, the system has not been recently initiated. The basic χ^2 test value is given as

$$\chi^2 = \sum_{i=1}^{c} \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i} \qquad [3.29]$$

where, the range of data is divided into c suitable groups, n_i is the number of observed events in the ith class and $\hat{n}_i$ is the expected number of events in the ith

class calculated from the fitted model. For testing purposes, the model has $c - \rho$ degrees of freedom, where, ρ is the number of estimated parameters. The chi-squared test statistic for analysing the fit of the delay time models constructed in this chapter is defined as

$$\chi^2 = \sum_{j=1}^{z} \frac{(N_f(I_j) - E[N_f(I_j)])^2}{E[N_f(I_j)]} + \sum_{n=1}^{l} \frac{(N_p(n) - E[N_p(n)])^2}{E[N_p(n)]} \qquad [3.30]$$

where, $N_p(n)$ is the number of observed fault removals in group n and $\mathbf{E}[N_p(n)]$ is the number predicted by the model. The number of degrees of freedom for this test procedure is $l + z - 1 - \rho$ where ρ is the number of model parameters. The χ^2-test can also be used as a selection criterion.

3.2.6 Optimal maintenance policies with fault injection at PM

Once the delay time distributions have been selected and the parameters estimated from the data and verified, a control function can be established. The downtime control function $D(T)$ represents the expected downtime per unit time and is of the same form as that prescribed for the basic delay time model. The downtime under a PM policy of perfect inspection on T hours is

$$D(T) = (\mathbf{E}[N_f(T)]d_f + d)/(T + d) \qquad [3.31]$$

where, d_f is the average duration of a failure repair, d is the average duration of a PM and $\mathbf{E}[N_f(T)]$ is given by equation [3.13]. The expression is minimised with respect to the decision variable T. The optimal policies and associated unit downtime can be compared for the current situation, incorporating substandard maintenance procedures, and a hypothetical situation where, the identified fault injections are removed thus, demonstrating the benefits of improving maintenance practice. It is obvious that the injection of faults at PM will make the inspection maintenance

process less desirable and this will be reflected in the minimisation of $D(T)$. As we discussed previously, an excessive number of fault injections could result in a situation where the optimal policy recommendation is to forgo inspections altogether in favour of a breakdown or 'contingency repair' process, if the issue cannot be resolved. The associated cost per unit time can be established by evaluating the limit of $D(T)$ as $T \to \infty$. If there are no inspections, both d and v are 0 and the expected downtime per unit time is

$$D(\infty) = \left(\frac{k}{T} \int_0^T F(T-u)du \right) d_f \Bigg|_{T \to \infty}$$

As $T \to \infty$, the cumulative density $F(T-u) \to 1$ and we have

$$D(\infty) = \left(\frac{k}{T} \int_0^T du \right) d_f \Bigg|_{T \to \infty} = \left(\frac{kT}{T} \right) d_f \Bigg|_{T \to \infty} = k d_f \qquad [3.32]$$

This value can then be compared with $D(T^*)$ obtained from the ideal scenario model with defect injections at PM removed from the process.

3.3 Numerical example 1

In this example we consider two separate cases. We assume that root cause evidence is not available upon the occurrence of each failure, the model is defined for situations where management or engineers suspect that the PM process is in some way fallible and specifically that fault injection may be taking place during inspection and repair procedures. For the first case, the delay time distributions associated with each fault type (i.e. injected at PM or arising naturally) are taken to be identical. We present a numerical comparison of the basic perfect inspection delay time model and the model incorporating fault injection for two different parameter sets. Selection criteria are applied and the fit of each model to the data is

analysed. Appropriate recommendations are then made utilising the long-term expected downtime per unit time control function. In the second case, we drop the restriction placed on the delay time distributions and allow the parameters for each fault type to differ whilst retaining the same distributional form. This presents a problem for the estimation process.

Case 1: $f(h) = g(h)$

Considering the basic delay time model with an exponential delay time distribution $f(h) = \lambda e^{-\lambda h}$ for $h \geq 0$, we have $\mathbf{E}[N_f(I_j)], j = 1, 2, \ldots, z$, and $\mathbf{E}[N_p(T)]$ as

$$\mathbf{E}[N_f(I_j)] = k\Delta - \frac{k}{\lambda}(e^{-\lambda(j-1)\Delta} - e^{-\lambda j\Delta}) \qquad [3.33]$$

$$\mathbf{E}[N_p(T)] = \frac{k}{\lambda}(1 - e^{-\lambda T}) \qquad [3.34]$$

For the delay time model incorporating the potential for fault injection at PM with identical exponential delay time distributions for both injected faults and those arising naturally during operation; $f(h) = g(h) = \lambda e^{-\lambda h}$ for $h \geq 0$, the equivalent expressions for $\mathbf{E}[N_f(I_j)]$ and $\mathbf{E}[N_p(T)]$ are

$$\mathbf{E}[N_f(I_j)] = k\Delta + \left(\nu - \frac{k}{\lambda}\right)(e^{-\lambda(j-1)\Delta} - e^{-\lambda j\Delta}) \qquad [3.35]$$

$$\mathbf{E}[N_p(T)] = \frac{k}{\lambda}(1 - e^{-\lambda T}) + \nu e^{-\lambda T} \qquad [3.36]$$

1 (a) We take the average fault arrival rate during standard operation to be $k = 0.08$ per hour, the delay time distribution is parameterised with $\lambda = 0.1$ and the average number of faults injected at an inspection is $\nu = 4$. Given that the parameters have been specified, the fault arrival and inspection process are simulated over a period of 5000 hours with a constant inspection interval of 100 hours.

To recapture the parameters we group the simulated output. The failure time information from the 50 cycles is organised into $z = 5$ non-overlapping intervals of duration $\Delta = 20$ hours. We also consider the total number of faults removed at PM producing an additional ($l = 1$) class/group. The resulting output for the 6 event types is given in table 3.1 below.

Event	Total Number
failures in (0, 20) over 50 cycles	227
failures in (20, 40) over 50 cycles	92
failures in (40, 60) over 50 cycles	87
failures in (60, 80) over 50 cycles	74
failures in (80, 100) over 50 cycles	77
fault removals at PM	38

Table 3.1 - The events simulated in each group/class for numerical example 1, case 1a

As we demonstrated earlier in the chapter, the basic delay time model can be parameterised using a simpler likelihood function with a reduction in the necessary grouping of the data. However, in this example, we opt to retain the same form for the likelihood function and apply the same grouping to the data, thus enabling a direct comparison of the maximum likelihood values produced by both models. In situations where a more complex distribution is required to reflect the behaviour of the delay times, the likelihood formulation in this example would probably be required anyway. Using equation [3.28], we can establish the likelihood function (applicable to both model formulations) for this case as

$$l = 38\log(\mathbf{E}[N_p(100)]) - 50(100k+\nu) + 227\log(\mathbf{E}[N_f(I_1)] + 92\log(\mathbf{E}[N_f(I_2)])$$

$$\ldots + 87\log(\mathbf{E}[N_f(I_3)]) + 74\log(\mathbf{E}[N_f(I_4)]) + 77\log(\mathbf{E}[N_f(I_5)]) - 262.875$$

where,

$$\mathbf{E}[N_f(I_j)] = 20k + \left(\nu - \frac{k}{\lambda}\right)(e^{-20(j-1)\lambda} - e^{-20j\lambda})$$

and $\nu = 0$ in the case of the basic delay time model. Using the Matlab© algorithm 'fmincon' from the optimisation toolbox, the parameter estimates obtained for case 1a are given in table 3.2. The constrained version of the optimisation algorithm is selected due to the fact that the parameters k, ν and λ can not be less than 0.

Parameter	Basic DTM	Fault Injection DTM
$\hat{k}$	0.1190	0.0777
$\hat{\lambda}$	0.3471	0.1047
$\hat{\nu}$		4.1278

Table 3.2 - The estimated parameters for numerical example 1, case 1a

To select the most appropriate model for the data set using the proposed likelihood function and taking into consideration the number of parameters used to characterise the relevant failure processes, we use the AIC as given in equation [3.20]. The results of the model selection process for case 1a are given in table 3.3. It is clear that the model incorporating fault injection at PM produces a substantially lower AIC and hence is the model that would be selected in a practical situation. However, the parameter values obtained are merely the best estimates for the particular models chosen. We are still required to validate the model by establishing the level of fit to the data.

	$\nu = 0$ (Basic DTM)	$\nu > 0$ (Fault Injection)
Max. Log-Likelihood	-450.89	-358.77
No. Parameters	2	3
AIC	905.78	723.54

Table 3.3 - Selecting between the two proposed models for numerical example 1, case 1a

Figure 3.4 illustrates the actual number of events observed for each category over the 50 cycles and compares them with the expected number predicted by the basic delay time model.

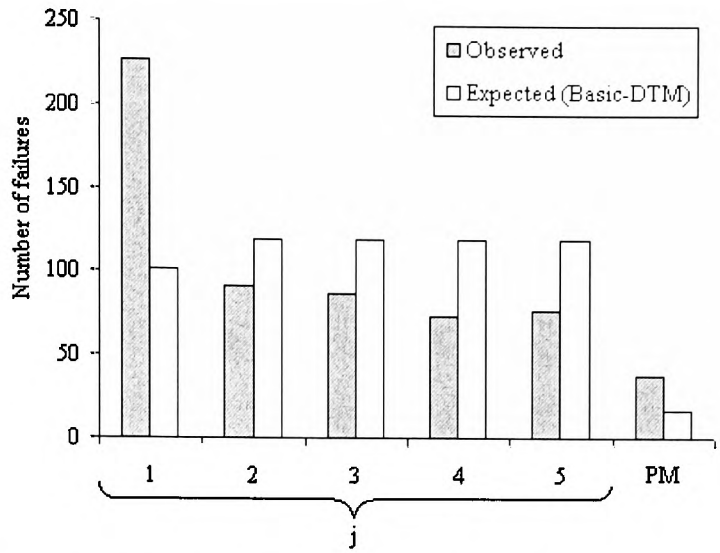

Figure 3.4 - Illustrating the observed number of failures and PM repairs against the basic delay time model predictions for numerical example 1, case 1a

Similarly, figure 3.5 compares the observed data and the associated predictions given by the delay time model incorporating fault injection.

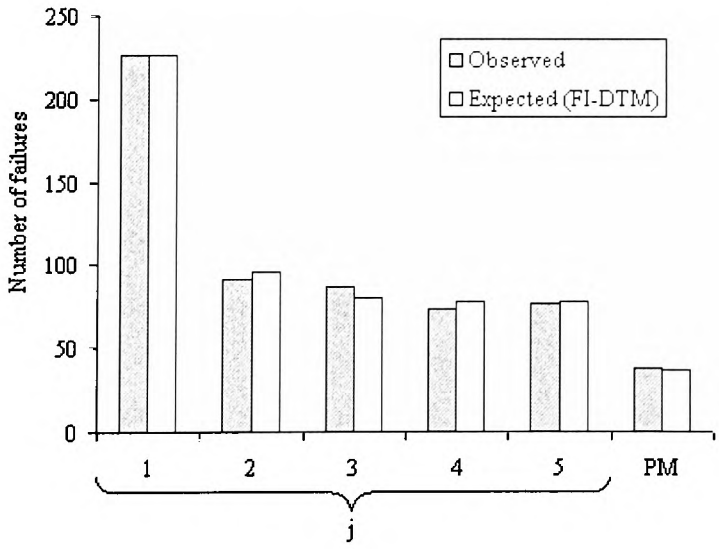

Figure 3.5 - Illustrating the observed number of failures and PM repairs against the predictions obtained from the delay time model with fault injection at PM for numerical example 1, case 1a

Figures 3.4 and 3.5 clearly demonstrate that the model incorporating fault injection produces the best fit to the data of the two proposed models by comparing the actual and predicted results. To statistically verify the fit of each model to the data, we use the χ^2- test given by equation [3.30]. For the basic model with 3 degrees of freedom, the χ^2 value of 225.7 is substantially larger than the test value of $\chi^2_{0.05}(3) = 7.815$ at the 5% significance level. As such, the basic delay time model of the failure process is rejected. The model incorporating fault injection has just 2 degrees of freedom and produces a χ^2 value of 1.0257 that is then compared with a test value of $\chi^2_{0.05}(2) = 5.991$. We can conclude that the model provides an adequate fit to the data as expected. Using equation [3.11] the downtime control function can be constructed as

$$D(T) = \frac{\left(kT + \left(\nu - \frac{k}{\lambda}\right)(1 - e^{-\lambda T})\right) d_f + d}{T + d} \qquad [3.37]$$

Only the model with fault injection is investigated here as the basic model would not be adopted in practice due to the lack of fit. For an average PM duration of $d = 1$ hours and an average failure repair time of $d_f = 2$ hours, the expected downtime per unit time under an inspection interval of T hours is

$$D(T) = \frac{0.1554T - 6.7714e^{-0.1047T} + 7.7714}{T + 1}$$

Figure 3.6 illustrates the potential benefits of improving maintenance practice and removing the artificial injection of defects at PM. The solid line represents the downtime per unit time against the PM interval T for the current existing process and the broken line represents the improved version of the same process with the impact of injected faults excluded.

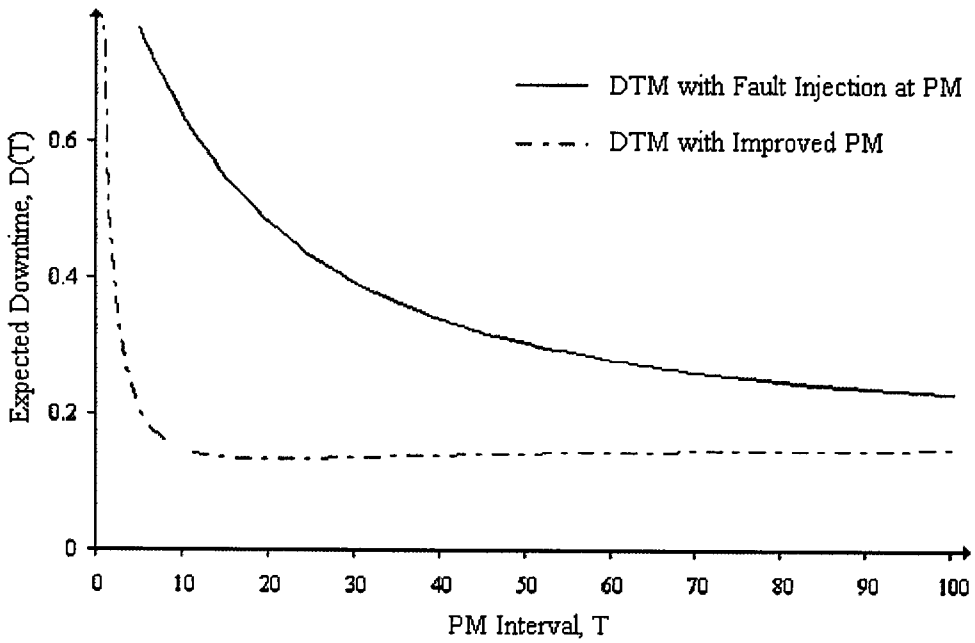

Figure 3.6 - Illustrating the benefits of improving PM for numerical example 1, case 1a

It is clear that improving maintenance procedures for this case and removing the injection of defects from the existing process would result in a reduction in the expected unit downtime. This will change the situation from one where an optimal maintenance policy on cycle T is not available and a contingency repair policy would be the appropriate recommendation, to a situation where the optimal policy would be to perform a PM approximately every $T^* = 18$ hours. This rather extreme case illustrates the potential benefits that can be achieved by modelling the injection of faults at maintenance interventions. Returning to the primary objective of this research, the example has demonstrated that the proposed methodology is capable of recapturing the necessary parameters from the data. The estimated values are very close to those used to simulate the data for this particular example. However, it should be noted that, a large number of cycles of data were generated to obtain this level of accuracy. To further verify the functional form of the proposed likelihood expression, a simulation test was conducted and the level of bias in the estimated values found to be insignificant for this case.

1 (b)

In case 1a, we adopted parameter values that characterised an extreme version of the scenario under consideration with defects attributable to injection at the previous PM contributing a third of the total number expected. In addition, a large number of cycles of data were simulated to obtain the results. In case 1b, we consider a less extreme injection process and simulate a limited number of cycles of data. We take the average fault arrival rate during standard operation to be $k = 0.1$ per hour, the delay time distribution is parameterised with $\lambda = 1/30$ and the average number of faults injected at an inspection is $\nu = 1.25$. The process is simulated for just $L = 20$ cycles of duration $T = 100$ hours. The fault detection process at inspection remains perfect.

Following the same process as case 1a, the parameter estimates obtained for case 1b are given in table 3.4.

Parameter	Basic DTM	Fault Injection DTM
$\hat{k}$	0.1125	0.0992
$\hat{\lambda}$	0.0462	0.0367
$\hat{\nu}$	-	1.3315

Table 3.4 - The estimated parameters for numerical example 1, case 1b

When comparing and selecting between the two models, the results in table 3.5 illustrate that the model with fault injection at PM is the best choice based upon the minimum AI criterion. However, as we expected, the difference is far less pronounced than that observed in case 1a. This is due to the fact that the injected defects contribute (proportionally) less to the total number of failures and PM removals and therefore, with faults arising naturally being the substantial contributor, a basic model representation of the process is almost good enough.

	$\nu = 0$ (Basic DTM)	$\nu > 0$ (Fault Injection)
Max. Log-Likelihood	-187.1163	-183.269
No. Parameters	2	3
AIC	378.233	372.538

Table 3.5 - Selecting between the two proposed models for numerical example 1, case 1b

In some situations it maybe that the proportion of injected defects is so small, we cannot differentiate and choose between the two proposed models. However, it is also likely that the impact of standard defects would make inspections a necessity and that the comparative effects of injected defects would be negligible. Optimisation of the existing process with respect to the interval T would then be the recommended policy. Figure 3.7 illustrates the observed failure and inspection repair data and the expected number of events obtained from both models.

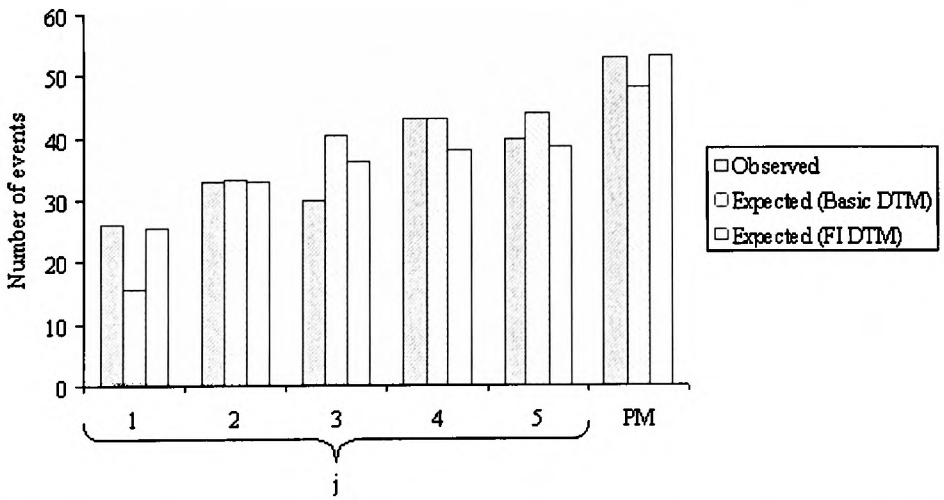

Figure 3.7 - Illustrating the observed number of events against the predictions obtained from the two proposed delay time models for numerical example 1, case 1b

From the histogram in figure 3.7, the model with fault injection appears to provide a better fit to the data, although the difference is marginal. The statistical assessment of model fit produces results similar to those obtained in 1a. The basic model is rejected at the 5% level with a χ^2 of 10.4355 being greater than the test value

of $\chi^2_{0.05}(3) = 7.815$. As with 1a, the model incorporating fault injection is deemed to be acceptable with a χ^2 of 1.7992 being less than the test value $\chi^2_{0.05}(2) = 5.991$. However, as is reflected by the AIC and the χ^2-test, the difference between the two models is far less substantial than that observed in case 1a. This is due to the decreased proportion of injected defects and the limited number of simulated cycles of data, thus producing a greater variability in the estimates from the actual underlying parameter values.

Case 2: $f(h) \neq g(h)$

In this second case, the delay time distributions for each fault type are allowed to differ from one another. This presents problems for parameter estimation but greater modelling flexibility. To assess the ability of the methodology when attempting to recapture the parameters, exponential distributions are used for both fault types as $f(h) = \lambda_1 e^{-\lambda_1 h}$ and $g(h) = \lambda_2 e^{-\lambda_2 h}$ for $h \geq 0$. The choice of distribution is by no means limited to exponential forms however a neat analytical solution is available for this case. The expressions for $\mathbf{E}[N_f(I_j)], j = 1, 2, \ldots, z$, and $\mathbf{E}[N_p(T)]$ are

$$\mathbf{E}[N_f(I_j)] = k\Delta - \frac{k}{\lambda_1}(e^{-\lambda_1(j-1)\Delta} - e^{-\lambda_1 j\Delta}) + \nu(e^{-\lambda_2(j-1)\Delta} - e^{-\lambda_2 j\Delta}) \quad [3.38]$$

$$\mathbf{E}[N_p(T)] = \frac{k}{\lambda_1}(1 - e^{-\lambda_1 T}) + \nu e^{-\lambda_2 T} \quad [3.39]$$

The failure and inspection process is simulated using an average rate of fault arrival during operation of k = 0.08 per hour and the associated exponential delay time distribution is parameterised with λ_1 = 1/30. The average number of fault injections at PM is taken to be $\nu = 3$ and the delay time distribution has parameter λ_2 = 1/15. The process is simulated for $L = 50$ regular cycles of duration $T = 100$ hours.

For parameter estimation purposes, the appropriate log-likelihood function is of the same form as equation [3.28]. However, convergence could not be achieved for the parameters under investigation as a range of optimal parameter combinations are available with each solution vector producing the same value for the maximum log-likelihood. The same problem was encountered with data sets simulated using many different parameter combinations. This relates to the identifiability problem, mentioned earlier, that is frequently encountered when modelling competing risks. To combat the problem, the causes of individual failures would need to be recorded or a subjective evaluation of the potential causes of faults undertaken.

In this case, it appears that a blend of subjective and objective estimation techniques is required to obtain the parameter estimates when the behaviour of the two fault types differs with regard to the delay time until failure. Subjective input could consist of expert or engineering opinion that is applicable to the particular application. In situations where the different originating fault types can be tagged upon failure of the system, the estimation techniques described in this chapter are unnecessary and the individual delay time distributions can be established independently. Subjective estimation of one or more of the model parameters could in some situations ease the estimation process for the other values. Christer & Waller (1984) and Wang (1997) discuss the use of subjective estimation in the form of expert opinion, failure mode and criticality analysis in the construction of delay time distributions with a discussion of the different means of combining the expert opinion. Snapshot modelling techniques (see Christer & Whitelaw (1983)) can also assist in establishing a rough characterisation of the current process. Here, we discuss a couple of options for subjective input that are applicable to this example and a number of other cases considered with exponential delay time distributions.

(i) A subjective estimate of the average rate of fault arrival during standard operation $\hat{k}$ is the first part of the process. Alternatively, for many cases, including this one, the following substitution can be used

$$\hat{k} = \frac{\sum_{i=1}^{L}\left(j_i + \sum_{j=1}^{z} m_{ij}\right)}{LT} - \frac{\nu}{T} \qquad [3.40]$$

thus reducing the number of parameters to be estimated. However, assumed knowledge of $\hat{k}$ or the suggested substitution cannot guarantee convergence of the optimisation algorithm with respect to the remaining parameters. We require a second stage to combine with $\hat{k}$ or the substitution. For the case considered here, the delay time model incorporating fault injection with identical exponential delay time distributions (as used in cases 1a and b) is initially applied giving parameter estimates of $\hat{k} = 0.0659$, $\hat{\lambda} = 0.0293$ and $\hat{\nu} = 4.3864$. The estimates clearly indicate a problem with the PM process however, given that it is simulated data under scrutiny, it is known that k has been under-estimated and that ν has been over-estimated to compensate for the underlying shorter delay times imposed upon the injected faults. The useful element is the estimate $\hat{\lambda} = 0.0293$ which is very close to the value of λ_1 used when simulating the data. The same closeness result has been obtained in other cases but not in all. Using a subjective estimate of $\hat{k}$ or the suggested substitution and the value $\hat{\lambda}_1 = \hat{\lambda}$, the remaining parameters can be established. In practical situations, the model can be compared with the results obtained from the identical distribution case using the AIC and chi-squared test procedures.

(ii) An alternative approach for utilising subjective input in the parameter estimation process for case 2 is to establish a relationship between the parameters of the two exponential delay time distributions. The parameter of an exponential delay time

distribution has the following property; $\lambda = 1/\bar{h}$ where, $\bar{h}$ is the mean delay time. In some situations, subjective estimation techniques could be utilised to reveal a relationship between the mean delay times for the two fault types, for instance; $\bar{h}_1 = \alpha \bar{h}_2$ the following relationship could be applied to the parameters; $\lambda_2 = \alpha \lambda_1$, therefore reducing the number of delay time parameters sought by estimation. For the case considered, the relationship; $\bar{h}_1 = 2\bar{h}_2$ provides the input necessary to obtain the following parameter estimates; $\hat{k}$ = 0.0809, $\hat{\lambda}_1$ = 0.0319, $\hat{v}$ = 2.8850 and $\hat{\lambda}_2$ = 0.0638. All the parameters are very close to the actual underlying values used to simulate the data however, the ability to subjectively estimate the relationship between the mean delay times is naturally specific to the particular case under consideration.

The subjective approaches suggested in (i) and (ii) are worth attempting if plots of the data imply that the PM process is problematic. The model fit procedures will establish the adequacy of the model and accurate estimation of the actual underlying process will naturally improve the accuracy of any resulting downtime models.

3.4 Aspects of human fallibility

3.4.1 Identification of the underlying process

All that would actually be observed of the processes represented in figure 3.2 is illustrated in figure 3.8 where, 'X' represents the removal of a defect at PM.

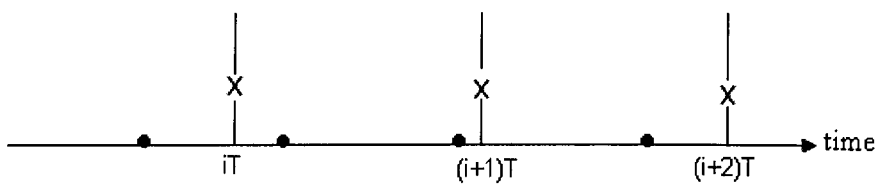

Figure 3.8 - The observed information associated with the processes depicted in figure 3.2

Failures that consistently arrive shortly after inspection may be attributable to faults injected at the previous inspection or to poor fault detection that would allow existing defects, at a more advanced stage of deterioration, to remain in the system at the start of the next operational interval.

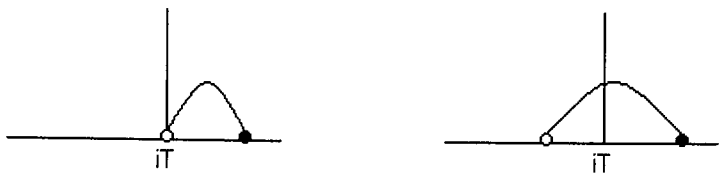

Figure 3.9 - Illustrating the potential origins of a fault leading to failure in the early stages of an inspection interval

Figure 3.9 demonstrates the potential origins of a failure occurring shortly after an inspection. The model specification and parameter estimation process is applied to select between the proposed delay time model that incorporates fault injection and a model incorporating imperfect detection of defects at PM. The issue is whether or not the proposed methodology can enable accurate identification of the actual underlying process and differentiate between the two types of human error. Initially, the delay time model incorporating imperfect detection of existing defects at PM is introduced.

3.4.2 Imperfect detection case ($\beta \leq 1$)

In this section, we discuss the modelling of an imperfect fault detection process at regular periodic inspections. The interval between inspections is again assumed to be of a constant duration T. Defining β as the probability that an individual fault is detected at a given inspection and $N_f(t)$ as the number of failures occurring over an interval $(0, t)$ after an inspection then, under steady-state conditions, we have

$\mathbf{E}[N_f((i-1)T,(i-1)T+t)] = \mathbf{E}[N_f(0,t)] = \mathbf{E}[N_f(t)]$. Under the specification given, $N_f(t)$ adheres to a non-homogenous Poisson process with expectation

$$\mathbf{E}[N_f(t)] = k\int_0^t F(t-u)du + k\sum_{i=1}^{\infty}\{1-\beta\}^i \int_0^T \{F(iT+t-u)-F(iT-u)\}du \qquad [3.41]$$

for $t \geq 0$, and failure rate function

$$r(t) = \frac{d}{dt}\mathbf{E}[N_f(t)] = k\int_0^t f(t-u)du + k\sum_{i=1}^{\infty}\{1-\beta\}^i \int_0^T f(iT+t-u)du \qquad [3.42]$$

As we are considering a system in a steady-state of operation with a constant interval between inspections of duration T, the expression for the expected number of breakdowns over a given interval can be arranged as

$$\mathbf{E}[N_f(T)] = kT\left(1-\frac{\beta}{T}\sum_{i=1}^{\infty}\{1-\beta\}^{i-1}\int_0^T (1-F(iT-u))du\right) \qquad [3.43]$$

see Christer & Waller (1984b) for details. The number of faults identified and subsequently rectified at an inspection under a constant inspection interval policy of duration T is defined as $N_p(T)$ and is also Poisson distributed with expectation

$$\mathbf{E}[N_p(T)] = k\beta\sum_{i=1}^{\infty}\{1-\beta\}^{i-1}\int_0^T (1-F(iT-u))du \qquad [3.44]$$

For parameter estimation, the structure of the likelihood function is the same, equation [3.28]. The same applies for the χ^2-test that is used to assess the quality of the model fit to the data. In addition, as both models use the same likelihood function for parameter estimation and the same groupings of the data apply, we could compare the two modelling approaches using the AIC criterion. When the same delay time distributions are used for both cases, the AIC is not required as the models have the same number of structural parameters and a direct comparison of the likelihood or log-likelihood is sufficient.

3.5 Numerical example 2

The objective of this numerical example is to demonstrate the ability of the proposed model selection and parameter estimation methodology when differentiating between a maintenance scenario with fault injection at PM and a scenario incorporating a fallible detection process. Using simulated data, the actual underlying process for this case is established with an average of $\nu = 3.2$ fault injections at PM. The average rate of fault arrival during standard operation is $k = 0.1$ per hour and both fault types are governed by the same exponential delay time distribution with parameter $\lambda = 0.05$. The process is simulated for $L = 50$ cycles of duration $T = 100$ hours.

We group the resulting failure data into $z = 5$ non-overlapping equidistant intervals of duration $\Delta = 20$ hours and figure 3.10 illustrates the simulated output.

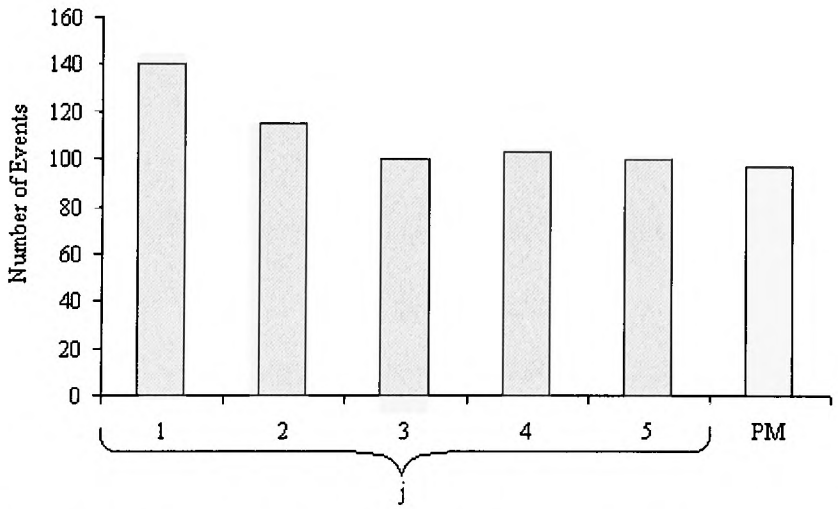

Figure 3.10 - Illustrating the grouped simulated failure and PM repair data for numerical example 2

From figure 3.10, it appears that PM has a negative influence and the issue is whether or not the model selection process can distinguish between fault injection at PM and the fallible detection of existing faults. For parameter estimation purposes, the same likelihood function is used for both variants of the model (see equation

[3.28]). With regard to the model incorporating fault injection, the expressions for the expected number of failures occurring in the jth increment, $\mathbf{E}[N_f(I_j)]$, and the number of repairs at PM, $\mathbf{E}[N_p(T)]$, with exponential delay time distributions are given by equations [3.35] and [3.36] respectively. In the imperfect detection case ($\beta \leq 1$) with an exponential delay time distribution, the expected number of failures in the steady state is

$$\mathbf{E}[N_f(T)] = kT\left(1 - \frac{\beta}{\lambda T}\left(\frac{1-e^{-\lambda T}}{1-(1-\beta)e^{-\lambda T}}\right)\right) \qquad [3.45]$$

The expected number of failures over increment j, ($\mathbf{E}[N_f(j\Delta)] - \mathbf{E}[N_f((j\text{-}1)\Delta)]$) is

$$\mathbf{E}[N_f(I_j)] = k\left(\Delta + \frac{\beta}{\lambda}\left(\frac{1-e^{-\lambda(j-1)\Delta}}{1-(1-\beta)e^{-\lambda(j-1)\Delta}}\right) - \frac{\beta}{\lambda}\left(\frac{1-e^{-\lambda j\Delta}}{1-(1-\beta)e^{-\lambda j\Delta}}\right)\right) \qquad [3.46]$$

The expected number of faults found and removed at an inspection with an imperfect detection process is

$$\mathbf{E}[N_p(T)] = \frac{k\beta}{\lambda}\left(\frac{1-e^{-\lambda T}}{1-(1-\beta)e^{-\lambda T}}\right) \qquad [3.47]$$

Using [3.28] and the BFGS optimisation algorithm, the parameter estimates given in table 3.6 are obtained for the two model variants.

	Fault Injection DTM	Imperfect Detection DTM
$\hat{k}$	0.0988	0.1311
$\hat{\lambda}$	0.0512	0.0012
$\hat{\nu}$	3.2159	
$\hat{\beta}$		0.0153

Table 3.6 - The estimated parameters for numerical example 2

The AIC is not required for model selection purposes as both of the models contain the same number of estimated parameters. The maximum log-likelihood is -408.036 for the fault injection model and -439.521 for the model with imperfect detection.

This demonstrates that for the case considered, the model with fault injection would correctly be selected in a practical situation. Figure 3.12 compares the predicted output from the two models with the actual data used for parameterisation.

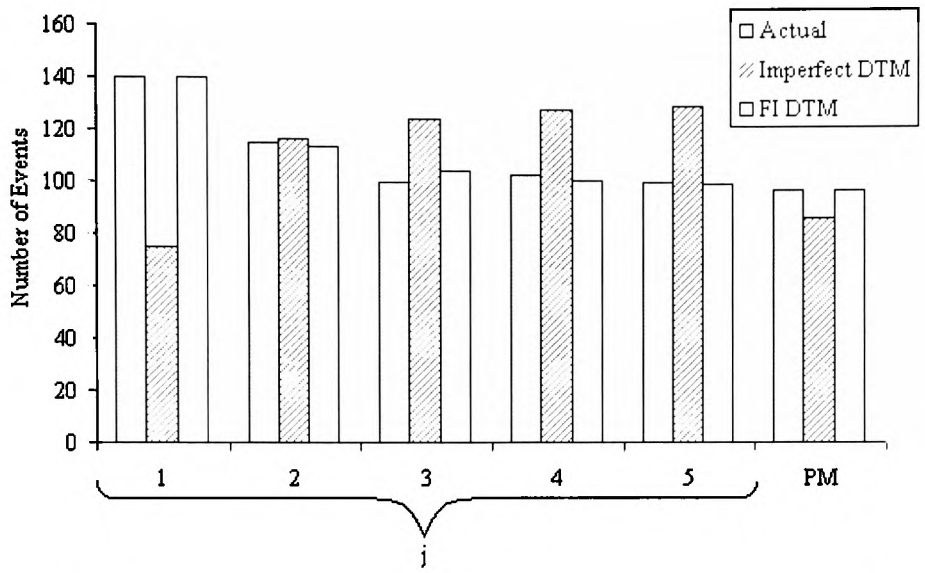

Figure 3.11 - Comparing the output from the two models with the actual observed data for numerical example 2

Figure 3.11 strongly indicates that the model incorporating fault injection provides the best predictions when compared with the actual data. This observation is confirmed by the χ^2-test results with 0.2349 for the fault injection model and 73.9755 for the imperfect detection model. When compared at the 5% significance level with a test value of $\chi^2_{0.05}(3) = 7.815$, the imperfect detection case is rejected whilst the fault injection model is accepted. The example demonstrates that it is possible to differentiate between the two scenarios when the impact of fault injection is sufficient and the injected faults and those arising naturally are assumed to behave in the same manner, i.e. have identical delay time distributions.

3.6 Combining both aspects of human error

Combining the effects of the potential for fault injection at PM and a fallible detection process results, we define the number of failures observed over an interval $(0, t)$ after any given PM as $N_f(t)$. Under steady-state conditions regarding plant performance and a substantial period of operation before the current interval between inspections, $N_f(t)$ follows a non-homogenous Poisson process with expectation

$$\mathbf{E}[N_f(t)] = k\int_0^t F(t-u)du + k\sum_{i=1}^{\infty}\{1-\beta\}^i\int_0^T\{F(iT+t-u)-F(iT-u)\}du$$

$$+\nu\sum_{i=0}^{\infty}(1-\beta)^i\{G(iT+t)-G(iT)\} \qquad [3.48]$$

for $0 \le t \le T$, with failure rate function

$$r(t) = \frac{d}{dt}\mathbf{E}[N_f(t)] = k\int_0^t f(t-u)du + k\sum_{i=1}^{\infty}\{1-\beta\}^i\int_0^T f(iT+t-u)du$$

$$+\nu\sum_{i=0}^{\infty}(1-\beta)^i g(iT+t) \qquad [3.49]$$

In the limiting steady-state case with a constant interval between inspections, of duration T, we have

$$\mathbf{E}[N_f(T)] = kT - k\beta\sum_{i=1}^{\infty}(1-\beta)^{i-1}\int_0^T(1-F(iT-u))du + \nu - \nu\beta\sum_{i=1}^{\infty}(1-\beta)^{i-1}(1-G(iT)) \qquad [3.50]$$

Similarly, the expected number of faults detected and removed at an inspection is

$$\mathbf{E}[N_p(T)] = k\beta\sum_{i=1}^{\infty}(1-\beta)^{i-1}\int_0^T(1-F(iT-u))du + \nu\beta\sum_{i=1}^{\infty}(1-\beta)^{i-1}(1-G(iT)) \qquad [3.51]$$

However, using data sets simulated according to the process described, it was found that accurate parameter estimation is not possible using the proposed methodology and the likelihood functions established earlier in the chapter without the use of

subjective input, see numerical example 1, case 2. This is another case of the 'identifiability' problem discussed earlier in that, a range of different parameter combinations produced the same maximum log-likelihood and a single optimal solution vector could not be established.

To investigate the use of prior subjective input, such as expert engineering opinion, a combination of the proposed likelihood formulation and the expectation-maximisation (EM) algorithm is one avenue that we considered. See Russell & Norvig (2003) for information on the EM algorithm. The algorithm is given in a single expression as

$$\underline{\hat{\theta}}^{c+1} = \max_{\underline{\theta}} \sum_{v} P\left(V = v \middle| M, J, \underline{\hat{\theta}}^{c}\right) \mathcal{L}\left(M, J, V = v \middle| \underline{\theta}\right) \qquad [3.52]$$

where, $\underline{\theta}$ represents the parameter set (excluding v) under investigation, M represents the information available on the failure process over all cycles, J represents the information available on the repairs undertaken at the inspections and $c = 0, 1, 2, \ldots$ is the index for the algorithm. A prior subjective distribution $p(v)$ is established for the parameter v over a range of candidate values (a_v, b_v) using expert opinion. The discrete distribution is then updated at subsequent iterations of the algorithm using equation [3.53] as

$$P\left(v \mid M, J, \underline{\hat{\theta}}^{c}\right) = \mathcal{L}\left(v, M, J, \middle| \underline{\hat{\theta}}^{c}\right) / \sum_{a_v}^{b_v} \mathcal{L}\left(v, M, J, \middle| \underline{\hat{\theta}}^{c}\right) \qquad [3.53]$$

The objective is to establish $\underline{\hat{\theta}}^{*}$ upon convergence of the algorithm and subsequently to evaluate the expression

$$\hat{v} = \max_{v} \left\{ L\left(M, J, \underline{\hat{\theta}}^{*} \middle| v\right) \right\} \qquad [3.54]$$

However, it was discovered that, when the mean of the prior subjective distribution is within the range of optimal solutions that maximise the standard likelihood function, the algorithm converges to the mean value. As such, simply inserting the mean of the prior distribution into the original likelihood function produces the same results for the parameter vector $\hat{\underline{\theta}}^*$. Similar results are observed when other system parameters are made the objective of the algorithm. When the underlying process contains both types of human error and the models incorporating fault injection exclusively and imperfect detection exclusively are applied to the data, the fault injection delay time model typically produces a better fit with an increase in v and a reduction in the standard arrival rate and the mean delay time. In a practical situation, the resulting conclusion that a review of maintenance procedures is necessary would still hold but, the modelling of potential gains would be hampered.

3.7 Discussion and further considerations

In this chapter, a number of modelling options have been proposed in an attempt to incorporate human fallibility. The principal developments include the incorporation of human error in the form of potential fault injection during the course of inspection based repairs with the major objective being the accurate identification of said process when it exists using the proposed model specification and parameter estimation techniques. Numerical example 1, cases 1a and b demonstrate that when the proposed functional form of the model effectively represents the actual underlying process used to generated the given data set, accurate estimation of the model parameters is achievable when fault injection is incorporated into the modelling process and both fault types behave in the same manner, i.e. we have $f(h) = g(h)$. The examples also demonstrate that with the proposed methodology for model selection and parameter estimation, accurate representation and estimation is

dependent on the ratio $\nu : kT$. We can correctly differentiate between the basic delay time process and a process incorporating fault injection at PM from objective data if the injected faults have a sufficient impact. In cases where the impact of fault injection is insignificant when compared with the effect of standard fault arrivals, a basic delay time representation will likely suffice. In numerical example 1, case 2, consideration was given to the parameter estimation problem when the delay time distributions for injected faults and those arising naturally are allowed to differ. The estimation process is much more complicated than the identical distribution case and due to a problem of identifiability, requires a blend of subjective and objective methods to achieve solutions.

In numerical example 2, the ability to differentiate between a scenario incorporating fault injection at PM and one incorporating imperfect detection at inspection was discussed. Again, when the impact of fault injection is substantial, the model selection methodology correctly identifies the underlying process. Further recommendations regarding model selection are now discussed.

The parameters values used in the numerical examples have been chosen to demonstrate the modelling and estimation process for the models in question. Analysis over a range of values revealed a number of behavioural patterns and some general recommendations can be made. When the failure process is not affected by human error (fault injection or imperfect detection at PM) plotting the observed failures over time since the last PM often reveals a pattern similar to that illustrated in figure 3.12. A basic delay time model representation will probably be sufficient in cases such as this however, an AIC comparison of the basic model and an imperfect detection model is the recommendation proposed here.

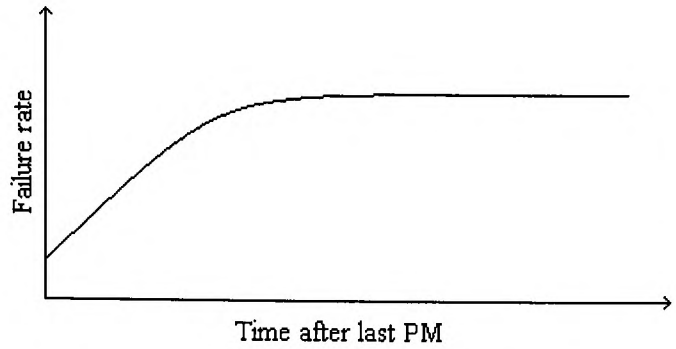

Figure 3.12 - Illustrating a typical failure process without human error at PM

Figure 3.13 illustrates a profile of the rate of occurrence of failures over time since the last PM that is often observed when the detection of existing faults at inspection is imperfect and some are allowed to remain in the system.

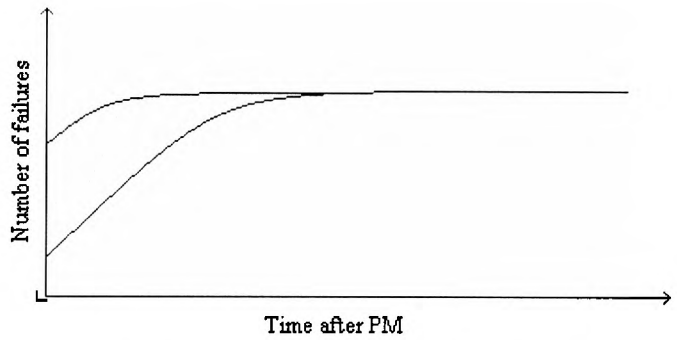

Figure 3.13 - Illustrating a typical failure process with imperfect detection of existing faults at PM

Figure 3.14 illustrates a failure process against time since the last PM that often occurs when fault injection is taking place during maintenance procedures.

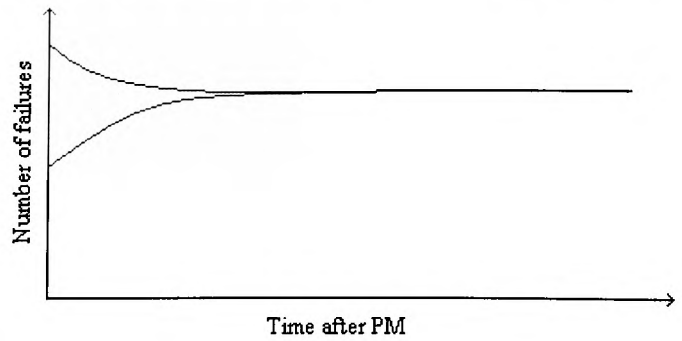

Figure 3.14 - Illustrating a typical failure process with human error based fault injection at PM

From figures 3.13 and 3.14 it is evident that in some situations, a process incorporating fault injection could be confused with a process with imperfect detection, however, both models would indicate that the maintenance process is problematic and a review of maintenance procedures would be the likely recommendation.

When the appropriate delay time model for a particular scenario cannot be parameterised using the techniques discussed due to a problem of identifiability, the following options must be considered. Firstly, we could consider a more direct approach to the estimation of the overall failure rate and provide recommendations on this basis. However, the approach would not enable quantification of the number of faults attributable to artificial injection and as such, the benefits of improving maintenance practice could not be ascertained. The second option for consideration would be the establishment of subjective Bayesian prior distributions for all or some of the parameters and application of a maximum a-posteriori (MAP) estimator. Alternatively, a version of the EM algorithm combined with the likelihood function and the Bayesian priors could produce better results than those obtained earlier in the chapter (for the combined fault injection and imperfect detection case) if the prior distributions have some kind of engineering (or expert) basis that is relevant to the particular application, unlike the uniform priors used in our unsuccessful investigation.

Consider the process incorporating fault injection at PM where, it is assumed that the injected faults have the same characteristics as the faults that arise naturally during production. An alternative means of representing the process and solving the problem of parameter estimation is to model the fault arrival process with a mixed

arrival rate distribution. Defining $a(u)$ as the fault arrival rate distribution for $u \geq 0$ we have the expected number of faults over an interval of duration T as

$$\mathbf{E}[N_f(T)] = KT\int_0^T a(u)F(T-u)du \qquad [3.55]$$

and when $a(u)$ is taken to be uniform, the expression reduces to the form utilised for the basic scenario without fault injection and $K = k$, see equation [3.1]. Now, considering mixed arrivals, the cumulative arrival rate density is

$$A(u) = 1-(1-P)(1-(u/T)) = P+(1-P)(u/T) \qquad [3.56]$$

for $P \in (0,1)$. Upon differentiating the arrival rate density becomes

$$a(u) = (1-P)/T \qquad [3.57]$$

and the number of failures expected over an interval $(0, T)$ is given by

$$\mathbf{E}[N_f(T)] = K(1-P)\int_0^T F(T-u)du \qquad [3.58]$$

Considering the behaviour of the failure process within a cycle, the number of failures expected over a sub-interval $((j-1)\Delta, j\Delta)$ is given by

$$\mathbf{E}[N_f((j-1)\Delta, j\Delta)] = K(1-P)\left(\int_0^{j\Delta} F(j\Delta-u)du - \int_0^{(j-1)\Delta} F((j-1)\Delta-u)du\right) \qquad [3.59]$$

and the expected number of faults found and removed at PM is

$$\mathbf{E}[N_p(T)] = KT\left(1-\left(\frac{1-P}{T}\right)\left(\int_0^T F(T-u)du\right)\right) \qquad [3.60]$$

For parameter estimation purposes, the structural form of the likelihood function is akin to that given by equation [3.26] with the modified forms for $\mathbf{E}[N_f((j-1)\Delta, j\Delta)]$ and $\mathbf{E}[N_p(T)]$ as given by [3.59] and [3.60]. The maximum likelihood estimates of the standard fault arrival rate k and the average number of artificial fault injections v are given by

$$\hat{k} = (1 - \hat{P})\hat{K} \qquad [3.61]$$

$$\hat{v} = \hat{K}T\hat{P} \qquad [3.62]$$

In terms of extending the current research, consideration could be given to modelling the potential impact of fault injection during the course of failure repairs, where the repairs are attributable to faults that arose or were injected earlier in the process. This would seem to be a plausible scenario as, the chances of poor quality maintenance would doubtlessly be increased when repairs are not scheduled or prepared for. Similarly, additional means of modelling the potentially negative impact of inspection maintenance include the application of a time based penalty for inspection repairs when establishing the downtime model of the system. For instance, defining d_p as the duration of an inspection based repair, the expected downtime per unit time becomes

$$D(T) = (\mathbf{E}[N_f(T)]d_f + d + \mathbf{E}[N_p(T)]d_p)/(T + d + \mathbf{E}[N_p(T)]d_p) \qquad [3.63]$$

In the next chapter, an alternative representation for the maintenance scenarios discussed here is presented. Using the representation, on-line real time applications of the models presented in this chapter are proposed. The ability to represent the number of fault injections at PM in a distributional form rather than a point estimate is also assessed.

Chapter 4. On-line modelling of fallible maintenance processes for complex systems using the delay time concept and probabilistic stochastic filtering

4.1 Introduction

As with the previous chapter, the problem scenario under consideration is that of a complex operational system subject to planned preventive maintenance (PM) where, the maintenance process is inspection based maintenance that incorporates a level of human error in the form of artificial fault injection. In this chapter, the delay time concept and the functional forms developed in chapter 3 are combined with a stochastic filtering approach to develop a state estimation and predictive decision model. The objective state of the system is the number of fault arrivals. Expressions for the evolution of the underlying fault arrival process and the stochastic relationship between the failure and fault arrival process are constructed for a discrete approximation of the continuous time system. The history of the failure process and the time that has elapsed since the last PM are used as input to the recursive stochastic filter. The filter is an application of the non-linear probabilistic framework given in chapter 2. However, the model in this chapter is developed for a discrete state space because the number of fault arrivals (including artificial injections) can only be a positive integer or zero. The form of the filter is that of a hidden Markov model (HMM) incorporating one-step transitions in the underlying state between increments of the PM cycle. HMM's are frequently used in image and speech processing as an application of approximate grid-based methods, see Arulampalam et al (2002), Forney (1973), Rabiner & Juang (1986) and Streit & Barrett (1990). There are two major advantages when modelling the process in this manner and these are now discussed.

1) Firstly, related to the research described in the previous chapter, the model formulation and parameter estimation process proposed in this chapter allow for the parameterisation of a prior distribution for the underlying state and provide greater flexibility when modelling the system. Figure 4.1 illustrates a typical process for a complex system incorporating fault injection at PM, where the black circles represent system failures. The actual inspection process is assumed to be perfect from the point of view of detecting existing defects whose delay time until failure has not yet run its course. The identified defects are then removed during the maintenance element of the PM.

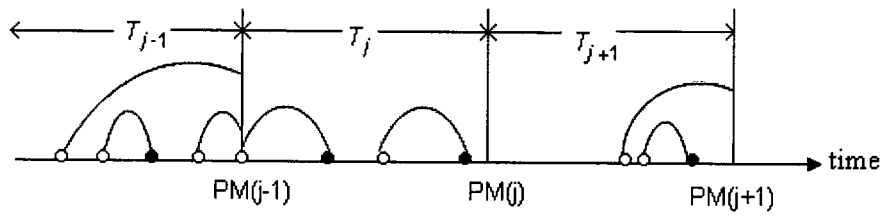

Figure 4.1 - An underlying fault arrival and injection process

However, all the information that we have available for estimating the underlying parameters for the case depicted in figure 4.1 consists of the failure times and the number of existing faults that are removed at PM. The failure and inspection data for the process is shown in figure 4.2, where the crosses represent faults removed at PM.

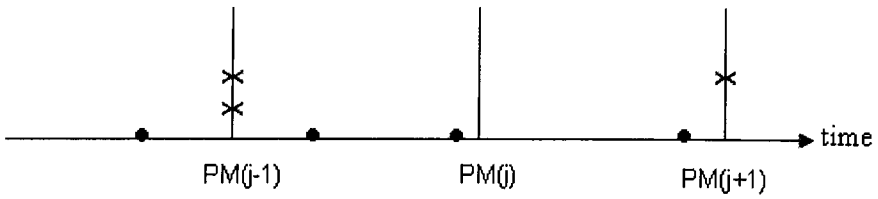

Figure 4.2 - The observed failure process for the scenario depicted in figure 4.1

When all the observed failure and PM data are processed for estimation using the recursive filtering equations, we obtain the optimal Bayesian posterior estimate of the parameter set that characterises the fault arrival, fault injection and failure

processes. Alternatively, in some scenarios, it may be more efficient to obtain the parameter estimates using the techniques described in chapter 3 as the functional form of the underlying process can be represented using both approaches. Naturally, the quality of the estimates under both approaches is dependent on both the suitability of the proposed functional form and the quantity of data that is available.

2) The second advantage of modelling the process in this manner is the ability to construct an adaptive decision model in order to maximise the impact when scheduling maintenance activities. The decision making associated with the models developed in chapter 3 is of a fixed interval nature whereby, an optimal recommendation is to maintain the system on a regular cycle of T^* units of time. With the model proposed in this chapter, information is obtained over time that is indicative of the underlying state. The current knowledge of the state is expressed in the form of a distribution that is conditional upon the observed CM history. The construction of said density can be extended to forecast the distribution into the future. Using this predictive density, an optimal time for the next PM can be established at each chosen time point or upon receiving new information. This has obvious benefits in terms of providing a complete representation of the likely status of the system and evaluating the most cost effective maintenance decision based upon all information that is currently available during the course of an operational cycle.

Initially, a continuous-time representation of the actual underlying fault process and the stochastically related failure information is presented. A recursive filtering algorithm is then developed for a discretised representation of the fault and failure processes. The issues of parameter estimation and the scheduling of maintenance activities are discussed and some examples are then given for a basic scenario

without human error at PM and for a scenario where fault injection is incorporated into the modelling process. To illustrate the application of the filter, the forms selected for the various elements that define the state and observation processes are relatively simple in the examples given however, the concept is readily adapted to more complex forms.

4.2 Preliminaries

The modelling of a system using the proposed methodology is subject to the following assumptions;

1. Defects arise naturally during operation of the system.
2. Defects may be injected during the course of a PM.
3. All faults have a delay time until failure that is governed by some density function.
4. Detection is perfect when; (a) upon failure, ascertaining the particular defect that has become the cause of a system failure, this diagnostic process is required before subsequent unscheduled maintenance can attend to the problem, (b) performing PM and identifying existing defects that have not yet resulted in a failure. An implication of this assumption is that each PM cycle can be treated independently when modelling fault arrivals and the occurrence of failures because there are no faults from previous cycles remaining in the system after PM. This is an assumption that could be relaxed in some situations in a similar manner to that employed to derive the imperfect inspection model of chapter 3.
5. Only the offending defect that has resulted in failure is attended to with unscheduled maintenance. All other defects in the system remain untouched and progress from the same level of degradation as just before the failure occurred.
6. Upon failure, the causal defect is removed completely.

The following notation is used to represent the various elements of the process;

- m is the number of observed cycles of data in an available sample.
- $PM(j)$ represents the time of the jth PM since the start of the observation period, $j = 1, 2, \ldots, m$.
- T_j is the duration of the operational period between $PM(j-1)$ and $PM(j)$, and clearly we have the relationship $PM(j) = \sum_{u=1}^{j} T_u$.

The observed information that is used to estimate the model parameters consists of the following;

- Q_j is the number of faults removed at the jth PM.
- $W_{j,t}$ is the number of failures observed over $\{PM(j-1), PM(j-1)+t\}$.

The underlying dynamics of the system are described using the following elements;

- $k_j(t)$ is the fault arrival rate function after t units of operational time into the jth cycle, for $t > 0$.
- $\alpha_{j,0}$ is the (unknown) number of faults artificially injected at $PM(j-1)$ that are then present in the system at the start of the jth operational cycle.
- v_j represents the expected number of faults injected at $PM(j-1)$.
- $\alpha_{j,t}$ is the total number of faults to have arisen over the interval $\{PM(j-1), PM(j-1)+t\}$ for $0 \le t < (PM(j) - PM(j-1))$. This includes those that are injected during the course of the previous PM, $\alpha_{j,0}$.
- h is the delay time until failure of a fault.
- All defects arising naturally have a delay time governed by some density $f(h)$ and uniform time of arrival u within an interval $(PM(j-1), PM(j))$. The

assumption of a uniform arrival distribution could be an issue for relaxation in some cases.

- All faults injected at PM have a delay time h governed by some density $g(h)$.

Note that, when considering a system that is in a steady state of operation, i.e. the assumption of perfect inspection detection capabilities holds and we have $k_j(t) = k(t)$ and $\nu_j = \nu$, we can assume that each operational cycle is statistically identical and independent. The impact of the steady state assumption is that we can analyse the failure pattern over m cycles of duration T_j where, the underlying dynamics are assumed to have the same properties for each cycle.

For the delay time models discussed in chapter 3, a key element of the modelling process is the function describing the expected number of failures over an interval where, the underlying dynamics are assumed to be in a steady state. The expected number is a function of the chosen delay time distributions with estimated parameters and the number of fault arrivals. Considering an age based fault arrival rate process, we have

$$\mathbf{E}[N_f(PM(j-1),\, PM(j-1)+t)] = (\int_0^t k_j(s)ds)b_j(t) + \nu_j G(t) \qquad [4.1]$$

as the expected number of failures over $\{PM(j-1), PM(j-1)+t\}$ where, equations [4.2] and [4.3] represent the proportion of the relevant fault type (natural or artificial) that will fail on an interval $(0, t)$ after the (j-1)th PM;

$$b_j(t) = \left(\int_0^t k_j(u)du\right)^{-1} \int_0^t k_j(u)F(t-u)du \qquad [4.2]$$

$$G(t) = \int_0^t g(h)dh \qquad [4.3]$$

The functions given by equations [4.2] and [4.3] are essential elements of the proceeding research when developing the required relationship between the observed

failure information and the underlying number of fault arrivals. It is clear that inserting equation [4.2] into equation [4.1] when $k_j(s) = k(s)$ produces the same expected number of failures over $(0, t)$ as we had in the previous chapter.

4.3 Continuous time problem statement

In this section, we consider a sampling period of m consecutive PM cycles and more specifically, the jth cycle within that period and we define the underlying dynamics of the fault arrival and failure process for a complex system. During the jth cycle, all that is observed of the system dynamics by time $PM(j-1)+t$ is the number of failures $\{W_{j,t} = 0, 1, 2, \ldots\}$ that have occurred since the start of the sampling period for any $t \in \{0, (PM(j) - PM(j-1))\}$. At the end of the cycle, we also observe Q_j, the number of faults that remain in the system until the jth PM and are subsequently preventatively removed with probability 1 under the assumption of perfect detection capabilities, see assumption 4. The state of the system, $\alpha_{j,t}$, is defined as the total number of faults to have arisen in the interval $\{PM(j-1), PM(j-1)+t\}$ and this includes faults that have occurred naturally during standard operation in the interval and faults that were injected via human fallibility at the last PM during routine inspection and maintenance procedures, denoted by $\alpha_{j,0}$. In a stochastic filtering context, the system equation for a discrete state, continuous time process describes the underlying dynamics of the fault arrival and potential PM fault injection processes and can be constructed as

$$\alpha_{j,t} = \alpha_{j,0} + \int_0^t k_j(s)ds + noise_{j,t} \qquad [4.4]$$

for $\alpha_{j,t} = 0, 1, 2, \ldots$ The number of failures in the interval $\{PM(j-1), PM(j-1)+t\}$ is described by the expression

$$W_{j,t} = \alpha_{j,0} G(t) + (\alpha_{j,t} - \alpha_{j,0}) b_j(t) + noise_{j,t} \qquad [4.5(a)]$$

where, $W_{j,0} = 0$ and the time dependent functions $b_j(t)$ and $G(t)$ are given by equations [4.2] and [4.3] respectively. Alternatively, equation [4.5(a)] can be expressed as

$$W_{j,t} = (\alpha_{j,t} - \int_0^t k_j(s)ds)G(t) + (\int_0^t k_j(s)ds) b_j(t) + noise_{j,t} \qquad [4.5(b)]$$

In the case of a basic scenario without any provisions for the inclusion of potential fault injection at PM, the initial state is set as $\alpha_{j,0} = 0$ in equation [4.4] and in equation [4.5(b)], we replace $\int_0^t k_j(s)ds$ with $\alpha_{j,t}$.

The objective of the stochastic filter is to provide the best conditional estimate $\mathbf{E}[\alpha_{j,t} \mid W_{j,t}]$ of the total number of faults $\alpha_{j,t}$ that have arisen after t units of time have elapsed in the jth cycle. The estimate is conditioned on the failure history that has been observed until this point. From this result, an estimate of the number of fault injections at the (j-1)th PM, $\alpha_{j,0}$, can also be established as

$$\mathbf{E}[\alpha_{j,t} \mid W_{j,t}] = \mathbf{E}[\alpha_{j,0} \mid W_{j,t}] + \mathbf{E}[\int_0^t k_j(s)ds \mid W_{j,t}] + \mathbf{E}[noise_{j,t} \mid W_{j,t}] \qquad [4.6]$$

where, the expected noise level is 0 and independent of $W_{j,t}$ and the rate of arrival $k_j(s)$ is also not a function of $W_{j,t}$, rather, the relationship applies in reverse because the number of failures is a function of the number of defect arrivals during operation as well as those that are artificially injected. We are therefore left with the result

$$\mathbf{E}[(\alpha_{j,0} \mid W_{j,t}] = \mathbf{E}[\alpha_{j,t} \mid W_{j,t}] - \int_0^t \hat{k}_j(s)ds \qquad [4.7]$$

where, the parameters of the pre-specified arrival rate function $\hat{k}_j(s)$ are estimated from the available failure data pertaining to previous PM cycles. Obviously, with the real-time, on-line operation based on a model established from prior data, the estimate of $\alpha_{j,0}$ is expected to get more accurate as t increases. An additional result that must be incorporated into the parameter estimation process is the number of faults that do not result in failure but are captured during the course of scheduled PM inspections. Defining Q_j as the number of defects identified and removed at *PM(j)*, we have

$$Q_j = (\alpha_{j,T_j} - \int_0^{T_j} k_j(s)ds)(1-G(T_j)) + (\int_0^{T_j} k_j(s)ds)(1-b_j(T_j)) + noise_{j,T_j} \quad [4.8]$$

for $Q_j = 0, 1, 2, \ldots$ where, the noise has a mean of 0. Given that we are employing the assumption of a perfect detection process for existing faults at PM, we have the following result for the *j*th cycle;

$$Q_j = \alpha_{j,T_j} - W_{j,T_j} \quad \rightarrow \quad \alpha_{j,T_j} = Q_j + W_{j,T_j}$$

which means that at time $PM(j-1)+T_j = PM(j)$ we have knowledge of α_{j,T_j}, the total number of faults (both injected at the previous PM and those arising naturally) that have occurred during the cycle and therefore, assuming that $k_j(s)$ has been appropriately defined and parameterised, we also have the best available estimate of the number of faults injected at the previous PM, $\alpha_{j,0}$.

Although the continuous time representation given in this section is an appropriate representation of the underlying processes and reflects the manner in which the observed information is obtained, a satisfactory stochastic filtering approach has not been found to provide accurate estimates of the underlying state when the information is obtained continuously. The general result when applying the

stochastic Ito calculus based approach described in chapter 2, section 2.5.3, is an infinite sequence of stochastic differential equations and approximate solutions are required. As a result, we now consider a discretised version of the same problem scenario and apply the probabilistic Bayesian filtering approach described in chapter 2, section 2.4.3.

4.4 Non-linear stochastic filtering (discrete-time, discrete-state case)

In terms of the manner in which systems of this nature are typically monitored, a discrete time representation may actually be more realsitic. Although a constant interval between discrete time points is used in the proceeding research, this is by no means a necessity. The model can easily be constructed to facilitate for additional updating of the proability density for the number of faults that have arrived upon the occurrence of each failure. In order to differentiate between the state and observation processes for the continuous and discrete time definitions of the problem scenario, we apply the following notation. The processes are defined at discrete intervals within an operational cycle and x_i represents the unknown number of faults that have arrived by the ith time point since the start of the current cycle with x_0 being the number injected at the previous PM. The number of failures that have occurred by the ith time point is denoted by N_i. All additional functions and parameters are unchanged from the continuous time problem definition.

4.4.1 The filtering equations

We assume the system to be in a steady state of operation and a perfect detection process to be in place. We therefore have $k_j(t) = k(t)$, $\nu_j = \nu$ and the inspection process is essentially treated as a renewal process in the sense that, current operational cycles are not affected by the events of previous cycles. Each interval

$(0,T_j)$ is divided into z_j equidistant sequential increments of length Δ, for $j = 1, 2, \ldots, m$. We can therefore express the system state dynamics (representing the fault arrival process) as a first-order Markov process;

$$x_i = x_{i-1} + \int_{(i-1)\Delta}^{i\Delta} k(s)ds + \varepsilon_i \qquad [4.9]$$

for $i = 1,2,\ldots, z_j$ on any PM cycle j where, ε_i is the error in the description of the ith transition. Representing the evolution of the state as a Markov process enables the construction of a discrete-time probabilistic stochastic filter for estimation of the state using stochastically related failure information. Using equation [4.5(b)], the observation dynamics for the failure process can be written as

$$N_i = (x_i - \int_0^{i\Delta} k(s)ds)G(i\Delta) + (\int_0^{i\Delta} k(s)ds)\,b(i\Delta) + \eta_i \qquad [4.10]$$

for $i = 1,2,\ldots, z_j$ where, η_i is the observation error at the ith discrete time point. As with the continuous problem statement, an additional and crucial observation (that must be incorporated into the parameter estimation process) is the number of existing defects removed at PM. Rewriting equation [4.8], we have

$$Q_j = (x_{z_j} - \int_0^{z_j\Delta} k(s)ds)(1 - G(z_j\Delta)) + (\int_0^{z_j\Delta} k(s)ds)(1 - b(z_j\Delta)) + \theta_j \qquad [4.11]$$

for $Q_j = 0, 1, 2, \ldots$ and $j = 1, 2, \ldots, m$ where, θ_j is the jth 0-mean estimation error. The initial state x_0 is assumed to be governed by a probability distribution $p(x_0)$ that is taken for convenience to be Poisson in the examples given later in this chapter however, other forms such as a binomial distribution may be more appropriate in some situations. In addition, we assume that the relationship between the observed N_i and the underlying x_i can be described by a distribution $p(N_i \mid x_i)$. The objective of the probabilistic filtering approach is to obtain an expression for this distribution and in the context under consideration, the distribution is utilised in a predictive manner

to schedule the next maintenance intervention. By the application of Bayes rule, the conditional density can be established as

$$p_i(x_i \mid \underline{N}_i) = p(x_i \mid N_i, \underline{N}_{i-1}) = \frac{p(x_i, N_i \mid \underline{N}_{i-1})}{p(N_i \mid \underline{N}_{i-1})} \qquad [4.12]$$

for x_i, $N_i = 0, 1, 2, \ldots$ and $x_i \geq N_i$. The numerator is established using the chain rule and hidden Markov one-step transition probabilities between potential realisations of the discrete-state process. In the context of hidden Markov modelling, this step of the derivation can be regarded as an application of the discrete Chapman-Kolmogorov equation;

$$p(x_i, N_i \mid \underline{N}_{i-1}) = p(N_i \mid x_i, \underline{N}_{i-1}) p(x_i \mid \underline{N}_{i-1}) = p(N_i \mid x_i) p(x_i \mid \underline{N}_{i-1})$$

$$= p(N_i \mid x_i) \sum_{x_{i-1}=N_{i-1}}^{x_i} p(x_i \mid x_{i-1}) p_{i-1}(x_{i-1} \mid \underline{N}_{i-1}) \qquad [4.13]$$

The denominator is obtained by taking the summation over all potential values of x_i considering that it is known that x_i must be greater than N_i;

$$p(N_i \mid \underline{N}_{i-1}) = \sum_{x_i=N_i}^{\infty} p(N_i \mid x_i) \sum_{x_{i-1}=N_{i-1}}^{x_i} p(x_i \mid x_{i-1}) p_{i-1}(x_{i-1} \mid \underline{N}_{i-1}) \qquad [4.14]$$

Substituting equations [4.13] and [4.14] into equation [4.12] produces

$$p_i(x_i \mid \underline{N}_i) = \frac{p(N_i \mid x_i) \sum_{x_{i-1}=N_{i-1}}^{x_i} p(x_i \mid x_{i-1}) p_{i-1}(x_{i-1} \mid \underline{N}_{i-1})}{\sum_{x_i=N_i}^{\infty} p(N_i \mid x_i) \sum_{x_{i-1}=N_{i-1}}^{x_i} p(x_i \mid x_{i-1}) p_{i-1}(x_{i-1} \mid \underline{N}_{i-1})} \qquad [4.15]$$

There are three separate functions included in the conditional density $p_i(x_i \mid \underline{N}_i)$ that require further explanation;

(i) $p(N_i \mid x_i)$ is a conditional density function that describes the probability of observing N_i failures in the interval $(0, i\Delta)$ given that an underlying number of faults x_i have arisen. In chapter 3 it was established that the number of failures in a given

interval follows a non-homogenous Poisson process (NHPP) with a time-dependent failure rate. As such, we have

$$p(N_i \mid x_i) = \mathbf{E}[N_i \mid x_i]^{N_i} \exp\{-\mathbf{E}[N_i \mid x_i]\} / N_i! \qquad [4.16]$$

when $N_i \le x_i$ and $p(N_i \mid x_i) = 0$ when $N_i > x_i$. We also have the expectation of N_i given x_i as

$$\mathbf{E}[N_i \mid x_i] = (x_i - \int_0^{i\Delta} k(s)ds)G(i\Delta) + (\int_0^{i\Delta} k(s)ds)\,b(i\Delta) \qquad [4.17]$$

where, $b(i\Delta)$ and $G(i\Delta)$ are given by [4.2] and [4.3] for $t = i\Delta$ and $k_j(s) = k(s)$.

(ii) $p_{i-1}(x_{i-1} \mid \underline{N}_{i-1})$ is given by the previous stage of the recursive filter and $p_0(x_0 \mid \underline{N}_0) = p(x_0)$. It is assumed that the number of faults injected at a given PM is governed by a Poisson distribution with mean ν as

$$p_0(x_0 \mid \underline{N}_0) = (\nu^{x_0} e^{-\nu})/x_0! \qquad [4.18]$$

for $x_0 = 0, 1, 2, \ldots$

(iii) $p(x_i \mid x_{i-1})$ are one-step Markov transition probabilities where naturally we must have $x_i \ge x_{i-1}$. From the system expression, equation [4.9], it can be established that the number of faults arriving over an interval $((i\text{-}1)\Delta, i\Delta)$ follows a non-homogenous Poisson process (NHPP);

$$x_i - x_{i-1} \sim Poisson(\int_{(i-1)\Delta}^{i\Delta} k(s)ds) \qquad [4.19]$$

The one-step transition probabilities are thus defined as

$$p(x_i \mid x_{i-1}) = \frac{(\int_{(i-1)\Delta}^{i\Delta} k(s)ds)^{(x_i - x_{i-1})} e^{-\int_{(i-1)\Delta}^{i\Delta} k(s)ds}}{(x_i - x_{i-1})!} \qquad [4.20]$$

when $x_i \ge x_{i-1}$ and 0 otherwise.

4.4.2 Parameter estimation

As with the delay time models of chapter 3 the technique of maximum likelihood estimation (MLE) is employed for parameter estimation purposes. Unlike the MLE approach in the previous chapter, the likelihood function for this model is developed as the product of conditional probabilities. At each discrete time point for the jth PM cycle, the probability that N_i failures have occurred in the interval $(0, i\Delta)$ conditional on the failure history over the interval $(0, (i\text{-}1)\Delta)$ is denoted by $p(N_{ji} \mid \underline{N}_{j,i-1})$. Considering the availability of m cycles of data, the likelihood of observing all the information is

$$L = \prod_{j=1}^{m}\left(p(Q_j \mid x_{z_j})\left(\prod_{i=1}^{z_j} p(N_{ji} \mid \underline{N}_{j,i-1})\right)\right) \qquad [4.21]$$

for $j = 1, 2, \ldots, m$. The functional form of $p(N_{ji} \mid \underline{N}_{j,i-1})$ is given by equation [4.14] and the conditional probability that Q_j defects survive until the jth PM is

$$p(Q_j \mid x_{z_j}) = \mathbf{E}[Q_j \mid x_{z_j}]^{Q_j} \exp\{-[Q_j \mid x_{z_j}]\}/Q_j! \qquad [4.22]$$

for $Q_j = 0, 1, 2, \ldots$ where,

$$\mathbf{E}[Q_j \mid x_{z_j}] = (x_{z_j} - \int_0^{z_j\Delta} k(s)ds)(1 - G(z_j\Delta)) + (\int_0^{z_j\Delta} k(s)ds)(1 - b(z_j\Delta)) \qquad [4.23]$$

The log-likelihood function can be optimised with respect to the parameters of interest using a BFGS quasi-Newton search algorithm, see chapter 2.

4.4.3 Scheduling maintenance activities

As discussed previously, one of the advantages of modelling a complex system in this manner is the ability to make adaptive decisions using the observed failure pattern. There are a couple of ways of establishing a decision model that are discussed here. The first involves an opportunistic PM process whereby, upon the

occurrence of a failure, an optimal estimate of the pre-inspection state of the system is used in determining

$$\hat{A}_i = \hat{x}_i - N_i = \mathbf{E}[x_i \mid \underline{N}_i] - N_i \qquad [4.24]$$

where, $\hat{A}_i$ represents the expected number of faults currently present in the system. The decision to perform PM is then based upon some threshold level for A that is a function of the associated costs. Upon performing the inspection phase of the PM, $A_z = Q$ becomes known information and can be used to update knowledge of the system parameters. The second approach to decision modelling in this context and the method employed in this chapter is to forecast or project the conditional density that is established using the stochastic filter at intervals of Δ. The predictive density is then used to establish the expected number of failures over the range of the projection. With the relevant cost information, an optimal PM time can be selected from prospective intervals of duration 0, Δ, 2Δ, ... The conditional distribution established at the ith time point for x_i given $\underline{N}_i$ is projected at intervals of $w\Delta$ as

$$p(x_{i+w} \mid \underline{N}_i) = \sum_{x_i = N_i}^{x_{i+w}} p(x_{i+w} \mid x_i) p_i(x_i \mid \underline{N}_i) \qquad [4.25]$$

for $w = 1, 2, \ldots$ where, $x_{i+w} \mid x_i \sim \text{Poisson}\left(\int_{i\Delta}^{(i+w)\Delta} k(s)ds \right)$. The conditional probability that $N_{i+w} - N_i$ failures will occur in the interval $(i\Delta, (i + w)\Delta)$ is established by evaluation of

$$p(N_{i+w} \mid \underline{N}_i) = \sum_{x_{i+w} = N_i}^{\infty} p(N_{i+w} \mid x_{i+w}) p(x_{i+w} \mid \underline{N}_i) \qquad [4.26]$$

Denoting c_f and c_{pm} as the average cost of a failure and a PM respectively, the cost function is defined as

$$C(i,w) = \{\mathbf{E}[N_{i+w} \mid \underline{N}_i]c_f + c_{pm}\}/(i+w)\Delta \qquad [4.27]$$

where, the expectation is given by $\mathbf{E}[N_{i+w} \mid \underline{N}_i] = N_i$ for $w = 0$ and

$$\mathbf{E}[N_{i+w} \mid \underline{N}_i] = \sum_{N_{i+w}=N_i}^{\infty} N_{i+w} p(N_{i+w} \mid \underline{N}_i) \qquad [4.28]$$

for $w = 1, 2, \ldots$. The optimal time at which to schedule a PM using all the information currently available is then established as $(i + w^*)\Delta$ where w^* is evaluated as $w^* = \min_w(C(i,w))$ for $w = 0, 1, 2, \ldots$

4.5 Case 1 - basic scenario, $x_0 = 0$

The objective of this first example is to illustrate the modelling and parameter estimation process using specific delay time distributions and parameters for the various component elements and to demonstrate the ability of the filtering approach when tracking the underlying number of faults that have arisen in the system. The scenario under consideration is a single PM cycle from a simple fault arrival process with no artificial fault injection occurring during maintenance interventions. In addition, it is assumed that there is perfect detection of existing faults at PM and a constant defect arrival rate during standard operation, $k(t) = k$.

4.5.1 Modelling the process

The state equation is given for the number of faults that occur in the interval $(0, i\Delta)$ and is derived from the discrete-time problem statement and specifically equation [4.9] in section 4.4.1. The state at the ith time point is described by

$$x_i = x_{i-1} + k\Delta + \varepsilon_i \qquad [4.29]$$

for $i = 1, 2, \ldots, z$. Due to the assumptions regarding perfect PM, the number of fault injections is $x_0 = 0$. The observation equation dictates the underlying dynamics of the process that results in the observed failures and is given by

$$N_i = x_i b(i\Delta) + \eta_i \qquad [4.30]$$

for $i = 1, 2, \ldots, z$, where, $N_0 = 0$ and the proportion of defects resulting in failure over $(0, i\Delta)$ for uniform arrivals is given by $b(t)$ for $t = i\Delta$, see equation [4.2]. The choice of an appropriate delay-time distribution is an important part of model building in this context however, in this case and with the subsequent example in mind, for simplicity, exponentially distributed delay times are used. Considering exponential delay time distributions and the assumption of a constant fault arrival rate, a more basic model could provide equivalent results for this case, however, the example serves to demonstrate the methodology.

We have $f(h) = \lambda e^{-\lambda h}$ for $h \geq 0$ and the proportion of defects resulting in failure becomes

$$b(i\Delta) = 1 - (1/i\Delta)(1 - e^{-i\lambda\Delta}) \qquad [4.31]$$

Completing the equation set, the number of defects identified and subsequently removed at PM is;

$$Q = x_z(1 - b(z\Delta)) + \theta \qquad [4.32]$$

The filtering estimate is provided by the discrete-state conditional density $p_i(x_i|\underline{N}_i)$ given in equation [4.15]. The first two component elements of $p_i(x_i|\underline{N}_i)$ are subject to adjustment as follows;

(i) The distribution $p(N_i \mid x_i)$ is still a standard Poisson distribution for $N_i \leq x_i$ given by equation [4.16] where, under the assumptions of the basic scenario with no

fault injection at PM, the conditional expectation becomes $\mathbf{E}[N_i \mid x_i] = x_i b(i\Delta)$ and $b(i\Delta)$ is given by equation [4.31].

(ii) The only change to $p_{i-1}(x_{i-1} \mid \underline{N}_{i-1})$ concerns the initialisation of the recursive filter. Under the assumptions of this basic system, we have $p_0\big((x_0 = 0) \mid \underline{N}_0\big) = 1$.

An expression for the conditional density $p_i(x_i|\underline{N}_i)$ is developed for the particular scenario described. An example is then presented using simulated data with specified parameters to demonstrate the use of the filter in tracking the underlying fault arrivals.

4.5.2 The filtering expression $p_i(x_i \mid \underline{N}_i)$

As already noted when discussing the component elements of the filtering expression, the initialisation of the filter requires modification under the assumptions of the basic scenario. For the first stage of the process we have a reduced form for the filter due to the fact that there are no faults present in the system at the start of operation. This is attributable to the assumptions regarding perfect PM. A closed form filtering expression is obtainable as

$$p_i(x_i \mid \underline{N}_i) =$$

$$\frac{(k\Delta)^{x_i} x_i^{N_i} e^{-b(i\Delta)x_i} \sum_{x_{i-1}=N_{i-1}}^{x_i} \frac{x_{i-1}^{N_{i-1}} e^{-b((i-1)\Delta)x_{i-1}}}{(x_i - x_{i-1})!} \left(\prod_{d=1}^{i-1} \sum_{x_{i-d-1}=N_{i-d-1}}^{x_{i-d}} \frac{x_{i-d-1}^{N_{i-d-1}} e^{-b(\Delta)x_{i-d-1}}}{(x_{i-d} - x_{i-d-1})!} \right)}{\sum_{x_i=N_i}^{\infty} (k\Delta)^{x_i} x_i^{N_i} e^{-b(i\Delta)x_i} \left(\prod_{d=0}^{i-1} \sum_{x_{i-d-1}=N_{i-d-1}}^{x_{i-d}} \frac{x_{i-d-1}^{N_{i-d-1}} e^{-b(\Delta)x_{i-d-1}}}{(x_{i-d} - x_{i-d-1})!} \right)} \qquad [4.42]$$

where, x_0 and N_0 are 0. Equation [4.42] is necessary for parameter estimation purposes (as discussed in the next section) however from a computational perspective the following recursive expression is more suitable;

$$p_i(x_i \mid \underline{N}_i) = \frac{(b(i\Delta)x_i)^{N_i} e^{-b(i\Delta)x_i} \sum_{x_{i-1}=N_{i-1}}^{x_i} \frac{(k\Delta)^{x_i - x_{i-1}} e^{-k\Delta}}{(x_i - x_{i-1})!} p_{i-1}(x_{i-1} \mid \underline{N}_{i-1})}{\sum_{x_i=N_i}^{\infty} (b(i\Delta)x_i)^{N_i} e^{-b(i\Delta)x_i} \sum_{x_{i-1}=N_{i-1}}^{x_i} \frac{(k\Delta)^{x_i - x_{i-1}} e^{-k\Delta}}{(x_i - x_{i-1})!} p_{i-1}(x_{i-1} \mid \underline{N}_{i-1})} \quad [4.43]$$

An alternative non-recursive formulation of a conditional density for this case is given as

$$p_i(x_i \mid N_i) = \frac{p(N_i \mid x_i)p(x_i \mid x_0 = 0)}{\sum_{x_i=N_i}^{\infty} p(N_i \mid x_i)p(x_i \mid x_0 = 0)} \quad [4.44]$$

where, $N_i \mid x_i$ is as before and $x_i \mid x_0 \sim \text{Poisson}\left(\int_0^{i\Delta} k(s)ds\right)$. However, this formulation does not utilise the failure pattern that has been observed until this point in the current cycle. The issue of parameterisation of the density for this case is now discussed.

4.5.3 Parameter estimation

For this basic scenario, only the interval and inspection data are needed for parameter estimation, the failure times within each cycle are not required. This is attributable to the fact that the behaviour of the failure process during the course of an inspection cycles is not of concern here with the reasoning being that a constant rate of fault arrival k is assumed and there is no defect injection being modelled in this initial trial case. As such we take the number of increments for the jth cycle, $z_j = 1$, and for simplicity we consider a constant interval between inspections for all available cycles of data, denoted by $\Delta = T_j = T$. N_j is defined as the number of failures observed in cycle j, Q_j as the number of faults found at inspection j and x_j as the number of faults that arise during cycle j where, $\{N_j, Q_j, x_j = 0, 1, 2, \ldots\}$. Using equation [4.21], the

likelihood function is modified for this simplified case to incorporate the minimum necessary information for parameter estimation as

$$\mathcal{L} = \prod_{j=1}^{m} p(N_j \mid N_{j0} = 0)\, p(Q_j) \qquad [4.45]$$

From equation [4.14] we have the conditional probability

$$p(N_j \mid N_{j0} = 0) = \sum_{x_j = N_j}^{\infty} p(N_j \mid x_j) p(x_j \mid x_{j0} = 0) p_0(x_{j0} = 0 \mid N_{j0} = 0) \qquad [4.46]$$

where, for the basic scenario with no fault injection at PM, we have the initial probability $p_0(x_{j0} = 0 \mid N_{j0} = 0) = 1$. We also have knowledge of the total number of faults to have arisen in each interval as $x_j = N_j + Q_j$. As a result, the conditional probability reduces to

$$p(N_j \mid N_{j0} = 0) = p(N_j \mid x_j) p(x_j \mid x_{j0} = 0) \qquad [4.47]$$

Naturally, it follows from the relevant definitions that $N_j, Q_j \leq x_j$ and $x_j \geq x_{j0}$.

The constituent elements of equations [4.45] and [4.47] are

$$p(N_j \mid x_j) = \left((b(T)x_j)^{N_j} e^{-b(T)x_j} \right) / N_j! \qquad [4.48]$$

$$p(x_j \mid x_{j0} = 0) = \left((kT)^{x_j} e^{-kT} \right) / x_j! \qquad [4.49]$$

$$p(Q_j \mid x_j) = \left(((1 - b(T))x_j)^{Q_j} e^{-(1-b(T))x_j} \right) / Q_j! \qquad [4.50]$$

Inserting these expressions into the likelihood function gives

$$L = \prod_{j=1}^{m} \left(\frac{(b(T)x_j)^{N_j} e^{-b(T)x_j} (kT)^{x_j} e^{-kT} ((1 - b(T))x_j)^{Q_j} e^{-(1-b(T))x_j}}{N_j! x_j! Q_j!} \right) \qquad [4.51]$$

where, $b(T) = 1 - (1/\lambda T)(1 - e^{-\lambda T})$. Re-arranging the likelihood function and cancelling some of the exponential terms produces

$$\mathcal{L} = \prod_{j=1}^{m} (1/(N_j! x_j! Q_j!))(b(T)x_j)^{N_j} ((1-b(T))x_j)^{Q_j} (kT)^{x_j} e^{-(x_j+kT)} \qquad [4.52]$$

We then simplify the likelihood function further by omitting all the terms that are not a function of the parameters under investigation and are merely multiplicative constants established from the known variables N_j, Q_j and x_j. The likelihood function reduces to

$$\mathcal{L} = \prod_{j=1}^{m} b(T)^{N_j} (1-b(T))^{Q_j} k^{x_j} e^{-kT} \qquad [4.53]$$

As with the maximum likelihood functions that we established in the previous chapter, taking natural logarithms of equation [4.53] can ease the estimation process considerably. The log-likelihood function is

$$l = \sum_{j=1}^{m} N_j \log(b(T)) + Q_j \log(1-b(T)) + x_j \log(k) - kT \qquad [4.54]$$

If we make the substitution $x_j = N_j + Q_j$ and expand the summation in the log-likelihood function, we obtain

$$l = (\sum_{j=1}^{m} N_j) \log(b(T)) + (\sum_{j=1}^{m} Q_j) \log(1-b(T)) + (\sum_{j=1}^{m} (N_j + Q_j)) \log(k) - mkT \qquad [4.55]$$

The function can now be easily maximised with respect to the system parameters under investigation using an optimisation algorithm. Alternatively, if we take the partial differential of the log-likelihood function with respect to k and equate with 0, we acquire the following estimate;

$$\hat{k} = (1/mT) \sum_{j=1}^{m} (N_j + Q_j) \qquad [4.56]$$

Note that the grouping of the failure time information over each interval j and equation [4.56] only apply on the assumption of a constant average fault arrival rate. Estimating the parameters of a non-constant fault arrival function $k(s)$ would require

the partitioning of each interval and consideration of the number of failures occuring in each sub-interval.

This is the same result obtained for the basic model in chapter 3 and is again an estimate that could have been derived easily as the total number of faults divided by the total time over all available cycles. We then would have been left with the basic task of solving either of

$$\mathbf{E}[N_j \mid x_j] = (1/m)\sum_{j=1}^{m} N_j \qquad [4.57]$$

$$\mathbf{E}[Q_j \mid x_j] = (1/m)\sum_{j=1}^{m} Q_j \qquad [4.58]$$

for the single remaining parameter $\hat{\lambda}$ using equations [4.17] and [4.23]. However, the objective of this basic scenario is to demonstrate and test the methodology of the stochastic filtering formulation and the associated maximum likelihood parameter estimation process. In this initial case, we use a simple exponential distribution for the delay time until failure of a given defect. However, the framework that we have developed allows for the use of more complex distributional forms.

4.5.4 Example

This example serves to demonstrate the application of the filter in estimating the underlying state under the assumptions of a basic inspection process. The process is simulated for 50 cycles of duration 200 hours with a constant arrival rate $k = 0.1$ per hour and exponential delay time distribution with parameter $\lambda = 0.025$.

Using equations [4.56] and [4.58] and applying a BFGS algorithm from Matlab© function 'fmincon', the parameters are recovered from the simulated data as $\hat{k} = 0.107$ and $\hat{\lambda} = 0.0262$. As simulated data is being used, the actual underlying fault arrival process is known for any available cycle of failure data. As such, a

further cycle of data is simulated with the same parameter set and the filter given by equation [4.42] is applied to track the progress of the fault arrival process. Figure 4.3 illustrates the recursive filtered estimates at increments of duration Δ = 20 hours against the actual underlying process.

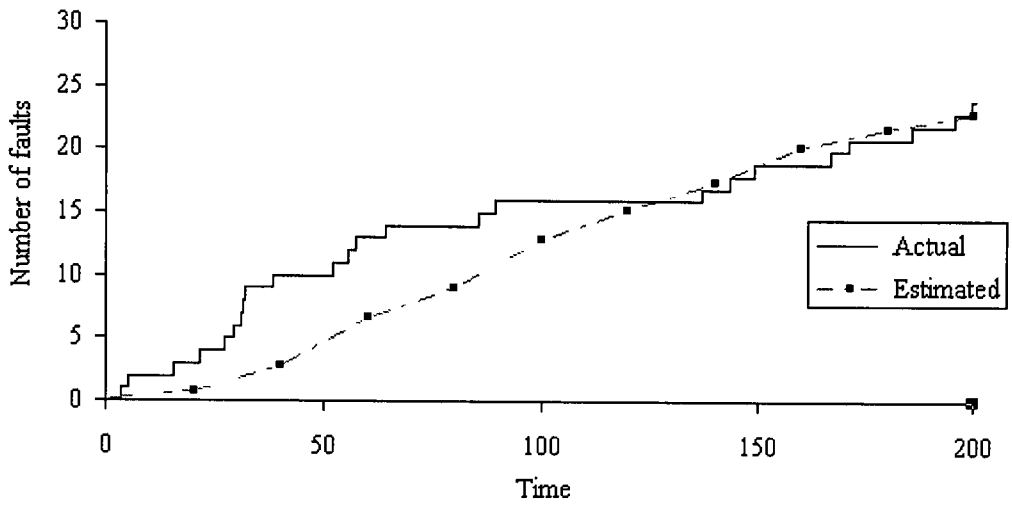

Figure 4.3 - Comparing the actual and filtered estimates of the fault arrival process

From figure 4.3, it is clear that the filter tracks the underlying fault arrival process reasonably effectively although an excessive number of faults over the range 25-40 hours complicated the estimation process. For this basic case, a simpler model could have provided the same results however, the application is intended to demonstrate the use of the filter and the process is extendable to more complex situations as is demonstrated in case 2. Figure 4.4 illustrates the evolution of the filtering conditional density over time for the simulated cycle of data and it is clear that the number of failures observed over time has the desired influence on the structure and range of the conditional density and this effect would ultimately carry forward to modify the replacement decision accordingly.

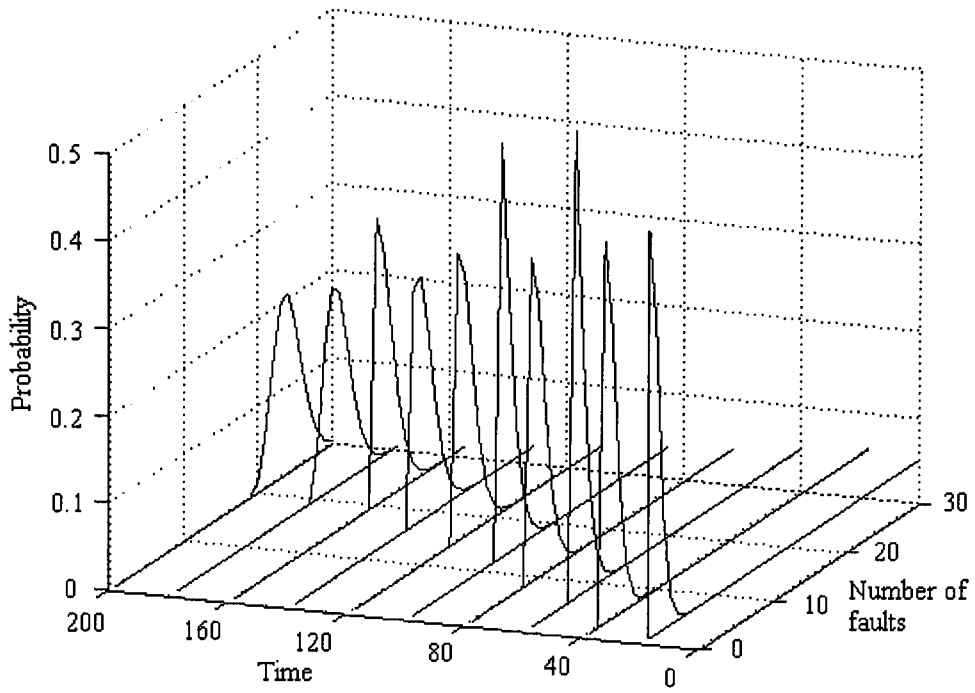

Figure 4.4 - Illustrating the conditional density obtained at each recursion of the filter

The analysis of the subsequent case demonstrates the adaptability of the replacement policy using the filtering approach.

4.6 Case 2 – fault injection scenario, $x_0 \geq 0$

The system described in this second case is subject to human error and the number of fault injections at PM is assumed to be governed by a Poisson distribution with mean v. As with the first case, a constant arrival rate k is assumed during the course of standard operation.

4.6.1 The filtering expression $p_i(x_i \mid \underline{N}_i)$

When programming the stochastic filter for on-line estimation, the conditional probabilities obtained at each recursion can be stored in an array and used directly as input at the next recursion. This approach is more efficient than the utilisation of a

closed form representation as it reduces the level of computation. Using equations [4.15] - [4.17] and [4.20], we have

$$p_i(x_i \mid \underline{N}_i) =$$

$$\frac{\left((x_i - ik\Delta)G(i\Delta) + (ik\Delta)b(i\Delta)\right)^{N_i} e^{-x_i G(i\Delta)} \sum_{x_{i-1}=N_{i-1}}^{x_i} \frac{(k\Delta)^{(x_i - x_{i-1})}\, p_{i-1}(x_{i-1} \mid \underline{N}_{i-1})}{(x_i - x_{i-1})!}}{\sum_{x_i=N_i}^{\infty} \left((x_i - ik\Delta)G(i\Delta) + (ik\Delta)b(i\Delta)\right)^{N_i} e^{-x_i G(i\Delta)} \sum_{x_{i-1}=N_{i-1}}^{x_i} \frac{(k\Delta)^{(x_i - x_{i-1})}\, p_{i-1}(x_{i-1} \mid \underline{N}_{i-1})}{(x_i - x_{i-1})!}} \qquad [4.59]$$

For parameter estimation purposes, we require a closed-form expression for the conditional distribution. The following general expression is available for the *i*th recursion;

$$p_i(x_i \mid \underline{N}_i) = \left\{ \left((x_i - ik\Delta)G(i\Delta) + (ik\Delta)b(i\Delta)\right)^{N_i} e^{-G(i\Delta)x_i} \right.$$

$$\times \sum_{x_{i-1}=N_{i-1}}^{x_i} \frac{(k\Delta)^{(x_i - x_{i-1})}\left((x_{i-1} - (i-1)k\Delta)G((i-1)\Delta) + ((i-1)k\Delta)b((i-1)\Delta)\right)^{N_{i-1}} e^{-G((i-1)\Delta)x_{i-1}}}{(x_i - x_{i-1})!}$$

$$\times \left(\prod_{d=1}^{i-1} \sum_{x_{i-d-1}=N_{i-d-1}}^{x_{i-d}} \frac{(k\Delta)^{(x_{i-d} - x_{i-d-1})} e^{-G((i-d-1)\Delta)x_{i-d-1}}}{(x_{i-d} - x_{i-d-1})!} \right.$$

$$\left. \left. \times \left((x_{i-d-1} - (i-d-1)k\Delta)G((i-d-1)\Delta) + ((i-d-1)k\Delta)b((i-d-1)\Delta)\right)^{N_{i-d-1}} \right) \frac{\nu^{x_0}}{x_0!} \right\} /$$

$$\left\{ \sum_{x_i=N_i}^{\infty} \left((x_i - ik\Delta)G(i\Delta) + (ik\Delta)b(i\Delta)\right)^{N_i} e^{-G(i\Delta)x_i} \right.$$

$$\times \left(\prod_{d=0}^{i-1} \sum_{x_{i-d-1}=N_{i-d-1}}^{x_{i-d}} \frac{(k\Delta)^{(x_{i-d} - x_{i-d-1})} e^{-G((i-d-1)\Delta)x_{i-d-1}}}{(x_{i-d} - x_{i-d-1})!} \right.$$

$$\left. \left. \times \left((x_{i-d-1} - (i-d-1)k\Delta)G((i-d-1)\Delta) + ((i-d-1)k\Delta)b((i-d-1)\Delta)\right)^{N_{i-d-1}} \right) \frac{\nu^{x_0}}{x_0!} \right\} \qquad [4.60]$$

4.6.2 Parameter estimation

The likelihood function given in structural form by equation [4.21] is used to estimate the parameters of the conditional distribution given by equation [4.60].

Attention is now given to the constituent elements of equation [4.21]. Considering a single history, we have the likelihood of observing the history of failures and the number of faults removed at PM as

$$\mathcal{L} = p(Q \mid x_z)\prod_{i=1}^{z} p(N_i \mid \underline{N}_{i-1}) \qquad [4.61]$$

Under the assumption of perfect detection of existing faults, the probability of removing Q faults at PM is conditioned on $x_z = N_z + Q$, the total number of fault arrivals (natural and injected) over the cycle, as

$$p(Q \mid x_z) =$$

$$(1/Q!)\big((x_z - kz\Delta)(1-G(z\Delta)) + (kz\Delta)(1-b(z\Delta))\big)^{Q} e^{-\left((x_z - kz\Delta)(1-G(z\Delta)) + (kz\Delta)(1-b(z\Delta))\right)} \qquad [4.62]$$

The estimation process for the product element of equation [4.61] is also simplified in this case with the availability of x_z. All summations to infinity with respect to x_i are converted to summations with the limit x_z as at no prior stage can the total number of arrivals have been greater. In addition, the final term in the product is simplified to incorporate knowledge of the total number of fault arrivals as

$$p(N_z \mid \underline{N}_{z-1}) = p(N_z \mid x_z = N_z + Q) \sum_{x_{z-1}=N_{z-1}}^{x_z} p(x_z \mid x_{z-1}) p_{z-1}(x_{z-1} \mid \underline{N}_{z-1}) \qquad [4.63]$$

As noted previously, the estimation techniques discussed in the previous chapter are simpler to apply and can also be employed here to obtain the parameters for the filtering expression as the same structural forms for the system apply with the only difference being, in this chapter, the modelling of a distribution for the number of fault injections at PM. However, for the cases discussed here, the initial distribution is assumed to be Poisson and the average number of injections estimated using the techniques documented in chapter 3 can be taken as the mean of the distribution. Similar means of parameterisation could be used if other distributional forms are selected.

4.6.3 Predictive equations

The cost function established by equations [4.27] and [4.28] is dependent on the predicted distribution $p(N_{i+w} \mid \underline{N}_i)$ given in structural form by equation [4.26]. For this particular case, with the assumption of a constant fault arrival rate during operation of the system, the constituent elements of equation [4.26] are

$$p(N_{i+w} \mid x_{i+w}) = \frac{\big((x_{i+w} - (i+w)k\Delta\big)G((i+w)\Delta) + \big((i+w)k\Delta\big)b((i+w)\Delta)\big)^{N_{i+w}}}{N_{i+w}!}$$

$$\times \quad e^{-\left((x_{i+w}-(i+w)k\Delta\right)G((i+w)\Delta) + \left((i+w)k\Delta\right)b((i+w)\Delta)\right)} \qquad [4.64]$$

and

$$p(x_{i+w} \mid \underline{N}_i) = \sum_{x_i=N_i}^{x_{i+w}} \frac{(wk\Delta)^{x_{i+w}-x_i} e^{-wk\Delta}}{(x_{i+w} - x_i)!} p_i(x_i \mid \underline{N}_i) \qquad [4.65]$$

The example presented in the next sub-section illustrates the estimation of the underlying state for the system described in case 2 and compares the estimates obtained at each discrete time point with the best estimate available when the failure information of the current cycle is not utilised. Also, the potential benefits of modelling the system in this manner are demonstrated using the cost function given by equation [4.27].

4.6.4 Example

In the context of the system described for case 2, we simulate a cycle of data with a constant arrival rate during standard operation of k = 0.1 per hour, and an average number of fault injections at PM of ν = 1.5. The delay time distribution is taken to be exponential for both fault types, $f(h) = \lambda e^{-\lambda h}$ for $h \geq 0$ with parameter λ = 0.05. In order to illustrate the forecasting of the predictive density and the adaptive maintenance scheduling policy, we assume costs of c_f = 8 and c_{pm} = 14 for a failure and a PM respectively.

The simulated data is grouped and the analysis conducted at equidistant intervals of duration $\Delta = 10$. As discussed previously, the modelling is easily modified to relax the assumption of equidistant intervals between recursions of the filtering process. For instance, upon failure of the system, the filter could be applied to evaluate whether or not an opportunistic PM could replace the impending failure repair process.

Iteration, i	Time, $i\Delta$	No. faults, x_i	No. failures, N_i
0	0	1	0
1	10	1	0
2	20	2	1
3	30	3	2
4	40	3	2
5	50	4	2
6	60	6	4
7	70	7	6
8	80	8	6
9	90	9	7
10	100	10	9

Table 4.1 - The simulated fault arrivals and failures

Using the cost model proposed by equations [4.25] - [4.28], the optimal PM time can be established at each discrete time point or upon receiving further information, such as the occurrence of a failure. As is illustrated in figure 4.5, with $i = 0$ and $\mathbf{E}[N_{i+w} \mid \underline{N}_i] = N_i$ for $w = 0$, the initial optimal PM time for the system upon commencing a new operational cycle is approximately 54 hours. This estimate can also be obtained using the techniques discussed in the previous chapter with the models adjusted to incorporate cost rather than downtime.

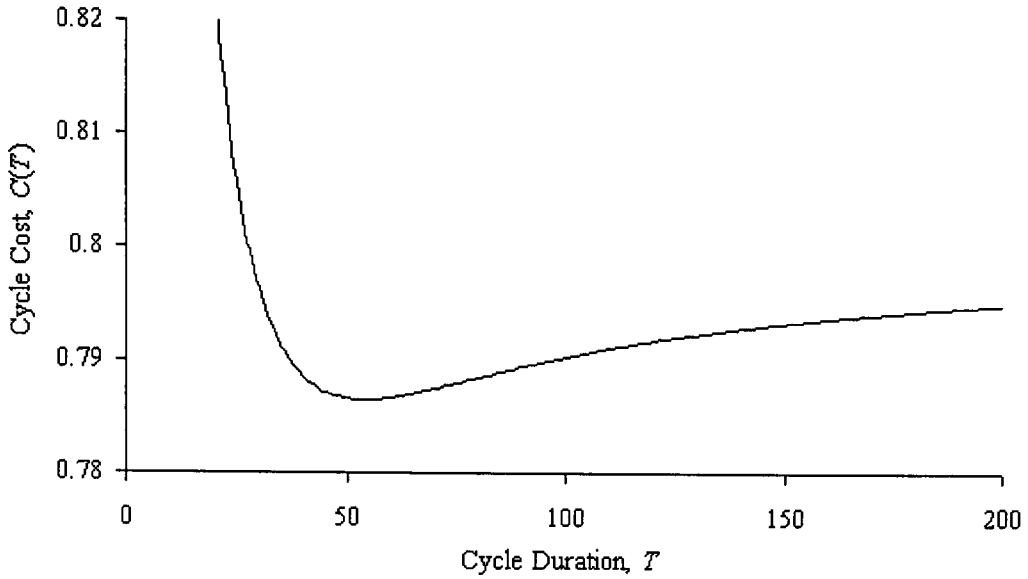

Figure 4.5 - Illustrating the expected cost of a cycle against the duration for the system described in case 2, with parameters k = 0.1/hour, λ = 0.05 and ν = 1.5.

When the dynamic failure information pertaining to cycles that utilise a model (rather than those that are used to establish the model) is not incorporated, the best estimate of the underlying state at a given recursion i within a cycle is simply $\mathbf{E}[x_i]$ rather than $\mathbf{E}[x_i|\underline{N}_i]$. For this example, the best estimate would be $\mathbf{E}[x_i] = \mathbf{E}[x_0] + ki\Delta = \nu + ki\Delta$. Figure 4.6 illustrates the actual underlying fault arrivals, the filtered estimate and the best estimate available without incorporating the failure information. It is clear from the figure that the filter provides a substantial improvement in describing the stochastic behaviour of the system for this cycle than would have otherwise been available.

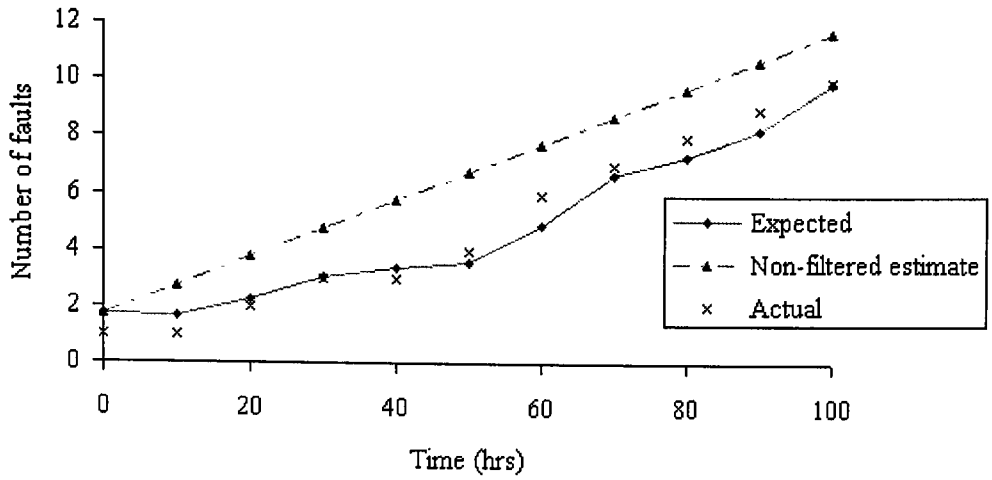

Figure 4.6 - Comparing the actual number of fault arrivals with the filtered estimates and estimates obtained without incorporating the failure history

The expressions for the predictive density and the cost function, equations [4.25] - [4.28], are established at intervals of identical duration to the interval between subsequent recursions of the filter, Δ. However, for this particular example, the state estimation process takes place at intervals of $\Delta = 10$ hours which is not convenient for forecasting purposes. As such the predictive density and cost function are established (at the ith time point) at intervals of duration $i\Delta + w\theta$ rather than $(i+ w)\Delta$ and θ is taken to be 1 hour. Therefore, at each recursion we can obtain the most cost effective time to schedule a PM to the nearest hour. Figure 4.7 compares the PM scheduling decisions obtained at each stage using the stochastic filtering process, the predictive equations and the cost model. As noted previously, when considering a static maintenance policy that does not incorporate the failure history corresponding to the current cycle, the optimal time at which to schedule the next PM is found to be a fixed $T^* = 54$ hours. As such, the maintenance decisions associated with this policy decrease linearly over time, as is illustrated in figure 4.7. The expected cost of this decision is $C(T^* = 54) = 0.78646$. In comparison, when using the adaptive policy that incorporates the stochastic filter, the optimal decision obtained at 60

hours into the cycle is to perform a scheduled PM in 6 hours and at no previous discrete time point is the decision to perform PM less than the interval between check points of 10 hours. As such, the total cost associated with applying the adaptive decision policy for this particular cycle is

$$C(i=6,T=66) = (6(8)+3)/66 = 0.77272$$

as 6 failures occur in the 66 hours.

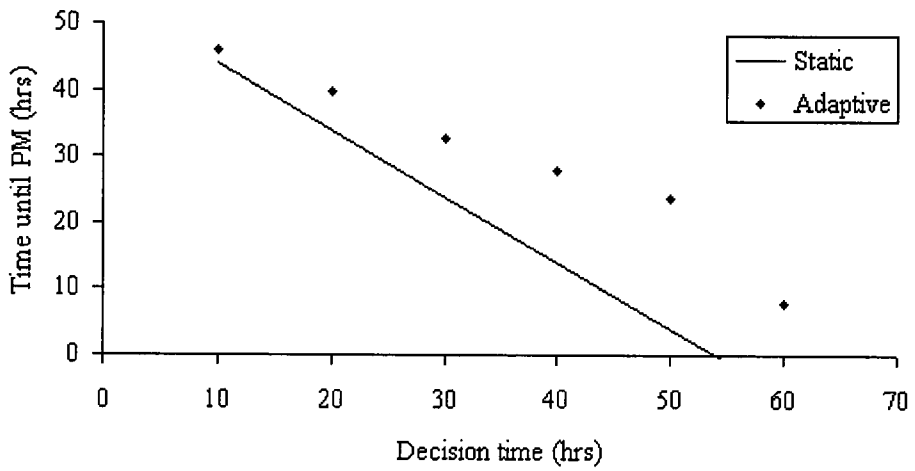

Figure 4.7 - Comparing the scheduling decisions that are available using the stochastic filter and a static PM policy

To conclude, for this second case with the example given, we have demonstrated that discretising the fault and failure processes and applying the stochastic filter incorporating hidden Markov state transitions can provide improved estimates of the state of a complex system (that incorporates fault injection during the PM process) than would have been available without incorporating the failure history attributable to the current cycle. In addition, we have demonstrated that the failure process can be used to provide replacement decisions that potentially result in additional operating time and/or improved costs than a static maintenance policy that does not cater for this information. The adaptive maintenance policy takes into consideration the current failure history and the number of faults that are estimated (using the filter)

to be currently present in the system. For the particular cycle simulated in the demonstration, fewer failures are observed during the cycle than the prior expected number and using the filter, fewer faults are expected to exist in the system at each check point than expected prior to commencing the cycle. As such, the adaptive policy enables an additional 22% operating time for the cycle at a reduced cost. However, it is important to note that the example given is just one potential realisation of the various events and processes.

4.7 A continuous time stochastic filtering representation

In this section, we return to the continuous time definition of the problem scenario given in section 4.3 and the form of an appropriate continuous time stochastic filter is derived in accordance with the modelling discussed in chapter 2, section 2.5.3. The filter is tailored for this particular application where, both the evolution of the underlying state and the continuous observation stream are described by positive integer counting processes. For simplicity, we assume the system to be in a steady state of operation. Considering an individual PM cycle, the number of fault arrivals (underlying state), over any interval $(0,t)$ after a PM, is defined as a counting process and represented as a semi-martingale of the form

$$\alpha_t = \alpha_0 + \int_0^t k(s)db_s + m_t \qquad [4.66]$$

where, b_s is a time dependent function and m_t is a martingale adapted to a σ-field $\{\beta_t\}$ that is generated by all the available processes. The observed number of failures is also a counting process defined as

$$W_t = \int_0^t h_s ds + M_t \qquad [4.67]$$

where, W_t is $\{\beta_t\}$-adapted and M_t is a martingale. A problem that arises from the continuous time definition of the problem scenario when attempting to apply the filtering process is now discussed. From equation [4.67], we require the function h_s to be an integratable function of the state α over the range $(0, t)$ as we are defining a failure rate that is proportional to the number of faults that have arisen. However, the relationship is not specified in this manner in the delay time based representation of section 4.3 when fault injection is incorporated. As a result, we propose the use of an alternative function such as

$$h_s = h + qe^{-\eta s}\alpha_s \qquad [4.68]$$

which can be parameterised to produce either a constant, increasing or decreasing failure rate over time as relevant to the particular application. However, the expression does not contain the rationale of the delay time approach when defining the relationship between the state and observation. This is an obvious drawback particularly with regard to defining and parameterising the constituent processes of the state and observation processes using the techniques of documented in chapter 3. The objective of the methodology outlined here is to obtain the conditional estimate or filter of the state, given the σ-field generated by the observation process $\Im_t^W = \sigma\{W_s, s \le t, t \in T\}$, as

$$\hat{\alpha}_t = \mathbf{E}[\alpha_t \mid \Im_t^W] \qquad [4.69]$$

where, $\Im_t \subset \beta_t$. The result that $M_t^2 - \langle M, M \rangle_t$ is a martingale is utilised in the derivation of the filter and in general, for a counting process, we have $\langle M, M \rangle_t = \int_0^t h_s ds$, i.e. the quadratic variation of the martingale M at time t is equal to the compensator of the process N. We also have the result $M_t m_t - \langle M, m \rangle_t$ is a

martingale and that typically $\langle M, m \rangle_t = 0$. The optimal estimate of the underlying state at time t is given by

$$\hat{\alpha}_t = \hat{\alpha}_0 + \int_0^t \hat{k}(s) db_s + \int_0^t \phi_s [dW_s - \hat{h}_s ds] \qquad [4.70]$$

where,

$$\phi_t = \hat{h}_t^{-1} \left[\mathbf{E}[\alpha_t h_t] - \mathbf{E}[\alpha_t]\mathbf{E}[h_t] + \frac{d}{dt}\langle M, m \rangle_t \right]_{t=t-} \qquad [4.71]$$

for $\hat{h}_t > 0$ and 0 otherwise. However, as noted in section 4.3, the output of the estimation or filtering process is an infinite sequence of stochastic differential equations and approximate solutions are required. This factor, combined with the necessity to approximate the dynamics of the function h_s, makes the probabilistic Bayesian approach, given in the previous section, more appealing for this particular problem. This conclusion is made on the assumption that the interval between the discrete time points is appropriately small to approximate the continuous time manner in which the observations are obtained.

4.8 Summary and discussion

In this chapter, we have investigated the ability to utilise the modelling processes of the previous chapter and develop adaptive maintenance scheduling models that utilise the observed failure information during the course of an operational cycle. The continuous time definition of the problem scenario is used to describe the underlying dynamics using a state space form and the delay time approach. However, as discussed in the last section, an appropriate continuous time estimation procedure could not be established. Discretising the dynamics of the fault and failure arrival processes enabled the construction of a probabilistic Bayesian stochastic filter. The filter utilises the failure history and the time that has elapsed in the construction of a

conditional density for the total number of fault arrivals since the last PM. The parameters of the proposed filter can be estimated using the techniques of chapter 3 however, alternative approaches are discussed that enable the parameterisation of a density function for the number of faults injected at PM. We then proposed an extension of the state estimation methodology in forecasting future failure patterns during the course of an operational cycle and using the forecasts and the relevant costs to optimally schedule the next PM. Examples are given, using simulated data, for a basic fault arrival process and then a process with fault injection at PM. In the first example, we demonstrate the state estimation approach and illustrate the construction of the conditional density. In the second example, the state estimation process is illustrated for the fault injection scenario and the scheduling of the next PM, using the failure pattern of the current operational cycle to date, is demonstrated. For the particular cycle of data considered, use of the PM scheduling model provides an increase in operational availability when compared with a fixed interval policy. Fewer failures than expected occur during the course of the particular cycle we consider and the adaptive model takes this into consideration and recommends prolonging the operational period at each discrete time point. Naturally, different cycles will produce different failure patterns and the stochastic filter is designed to adapt accordingly. The examples are given for relatively simplistic versions of the constituent processes however the approach is the same for more complex systems.

Chapter 5. Condition monitoring and condition-based maintenance

5.1 Introduction to condition-based maintenance

The use of condition monitoring (CM) information in industrial and technological applications is continuously increasing to provide estimates and predictions regarding the condition or state of dynamic systems that are stochastic by nature and subject to some form of random deterioration. Condition-based maintenance (CBM) involves the utilisation of monitoring information in guiding decision making when scheduling maintenance activities. The type of system under consideration in this and the subsequent three chapters is a single working component or system with a dominant failure mode that is monitored using one or more of the CM techniques discussed in the next section. The key topics of interest in CBM applications can be loosely divided into two categories with the first including the identification of any available indicatory condition information and the subsequent monitoring and interpretation of said information. In some cases, this first category could also include a fault diagnosis process designed to indicate that, although the system has not failed, it is operating at a substandard level of performance. The second category of topic in CBM involves the estimation and prediction of the underlying state of a system and the use of this information in associated maintenance decision making activities. In many applications, there will not be a clear distinction between these two categories.

In this chapter, the concept of the state or condition of a single working component and the means by which this information can be inferred from condition data are discussed. The issue of initial fault detection and the various techniques that are

available using CM information are then addressed. The chapter concludes with a review of the relevant CBM literature.

Defining the state is a complex issue in CBM applications and is dependent on the particular system and the available condition information. When considering an initial phase of operation and any fault detection problems, the component can be defined as either in a normal state or defective. However, it is beyond this point that defining the state is non-trivial. Also, some systems or types of monitoring information are not modelled using a two stage process. A measure of the underlying condition is required. Typical applications of many of the state estimation techniques discussed in this report include financial issues or tracking objects through 3D spaces etc. In contexts such as those, the state that we may wish to estimate can be defined as a monetary value or a set of co-ordinates, respectively. However, in the context of CBM, the state/condition of a component is not easily given a value. As such, we consider variables that are related to or are functions of the quality and efficiency of the components operational state at a given moment in time after conception. In chapters 6 – 8, we consider the remaining useful life of a defective component before failure. Many of the systems are also subjected to regular maintenance activities, with the typical scenario being; at planned maintenance checks or inspections, the necessary restorative action is diagnosed and subsequently implemented. The usual assumption is that maintenance produces only a partial restoration of the component/system and that the useful application of maintenance to the same system has a finite duration. As such, when maintenance can no longer restore the unit to a satisfactory workable condition, a replacement is scheduled.

5.2 Condition monitoring

Condition Monitoring refers to techniques that are used to determine or indicate the condition or state of a component. On the subject of CM there is a substantial amount of information available. Of particular use for reference purposes are Collacott (1977) and Tandon & Choudry (1999). Starr & Rao (2001) and the COMADEM proceedings (2001, 2002) provide an extensive account of available monitoring techniques. Condition monitoring information can be classified accordingly; 'Direct information' consists of variables which directly determine the condition or state of the system, such as the wear of a component. The observed information is usually contaminated by noise and as such the actual condition of the component must be inferred from the data. 'Indirect information' is a condition output that is not a direct deterministic measure of the system but is assumed to be stochastically correlated with the unknown underlying state of the system. This includes techniques such as oil analysis or vibration monitoring. A typical approach in this scenario is to model the hazard rate thereby incorporating the condition information into the modelling process. Available CM techniques include;

- vibration analysis
- thermography
- spectometric oil analysis and ferrous debris quantification
- optical microscopy

  ultrasonics and x-ray analysis
- motor current analysis

amongst many others. More recent techniques include the use of larger ranges upon the spectra of a signal, such as the analysis of acoustic emissions however, processing difficulties are common with large amounts of input data requiring

reduction before a fault detection technique could be applied to extract the necessary features in an attempt to inform the user of the likely condition of the component. Techniques such as infra-red imaging for electrical components experience the same kind of processing problems. For many of the techniques described in later sections concerning both fault detection problems and the estimation of the condition or the remaining useful life of a component, the particular monitoring technique is irrelevant as the modelling process is the same for different types of CM input and only differs according to the type of data employed. For instance, whether the information is of a discrete or continuous nature or is stationary etc. In most cases, a direct or indirect/stochastic relationship must be established between the state or remaining life of a component and the indicator information. With many components involved in a system, it becomes necessary to consider which of the components require specific types of monitoring. Techniques such as Fault-tree Analysis or Failure Mode and Effect Analysis can be used to assign monitoring priorities to different components. Improved condition monitoring is likely to develop with the introduction of on-line continuous monitoring of components e.g. with smart sensors and built in vibration sensors being more prevalent on key equipment. This will enable the increased use of CM indicator information in ascertaining or estimating the reliability of components, (as is demonstrated in the next two chapters) in addition to the more commonplace usage in defect identification.

5.3 Initial fault detection

Applying a fault detection process to the monitored indicator information is often the first stage in the CBM modelling process. The detection of a component in a defective state can be a complex task and the sensitivity of 'alarm' systems varies

according to the importance of the individual component or the equipment/unit within which it resides. Many factors may influence both the performance and the condition of a component such as;

- different types of fault
- the type or quality of the installed component
- the severity of the operating cycle
- environmental conditions

Much of the research into fault detection problems has used vibration monitoring or oil-based information as the primary indicators of the state of a component. The vibration monitoring of bearings is most commonplace with rolling element or ball-bearings being amongst the more abundant components in rotating machinery. When considering the vibration monitoring of a bearing, the type of fault may be misalignment, poor installation, rotor imbalance or flow-induced vibration or one of a number of other fault types. The type of bearing installed will obviously have an impact as will the type of lubrication and the environmental conditions such as the ambient temperature. An idealistic fault detection process would incorporate all available factors including the age of the component and the considerations already discussed into the modelling and decision making process in order to asses whether the condition indicators are consistent and whether irregularities or rapid increases in a signal that may indicate the presence of a fault are natural occurrences, such as normal periodic behaviour or increasing trends over time as the component ages, or whether the effects are indeed attributable to the arrival of a defect in the component. More advanced techniques are required to distinguish between said scenarios.

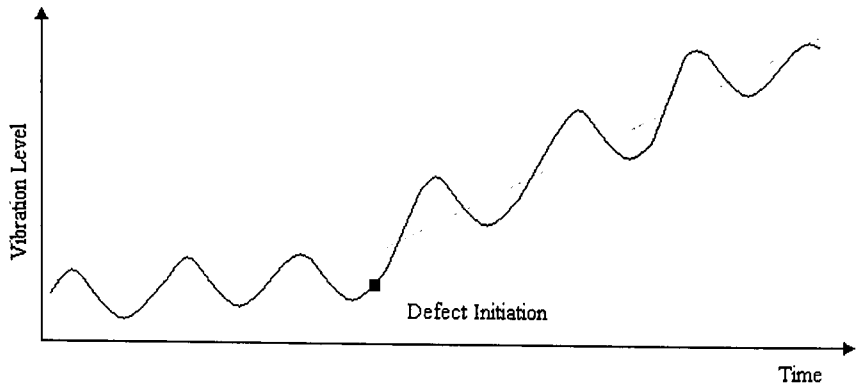

Figure 5.1 - Illustrating a change in the vibration level of a monitored component after the initiation of a defect

Techniques for analysing and deconstructing the vibration signal in order to identify defective components include the following; (i) Spectral analysis using orthogonal decomposition. This essentially involves comparing and testing for significant differences between the spectra of the vibration signal for the two contrasting states; normal operation and defective. (ii) Statistical signal processing and statistical process control (SPC) includes techniques from simple control charts to more complex methods of moment and kurtosis monitoring. Many detection processes monitor the mean^2 level of the signal or examine specific frequencies. Significant changes are assumed to indicate the existence of a defect. (iii) Bispectral analysis provides more information than basic spectral analysis and although it may be more complicated to implement, it is capable of extension to signals that are non-linear and non-Gaussian by considering the phase relations between elements of the signal.

Most approaches to detection of defects using a vibration signal, including those given above (excluding Bispectral analysis), are based upon assumptions regarding the stationarity of the signal however, the signal often includes modulation and there are a variety of signal processing algorithms available that can be used to extract key indicative features of a signal. The algorithms incorporate techniques such as higher-order statistics (extensions of 2nd-order measures e.g. autocorrelation and variance)

such as skewness and kurtosis, cyclostationary statistics, time-frequency analysis and wavelet transforms. These methods can be useful in providing additional information, when compared with standard spectral analysis, as to the structure of the input information. Principal components analysis (PCA) and autoregressive algorithms are useful techniques that can be used to reduce or compress the dimensions of observed temporal and spectral indicator information in order for feature extraction to be implemented. The method can be applied at the start of the process to ascertain the factors that have greater potential as indicators of the condition or alternatively, PCA could be applied iteratively, in a more dynamic manner, at each stage in an attempt to focus on the effects that may currently be indicative of the condition/state of the component. This is particularly useful in situations where different fault types produce irregularities in different monitoring processes. See the application documented in chapter 7 for more information on the use of PCA for reducing the dimensionality of multivariate data.

Advances in condition monitoring and fault detection are constantly arriving via the use of improved fuzzy logic, neural network techniques, genetic algorithms and the development of "expert" systems that interpret and fuse sensor data with prior and expert knowledge to improve condition classification and automation opportunities. Genetic Algorithms (GA) are becoming more popular as a means of selecting features of inputs that are currently indicative of the state of the monitored equipment, see Man et al (1999) for details of the available GA's.

Artificial neural networks are increasingly common as regards general pattern recognition with many architectures and training algorithms now available to classify high-dimensional non-linear indicator information, examples of these include; Multi-layered Perceptron's (MLP), Tree-structured Self-organising Map's (TS-SOM) and

Self-organizing Map's (SOM). Support Vector Machines (SVM) now exist for pattern recognition via recent developments that now provide improved classification using modified kernels and feature selection (GA's). See the DAME project and Quick system for further information on the use of these techniques for CM diagnosis and prognosis applications, Austin et al (2003) and Nairac et al (1999).

Further recent developments in condition monitoring and fault diagnosis include; the use of a Hidden Markov modelling (HMM) approach to the problem of defect detection, see Wang (2004), and Morphological Signal Processing that encompasses a broad collection of concepts and tools for signal analysis and non-linear operations that are derived through mathematical morphology which draws upon results in set theory lattice algebra and stochastic geometry.

5.4 CBM background

In this section, some of the notable contributions in the CBM literature are discussed. As noted in section 5.2, the choice of CBM technique is dependent on the type of condition information being utilised in the model, i.e. whether the monitored information is of a direct or indirect nature. The direct monitoring of actual condition has been prevalent in condition based maintenance literature to date, see Christer & Wang (1995), however, the issue of indirect monitoring is far more complex from the point of view of constructing reliable decision models. With direct CM information such as the analysis of the wear of a unit or measuring crack depth, the cumulative component degradation is often subjected to state space discretisation and subsequently modelled as a jump process, see for instance Easry et al (1973) or Mercer (1961) for examples. Studies incorporating direct monitoring where the cumulative component wear is modelled as a continuous process, include Abdel-Hameed (1975), Gilgmayr (1987), Park (1988) and Christer & Wang (1992, 1995).

Aven (1996) and Heinrich & Jensen (1992) developed stochastic counting process models for degradation applied to directly observed systems and Wang (2000) presents a random co-efficient growth model for state estimation and prediction. With indirect CM information, a common approach is to model the hazard; Newby (1993) presents an overview of hazard based models. The most widely used hazard-based technique is the proportional hazards model (PHM). The proportional hazards model has seen frequent applications in a medical context for quite some time, see Cox (1973), but has only been applied to industrial maintenance problems relatively recently. Studies using the PHM in a maintenance or replacement based context include Ansell & Philips (1989), Bendell et al (1986), Makis & Jardine (1991a and b) and Kumar & Westberg (1997). Banjevic et al (2001), Banjevic & Jardine (2004) and Vlok et al (2002) use a proportional hazards model where the covariate process is approximated by a discrete state Markov chain. However, the PHM and a Markovian model developed by Gong & Tang (1997) amongst others, only use the current CM observation and not the entire component history when estimating the expected residual life of the component. This issue is discussed in more depth for the PHM in the subsequent chapters. Coolen & Dekker (1995) explore the sensitivity analysis of a 2-stage approach to maintenance decision making however, the paper does not include any recommendations regarding parameter estimation which forms a substantial part of the problem. Scarf (1997) presented a review of papers in condition monitoring until that point in time and notes that many more models and applications are required in the area of CBM in order for industry to utilise the CM information that is often obtained at some expense, effectively. Wang et al (2000) use a gamma process and vibration monitoring for state estimation and CBM applied to plant.

It is vibration monitoring that provides the stochastically related condition information for the analysis in the next chapter. The focus of the investigation is on the ability of different CBM techniques to predict the expected time remaining before a component fails. The state of the system is defined as the residual life and the objective is to make optimal recommendations regarding the scheduling of replacements. The vibration information is obtained from the irregular monitoring of components in a laboratory fatigue experiment and the techniques employed and compared are the aforementioned PHM, see Makis & Jardine (1991), and a probabilistic stochastic filtering approach. Chapter 7 provides a similar comparison of the two techniques using the level of metal wear particles found in oil samples as the indicator information. The modelling process is modified slightly for both the PHM and the filter and the issue of multiple indicators of condition is addressed. See chapter 2 for details on the filtering methodology. Both modelling approaches are established in the CBM literature. The first paper to utilise a stochastic filtering approach and the concept of 'conditional residual time' is Christer et al (1997) where a linear state space model and the Kalman filter are developed and applied to a case study on an industrial furnace erosion problem. In the paper by Wang & Christer (2000), the convenient assumptions of linearity and Gaussian distributed disturbances are relaxed in an attempt to propose a general probabilistic stochastic filtering model for CBM applications that incorporates the partially restorative effects of maintenance interventions. The model is subsequently applied to a specific case using vibration monitoring in Wang (2002). The topics of appropriately selecting monitoring intervals and replacement times for CBM applications are discussed in Wang (2003). Lin & Makis (2002) present a stochastic filtering process for a continuous-discrete model using oil analysis and more recently, Wang (2004)

presented a hidden Markov model for state estimation with an emphasis on the identification of the different operational phases for a 2-stage monitoring process.

Chapter 6. The proportional hazards model and a stochastic filter for condition based maintenance applications

6.1 Introduction

In this chapter, we discuss two established techniques for estimating the probability of failure and the expected time until failure (or residual life) of a component given stochastically related condition monitoring (CM) parameters. The CM parameters are collected at regularly or irregularly spaced discrete time points. The techniques that we consider for establishing the conditional probability densities for the residual life are stochastic filtering and the proportional hazards model (PHM). We consider a probabilistic approach to the filtering problem (see chapter 2, Jazwinski (1970) and Aoki (1967)) whereby the initial condition and the relationship between the observed monitoring parameter and the condition of the component are both modelled using a probability density function. The PHM was introduced in Cox (1972) and the form of the general PHM as used throughout chapters 6 -8 is given in Cox & Oakes (1984). The issue of parameter estimation is addressed for the two approaches before a brief section on the selection of an appropriate measure of the underlying state or condition. Before discussing the modelling process for the two state estimation techniques, some general assumptions and notation that are applicable to both approaches are presented;

- CM can be regular or irregular and takes place at discrete time points.

  There is no maintenance or preventative means other than replacement.

When considering the estimation of the state and the decision model for an individual component, the following notation is applied;

- T is the failure or suspension time of the component,
- t_i is the time of the ith CM point,
- $\underline{y}_i = \{y_{1i}, y_{2i}, \ldots, y_{ri}\}$ is a vector of CM information parameters observed at

time t_i where y_{ki} is the kth element of $\underline{y}_i$ for $k = 1, 2, \ldots, r$.

- $\underline{Y}_i = \{\underline{y}_i, \underline{y}_{i-1}, \ldots, \underline{y}_1\}$ is the CM history of the unit that is available at the ith monitoring point.

When analysing a given data set consisting of m components (j = 1, 2,..., m) and estimating the necessary parameters for both models, it is necessary to define a distinction between the individual components. An additional subscript j is added to the notation given above and n_j is the number of CM points for the jth component before failure/replacement.

6.2. The proportional hazards model

In this section, we discuss the proportional hazards model (PHM) where the evolution of a covariate process is approximated by a continuous-time, discrete-state Markov chain. The use of a conditional reliability function in establishing an iterative failure-time distribution is addressed. For the failure time distribution, we consider the expected failure time and optimal replacement times at monitoring points throughout the lifetime of the component.

6.2.1 The hazard and reliability function

The hazard rate is a common feature of reliability analysis and is usually defined as the instantaneous failure rate. The hazard rate for the PHM is given by

$$h(t_i, \underline{y}_i) = h_0(t_i)\lambda(\underline{\gamma}, \underline{y}_i) \qquad [6.1]$$

where, $h_0(t)$ is the baseline hazard that is dependent only on the operational age of the unit and $\lambda(\underline{\gamma}, \underline{y}_i)$ is an adjusting functional term consisting of a vector of coefficients $\underline{\gamma}$ and the time dependent covariates $\underline{y}_i$. The function $\lambda(\underline{\gamma}, \underline{y}_i)$ can take a number of different forms with the most common choice being the exponential form;

$\lambda(\underline{\gamma}, \underline{y}_i) = \exp\{\underline{\gamma}' \underline{y}_i\}$ utilised in the next two chapters. When the baseline hazard is left unspecified, the model utilising the hazard rate given by equation [6.1] is known as the 'semi-parametric Cox PHM'. There are two kinds of covariate that the PHM can incorporate into the estimation process with the first being 'concomitant' factors. This categorisation applies to factors or variables that are not functions of the particular unit but have a bearing upon the condition and performance of the component, such as the environmental conditions. The second type of covariate catered for by the PHM is the category of 'diagnostic variables' consisting of covariates that indicate the likely state or expected time until failure of the component. To ascertain the reliability of a component at time t with hazard rate given by equation [6.1], we are required to approximate a continuous vector sample path $\underline{y}(s)$ from the observations taken at discrete time points. The monitoring points are denoted by $(0 = t_0) < t_1 < \ldots < t$. Using the approximated $\underline{y}(s)$ the reliability can be evaluated as

$$R(t, \underline{Y}) = \exp\{-U(t, \underline{Y})\} = \exp\left\{-\int_0^t h(s, \underline{y}(s))ds\right\} \qquad [6.2]$$

where $U(t, \underline{Y})$ represents the cumulative hazard until time t. Makis & Jardine (1991) advocate the use of a constant value for $\underline{y}(s)$ between the inspection points due to the variability that is typically observed in monitored information however, linear interpolation between monitoring points is also an option.

A similar model to the PHM is the accelerated life model, see Cox & Oakes (1984), however, the vector of covariates are uilised within the hazard as

$$h(t_i, \underline{y}_i) = h_0(t_i, \lambda(\underline{\gamma}, \underline{y}_i))$$

rather than in the form given in equation [6.1] for the PHM. In accelerated life

testing, stress covariates are increased over time until a failure occurs. See Lawless (1982) for an example of accelerated life testing using voltage as a covariate.

6.2.2 Modelling covariate behaviour

There are a number of approaches that can be used in order to model and predict covariate behaviour including regression models and time series. In the research documented here, an approach outlined by Makis and Jardine (1991) is used whereby the PHM is combined with a discrete Markov process for estimation and prediction of the covariates in question. The Markov process is designed to predict the covariate development and also indicate the probability that the component will move to a failure state over specified intervals, i.e. prediction of residual life expectancy. The covariate process is subjected to discrete approximation thereby reducing the condition information to a finite number of states and a state transition probability matrix is established. Differing behaviour of covariates over time can be modelled using a non-homogenous Markov chain. The non-homogeneity can be established using a number of different options such as a time dependent transition matrix or the time scale can be divided into intervals of approximately identical transition probabilities, see Therneau & Grambsch (2000) and Fisher & Lin (1999). The following results for the transition rates or probabilities are presented for an individual element of the CM process $\underline{y}$ on the assumption that the individual elements are independent. In practical situations this is often not the case and some suggestions for resolving the issue are discussed later in this chapter and applied in chapter 8. For cases involving regular (identically spaced) condition monitoring, the interval between inspections is denoted by Δ. Considering an individual element of the CM covariate process, the transition probabilities

$$P_{k,ab}(\Delta) = P(y_{k,(i+1)\Delta} = b \mid y_{k,i\Delta} = a) \qquad [6.3]$$

are easily estimated from the data on a classical frequency basis for $k = 1, 2, \ldots, r$. With irregular condition monitoring, the rates of transition must be estimated first and the transition probabilities established from these. If we define the CM observation process $\{y_k(t), t \geq 0\}$ as a continuous-time, discrete-state Markov process with state space $S_k = \{1, 2, \ldots, s_k\}$ and establish the probability matrix

$$\underline{P}_k(t) = P_{k,ab}(t) = P(y_k(t) = b \mid y_k(0) = a) \qquad [6.4]$$

from the matrix of transition rates $\underline{\Lambda}_k = [\lambda_{k,ab}]$, the transition rates between the states of $y_k(t)$ can be expressed as

$$\lambda_{k,a} = \lim_{h \to 0}\left(\frac{1 - P_{k,a\cdot}(h)}{h}\right) \qquad [6.5]$$

$$\lambda_{k,ab} = \lim_{h \to 0}\left(\frac{P_{k,ab}(h)}{h}\right) \qquad [6.6]$$

for $a, b \in S_k$ where $P_{k,a\cdot}(h)$ is the total probability of moving to any state other than a. To estimate these transition rates from data we use the maximum likelihood approach. The log-likelihood of observing the m independent sample paths of $y_k(t)$ is given by

$$l_k = C + \sum_{a}^{s_k}\sum_{b}^{s_k} N_k(a,b)\log(\lambda_{k,ab}) - \sum_{a}^{s_k} W_k(a)\lambda_{k,a} \qquad [6.7]$$

where, C is a constant that is omitted from estimation, $N_k(a,b)$ is the number of transitions from a to b during all m sample paths and $W_k(a)$ is the total time spent in state a during all m sample paths. Using a result

$$-\lambda_{k,a} = -\sum_{b \neq a} \lambda_{k,ab} \qquad [6.8]$$

from Doob (1953), we can derive the maximum likelihood estimate for each rate of transition between states as

$$\hat{\lambda}_{k,ab} = \begin{cases} \dfrac{N_k(a,b)}{W_k(a)} & \text{if } a \neq b \quad \text{and} \quad W_k(a) > 0, \\ 0 & \text{if } a \neq b \quad \text{and} \quad W_k(a) = 0. \end{cases} \qquad [6.9]$$

When converting the transition rates to probabilities over a suitably short interval ω we assume that, at most, one transition between states could occur. The transition probabilities over ω are derived as

$$p_{k,aa}^{(\omega)} = \exp\left(-\left(\sum_{b \neq a} \lambda_{k,ab} \right) \omega \right) \qquad [6.10]$$

$$p_{k,ab}^{(\omega)} = \left\{ 1 - \exp\left(-\left(\sum_{c \neq a} \lambda_{k,ac} \right) \omega \right) \right\} \left(\frac{\lambda_{k,ab}}{\sum_{c \neq a} \lambda_{k,ac}} \right) \qquad [6.11]$$

The second component of $p_{k,ab}^{(\omega)}$ is for situations where more than one state is directly communicable from the current state. Note that c represents the directly communicable states from the current state and also that $b \in c$. For non-homogenous processes, the transition probabilities must be calculated separately for each phase of homogenous behaviour.

6.2.3 Parameter estimation

Conditional or partial likelihood techniques (developed by Cox) can be used to estimate the parameters of the PHM. This technique can also be used to estimate the parameters of the functional term when the form of the baseline hazard is not specifically defined in a functional form. However, the subsequent non-parametric estimation of the baseline hazard can be complex. Other approaches such as extreme value transformation can be used to obtain the parameter estimates however, for the proceeding research we follow an approach recommended in a CBM context in Vlok et al (2002). For the model developed in Vlok et al (2002), parameter estimation can

be undertaken with a distinction between units operated until failure and those that operated until a 'suspension' (censored data) that could be the end of a test period (calendar suspension) or a preventive replacement. Failures are indicated by $c_j = 1$ and suspensions by $c_j = 0$. The number of failures in a data set is given by $\phi = \sum c_j$ and the number of suspensions is $m - \phi$ as the data consist of m independent lifetimes with each unit denoted by $(T_j, \underline{Y}_{j,n_j}, c_j)$, for $j = 1, 2, \ldots, m$, where T_j is the failure or suspension time. Maximum likelihood estimation can be used to obtain the parameters of the PHM. The likelihood function includes the hazard and conditional reliability functions and is constructed as

$$\mathcal{L} = \prod_l h(T_l, \underline{y}_{l,T_l}) \prod_j R(T_j, \underline{Y}_{j,n_j}) \qquad [6.12]$$

where, l indexes failure times only, j represents both failures and suspensions and $h(T_l, \underline{y}_{l,T_l})$ is given by equation [6.1]. In the reliability function $R(t, \underline{Y})$ at time t, equation [6.2], the cumulative hazard is represented by $U(t, \underline{Y})$ and in the context of establishing an approximate reliability function for parameter estimation purposes, we have

$$R(T_j, \underline{Y}_j) = \exp\{-U(T_j, \underline{y}_j(T_j))\} \qquad [6.13]$$

Finally, we construct r continuous-time sample paths for the jth component that are collectively represented by $\underline{y}_j(t)$ for $t \geq 0$. The individual path for unit j, principle component k, ($k = 1, 2, \ldots, r$), is a constant stepwise function that is constructed from the CM parameters available at the discrete monitoring points. It is denoted by $y_{jk}(t)$ and defined on the discrete state space $S_k = 1, 2, \ldots, s_k$ where, s_k represents the number of states. When there is a transition in the state-space of any of the

paths $y_{jk}(t)$, we have a transition in $\underline{y}_j(t)$, i.e. when any of the r elements in the continuously evolving vector $\underline{y}_j(t)$ changes value, we have a transition in the overall state. To obtain an approximation of the cumulative hazard, we convert the continuous $\underline{y}_j(t)$ into a discrete semi-Markov process $\{\underline{A}_{jd}, \tau_{jd}\}$ where, τ_{jd} is the elapsed time in the dth state, $\underline{A}_{jd}$ is a vector representing the dth state of the observed path $\underline{y}_j(t)$ for $d = 1, 2, \ldots, q_j$ and q_j is the number of consecutive states observed for the jth component. Semi-Markov processes are used frequently in queuing and reliability theory. Markov renewal processes are related to semi-Markov processes and describe the number of times a process is in a given set of states during a period of time. The cumulative hazard at the failure or suspension time of the jth unit, T_j, is approximately

$$U(T_j, \underline{y}_j(T_j)) = \sum_{d=1}^{q_j} \left(\int_{V_{j,d-1}}^{V_{jd}} h(s, \underline{A}_{jd}) ds \right) \qquad [6.14]$$

where, V_{jd} is the time of the dth transition and we have $V_{j0} = 0$ and $V_{jq_j} = T_j$. The transition times are established as $V_{jd} = \sum_{a=1}^{d} \tau_{ja}$.

6.2.4 The conditional reliability function

In this section, we consider the incremental approximation of the conditional reliability function as a discrete semi-Markov process at discrete time points throughout the life of a component given the monitored history to date. Using the conditional reliability function, a conditional failure time distribution (or residual life distribution) is established at each of the discrete time points on the assumption that, within a very small increment, only one transition between covariate states may occur

with an associated probability. From equation [6.2], the reliability at time t is defined as $R(t,a) = P(T > t, \underline{y}(t) = a)$ where a represents a vector state of the covariate Markov process and $\underline{y}(t_i)$ is an approximation of the discrete observation vector $\underline{y}_i$. Given that at time t_i, the observed CM information is within covariate state 'a', the conditional reliability at time $t_i + x_i$ with a covariate vector that is within state 'b' can be expressed as

$$R(t_i + x_i, b \mid t_i, a) = P(T > (t_i + x_i), \underline{y}(t_i + x_i) = b \mid T > t_i, \underline{y}(t_i) = a) \qquad [6.15]$$

where, $R(t_i, b \mid t_i, a) = \lim_{x_i \downarrow 0} R(t_i + x_i, b \mid t_i, a) = \partial_{ab}$, and ∂_{ab} represents a Kronecker-delta. At time t_i, with the continuous-space CM output from the component, $\underline{y}_i$, converted to the discrete-space, continuous-time series $\underline{y}(t_i)$ and taking ω to be a suitably small increment, we have

$$R(t_i + \omega, b \mid t_i, a) =$$

$$P(\underline{y}(t_i + \omega) = b \mid T > (t_i + \omega), \underline{y}(t_i) = a)\ P(T > (t_i + \omega) \mid T > t_i, \underline{y}(t_i) = a) \qquad [6.16]$$

As noted above, within $(t_i, t_i + \omega)$, we assume that one or no transitions between states of the covariate process can occur, the conditional reliability after a single increment is given by

$$\widetilde{R}(t_i + \omega, b \mid t_i, a) = \exp\left\{ - \int_{t_i}^{t_i + \omega} h(s,b) ds \right\} \widetilde{P}_{ab}(\omega) \qquad [6.17]$$

where, $\widetilde{P}_{ab}(\omega)$ is the probability of a transition from covariate state vector a to state vector b within an increment ω. As discussed previously, the kth covariate in the vector $\underline{y}(t)$ can be in any one of s_k prescribed states, $y_k(t) \in S_k = \{1, 2, ..., s_k\}$ at time t. Now, considering this discretised observation process, the Chapman-Kolmogorov equation for a discrete-time, discrete-state process is a recursive formula for the

derivation of n-step transition probabilities using the associated one-step probabilities between all known states. Taking y to be a general state and $h < u$, the Chapman-Kolmogorov equation is

$$p(y_u \mid y_h) = \sum_{y_{u-1}} p(y_u \mid y_{u-1}) p(y_{u-1} \mid y_h) \qquad [6.18]$$

We can now define an approximation to the conditional reliability function at time t_i given the vector of CM information parameter $\underline{y}_i$. For the conditional reliability function constructed here, the cumulative hazard approximately incorporates the transitions between states and the potential for multiple transitions. From equations [6.16] and [6.17], the conditional reliability after a single increment of duration ω is given by

$$R(t_i + \omega, \underline{y}(t_i + \omega) \mid t_i, \underline{y}(t_i)) = \exp\left\{ - \int_{t_i}^{t_i+\omega} h(s, \underline{y}(t_i + \omega)) dt \right\} p(\underline{y}(t_i + \omega) \mid \underline{y}(t_i)) \qquad [6.19]$$

Generalising equation [6.19] using equation [6.18], we can obtain an expression for the approximate conditional reliability for the vector $\underline{y}(t_i + u\omega)$ after u increments, $u = 1, 2, 3\ldots$, as

$$R(t_i + u\omega, \underline{y}(t_i + u\omega) \mid t_i, \underline{y}(t_i)) = \exp\left\{ - \int_{t_i+(u-1)\omega}^{t_i+u\omega} h(s, \underline{y}(t_i + u\omega)) ds \right\}$$

$$\times \sum_{\underline{y}(t_i+(u-1)\omega)} p(\underline{y}(t_i + u\omega) \mid \underline{y}(t_i + (u-1)\omega)) R(t_i + (u-1)\omega, \underline{y}(t_i + (u-1)\omega) \mid t_i, \underline{y}(t_i)) \qquad [6.20]$$

where, for $u = 1$, we have $R(t_i + (u-1)\omega, \underline{y}(t_i + (u-1)\omega) \mid t_i, \underline{y}(t_i)) = 1$. The joint transition probabilities are

$$p(\underline{y}(t_i + u\omega) \mid \underline{y}(t_i + (u-1)\omega)) = \prod_{k=1}^{r} p(y_k(t_i + u\omega) \mid y_k(t_i + (u-1)\omega)) \qquad [6.21]$$

for u = 1, 2, ... Equation [6.21] is established on the assumption that the individual CM parameters are independent and therefore their joint transition probability is simply the product of the individual transition probabilities. The transition probabilities for the CM parameters are derived from the transition rates between the discrete states, see equations [6.10] and [6.11]. In reality, it is likely that the parameters will be correlated with one another, which will have a negative impact upon the parameter estimation process. A number of different means of addressing this problem, using data transformation techniques, are discussed later in this chapter and applied to real case data in chapter 8. Defining the residual life at the ith monitoring point as $x_i = u\omega$, the conditional reliability at time $t_i + x_i$ can be discretely approximated as

$$R_i(x_i \mid \underline{y}_i) = R(t_i + u\omega \mid t_i, \underline{y}(t_i)) = \sum_{\underline{y}(t_i+u\omega)} R(t_i + u\omega, \underline{y}(t_i + u\omega) \mid t_i, \underline{y}(t_i)) \qquad [6.22]$$

Accuracy could potentially be improved by adjusting the model so that transitions occur at the mid-point of an increment. Further possibilities include applying the same probabilistic approach to the cumulative hazard as

$$H_m^{(\underline{y}(t_i+u\omega))} = \sum_{\underline{y}(t_i+(u-1)\omega)} \left\{ \int_{t_i+(u-1)\omega}^{t_i+u\omega} h(s, \underline{y}(t_i + u\omega))dt \; p(\underline{y}(t_i + u\omega) \mid \underline{y}(t_i + (u-1)\omega)) + H_{u-1}^{(\underline{y}(t_i+(u-1)\omega))} \right\} \qquad [6.23]$$

where, $H_0^{(t_i)} = 0$. The conditional reliability at time $t_i + x_i$ is

$$R_i(x_i \mid \underline{y}_i) = R(t_i + u\omega \mid t_i, \underline{y}(t_i)) = \exp\left(- \sum_{\underline{y}(t_i+u\omega)} H_u^{(\underline{y}(t_i+u\omega))} \right) \qquad [6.24]$$

with $x_i = u\omega$ for u = 1, 2, 3...

6.2.5 The conditional failure-time distribution

At time t_i, with x_i as the time remaining before the component fails, we are able to establish the conditional failure time distribution $p_i(x_i \mid \underline{Y}_i)$ at the ith CM point as

$$p_i(x_i \mid \underline{Y}_i) = p_i(x_i \mid \underline{y}_i) = -\frac{d}{dx_i}(R_i(x_i \mid \underline{y}_i)) \qquad [6.25]$$

which is a standard result, see Cox & Oakes (1984). As the conditional reliability function is defined at discrete intervals ($x_i = \omega, 2\omega, 3\omega...$) and incorporates transitions between states of the CM covariate processes, we normalise the conditional density given by equation [6.25] as

$$p_i(x_i \mid \underline{Y}_i) = \frac{\frac{d}{dx_i}(R_i(x_i \mid \underline{y}_i))}{\int_0^{\infty} \frac{d}{du}\left[R_i(u \mid \underline{y}_i)\right]du} \approx \frac{\frac{d}{dx_i}(R_i(x_i \mid \underline{y}_i))}{\sum_{k=1}^{D} \frac{d}{du}\left[R_i(u \mid \underline{y}_i)\right]_{u=k\omega}} \qquad [6.26]$$

for suitably large D. Alternatively, the denominator can be approximated using a numerical integrator such as a trapezium approximation, Simpsons quadrature or Romberg integration which is an adaptive routine using refinements of the extended trapezium approximation, see Stoer & Bulirsch (1980). In situations where the choice of baseline hazard prevents an analytical approach to establishing the conditional failure time distribution, the following approximation can be used for suitably large c;

$$R_i(x_i \mid \underline{y}_i) = -\lim_{h \downarrow 0}\left(\frac{R_i(x_i + h \mid \underline{y}_i) - R_i(x_i \mid \underline{y}_i)}{h}\right) \approx \frac{R_i(x_i \mid \underline{y}_i) - R_i(x_i + 10^{-c} \mid \underline{y}_i)}{10^{-c}} \qquad [6.27]$$

6.3 Stochastic filtering

6.3.1 Introduction

The stochastic filter explored here is constructed on the following premise. The delay-time (Christer& Waller (1984)) is the time that elapses between the origin of a

defect in the component and the time when the defective component fails and is no longer operational, x_0. As is illustrated in figure 6.1, the residual delay-time at time t_i after the origination of a defect is denoted by x_i.

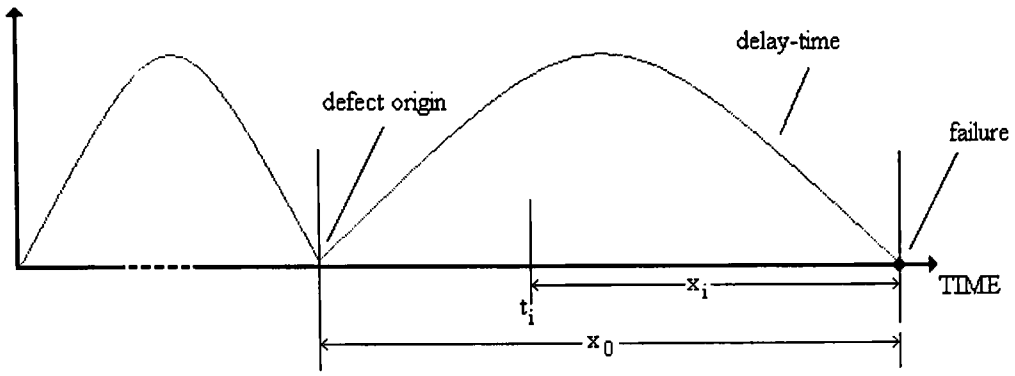

Figure 6.1 - Illustrating the residual delay-time until failure of a component

The illustration given in figure 6.1 is particularly relevant to vibration monitoring applications where distinct phases of the components life are identifiable, as demonstrated in the following chapter. When modelling oil-based CM information, as in chapter 8, these distinct phases can not be recognised from the data and as such, we begin the modelling of the residual life at the start of the components life. The standard delay-time distribution $p(x_0)$ is estimated from failure time data pertaining to similar components and provides the initial estimate of the delay-time. If it were found that the condition readings shared no correlation with the residual delay-time or indeed in cases where the CM information is not available, the probability distribution $p(x_0)$ and the time t_i would be the only information to guide decision making. Otherwise, at the ith checkpoint at time t_i, the posterior distribution for the residual delay-time, x_i, is updated using both the CM information $\underline{y}_i$ and the age of the component t_i. The filtering model prescribed here is subject to the following assumptions;

- The residual delay time x_i is a random variable with its posterior estimate

conditional on the observed condition monitoring history $\underline{Y}_i$.

- The CM information $\underline{y}_i$ is assumed to be a random vector that is a function of the residual delay time with random noise.

The following notation applies in a filtering context throughout this chapter;

- x_i is the residual delay time at t_i,
- $p_i(x_i \mid \underline{Y}_i)$ is the posterior conditional probability distribution for the residual delay time at time t_i given the monitoring history to date,
- $p(\underline{y}_i \mid x_i)$ is the conditional probability distribution for the condition monitoring vector, $\underline{y}_i$, at time t_i given that the residual delay time is x_i,

6.3.2 The residual delay time distribution

Given that the unit is in a defective state and has survived until the ith CM point at time t_i, the residual delay time can be expressed as the residual delay time at the previous CM point, time t_{i-1}, minus the interval between the two points as

$$x_i = x_{i-1} - (t_i - t_{i-1}) \qquad [6.28]$$

If the delay time is not sufficient for a unit in working condition at the previous inspection to survive until the current inspection at time t_i, then the delay time is non-existent and hence is not defined at t_i. The objective of the stochastic filtering approach is to obtain an expression for $p_i(x_i \mid \underline{Y}_i)$, the posterior conditional probability distribution for the residual delay time of a defective unit given the monitored condition history to date. A key element of the recursive filter is $p(\underline{y}_i \mid x_i)$, the conditional distribution of the condition vector $\underline{y}_i$ given that a particular underlying state (residual delay time) exists. If the individual observations y_{ki} are independent from one another (for each j) at time t_i, then the joint conditional

distribution can easily be established as

$$p(\underline{y}_i \mid x_i) = \prod_{k=1}^{r} p(y_{ki} \mid x_i) \qquad [6.29]$$

where, $p(y_{ki} \mid x_i)$ is constructed according to the nature of the observed data and the parameters are estimated from available CM data pertaining to analogous units. If the individual observations y_{ki} are not independent, the construction of $p(\underline{y}_i \mid x_i)$ is more complex. One option is to directly employ a multivariate probability distribution such as a multi-gamma or multi-exponential distribution, see Johnson & Kotz (1972). The other option is to use a data reduction technique as discussed later in this chapter in section 6.4.

For a complete description of the general formulation for the probabilistic stochastic filter including the provision for control functions such as component maintenance as a data contaminant, the reader is referred to Wang & Christer (2000). The specific case of the stochastic filter used here and a more complete description of the derivation can be found in Wang (2002). The conditional distribution is given by

$$p_i(x_i \mid \underline{Y}_i) = p(x_i \mid \underline{y}_i, \underline{Y}_{i-1}) = \frac{p(x_i, \underline{y}_i \mid \underline{Y}_{i-1})}{p(\underline{y}_i \mid \underline{Y}_{i-1})} \qquad [6.30]$$

where, the numerator is

$$p(x_i, \underline{y}_i \mid \underline{Y}_{i-1}) = p(\underline{y}_i \mid x_i, \underline{Y}_{i-1}) p(x_i \mid \underline{Y}_{i-1}) = p(\underline{y}_i \mid x_i) p(x_i \mid \underline{Y}_{i-1}) \qquad [6.31]$$

and integrating over all potential values of the residual delay time x_i at time t_i gives the denominator of $p_i(x_i \mid \underline{Y}_i)$ as

$$p(\underline{y}_i \mid \underline{Y}_{i-1}) = \int_0^{\infty} p(\underline{y}_i \mid x_i) p(x_i \mid \underline{Y}_{i-1}) \, dx_i \qquad [6.32]$$

Using the relationship established in equation [6.28] we obtain the following one-step predictive distribution,

$$p(x_i \mid \underline{Y}_{i-1}) = \frac{p_{i-1}(x_i + t_i - t_{i-1} \mid \underline{Y}_{i-1})}{\int_{t_i - t_{i-1}}^{\infty} p_{i-1}(u \mid \underline{Y}_{i-1}) du} \qquad [6.33]$$

where the denominator normalises the distribution. As noted above, the expression that we are interested in evaluating is $p_i(x_i \mid \underline{Y}_i)$ given by equation [6.30]. If the initial standard delay-time distribution $p_0(x_0 \mid \underline{Y}_0) = p(x_0)$ and the conditional distribution $p(\underline{y}_i \mid x_i)$ are known, then equation [6.30] can be evaluated recursively.

6.3.3 Parameter estimation

The first consideration when fitting the filtering model to a data set is the selection of appropriate forms for $p(x_0)$ and $p(\underline{y}_i \mid x_i)$ which are dependent on the specific data set and estimation of the relevant parameters. The distribution $p(x_0)$ is the standard failure delay time distribution for the components and the parameters can be estimated from prior objective failure data pertaining to similar components using the maximum likelihood approach. To estimate the parameters of $p(\underline{y}_i \mid x_i)$ it is necessary to apply the maximum likelihood technique as a product of conditional probabilities. According to Wang (2002), the likelihood function for an individual component is given as

$$\mathcal{L} = \left(\prod_{i=1}^{n} p(\underline{y}_i \mid \underline{Y}_{i-1}) P_{i-1}(x_{i-1} > t_i - t_{i-1} \mid \underline{Y}_{i-1}) \right) p_n(x_n = T - t_n \mid \underline{Y}_n) \qquad [6.34]$$

where a lower-case p represents a density function, an upper-case P represents a probability and

$$P_{i-1}(x_{i-1} > t_i - t_{i-1} \mid \underline{Y}_{i-1}) = \int_{t_i - t_{i-1}}^{\infty} p_{i-1}(x_{i-1} \mid \underline{Y}_{i-1}) dx_{i-1}$$

Extending equation [6.34] to consider multiple component histories, the likelihood

function for m components is

$$\mathcal{L} = \prod_{j=1}^{m}\left(\prod_{i=1}^{n_j} p(\underline{y}_{ji}|\underline{Y}_{j,i-1})P_{j,i-1}(x_{j,i-1}>t_{ji}-t_{ji-1}|\underline{Y}_{j,i-1})\right)p_{jn_j}(x_{jn_j}=T_j-t_{jn_j}|\underline{Y}_{jn_j}) \quad [6.35]$$

6.4 Multiple indicators of condition

In sections 6.2 and 6.3, the PHM and the filter are presented for a condition based maintenance scenario involving multiple indicators of condition. Much of the modelling for both techniques has been presented on the assumption that the individual condition monitoring parameters are independent from one another. However, in reality, multiple indicators of condition are potentially correlated and this can affect the estimation of the relevant parameters for both techniques. In this section, we discuss a number of different means of removing any correlation between the condition indicators and potentially reducing the dimensions of the CM information used in the models. This is achieved in the case studies of chapter 8 by classifying, at time t_i, the condition input to both models, $\underline{y}_i$, as a vector of linearly independent transformed data obtained from the original condition information vector, represented by $\underline{z}_i$.

6.4.1 Principle components analysis (PCA)

6.4.1.1 Establishing significant principle components

Principle components analysis (PCA) is a well-known linear transformation technique that is used for reducing the dimensionality of multivariate data whilst still preserving most of the variance, and has seen frequent use in multiple regression applications. See for example, Chatfield & Collins (1980) and Jolliffe (1986). PCA can be applied to remove any collinearity between CM indicators and it potentially reduces the number of inputs to the next stage of the modelling process. This is achieved for a

vector of p random variables $\underline{z} = (z_1, z_2, ..., z_p)'$ by creating p uncorrelated linear combinations of the variables. The combinations are established such that the kth linear combination has the kth largest variance amongst all possible linear combinations; $k = 1, 2, \ldots, p$. The kth combination is then labelled the 'kth principle component'. In the majority of cases, most of the variation contained in the data for the p random variables can be described by the first few linear combinations and as such, the number of variables used in modelling can be reduced without the loss of too much information. Now, considering the information available from a single history, the observation matrix $\underline{Z}$ consists of n observations for each of the p random variables and therefore has dimension $n \times p$. In theory, if we know the values of the actual covariance matrix $\underline{\Sigma}$ corresponding to the vector $\underline{z}$, the principle components can be calculated as

$$y_k = \underline{a}_k' \underline{z} \qquad [6.36]$$

for $k = 1, 2, \ldots, p$ where, $\underline{a}_k$ is the eigenvector of $\underline{\Sigma}$ that has the kth largest associated eigenvalue. However, we are rarely in possession of this knowledge. In our scenario, the CM data represents a sample from a population and as such, we are required to use the sample covariance matrix $\underline{S}$. For a sample of data containing the p random variables observed at n consecutive condition monitoring points, the sample covariance matrix is given by

$$\underline{S} = \underline{Q}'\underline{Q} / (n\text{-}1) \qquad [6.37]$$

where, $\underline{Q}$ represents the centred observation matrix that has the same dimensions as the original data set, $n \times p$. The centred observation matrix is obtained by subtracting the mean for each of the p random variables in the vector $\underline{z}$ from n data entries. An alternative approach for obtaining the principle components of a multivariate data set

is singular value decomposition (SVD). SVD is applied to the centred observation matrix $\underline{Q}$ and solutions are obtained for the following matrices

$$\underline{Q}/\sqrt{n-1} = \underline{A}\underline{B}\underline{C}' \qquad [6.38]$$

where, $\underline{A}$ and $\underline{C}$ are unitary matrices and $\underline{B}$ is a diagonal matrix of the same dimensions as the covariance matrix with non-negative diagonal elements of decreasing magnitudes. The matrix $\underline{C}$ contains the principle component coefficients and the squared diagonal elements of $\underline{B}$ are the eigenvalues of the covariance matrix. Principle components analysis has been found to be particularly useful in multivariate applications when there is an excessive number of variables and a large amount of cross-correlation between some or all them. It is true that some information is lost from the original data set when reducing the content of the data in this manner. When considering the application of PCA to data obtained in the context of condition monitoring, there are a number of problems that must be addressed. Firstly, when applying PCA, the samples from a population are usually assumed to be independent however, in CM applications the sample data will likely contain a large amount of auto-correlation implying that the data corresponding to each variable (element of $\underline{z}$) may be time-dependent and, as a result of this, the entries in the sample may be dependent upon one another. However, this issue is not particularly important in our setting as we are not utilising PCA for inferential purposes, but as a data transformation technique that merely prepares the data for input to the CBM model. Another problem is that the cross-correlation structure between the p variables contained in $\underline{z}$ could change over time. This is an issue that when present, can be tackled using dynamic principal components analysis (DPCA) that typically utilises time series such as autoregressive models and applies the techniques to the principal components histories. DPCA is also useful when modelling time-dependent variables.

A further problem that must be addressed when applying PCA is the relative magnitude of the data pertaining to the p random variables in the original vector $\underline{z}$. If the variance of one variable is substantially greater than another, it will have a greater weighting in the first few principle components. To address this problem, the elements of $\underline{z}$ can be standardised by subtracting the mean from the associated data and dividing by the standard deviation. The correlation matrix can then be established in preference to the covariance matrix and the eigenvectors evaluated. PCA can then be applied to the standardised data set. In addition, the number of observations included in each history will be different and the determination of principle components must consider this. The final issue that we should consider is that, although PCA can be applied for the removal of collinearity between the original variables and as such, produces linearly independent components, a higher-order of dependence may still exist. A technique called independent components analysis (ICA) has recently been developed to consider this problem. The use of ICA for dimension reduction is discussed later in this section.

6.4.1.2 Using PCA when modelling CBM

There are a number of options that can be considered when using PCA as a means of transforming data for use in CBM applications. Option A involves the use of PCA for dimension reduction but, still using some of the original untransformed data as the input to the CBM model. Option B involves using the transformed data only i.e. the principal components, directly in the model.

A1. After applying PCA to the original data set, we may find that some of the original elements of $\underline{z}$ contribute the majority of the variance contained in the evaluated principal components. These elements could then be retained and used as the input to the CBM model. The other elements of $\underline{z}$ could then be ignored resulting in a

reduction in the input to the model and a reduction in the amount of actual monitoring required. However, the problem of cross-correlation between the selected original CM variables still exists.

A2. When a data set has been transformed using PCA, the transformed data can be converted back to the original data set using the inverse matrix. However, the original data set can only be retrieved in its entirety if all the principle components are retained as having a significant impact. When, some of the principle components are deemed to be insignificant and as such, are omitted, back-transformation of the data will produce a reduced version of the original data set. This reduced data set can then be used as the input to the state estimation models.

B1. An alternative approach involves using PCA to remove the collinearity between the variables in $\underline{z}$ and to use all the principle components as a direct input to the CBM model. Although the dimensions of the data used to establish the parameters and build the CBM model have not changed ($r = p$), and the problem of an excessive joint sample space may still exist, the fact that the individual elements of $\underline{y}$ are linearly independent from one another, should improve the parameter estimation capabilities.

B2. The final option for incorporating PCA into CBM model building is to include the principle components deemed as significant in the model and exclude the others ($r \leq p$). This option solves the problems of collinearity and dimension reduction however, a new question is raised; how do we evaluate the significance of a principal component? Should acceptance of a principal component as an input to the CBM model be based upon the variance, the p-value or both? Another option is to introduce the principle components one by one in a stepwise manner and assess the model fit until no further improvement is achieved.

6.4.2 Dynamic principal components analysis (DPCA)

As has been discussed, the PCA approach can be applied to consider any potential collinearity between the different types of CM observations. However, PCA does not tackle the issue of potential auto-correlation between observations of the same variable at successive monitoring points. When the original data set;

$$\underline{z}_i = (z_{1i},...,z_{pi})' \quad ; \; i = 0, \pm 1, \pm 2,...$$

is auto-correlated, dynamic principle components analysis (DPCA) can be used to reduce the dimensionality of the data and reduce both cross-correlation between the elements of $\underline{z}$ and any auto-correlation between observations. Typical applications of DPCA utilise time series in the evaluation of principle components. One option is to use an auto-regressive model of order w, AR(w). The PCA approach is then applied to the covariance matrix (of lag w) of the observed CM process. An AR(w) model represents the process $\underline{z}$ as;

$$\underline{z}_i - \sum_{v=1}^{w} \underline{\Phi}_v \underline{z}_{i-v} = \underline{\partial} + \underline{\varepsilon}_i \qquad [6.39]$$

The constituent elements of equation [6.39] are $\underline{\partial} = (\underline{I} - \underline{\Phi}_1 - ... - \underline{\Phi}_w)\bar{\underline{z}}$ where $\underline{I}$ is an identity matrix and $\bar{\underline{z}}$ is the mean CM observation vector, $\underline{\Phi}_v$ (v = 1, 2,..., w) is a $p \times p$ matrix of coefficients and $\underline{\varepsilon}_i$ is a p-dimensional white noise process with covariance matrix $\underline{\Sigma}$. The matrix $\underline{Z} = [\underline{z}_{w+1},...,\underline{z}_n]'$ of dimension $(n-w) \times p$ and $\underline{\beta}' = [(\underline{I} - \underline{\Phi}_1 - ... - \underline{\Phi}_w)\bar{\underline{z}}, \underline{\Phi}_1,...,\underline{\Phi}_w]$ are defined for the process $\underline{z}$ that consists of n sets of observations over time. The objective is to establish the following general linear model;

$$\underline{Z} = \underline{a}\underline{\beta} + \underline{\Xi} \qquad [6.40]$$

where, $\underline{\Xi} = [\underline{\varepsilon}_{w+1}, \underline{\varepsilon}_{w+2}, ..., \underline{\varepsilon}_n]'$ is the matrix of errors for each CM parameter over the observations between the $(w+1)$th and nth time point, and $\underline{a}$ is a matrix of dimension $(n-w)\times(wp+1)$ with rows $\underline{a}'_i = [1, \underline{z}'_i, ..., \underline{z}'_{i-w}]$ for $i = w+1, ..., n$. The parameters are estimated using the method of least squares minimisation as

$$\hat{\underline{\beta}} = (\underline{a}'\underline{a})^{-1}\underline{a}'\underline{z} \qquad [6.41]$$

with the covariance matrix given by

$$\hat{\underline{\Sigma}} = \frac{1}{(n-w)-(pw+1)}\hat{\underline{\Xi}}'\hat{\underline{\Xi}} \qquad [6.42]$$

A number of test and model validation procedures are given for the AR(w) model in Makis et al (2005), pages 7-9. The order w of the chosen AR model dictates the lag incorporated when applying DPCA. The principal components for the DPCA approach are obtained using the expression;

$$\underline{Y}_i = \underline{U}'\underline{O}_i \qquad [6.43]$$

where, $\underline{O}_i = (\underline{z}'_i, \underline{z}'_{i-1}, ..., \underline{z}'_{i-w})'$ and $\underline{U} = (u_1, u_2, ..., u_p)'$ are the eigenvectors of the sample covariance matrix $\underline{S}$. $\underline{S}$ is constructed as a block matrix with $(w+1)\times(w+1)$ blocks of dimension $p\times p$. As such, the data vector considered when using DPCA is $(\underline{z}'_i, \underline{z}'_{i-1}, ..., \underline{z}'_{i-w})$ rather than just $\underline{z}'_i$ with the standard PCA approach where the observations are assumed to be independent over time. A Scree test can be applied to assess the suitability of the DPCA approach when applied to specific cases. However, it should be noted that the application of an AR(w) model would require the observations in the data set to be spaced equidistantly. In cases where the data set consists of irregular CM readings, the data could be scaled according to the wear increment and the duration over which the wear is accrued;

$(\underline{z}_i - \underline{z}_{i-1}) / (t_i - t_{i-1})$. Alternatively, some data points that are of an inconsistent duration could be discarded in an attempt to approximate a regular process.

6.4.3 Independent components analysis (ICA)

As discussed previously in this chapter, PCA is limited in some applications as the resulting components are only linearly independent. As with PCA, independent components analysis (ICA) is a statistical technique for revealing hidden factors that underlie sets of random variables however, with the ICA approach, a non-Gaussian representation using a statistical 'latent variables' model produces components that are as statistically independent as possible, see Jutten & Herault (1991). At each CM point, it is assumed that we observe p linear mixtures $z_1,\ldots, z_p$ of p independent components as

$$z_l = a_{l1}s_1 + a_{l2}s_2 + \ldots + a_{lp}s_p$$

for $l = 1,\ldots, p$. At the ith CM point, the observed reading $z_l(t_i)$ is a sample of the random variable z_l. Defining $\underline{z}$ as a vector with elements $z_1,\ldots, z_p$, $\underline{s}$ as a vector with the elements being the independent components $s_1,\ldots,s_p$ and $\underline{A}$ as a matrix containing the co-efficients a_{lk}, we have $\underline{z} = \underline{A}\underline{s}$ and are required to compute $\underline{A}$ and $\underline{s}$ from $\underline{z}$. As with PCA, the independent components s_k are latent variables that cannot be observed directly. With ICA, the components are assumed to be statistically independent and governed by non-Gaussian distributions. If the actual functional form of the distribution is known, the computation of the independent components is much simpler. The matrix $\underline{A}$ is known as the 'mixing matrix' and is usually assumed to be square although, this assumption can be relaxed in some cases. After computation of $\underline{A}$, see Comon (1994), the independent components can be obtained as $\underline{s} = \underline{A}^{-1}\underline{z}$. For the specific cases investigated in chapter 8, only the first principle component is found

to be significant. With only a single component being utilised as the CM input to the CBM models, there cannot be any correlation between the components and as a result, the use of ICA is not a necessity.

6.5 Failure time analysis and replacement decisions

6.5.1 MTTF and MSE analysis

Using both the proportional hazards model and the stochastic filtering approach we are able to construct the conditional probability distribution $p_i(x_i \mid \underline{Y}_i)$ where, $\underline{Y}_i$ represents the history of condition information and x_i represents the time remaining before failure. In the case of the stochastic filter, we call this value the residual delay-time, Wang & Christer (2000), and in the case of the PHM the distribution reduces to $p_i(x_i \mid \underline{Y}_i) = p_i(x_i \mid \underline{y}_i)$ as only the current CM vector is used in its construction. However, with the conditional distribution $p_i(x_i \mid \underline{Y}_i)$ in hand we are still required to select an appropriate measure or point-estimate of the time remaining before failure, $\hat{x}_i$. For example, this estimate could be the mean, mode, median or some other measure. The prediction error, at time t_i, for a point-estimate $\hat{x}_i$ can be expressed as

$$\tilde{x}_i = x_i - \hat{x}_i \qquad [6.44]$$

Define $\rho(\varepsilon)$ as a real-valued, non-negative convex function and $L(\tilde{x}_i)$ as a loss/criterion function with the following properties;

$$L(0) = 0, \qquad \rho(\varepsilon^2) \geq \rho(\varepsilon^1) \geq 0 \rightarrow L(\varepsilon^2) \geq L(\varepsilon^1) \geq 0.$$

A typical choice of criterion function is

$$\rho(\varepsilon) = |\varepsilon| = (\varepsilon^{\mathrm{T}}\varepsilon)^{1/2}$$

According to Jazwinski (1970, Theorem 5.2 and 5.3), the conditional mean is the minimum variance estimate for all filtering and prediction problems, regardless of the

properties of the conditional density function. The conditional mean is the expectation taken with respect to the observed condition history. At time t_i, the expected time remaining before failure given monitored condition information to date is given by

$$\hat{x}_i = \mathbf{E}[x_i \mid \underline{Y}_i] \qquad [6.45]$$

In the stochastic filtering case, we are required to approximate the following integral to obtain an estimate of the conditional mean as

$$\mathbf{E}[x_i \mid \underline{Y}_i] = \int_0^\infty x_i p_i(x_i \mid \underline{Y}_i) dx_i \qquad [6.46]$$

with a variance about the mean estimate given by

$$\mathbf{Var}[x_i \mid \underline{Y}_i] = \int_0^\infty (x_i - \hat{x}_i)^2 p_i(x_i \mid \underline{Y}_i) dx_i \qquad [6.47]$$

In the PHM case, the following expression is evaluated for an approximate conditional mean;

$$\mathbf{E}[x_i \mid \underline{Y}_i] = \omega \sum_{x_i} x_i p_i(x_i \mid \underline{Y}_i) \qquad [6.48]$$

because, for $x_i = \omega, 2\omega, 3\omega \ldots$ we have $\sum_{x_i} p_i(x_i \mid \underline{Y}_i) = (1/\omega)$, and naturally it is a requisite that the cumulative probability density to tend towards 1 as x_i tends towards ∞. As with the establishment of the denominator of the conditional failure time distribution, equation [6.26], the expectation of equation [6.48] can be approximated using an alternative numerical integrator. The variance about the estimate in the PHM case is given by

$$\mathbf{Var}[x_i \mid \underline{Y}_i] = \omega \sum_{x_i} (x_i - \hat{x}_i)^2 p_i(x_i \mid \underline{Y}_i) \qquad [6.49]$$

To ascertain the fit of the two techniques when applied to data, two loss functions based on the residual errors obtained for each point estimate at all available monitoring points for all components are

$$\sum_j \sum_i |\varepsilon_{ji}| = \sum_j \sum_i ((\hat{x}_{ji} - x_{ji})^2)^{1/2} \qquad [6.50]$$

$$\sum_j \sum_i \left(\frac{|\varepsilon_{ji}|}{x_{ji}} \right) = \sum_j \sum_i \frac{((\hat{x}_{ji} - x_{ji})^2)^{1/2}}{x_{ji}} \qquad [6.51]$$

However, we have more information contained in the conditional density $p_i(x_i \mid \underline{Y}_i)$ than a point-estimate and as such we are able to construct a measure of the mean-square error (MSE) at time t_i. The actual time remaining before component failure, that is only available after failure, is $T - t_i$ and the MSE for the ith estimated conditional distribution about the actual value in the stochastic filtering case is

$$MSE_i = \int_0^\infty (x_i + t_i - T)^2 \, p_i(x_i \mid \underline{Y}_i) dx_i \qquad [6.52]$$

In the PHM case, for $x_i = \omega, 2\omega, 3\omega...$, we have

$$MSE_i = \omega \sum_{x_i} (x_i + t_i - T)^2 \, p_i(x_i \mid \underline{Y}_i) \qquad [6.53]$$

The resulting loss function that enables a direct comparison of the fit of the two modelling approaches to all available data can then be established simply as the sum of the MSE evaluated for all available monitoring points for each component under scrutiny. We have

$$Total\ MSE = \sum_j \sum_i MSE_{j,i} \qquad [6.54]$$

The decision to select an appropriate modelling approach for future monitoring and control purposes is then simply the model producing the lowest total mean-square error given above. If both models are seen to exhibit desirable elements, a

combination of model estimates could be used based on a weighting scheme around the mean-square error loss function, assuming that sufficient computational power is available to run more than one model in parallel.

6.5.2 Replacement policies

A further advantage in establishing a conditional probability distribution over the evaluation of a single point estimate of residual life is the availability of the cumulative density function enabling the construction of decision models that incorporate the probability of failure before a particular instant conditional upon the monitoring history to date. A replacement policy can be established using renewal-reward theory and the long-run 'expected cost per unit time', see Ross (1996). Initially, some further notation is defined;

- T_R is the planned replacement time (to be optimised),
- C_P is the cost of a preventive replacement,
- C_F is the replacement cost due to a failure of the component.

The general form of the cost model is presented first. At time t, the expected cost per unit time is given by

$$C(t,T_R) = \frac{E(Cycle\ Cost\ |T_R)}{E(Cyele\ Length\ |T_R)} = \frac{C_P + (C_F - C_P)P(Failure\ |T_R)}{E(Cycle\ Length\ |T_R)} \qquad [6.55]$$

where, $C(t,T_R)$ is to be optimised with respect to T_R. The probability of failure given a specified replacement time is given by; $P(\text{Failure}|T_R) = P(x < T_R - t)$. The replacement decision at the ith CM point is then obtained via the optimisation of

$$C(t_i,T_R) = \frac{C_P + (C_F - C_P)P_i(x_i < T_R - t_i \mid \underline{Y}_i)}{t_i + (T_R - t_i)(1 - P_i(x_i < T_R - t_i \mid \underline{Y}_i)) + \int_0^{T_R - t_i} x_i\, p_i(x_i \mid \underline{Y}_i)dz} \qquad [6.56]$$

where, an upper-case P represents a probability and a lower-case p denotes a

conditional density. In the filtering case, the integrals are solved by approximation and for the PHM we have

$$\int_0^{T_R - t_i} x_i \, p_i(x_i \mid \underline{Y}_i) dx_i \;=\; \omega \sum_{x_i=\omega}^{T_R - t_i} x_i \, p_i(x_i \mid \underline{y}_i) \qquad [6.57]$$

for $x_i = \omega, 2\omega, 3\omega \ldots (c\omega = T_R - t_i)$. Again, this approximation may be improved using a numerical integrator.

6.6 Discussion

In this chapter, we have introduced two techniques that are available for condition-based maintenance applications. The techniques are proportional hazards modelling and probabilistic stochastic filtering. The models are presented in the context of a scenario involving the prediction of residual life at condition monitoring points throughout the lifetime of a component. The modelling and associated parameter estimation methodologies have been presented for both models and issues regarding the handling of multiple indicators of condition have been addressed. In addition, we have considered different means of comparing the two modelling approaches and establishing replacement decisions when the models are applied to case data. Additional tests of fit, for both models, are a topic for future research. In the research documented in the next two chapters, we focus on comparing the performance of the two models using the MSE criterion and the replacement cost model introduced in the previous section of this chapter. We establish specific cases of the PHM and the filter and apply the models to case scenarios involving vibration monitoring and subsequently, oil-based condition information.

Chapter 7. A case comparison of the PHM and a stochastic filter for a CBM application using vibration monitoring

7.1 Introduction

The type of CM parameter considered in this chapter is the overall vibration level of a single working component. In order to provide a fair comparison, the two techniques are implemented on the same data set with the same two-stage approach to modelling where, the first stage corresponds to the normal operation of a component and the second stage is the defective phase of the components life, see figure 6.1. On starting this comparative investigation, one perceivable modelling advantage of the proposed stochastic filter over the PHM is the fact that discrete approximation is not required in the filtering case to handle the observed information parameter and the entire component history is utilised rather than just the current CM reading. There are other fundamental issues that mark the key difference between the two approaches particularly regarding whether or not the CM information influences the state of the component, which it does not in the case of vibration monitoring. The PHM is not particularly appropriate when modelling situations where the covariates are response variables, see Moore & McCabe (2003). In a filtering context, the CM parameters are taken to be a function of the residual life but the residual life is not a function of the CM reading. However, when applying the PHM, the hazard is a function of the CM parameters. In the former the CM parameters are treated as random variables that are assumed to be correlated stochastically with the residual life and in the latter the CM parameters are covariates that change according to a separate stochastic process. These issues are rather important as they are related to the fundamental principle in condition monitoring; in most CM applications, the observed parameters function merely as indicator information governed by the underlying system state, but this

relationship does not hold in reverse. This is particularly true in cases involving vibration based monitoring where an extreme elevation in the amplitude of the vibration level is usually triggered by a hidden defect, but one cannot say that such a high reading caused the defect. The stochastic filtering approach follows this principle but the proportional hazards model does not. As has already been mentioned, the life of a component maybe divided into two phases namely, normal operation and defective. This classification of the different phases of component life is particularly appropriate for vibration monitoring applications as the vibration signal does not typically display trend in the initial normal phase of operation but begins to increase rapidly upon the commencement of the defective phase of its usable lifetime. As such, the vibration level is used to ascertain whether or not the component has become defective. To establish an appropriate threshold level between the two phases, a statistical process control (SPC) approach known specifically as a Shewharrt chart is employed.

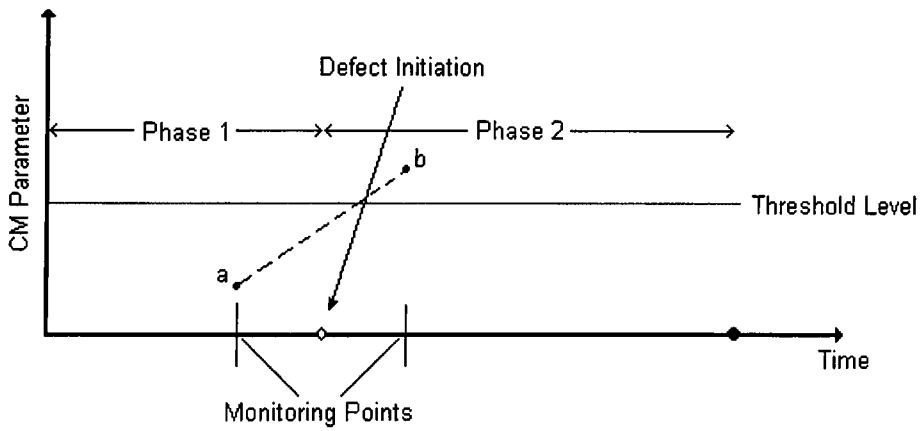

Figure 7.1 - Illustrating the two phases of component operation and the initiation of a defect

When utilising a Shewharrt chart, a critical level for the vibration reading can be established using assumptions on the stationarity and the variance of the vibration signal. Points *a* and *b* in figure 7.1 represent CM parameter readings at monitoring points before and after the threshold is breached. The initiation point is then assumed

to be the mid-point between the two monitored readings and estimation of the time remaining before a failure occurs commences from this point, see Wang (2002) for more details. The SPC approach to defect detection is quite crude but it enables a direct comparison of the two techniques. Recently developed hidden Markov modelling techniques in Wang (2004) would provide a more precise estimate of the origin of a defect in the filtering case. A further alternative to defect detection is the use of subjective expert opinion that is particularly useful when the signal during the normal phase of operation is erratic.

7.2 The data

A case comparison was conducted on the life histories of six bearings tested until failure in a laboratory fatigue experiment. As failure times were not recorded, ten hours is added on to the final monitoring time to give an approximate failure time, T_j, for each bearing. For more information on the type of data see Wang (2002). The six life histories comprising of the monitoring times with associated condition information (vibration level) are illustrated in the figure below.

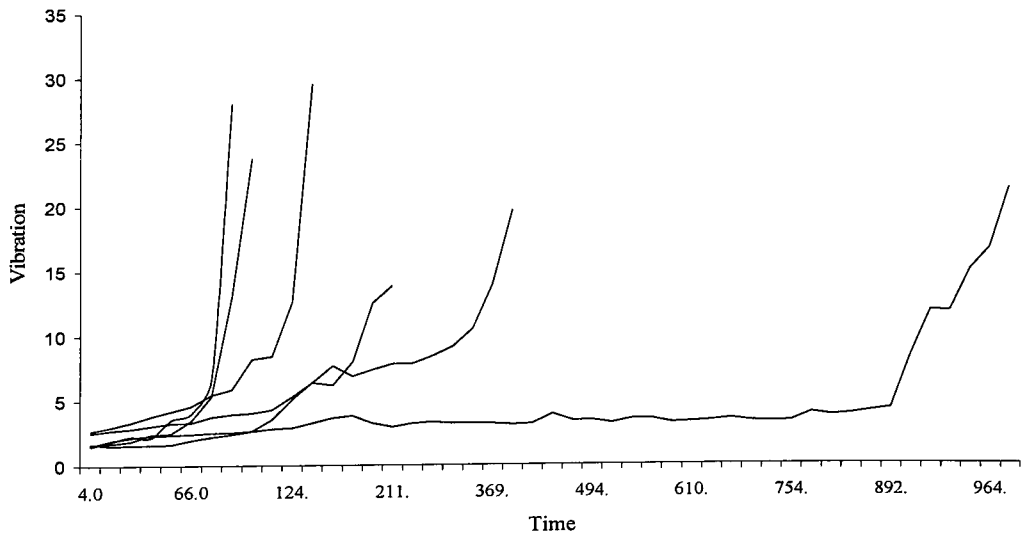

Figure 7.2 - Illustrating the vibration data for the 6 bearings

The data in figure 7.2 clearly lends itself to a two-stage approach to analysis as is illustrated in the introduction to this chapter in figure 7.1. A threshold vibration level of five signals the start of the second phase of the analysis; see Wang (2002) for details. Defects are assumed to initiate/originate at the mid-point between the CM point at which the vibration level is noticed to have breached the action limit and the previous CM point.

7.3 The proportional hazards model

7.3.1 The Weibull PHM

In the hazard given by equation [6.1], we employ an exponential form for the functional term as $\lambda(\gamma, y_i) = \exp\{\gamma y_i\}$ in the scalar case as we are only considering a single CM parameter in the form of the overall vibration level. The Weibull distribution is often used in survival analysis applications and is a suitable choice for the baseline hazard rate because of its adaptability. The hazard rate for a Weibull PHM incorporating an exponential function of the covariates is

$$h(t_i, y_i) = \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1} \exp\{\gamma y_i\} \qquad [7.1]$$

where, β is the shape parameter, η is the scale parameter and $\beta, \eta > 0$.

7.3.2 Parameter estimation

In the case study explored in this chapter, there are no suspensions contained in the data. For cases such as this, we consider m independent lifetimes $(T_j, \underline{Y}_j)$ where $j = 1, 2, \ldots, m$ and $\underline{Y}_j$ is the monitoring history for component j. Now, denoting $y_j(s)$ as an approximate continuous sample path for the discrete observations, the likelihood function given by equation [6.12] becomes

$$\mathcal{L}(\beta,\eta,\gamma) = \prod_{j=1}^{m} h(T_j, y_j(T_j)) R(T_j, \underline{Y}_{j,n_j})$$

From equation [7.1], the hazard at the failure or suspension of the jth component is

$$h(T_j, y_j(T_j)) = \frac{\beta}{\eta}\left(\frac{T_j}{\eta}\right)^{\beta-1} \exp\{\gamma y_j(T_j)\} \qquad [7.2]$$

and $R(T_j, \underline{Y}_j)$ is given by equations [6.13] and [6.14]. The likelihood expression must be evaluated numerically and taking logarithms can ease this process. Using equation [7.2], the log-likelihood function is

$$l(\beta,\eta,\gamma) = \sum_{j=1}^{m}\left(\log(\beta) - \beta\log(\eta) + (\beta-1)\log(T_j) + \gamma y_j(T_j) - U(T_j, y_j(T_j))\right) \qquad [7.3]$$

where, from equation [6.14], the cumulative hazard at the failure time of the jth unit, T_j, is approximately

$$U(T_j, y_j(T_j)) = \sum_{d=1}^{q_j}\left(\int_{V_{j,d-1}}^{V_{jd}} \frac{\beta}{\eta}\left(\frac{s}{\eta}\right)^{\beta-1} \exp\{\gamma A_{jd}\} ds\right) \qquad [7.4]$$

where, A_{jd} is the dth state of the discretised covariate process $y_j(t)$ and V_{jd} is the time of the dth transition. Expanding the log-likelihood, re-arranging and inserting the expression for the cumulative hazard gives

$$l(\beta,\eta,\gamma) = m\log(\beta) - m\beta\log(\eta) + (\beta-1)\sum_{j=1}^{m}\log(V_{jq_j}) + \gamma\sum_{j=1}^{m} y_j(V_{jq_j})$$
$$-\eta^{-\beta}\sum_{j=1}^{m}\sum_{d=1}^{q_j}(e^{\gamma A_{jd}}(V_{jd}^{\beta} - V_{j,d-1}^{\beta})) \qquad [7.5]$$

Vlok et al (2002) recommend the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method for maximisation of the log-likelihood function given by equation [7.5], another possibility is the Davidon-Fletcher-Powell update. The log-likelihood function can also be used to estimate the variance-covariance matrix that in turn can be used to estimate the standard errors of the parameter estimates. As noted

previously, the reliability function given by equation [6.2] can be expressed as $R(t,y) = \exp\{-U(t,y)\}$ where, $U(t,y)$ is the cumulative hazard. If the Weibull PHM is the correct choice of model for the system, the cumulative hazard at the termination points of the sample data should follow an exponential distribution, i.e. $U(T_j)$ are the model fit residuals. See Vlok et al (2002) and associated references for details on assessing the model fit. Major indicators that the model is appropriate for the data are the mean time to failure (MTTF) at the start of the second phase with a vibration reading in the first covariate bracket, the MTTF should be close to the mean failure time for the component histories that are used to estimate the parameters, also, the fit residuals obtained at each CM point when attempting to predict the failure time will give us some measure of the accuracy when compared with other models.

7.3.3 The conditional failure time distribution

Using equations [6.20], [6.22] and [6.26] for a Weibull baseline hazard and exponential function of the CM covariates, the conditional failure time distribution can be established for this case as

$$p_i(x_i \mid y_i) =$$

$$\left(\sum_{y(t_i+u\omega)} e^{\gamma y(t_i+u\omega)} \beta \eta^{-\beta} (t_i + x_i)^{\beta-1} \exp\left\{ - e^{\gamma y(t_i+u\omega)} \eta^{-\beta} \left((t_i + x_i)^{\beta} - (t_i + (u-1)\omega)^{\beta} \right) \right\} \right.$$

$$\left. \times \sum_{y(t_i+(u-1)\omega)} p(y(t_i + u\omega) \mid y(t_i + (u-1)\omega)) R(t_i + (u-1)\omega,\ y(t_i + (u-1)\omega) \mid t_i, y(t_i)) \right) \Big/$$

$$\left(\sum_{x_i} \sum_{y(t_i+u\omega)} e^{\gamma y(t_i+u\omega)} \beta \eta^{-\beta} (t_i + x_i)^{\beta-1} \exp\left\{ - e^{\gamma y(t_i+u\omega)} \eta^{-\beta} \left((t_i + x_i)^{\beta} - (t_i + (u-1)\omega)^{\beta} \right) \right\} \right.$$

$$\left. \times \sum_{y(t_i+(u-1)\omega)} p(y(t_i + u\omega) \mid y(t_i + (u-1)\omega)) R(t_i + (u-1)\omega, y(t_i + (u-1)\omega) \mid t_i, y(t_i)) \right) \qquad [7.6]$$

for $x_i = u\omega$, $(u = 1, 2, 3, ..)$.

7.4 The stochastic filter

7.4.1 The probability distributions

As noted in chapter 6, the CM information is assumed to be a random vector, or variable in the scalar case, that is a function of the residual delay time with random noise. There are many ways to model this relationship but the set-up proposed in this chapter is simple and works well. Modelling the impact of the residual delay time on the CM parameter reading and recognising the lack of impact on the delay-time from the CM information is definitely appropriate when considering the vibration level of a component and forms the key difference between the two approaches compared here. In the PHM case, covariates of any kind are seen to influence the hazard or failure rate. For the case study in this chapter, both $p(y_i \mid x_i)$ and $p_0(x_0)$ are taken to be Weibull distributions as

$$p_0(x_0) = \alpha\beta(\alpha x_0)^{\beta-1} e^{-(\alpha x_0)^{\beta}} \qquad [7.7]$$

$$p(y_i \mid x_i) = \rho\eta(\rho y_i)^{\eta-1} e^{-(\rho y_i)^{\eta}} \qquad [7.8]$$

The scale parameter in $p_0(x_0)$ is taken to be α, whereas in the case of the PHM we use the equivalent $1/\alpha$ as this simplifies the PHM parameter estimation process, see Vlok et al (2002). The following set-up is used to establish a relationship between y_i and x_i in equation [7.8];

$$\rho = 1/(A + Be^{-Cx_i}) \qquad [7.9]$$

The set-up given in equation [7.9] enables the relationship $E(y_i \mid x_i) \propto A + Be^{-Cx_i}$ that produces a negative correlation between y_i and x_i as required. See Wang (2002) for details on the selection of appropriate distributional forms.

7.4.2 Parameter estimation

We initially consider the estimation of the parameters of the Weibull initial delay-time

distribution $p_0(x_0)$, given by equation [7.7], using the lifetimes of m observed components. The likelihood function is

$$\mathcal{L}(\alpha,\beta) = \prod_{j=1}^{m}\left(\beta\alpha^{\beta}x_{j0}{}^{\beta-1}e^{-(\alpha x_{j0})^{\beta}}\right) \qquad [7.10]$$

and the log-likelihood function is

$$l(\alpha,\beta) = m\log\beta + m\beta\log\alpha + (\beta-1)\sum_{j=1}^{m}\log x_{j0} - \alpha^{\beta}\sum_{j=1}^{m}x_{j0}{}^{\beta} \qquad [7.11]$$

Now we consider the estimation of the parameters of $p(y_i \mid x_i)$, given by equation [7.8]. To establish the likelihood function and develop closed-form analytical solutions for the filtering equations, we define the function

$$\psi_h(u,t_i) = \frac{e^{-\left(\frac{y_h}{A+Be^{-C(u+t_i-t_h)}}\right)^{\eta}}}{(A+Be^{-C(u+t_i-t_h)})^{\eta}} \qquad [7.12]$$

From equation [6.32] the probability of observing the vibration reading y_i at time t_i, given the prior monitoring history, can be established as

$$p(y_i \mid \underline{Y}_{i-1}) = \begin{cases} \left(\dfrac{\int_0^{\infty}\eta y_i^{\eta-1}(u+t_i)^{\beta-1}e^{-(\alpha(u+t_i))^{\beta}}\prod_{h=1}^{i}\psi_h(u,t_i)\,du}{\int_{t_i-t_{i-1}}^{\infty}(u+t_{i-1})^{\beta-1}e^{-(\alpha(u+t_{i-1}))^{\beta}}\prod_{h=1}^{i-1}\psi_h(u,t_{i-1})\,du}\right) & i>1 \\ \dfrac{\int_0^{\infty}\eta y_1^{\eta-1}(u+t_1)^{\beta-1}e^{-(\alpha(u+t_1))^{\beta}}(A+Be^{-Cu})^{-\eta}e^{-\left(\frac{y_1}{A+Be^{-Cu}}\right)^{\eta}}du}{\int_{t_1}^{\infty}u^{\beta-1}e^{-(\alpha u)^{\beta}}du} & i=1 \end{cases} \qquad [7.13]$$

Using equation [6.35], the likelihood function for Weibull $p(y_i \mid x_i)$ is

$$\mathcal{L} = \alpha^{\beta}\beta\prod_{j=1}^{m}\left(T_j^{\beta-1}e^{-(\alpha T_j)^{\beta}}\prod_{i=1}^{n_j}\eta y_{ji}^{\eta-1}(A+Be^{-C(T_j-t_{ji})})^{-\eta}e^{-\left(\frac{y_{ji}}{\left(A+Be^{-C(T_j-t_{ji})}\right)}\right)^{\eta}}\right) \qquad [7.14]$$

Taking the natural logarithm of both sides of the likelihood function gives

$$l(A,B,C,\eta) = \log(\alpha^{\beta}\beta) + \sum_{j=1}^{m}\left\{ (\beta-1)\log(T_j) - (\alpha T_j)^{\beta} + n_j \log(\eta) + \sum_{i=1}^{n_j}(\eta-1)\log(y_{ji}) \ldots \right.$$

$$\left. \ldots \; -\eta\log(A+Be^{-C(T_j-t_{ji})}) - \left(\frac{y_{ji}}{(A+Be^{-C(T_j-t_{ji})})}\right)^{\eta} \right\} \qquad [7.15]$$

and given that α, β and T_j are known, some of the terms are not required for maximisation with respect to the parameters A, B, C and η.

7.4.3 The conditional failure time distribution

Using equations [6.30] – [6.33], the conditional residual delay-time distribution at time t_i can be derived as

$$p_i(x_i \mid \underline{Y}_i) = \begin{cases} \left(\dfrac{(x_i+t_i)^{\beta-1} e^{-(\alpha(x_i+t_i))^{\beta}} \prod_{h=1}^{i} \psi_h(x_i,t_i)}{\int_0^{\infty} (u+t_i)^{\beta-1} e^{-(\alpha(u+t_i))^{\beta}} \prod_{h=1}^{i} \psi_h(u,t_i)\,du} \right) & i>0 \\ \alpha\beta(\alpha x_0)^{\beta-1} e^{-(\alpha x_0)^{\beta-1}} & i=0 \end{cases} \qquad [7.16]$$

using the function given in equation [7.12].

7.5 Results

We consider two separate cases for comparing the performance of the proportional hazards model and the probabilistic filtering approach; case 1 is a comparison of the models fit to all six of the available component life histories and in case 2, three of the bearings are used for estimation purposes as a training set and the remaining three are used to test the prediction ability of the two techniques when the models are effectively applied to new data. To compare the models we use the MSE criterion

given by equations [6.52] and [6.53] for the filter and the PHM respectively. As discussed in chapter 6, the replacement policy given by equation [6.56] is used to compare the decisions produced by the models at each stage of the monitoring process where the time of origin of the defective phase of the components life is added to the denominator. Cost values are obtained from Wang (2002) and are C_F = £6000 for the average failure cost and C_P = £2000 for a preventive replacement.

7.5.1 Case 1

7.5.1.1 The proportional hazards model

In both cases, the individual rates of transition between states λ_{ab} are assumed to be homogenous. In order to discretise the Markov process, we assign ranges of the vibration signal to discrete covariate states as given in table 7.1

State	Vibration Level	Value
1	5 - 10	7.5
2	10 - 15	12.5
3	15 - 20	17.5
4	20 +	25

Table 7.1 – Covariate bandings for the PHM

Although the actual vibration reading at the time of failure is unknown, it is assumed to be the same as at the last available monitoring point. As such, a final measurement is created at $V_{jq_j} = T_j$. Table 7.2 documents the time that each component spent in normal operation and the time the vibration level spent within the specified discrete covariate states after the initiation of a defect.

$j = 1$		2		3		4		5		6	
State	Durat	State	Durat	State	Durat	State	Durat	State	Durat	State	Durat
Norm	88.5	Norm	74.25	Norm	79.5	Norm	149.3	Norm	119.3	Norm	904
1	19.5	1	54.75	1	22.5	1	66.25	1	136.8	1	24
4	10	2	16.5	2	3	2	24.5	2	24.5	2	36
		4	10	4	10			3	10	3	10
										4	10

Table 7.2 – The duration that the vibration reading spends in particular covariate states for each component

Using the likelihood function given by equation [7.5] and Matlab© v6.5 function 'fmincon' from the optimisation toolbox, the following parameter estimates are obtained for case 1.

Parameter	Estimate
β	2.0857
γ	0.2565
η	707.2768

Table 7.3 – The estimated parameters for the PHM, case 1

The transition probabilities can be defined over an interval ω using equations [6.10] and [6.11]. For a given state a, the total amount of time spent in state a during the lifetime of all six bearings is shown by $W(a)$ in table 7.4.

a	$W(a)$
0	1414.75
1	323.75
2	104.5
3	20
4	40

Table 7.4 – The total time spent in each state over all bearings

Using equation [6.9] and the numbers of transitions between the discrete states of the covariate vibration process, the matrix of transition rates is given in table 7.5.

	1	2	3	4
1	-	5/323.75	0	1/323.75
2	0	-	2/104.5	2/104.5
3	0	0	-	1/20
4	0	0	0	-

Table 7.5 – The matrix of transitions rates

The conditional failure time distribution and mean time to failure (MTTF) can now be determined at each monitoring point to establish the fit of the proportional hazards model to all the available data. An approximation interval of $\omega = 0.25$ hours is used throughout both cases for the PHM element of this case study. When the critical level is breached, the average time remaining before failure for the six bearings is 81.375 hours. For a vibration level within covariate state 1, the MTTF given by the model is approximately 80.11 hours. At the start of the second stage of the analysis, the distribution of the time remaining before failure is identical for all the components and is illustrated in figure 7.3. The MTTF obtained after 50 hours into the second stage is 69.32 hours (state 1), 32.62 hours (state 2), 19.52 hours (state 3) and 8.58 hours (state 4).

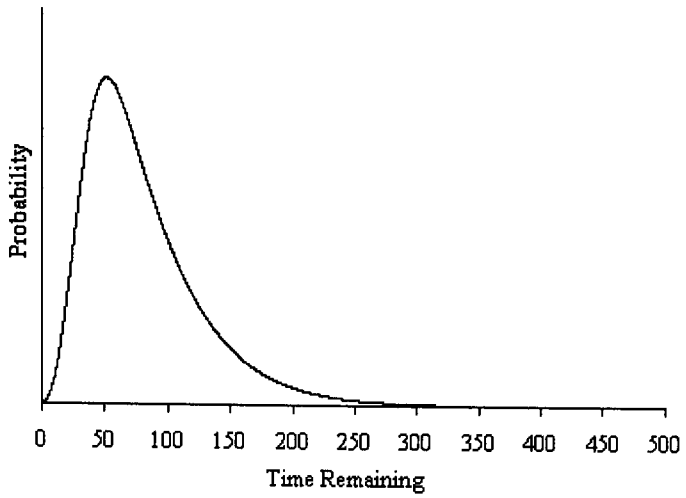

Figure 7.3 - Illustrating the conditional failure-time distribution obtained at time 0 hours for a vibration level within covariate band 1.

The MTTF after 150 hours into the second stage is 63.38 hours (state 1), 26.96 hours (state 2), 14.11 hours (state 3) and 4.57 hours (state 4). As expected, the distribution obtained at 150 hours (for any given state) has a reduced mean value when compared with the equivalent distribution obtained at time 50. The results of the MTTF and MSE loss function analysis are given for the PHM in table 7.6 for all monitoring points within the second-stage.

Bearing 1

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
80.5	4.0744	-	-	-	-		-
96.5	6.6828	21.5	77.83	2225	5397.6	18.5	18.62
108	27.9877	10	16.10	139.7	176.9	0.25	18.82
118	F						

Bearing 2

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
68	4.6055						-
80.5	5.4781	75	78.30	2219.9	2230.7	21	21.47
92.5	5.8982	63	75.33	2238.5	2390.5	16	19.83
104	8.2242	51.5	72.92	2230.6	2689.3	12	18.26

116.5	8.4698	39	70.62	2205.7	3205.4	8.5	16.69
129	12.6383	26.5	31.93	511.3	540.7	0.25	15.51
145.5	29.5518	10	6.30	32.9	46.7	0.25	14.73
155.5	F						

Bearing 3

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
74	3.4853			-	-	-	-
85	5.3865	30	78.50	2217.4	4569.8	20.5	20.50
102	13.0421	13	37.95	555.1	1177.3	3	19.43
105	23.7722	10	13.90	116.1	131.3	0.25	19.47
115	F						

Bearing 4

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
142	3.5474	-	-		-		-
156.5	5.1302	83.5	78.03	2222.9	2252.9	14.25	12.13
176.5	6.4198	63.5	73.42	2233.8	2332.1	7.75	11.13
188	6.1929	52	71.24	2213.9	2383.9	5.25	10.54
203	8.0022	37	68.72	2173.7	3179.6	2.75	9.82
215.5	12.549	24.5	30.42	491.6	526.6	0.25	9.30
230	13.9129	10	28.79	466.6	819.6	0.25	8.72
240	F						

Bearing 5

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
106	4.3209	-	-	-	-		-
132.5	5.3062	158	76.50	2235	8877.6	13.75	14.29
142	6.3926	148.5	76.35	2237.7	7736.7	10.5	13.61
154	7.6985	136.5	71.97	2222.2	6387.1	7.5	12.75
173.5	6.908	117	68.64	2172.1	4511	3.75	11.46
183	7.3993	107.5	67.19	2141.5	3766.3	2.5	10.90
197.5	7.8458	93	65.15	2090.5	2866.3	1	10.12
209	7.8588	81.5	63.65	2048.1	2366.9	0.25	9.37
222.5	8.4337	68	61.99	1997.1	2033.2	0.25	8.99
245	9.1468	45.5	59.45	1911.7	2106.1	0.25	8.17
256	10.5077	34.5	24.17	379.4	486.1	0.25	7.86

269.5	13.9153	21	23.31	361.2	366.6	0.25	7.47
280.5	19.6412	10	10.69	86.6	87.0	0.25	7.33
290.5	F						

Bearing 6

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
892	4.2867	-	-	-	-	-	-
916	8.3994	68	76.81	2233.2	2310.7	2.25	2.18
928	11.8194	56	37.57	554.0	893.5	0.25	2.16
940	11.8065	44	34.99	540.7	621.9	0.25	2.13
960	14.9804	24	31.75	509.2	569.2	0.25	2.09
964	16.6054	20	18.08	181.5	185.2	0.25	2.09
974	21.2831	10	6.40	33.9	46.9	0.25	2.20
984	F						

Table 7.6 – Results of the MSE and replacement decision analysis for case 1 using the PHM

The total MSE about the actual residual life for the PHM is 80269.2 and per observation the average MSE is 2293.41 for 35 observations. The total variance about the MTTF given by the PHM is 48430.3 and per observation the average variance is 1383.7.

7.5.1.2 The stochastic filter

In the case of the stochastic filter we initially establish the delay-time distribution $p_0(x_0)$. Upon defect detection and commencement of the second-stage, the failure (delay) times for the six bearings are 29.5, 81.25, 35.5, 90.75, 171.25 and 80 hours respectively. Using the log-likelihood function, given by equation [7.11], we obtain estimates of α = 0.0109 and β = 1.8691. The next task is the estimation of the parameters in the conditional density $p(y_i \mid x_i)$ given by equation [7.8] and the relationship between the delay-time and the vibration level that was established by equation [7.9]. Using the log-likelihood function, equation [7.15], and Matlab© v6.5 function 'fmincon' from the optimisation toolbox, the parameter estimates obtained

from the total sample of 35 observations are given in table 7.7.

Parameter	Estimate
A	7.3893
B	29.9213
C	0.0632
η	4.7060

Table 7.7 – The parameter estimates for the filter, case 1

The results of the MTTF and loss function analysis under the stochastic filter with an approximation increment of 0.05 hours are given in the following tables for all monitoring points within the second stage of component life.

Bearing 1

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
80.5	4.0744	-	-	-	-	-	-
96.5	6.6828	21.5	92.79	1578.8	6661.2	36	16.07
108	27.9877	10	8.20	11.1	14.4	0	18.54
118	F						

Bearing 2

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
68	4.6055	-	-	-		-	-
80.5	5.4781	75	96.19	1575.8	2024.8	40.5	17.87
92.5	5.8982	63	94.91	1388	2406.1	41.05	15.84
104	8.2242	51.5	84.57	1269.1	2362.4	34.3	15.04
116.5	8.4698	39	73.52	1109.2	2300.8	28.3	14.23
129	12.6383	26.5	30.84	56.1	74.9	14.45	14.19
145.5	29.5518	10	6.78	8.7	19.1	0	13.77
155.5	F						

Bearing 3

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
74	3.4853	-	-		-	-	-
85	5.3865	30	96.64	1582.2	6023.4	40.35	17.19

102	13.0421	13	29.83	59.7	342.8	13.20	17.83
105	23.7722	10	13.11	11.9	21.6	3.60	18.69
115	F						

Bearing 4

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
156.5	5.1302	83.5	96.18	1561.8	1722.4	35.15	10.87
176.5	6.4198	63.5	89.40	1330	2000.6	32.75	9.84
188	6.1929	52	88.26	1199.6	2514.1	38.95	9.23
203	8.0022	37	77.65	1066.2	2718.3	28.20	8.82
215.5	12.549	24.5	32.09	62.4	120	13.85	8.81
230	13.9129	10	20.17	35	161.2	5.85	8.55
240	F						

Bearing 5

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
106	4.3209	-	-	-	-	-	-
132.5	5.3062	158	93.46	1504.2	5669.6	35.05	12.52
142	6.3926	148.5	93.19	1341.2	4400.1	37.50	11.54
154	7.6985	136.5	85.58	1227	3819.9	33.25	10.97
173.5	6.908	117	78.96	1099.2	2546.4	28.35	10.16
183	7.3993	107.5	77.31	1000.5	1911.8	29.20	9.63
197.5	7.8458	93	70.35	889.7	1402.6	25.40	9.14
209	7.8588	81.5	66.78	802.7	1019.5	24.15	8.72
222.5	8.4337	68	59.38	672.4	746.6	20.70	8.34
245	9.1468	45.5	46.57	460.7	461.8	12.90	7.86
256	10.5077	34.5	35.09	147.9	148.3	11.65	7.55
269.5	13.9153	21	21.82	37	37.7	5.30	7.33
280.5	19.6412	10	12.13	17.2	21.7	0	7.13
290.5	F						

Bearing 6

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
892	4.2867	-	-	-	-	-	-
916	8.3994	68	80.94	1540	1707.4	15.15	2.16
928	11.8194	56	34.18	120.3	596.3	8.85	2.14
940	11.8065	44	25.36	47	394.3	5.25	2.12

960	14.9804	24	12.42	31.4	165.4	0	2.09
964	16.6054	20	12.28	21.5	81.1	0	2.08
974	21.2831	10	6.19	10.7	25.2	0	2.06
984	F						

Table 7.8 – Results of the MSE and replacement decision analysis for case 1 using the filter

The total MSE about the actual residual delay-time for the stochastic filter is 56643.8 and per observation, the average MSE is 1618.4 over 35 observations. The total variance about the MTTF given by the filter is 24876.2 and per observation the average variance is approximately 710.7. Although the point predictions of residual life are often similar under the two approaches, it is clear from the MSE comparison that the distribution produced by the stochastic filter provides a substantially better fit to the data. This is reflected in the relevant replacement decisions where the more dispersed curve of the PHM produces an optimal replacement time that is invariably sooner and more expensive than that obtained with the filter. The reasoning for this is the lack of confidence in the conditional distribution represented by a flatter curve over a much longer range than the filter. As a result, the cumulative probabilities in the PHM case rise steadily from the start of the projection interval and affect the replacement decisions. However, in the filtering case the distribution is much tighter in range. The conditional probability distributions obtained using both the proportional hazards model and the stochastic filtering approach are illustrated for case 1, bearings 2 and 6 at each condition monitoring point.

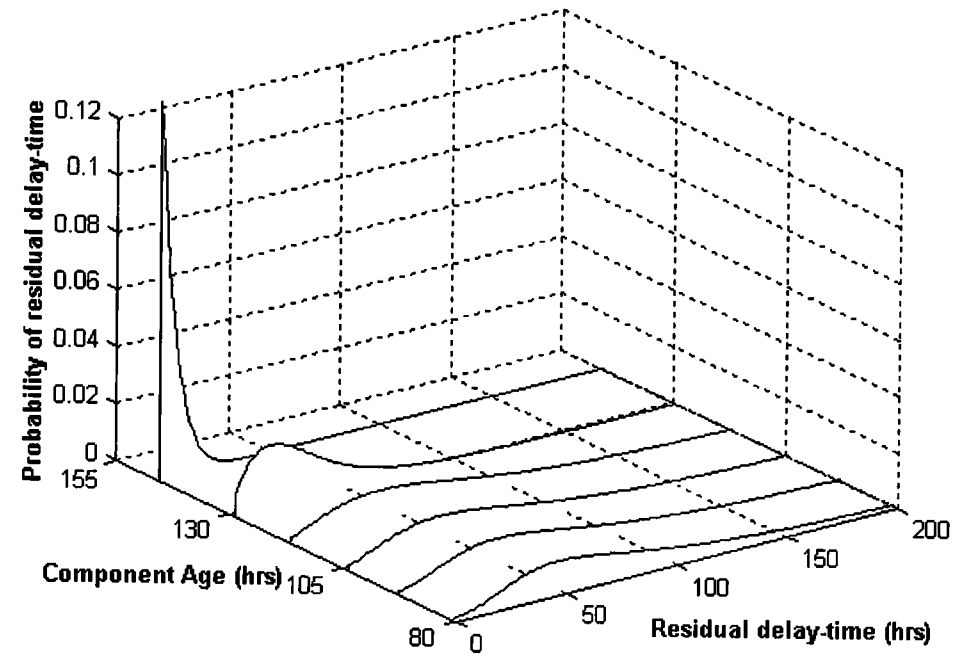

Figure 7.4 - Illustrating the conditional density $p_i(x_i|y_i)$ obtained using the proportional hazards model at each monitoring point for bearing 2.

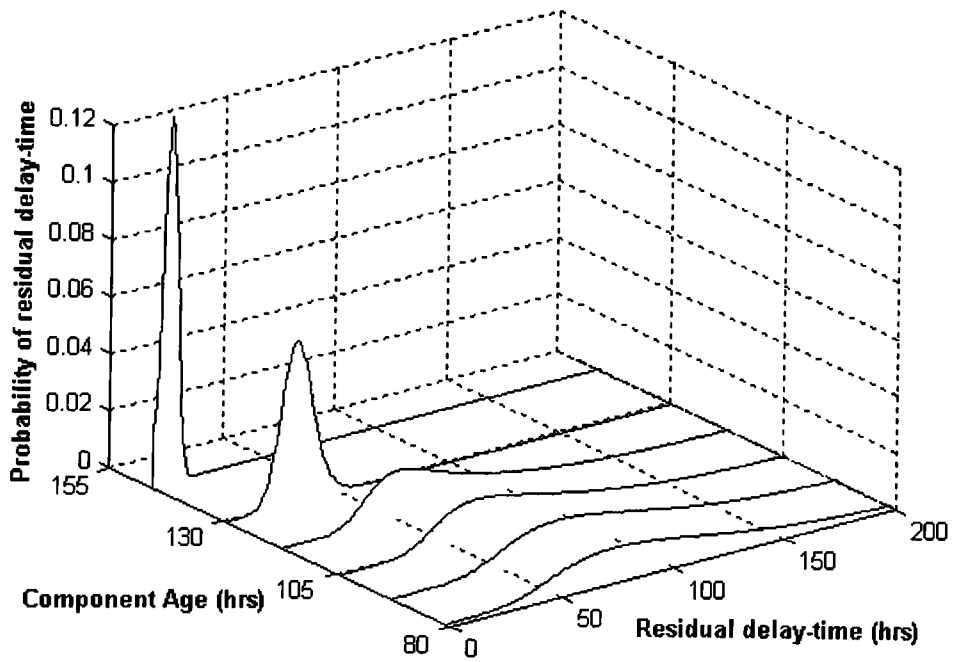

Figure 7.5 - Illustrating the conditional density $p_i(x_i|\underline{Y}_i)$ obtained using the stochastic filter at each monitoring point for bearing 2.

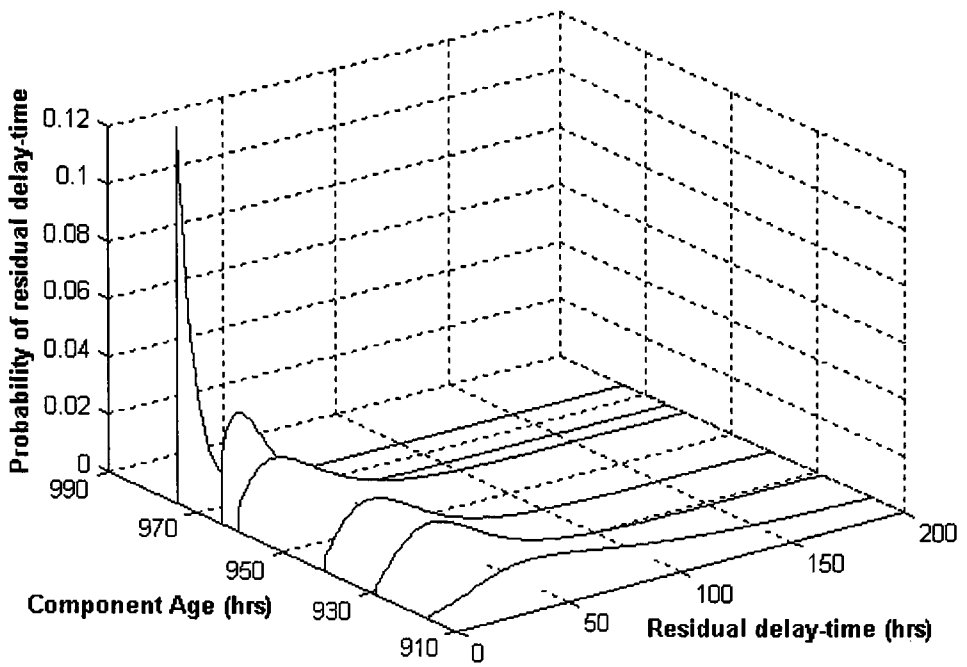

Figure 7.6 - Illustrating the conditional density $p_i(x_i|y_i)$ obtained using the proportional hazards model at each monitoring point for bearing 6.

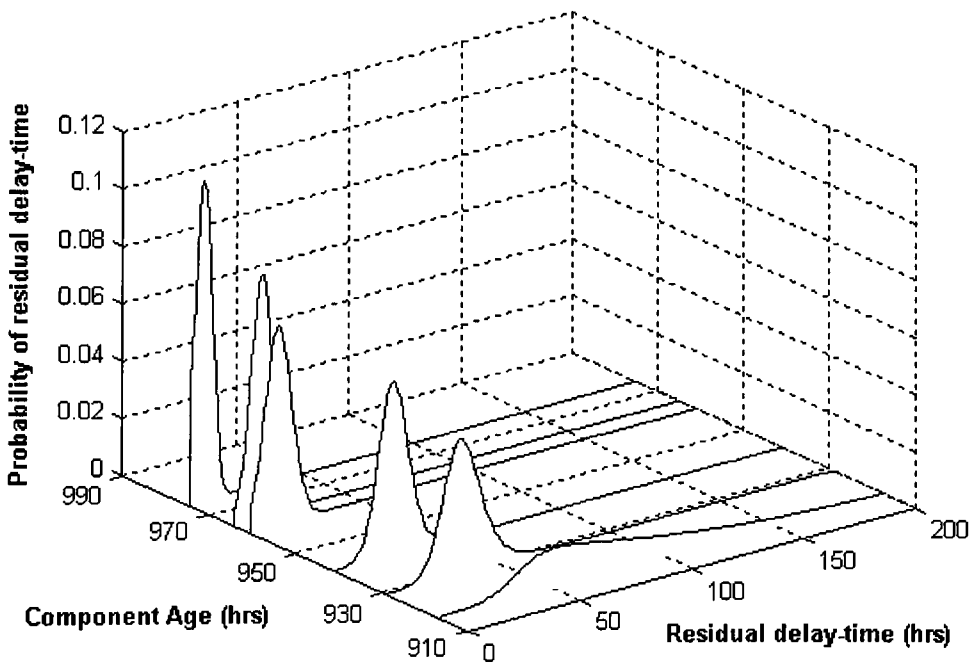

Figure 7.7 - Illustrating the conditional density $p_i(x_i|\underline{Y}_i)$ obtained using the stochastic filter at each monitoring point for bearing 6.

7.5.2 Case 2

For case 2, the vibration information from bearings 1, 4 and 6 is used to estimate the required parameters for both models and the fit of the model is established in the same manner as case 1. The model is then applied to the remaining bearings (2, 3 and 5) to evaluate how each model handles the new data.

7.5.2.1 The proportional hazards model

The parameter estimates for case 2 are given in table 7.9 and the matrix of transition rates in table 7.10.

Parameter	Estimate
β	2.2738
γ	0.1935
η	306.1257

Table 7.9 – The estimated parameters for the PHM, case 1

	1	2	3	4
1	-	2/109.75	0	1/109.75
2	0		1/60.5	0
3	0	0	-	1/10
4	0	0	0	-

Table 7.10 – The matrix of transition rates for the PHM, case 1

When the critical level is breached and the second stage begins, the average time remaining before failure for bearings 1, 4 and 6 is 66.75 hours. With a vibration level within covariate state 1, the MTTF given by the model is 67.68 hours. The results of the PHM; MTTF and MSE analysis for case 2 are given tables 7.11 and 7.12 for the model fit data and new data respectively.

Bearing 1

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
80.5	4.0744	-		-	-	-	-
96.5	6.6828	21.5	63.81	1237.7	3027.9	14	19.20
108	27.9877	10	18.06	156.5	221.4	0.25	18.74
118	F						

Bearing 4

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
142	3.5474	-	-		-	-	-
156.5	5.1302	83.5	64.15	1237.9	1612.2	10.50	12.34
176.5	6.4198	63.5	56.22	1187.6	1240.6	2.50	11.29
188	6.1929	52	52.53	1133.7	1134	0.25	10.64
203	8.0022	37	48.39	1054.4	1184.2	0.25	9.86
215.5	12.549	24.5	33.58	602.5	684.9	0.25	9.33
230	13.9129	10	30.34	530	943.8	0.25	8.75
240	F						

Bearing 6

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
892	4.2867		-	-	-	-	-
916	8.3994	68	62.05	1234	1269.4	0.25	2.18
928	11.8194	56	46.90	830.9	913.7	0.25	2.16
940	11.8065	44	42.36	769.3	772	0.25	2.13
960	14.9804	24	36.21	656.7	806.7	0.25	2.09
964	16.6054	20	12.63	80.3	134.6	0.25	2.11
974	21.2831	10	6.42	33.4	46.2	0.25	2.20
984	F						

Table 7.11 – The results for case 2 of the MSE and replacement decision analysis for the components used for model fitting under the PHM

Bearings 2, 3 and 5 are used to test the prediction ability of the model when applied to new data and the results are given in table 7.12.

Bearing 2

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
68	4.6055	-	-	-	-	-	
80.5	5.4781	75	64.62	1237.8	1345.6	16.5	22.23
92.5	5.8982	63	59.51	1220.1	1232.2	10	20.64
104	8.2242	51.5	55.37	1176.8	1191.8	5.25	18.92
116.5	8.4698	39	51.51	1115.8	1272.2	2	17.11
129	12.6383	26.5	36.56	664.6	765.6	0.25	15.55
145.5	29.5518	10	6.3	32.3	46	0.25	14.71
155.5	F						

Bearing 3

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
74	3.4853	-		-	-	-	
85	5.3865	30	64.97	1237.5	2466.2	16.5	21.16
102	13.0421	13	47.52	837.8	2029.7	4.75	19.42
105	23.7722	10	15.42	129.9	159.3	0.25	19.38
115	F						

Bearing 5

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
106	4.3209	-		-	-	-	
132.5	5.3062	158	61.52	1231.9	10539	9	14.64
142	6.3926	148.5	57.82	1205.3	9429	5	13.92
154	7.6985	136.5	53.76	1153.6	7999.7	1.75	12.96
173.5	6.908	117	48.26	1051.7	5776.5	0.25	11.53
183	7.3993	107.5	45.96	1000.4	4787.6	0.25	10.94
197.5	7.8458	93	42.81	923.9	3442.5	0.25	10.14
209	7.8588	81.5	40.59	866.2	2540.1	0.25	9.59
222.5	8.4337	68	38.22	802.6	1689.2	0.25	9.02
245	9.1468	45.5	34.79	707.2	821.9	0.25	8.20
256	10.5077	34.5	21.50	321.6	490.5	0.25	7.92
269.5	13.9153	21	19.98	286.2	287.3	0.25	7.54
280.5	19.6412	10	6.17	27	41.6	0.25	7.50
290.5	F						

Table 7.12 – The results for case 2 of the MSE and replacement decision analysis for the new components under the PHM

Considering the fit of the PHM developed for case 2, when the model is applied to the data used to establish the parameters (bearings 1, 4 and 6), the total MSE is 13991.6 with a total variance about the mean estimates of 10744.9. When investigating the application of the PHM to new data (bearings 2, 3 and 5) we obtain a total MSE of 58353.5 and a total variance of 17230.2.

7.5.2.2 The stochastic filter

The failure times for bearings 1, 4 and 6 are 29.5, 90.75 and 80 hours respectively and estimating the parameters of the density $p(x_0)$ we obtain $\alpha = 0.0133$ and $\beta = 2.925$. The estimated parameters for $p(y_i \mid x_i)$ are given in table 7.13.

Parameter	Estimate
A	7.6566
B	25.2323
C	0.0555
η	3.9743

Table 7.13 – The parameter estimates for the filter, case 1

The results of the MTTF and MSE analysis under the stochastic filter are given for case 2 in the following tables where bearings 1, 4 and 6 are used to ascertain the fit of the model to existing data.

Bearing 1

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
80.5	4.0744	-	-	-			-
96.5	6.6828	21.5	69.09	454.6	2719.5	31.8	16.49
108	27.9877	10	8.64	16.3	18.1	0	18.55
118	F						

Bearing 4

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
142	3.5474	-	-	-	-	-	

156.5	5.1302	83.5	71.28	450.4	599.7	30.55	11.08
176.5	6.4198	63.5	61.37	340.8	345.3	25.05	10.19
188	6.1929	52	58.44	279.6	321.1	24.85	9.60
203	8.0022	37	50.15	224	396.8	19.65	9.14
215.5	12.549	24.5	33.02	97.5	170.1	11	8.95
230	13.9129	10	20.99	53.1	173.8	2.70	8.66
240	F						

Bearing 6

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
892	4.2867	-		-	-		-
916	8.3994	68	62.32	442	474.3	13.50	2.16
928	11.8194	56	40.57	200.5	438.5	8.05	2.14
940	11.8065	44	29.96	100.3	297.4	3.95	2.12
960	14.9804	24	15.50	51.6	123.8	0	2.08
964	16.6054	20	14.04	32.9	68.4	0	2.08
974	21.2831	10	7.21	15.9	23.7	0	2.06
984	F						

Table 7.14 – The results for case 2 of the MSE and replacement decision analysis for the components used for model fitting under the filter

Bearings 2, 3 and 5 are used to test the model when applied to new data.

Bearing 2

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
68	4.6055	-	-	-	-	-	-
80.5	5.4781	75	71.67	455.6	466.7	35.25	18.48
92.5	5.8982	63	67.70	365.4	387.5	33.70	16.65
104	8.2242	51.5	60.03	302.6	375.3	28.70	15.65
116.5	8.4698	39	52.23	247.1	422.1	23.45	14.73
129	12.6383	26.5	33.43	99.6	147.6	13.15	14.38
145.5	29.5518	10	6.54	12.2	24.1	0	13.80
155.5	F						

Bearing 3

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
74	3.4853	-	-	-		-	-

85	5.3865	30	72.23	458.2	2241.9	35.35	17.72
102	13.0421	13	36.18	146	683.1	14.15	17.80
105	23.7722	10	13.85	20.6	35.5	2.05	18.92
115	F						

Bearing 5

Time	Vibration Level	Actual Time Remaining	MTTF	Variance	MSE	Replace in __	Expected Cost
106	4.3209	-	-	-	-	-	-
132.5	5.3062	158	67.37	427.3	8641.1	28.85	12.93
142	6.3926	148.5	65.43	343.6	7244.1	30.10	12.00
154	7.6985	136.5	58.82	284.8	6319.1	26.35	11.37
173.5	6.908	117	49.39	235.6	4806.9	18.80	10.63
183	7.3993	107.5	47.71	197.7	3772	19.10	10.09
197.5	7.8458	93	41.63	165.4	2804.5	14.70	9.58
209	7.8588	81.5	38.49	142.2	1992.4	12.90	9.15
222.5	8.4337	68	33.56	121.2	1307	9.25	8.75
245	9.1468	45.5	23.39	103.5	592.4	0	8.17
256	10.5077	34.5	21.81	80.4	241.5	0	7.81
269.5	13.9153	21	15.38	49.6	81.2	0	7.43
280.5	19.6412	10	8.82	23.1	24.4	0	7.15
290.5	F						

Table 7.15 – The results for case 2 of the MSE and replacement decision analysis for the new components under the filter

Considering the fit of the stochastic filter when the model is applied to the data used to establish the parameters (bearings 1,4 and 6), the total MSE is 6170.5 with a total variance about the mean estimates of 2759.5. When investigating the application of the filter to new data (bearings 2, 3 and 5) we obtain a total MSE of 42610.4 and a total variance of 4281.7. In the comparison provided by case 2, the same behaviour and results are observed as those seen for case 1 with the stochastic filter outperforming the PHM with regards to the both the fit of the models to the existing data and the application of the models to new data. The total MSE and the variance about the mean both demonstrate that for this particular case the stochastic filter gives

a tighter distribution about both the mean and the actual residual life. The replacement policy follows the same pattern as case 1 with the PHM producing much more conservative replacement decisions due to the uncertainty observed in the conditional density. In the following section, some of the reasons why the recursive filtering approach fits the data better in the cases studied here are discussed.

7.6 Discussion

In this chapter, we explored two cases using the vibration level data observed from six bearings; the first case exploring the fit of both models to all given data and the second examining the ability of the models to adapt to new input when only some of the available data is used to construct the actual model. Using both techniques we look to obtain an expression for the conditional distribution for the residual life at CM points throughout the life of a component. Confident predictions at all stages of the defective phase of component life are important however, one might argue that in practical on-line scenarios, a confident prediction in the later stages of component life (given reasonably spaced monitoring intervals) is indeed more important from the point of view of failure prevention than in the initial stages of defective operation, although admittedly this is a problem-specific issue. From figures 7.6 and 7.7 illustrating the evolution of the conditional density for case 1, and figures 7.8 and 7.9, it is clear that for this case, the distribution obtained using the PHM is typically flatter indicating a greater variance or reduced confidence in the actual estimate of expected residual life.

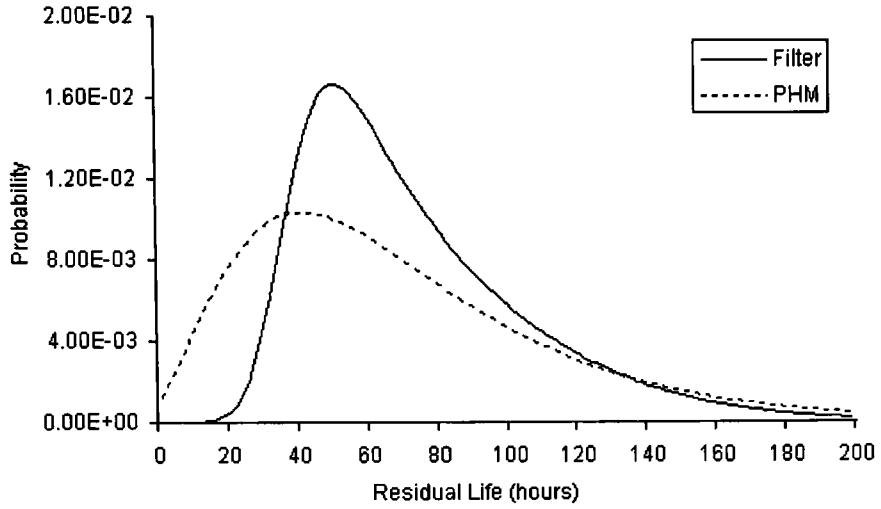

Figure 7.8 - A single stage comparison of the conditional density obtained using the PHM and the stochastic filter for bearing 2, CM checkpoint 4.

This is particularly evident in figure 7.9 where the probability associated with the mode of the stochastic filters conditional distribution is literally three times that of the PHM;

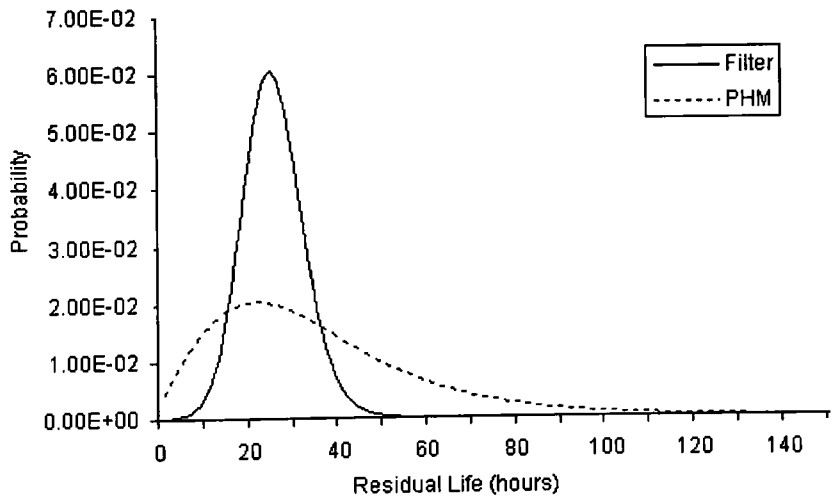

Figure 7.9 - A single stage comparison of the conditional density obtained using the PHM and the stochastic filter for bearing 6, CM checkpoint 3.

The MTTF and MSE analysis for case 1 confirm the situation observed in the diagrams where, although point predictions from the two techniques are similar, the filtering approach offers increased confidence in the predictions than the PHM. A

model that fits existing data well but is poor when applied to new data has little merit from an application perspective. This was the focus of case 2 where again the filtering approach provided a substantially better fit from an MSE perspective when applied to both the data used to construct the model and new data. As discussed in the case study, the flatter more dispersed curve of the PHM naturally leads to a more cautious replacement decision than the sharp, tight curve of the stochastic filter. This is because the probability of component failure before a particular instant is likely to be greater with the PHM, particularly in the early stages of monitoring. In addition to this, the PHM decisions tend to have a larger associated expected cost per unit time in both cases. Therefore, for the particular case studies investigated in this chapter, the decision model constructed around the PHM leads to less operational availability than the filtering model and at a greater expected cost. It is interesting to note that, when comparing the results from the two cases, the overall fit of the stochastic filter when applied to all the data used in case 2 is better than the fit of the filter to case 1 with less than half the data used for estimation. Clearly, the outlying case of bearing 5 has a greater than desired impact on parameter estimation and when excluded from the estimation and model fitting altogether, the fit of the stochastic filter to all the remaining data is much improved.

However, the key observation to be drawn from the research in this chapter pertains to the differing shapes of the conditional distributions produced by the two techniques. Greater variance in the PHM curve indicates a greater level of uncertainty in the expected residual life prediction. This could be attributable to the lack of monitoring history included in the model at each stage of the process with only the monitoring time and the current CM reading featuring in the conditional mean of the density and the true underlying value. In addition, the violation of the relationship between the

observed information and the residual life makes the PHM appear inferior to the stochastic filter in a vibration monitoring context, particularly when considering the uncertainty observed in the conditional density. However, it is important to note that a conclusive comparison and definition of the stochastic filter as a more efficient estimator for vibration monitoring applications would be dependent on both a greater number of cases and larger sample sizes. We should also mention that there are also a number of ways in which both techniques could be further tailored to the particular cases described in this chapter.

Chapter 8. A case comparison of the PHM and a stochastic filter for CBM applications using multiple oil-based condition monitoring parameters

8.1 Introduction

In this chapter, we continue the comparison of the proportional hazards model (PHM) and the probabilistic stochastic filtering approach that we introduced in the previous chapter. In the case study in chapter 7, both models were developed to consider the overall vibration level as a single indicator of a unit's condition. In this study we are investigating and comparing the prediction capabilities of the two approaches using multi-dimensional oil-based CM data. The information obtained from oil analyses will typically consist of the associated concentrations of worn metals and a number of different contaminants in an oil sample that are often called condition indicators. This means that when considering oil data, both models should be constructed to handle multiple parameters obtained simultaneously at CM points as described in chapter 6. As with the vibration information in the previous chapter, we assume that a stochastic relationship exists between the monitored information and the actual condition of the unit. Note that, in this chapter, a single piece or part of equipment that is subject to condition monitoring (CM), will be referred to as a 'unit' and not a 'component', in order to differentiate between equipment and the principle components established using PCA for data reduction purposes. The probabilistic filtering model, originally developed in Wang & Christer (2000) for vibration information, assumes the observed CM information be a function of the underlying residual life and this relationship is not assumed to hold in reverse. This assumption is appropriate in the case of vibration monitoring, as was discussed in length in chapter 7. However, when considering oil-based wear information, the standard assumption is that more wear

metals means less residual life remains. This is due to the fact that oil-based monitoring is primarily designed for detecting wear related failures. Such wear is typically correlated with the underlying residual life in a negative manner, that is, more wear implies less residual life. On the other hand, the observed wear metals in the oil samples are good indicators of the underlying wear, that is, more wear results in more observed metals, but the relationship between them is complicated and requires a careful examination. On the one hand, observed metal concentrations are caused by the underlying wear, but these metal particles in the oil may also accelerate the wear process and generate more particles. It appears that the PHM may be a suitable candidate for this case since it treats the observed oil metal information as covariates that alter the hazard, and indeed the residual life. A recent development, Wang & Zhang (2002), used a concept called 'proportional residual', which adopted a recursive nature like the filtering approach, but assumed that the observed wear metals change the residual life. On a purely theoretical ground, we argue a simple but appropriate way to model the relationship between wear and observed wear metals is to treat these metal concentrations as random variables that are caused by the underlying wear. Due to the frequency of oil top up's and changes, the influence of metal particles in the oil on the wear process can be negligible as wear is directly related to the residual life, the same relationship exists between the wear metals and the residual life. This simplifies the relationship between the wear and the concentration of metallic contamination as we only need to model; wear caused observed metal particles. It is noted that the PHM uses a different modelling principle and when fitting the two models to the same data set, it is interesting to know which method provides a good fit. This is the aim of this investigation. An alternative is to

use both modelling techniques in this situation but this approach will be discussed in future work.

The chapter is laid out as follows; firstly, we discuss in more depth the type of oil based CM data that we are using as the input to our models and the means by which the information might be obtained. Two case comparisons are then presented where significant principle components are used as the input to the models. In the filtering case, an alternative set-up is required to assist in establishing the relationship between each observed CM indicator or principle component and the underlying residual life of the unit. For the second case, we are required to address some of the issues associated with parameterising the models using censored CM histories with unknown failure times. At the end of the chapter, we discuss the respective merits of the two approaches for the particular cases considered.

8.2 Oil data

There are a variety of oil-based monitoring techniques that can be employed to ascertain the volume and type of foreign wear particles in an oil sample such as the spectrometric oil analysis program (SOAP), optical microscopy and ferrous debris quantification. Refer to chapter 5 for more details on oil-based monitoring. The cumulative metal concentrations obtained using SOAP will typically display a trend throughout the lifetime of the unit. This is in contrast to the information obtained using vibration monitoring, where the signal is usually flat and relatively stable in the initial stages of the unit's life and only begins to display an increasing trend upon the arrival of a defect. The foreign wear particles measured in SOAP are usually metallic elements that are measured in parts-per-million (ppm) and other contaminants. We note that the cumulative metal concentrations collected from SOAP, which is directly related to the wear process, usually display a naturally increasing process as the

amount of wear cannot decrease. Figure 8.1 illustrates a typical oil-based wear measurement process over time.

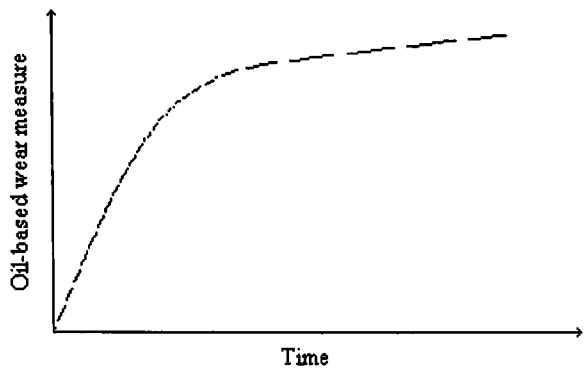

Figure 8.1 – Illustrating the cumulative wear metal process for a typical oil-based condition monitoring measure

When utilising oil-based information and considering the typical wear patterns that we observe, we may (as is demonstrated in the stochastic filtering element of this chapter) choose to model the incremental cumulative wear process; i.e. model the level of total wear that is accrued over the interval between monitoring points in order to provide more information on the evolution of the wear process. In cases where irregular CM is employed, the prediction of wear over a particular increment should reflect the duration of that increment. For the case studies explored in this chapter, the issues of multiple (potentially correlated) indicators of condition and dimension reduction are addressed using principle components analysis, as discussed in chapter 6. The CM vector $\underline{y}_i$ at time t_i that is used directly as input into the models is a vector of significant principle components obtained from the original oil-based information. The principle components are chosen based upon the variance and the p-values.

8.3 Case study 1

8.3.1 Introduction

For this initial case, we only have a single history of multivariate CM data available to establish the models. To compare the performance of the PHM and the filter in terms of assessing the ability to estimate the residual life, we again employ the total MSE criterion given by equation [6.54] using equations [6.52] and [6.53] for the filter and PHM respectively.

8.3.2 The data

The data set under consideration consists of the parts per million (ppm) of 5 types of metallic contaminant in oil samples that are obtained at irregular CM points. The metallic elements in the set are iron (Fe), copper (Cu), aluminium (Al), chromium (Cr) and nickel (Ni). Figure 8.2 illustrates the data set applicable to the single history under consideration for parameter estimation purposes.

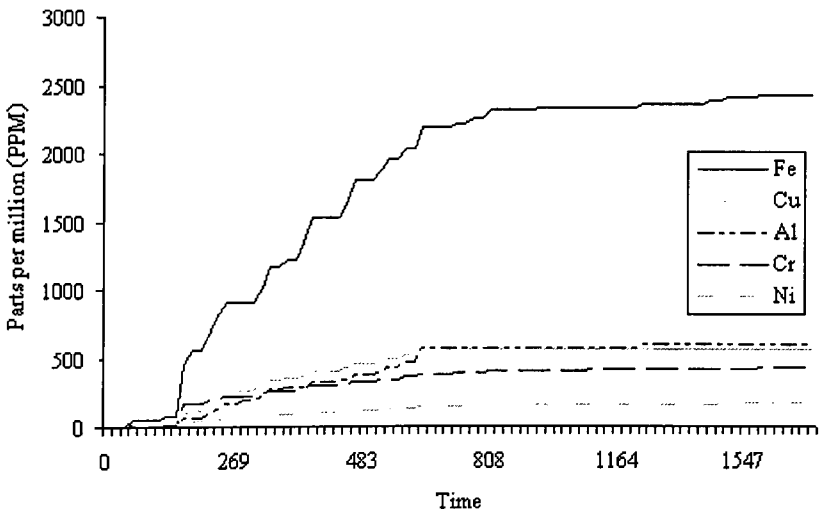

Figure 8.2 - The CM history for the 5 metal elements in parts per million

Principal components analysis is applied to reduce the dimensionality of the condition information and to remove any collinearity. Only the first principle component was

found to be significant and is illustrated in figure 8.3. As is evident from figures 8.2 and 8.3, the first element iron dominates the principal component.

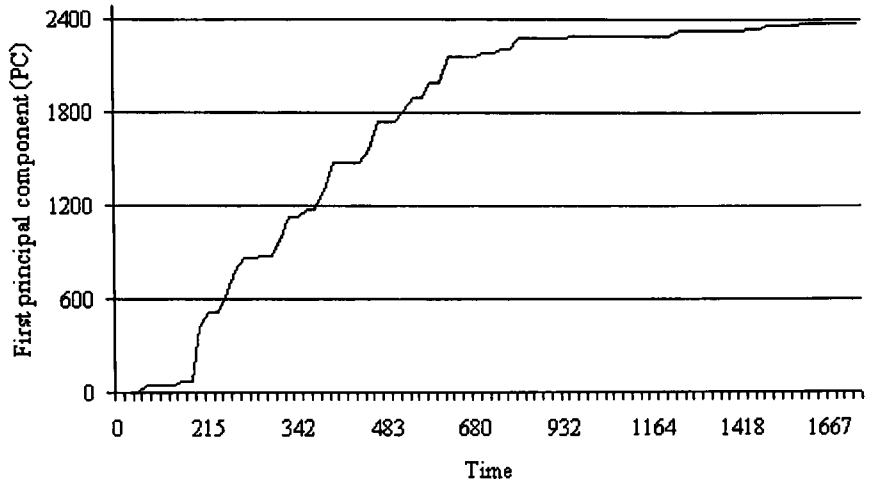

Figure 8.3 - The observed first principal component history

8.3.3 The proportional hazards model

When using the approximated Markov process to predict the future covariate state, we discretise the potential state-space for the CM parameter in question. The discretisation process is designed to approximate and enable efficient computation of the probabilities associated with particular transitions over the range of the parameters state-space and is necessary as there is not a continuous Markov model available for this case. However, even with the reduction in the number of potential states, the number of state combinations grows exponentially with the number of covariates included in the PHM and naturally this can lead to excessive computation and over fitting. Another issue is the potential collinearity between the covariates.

As discussed in chapter 6, some covariates are likely to be highly correlated in a CM context and this would result in inaccurate estimation of the remaining model parameters. The issues discussed are the reasons that, in the context of proportional hazards modelling, some CM scenarios with multiple information parameters may

require a means by which the removal of collinearity and data reduction can be achieved. As discussed in chapter 6, the technique of principle components analysis is appropriate for a problem of this nature. Lin et al (2002) state that the PCA approach is useful when combined with proportional hazards modelling, particularly when there is a large amount of correlation between the covariates, and when the number of covariates exceeds the number of histories used to estimate the parameters of the PHM. The approach we are utilising involves treating the significant principle components as actual covariates, discretising their respective ranges and inserting them directly into the parameter estimation and prediction processes.

Considering the fact that we only have a single significant principle component for this case, the PHM is identical to the model proposed in the previous chapter with the hazard given by equation [7.1] for a Weibull baseline hazard and exponential function of the single condition monitoring input. The conditional failure time distribution is given by equation [7.6] and defined at discrete intervals of duration ω. The only difference is that in this case, the condition input at the ith monitoring point, y_i, is the first principle component obtained from the original observation vector $\underline{z}_i$. Table 8.1 defines the ranges for the first principal component over which the discrete states are defined for the proportional hazards model. The table documents the mid-range value that is used as the covariate reading when the first principle component is within the appropriate range specified for the discrete state and the elapsed time that the unit spent in each state. Figure 8.4 illustrates the elapsed time and the transitions between the discretised states that are used for estimating the parameters of the PHM for this case. The failure time of the unit is 1722 hours and the transitions between states are assumed to occur at the mid-point between the relevant monitoring points.

State	PC range	Mid-value	Duration
1	(0, 600)	300	223.5
2	(600, 1200)	900	168
3	(1200, 1800)	1500	97
4	1800+	2100	1233.5

Table 8.1 - Documenting the ranges, values and elapsed times for the states of the PHM

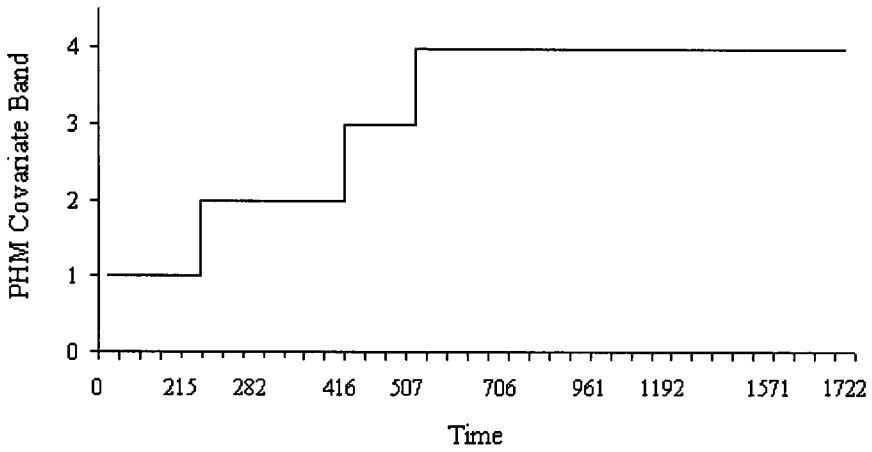

Figure 8.4 - Illustrating the transitions between the discrete PHM covariate states

Using equations [7.3] – [7.5] from the example given in the previous chapter, the likelihood function for a single unit history, $m = 1$, is

$$l(\beta,\eta,\gamma) = \log(\beta) - \beta\log(\eta) + (\beta-1)\log(T) + \gamma y(T) - \eta^{-\beta}\sum_{d=1}^{q}(e^{\gamma A_d}(V_d^{\beta} - V_{d-1}^{\beta}))$$

where, $y(t)$ is the approximate continuous-time sample path for the discretised states of the principle component and all other notation is consistent with the formulation in chapter 7. Using the Matlab© optimisation algorithm 'fmincon', the parameter estimates obtained for the PHM are given in table 8.2. As we are only utilising a single unit history for model parameterisation in this initial trial case and the range of the input information is discretised for use in the PHM, the amount of information included in the estimation process is minimal. As a result, the sample size is too small to enable accurate determination of the covariance matrix as described in chapter 2.

Parameter	Estimate
$\hat{\beta}$	26.7216
$\hat{\eta}$	3.51×10^{11}
$\hat{\gamma}$	0.2432

Table 8.2 - The PHM parameter estimates

Using an approximation increment of ω = 1 hour, the mean time until failure (MTTF) given by the PHM with the estimated parameters is 1721 hours. This expected residual life is very close to the failure time for the unit of 1722 hours. This result goes some way towards establishing the validity of the PHM and the estimated parameters for case 1 however we would expect the fit to be good with only a single data set being used for parameterisation.

8.3.4 The stochastic filter

As with the PHM, we are using the first principle component, y_i, (obtained from the original observation vector $\underline{z}_i$ at time t_i) as our direct CM input into the recursive stochastic filter. However, for reasons discussed in section 8.2, we choose to model the increment $\Delta y_i = y_i - y_{i-1}$. The form of the recursive stochastic filter is given by equations [6.30] – [6.33] with Δy_i replacing $\underline{y}_i$. The initial residual life distribution $p_0(x_0)$ is taken to be Weibull with parameters α and β as given by equation [7.7] for $x_0 \geq 0$. Given the fact that the first principle component is obtained from wear metals and considering an irregular monitoring process, the following relationship is appropriate when modelling the expected change in the first principle component between successive CM points;

$$\mathbf{E}[\Delta y_i \mid x_i] \propto \left(\frac{A + (B/x_i)}{C + Dt_i} \right) (t_i - t_{i-1}) \qquad [8.1]$$

The relationship given by equation [8.1] is a reflection of the age, the underlying residual life and the duration of the increment over which the change has been observed. The numerator in the first bracket enables the expected value to increase over time as the residual life x_i decreases. This is countered to some extent (defined by the appropriate parameters) by the increasing denominator over time enabling the curve of expectation against time to level off as indicated by our data. The second bracket $(t_i - t_{i-1})$ in equation [7.1] enables consideration of the fact that the expected wear over a particular interval should reflect the duration of that interval. One suggested distribution, that is subsequently utilised in the case study and maintains the required relationship between the condition information and the residual delay time, is a 2-parameter Weibull distribution given by

$$p(\Delta y_i \mid x_i) = \rho_i \eta (\rho_i \Delta y_i)^{\eta-1} e^{-(\rho_i \Delta y_i)^{\eta}} \qquad [8.2]$$

where, $\rho_i = \dfrac{C + Dt_i}{(A + (B/x_i))(t_i - t_{i-1})}$, which is essentially a 3-parameter Weibull distribution with y_{i-1} as the location parameter. Equation [8.2] enables the realisation of the relationship in expression [8.1] since we have $\mathbf{E}[\Delta y_i \mid x_i] \propto \rho$. At the ith monitoring point the conditional residual life distribution is

$$p_i(x_i \mid \underline{Y}_i) = \frac{(x_i + t_i)^{\beta-1} e^{-(\alpha(x_i+t_i))^{\beta}} \prod_{h=1}^{i} \phi_h(x_i, t_i)}{\int_0^{\infty} (u + t_i)^{\beta-1} e^{-(\alpha(u+t_i))^{\beta}} \prod_{h=1}^{i} \phi_h(u, t_i)\, du} \qquad [8.3]$$

for which we define the function

$$\phi_h(u, t_i) = \frac{e^{-\left(\frac{(C+Dt_h)\Delta y_h}{(A+B(u+t_i-t_h)^{-1})(t_h - t_{h-1})}\right)^{\eta}}}{(A + B(u + t_i - t_h)^{-1})^{\eta}} \qquad [8.4]$$

For parameter estimation purposes, the likelihood function for this case is given by equation [6.34] for a single component history with Δy_i replacing $\underline{y}_i$. It is evident from the derivation given in Appendix 1 that the likelihood function for this particular case with the chosen distributions is simply

$$\mathcal{L} = p(x_0 = T)\prod_{i=1}^{n} p(\Delta y_i \mid (x_i = T - t_i)) \qquad [8.5]$$

As with previous likelihood functions, optimisation with respect to the parameters of interest is made easier by taking logarithms of the expression giving

$$l(\alpha,\beta,A,B,C,D,\eta) = \beta\log\alpha + \log\beta + (\beta-1)\log T - (\alpha T)^{\beta} + n\log\eta + \ldots$$

$$\sum_{i=1}^{n}\left(\eta\log\left(\frac{C+Dt_i}{(A+B/(T-t_i))(t_i-t_{i-1})}\right)+(\eta-1)\log(\Delta y_i)-\left(\frac{(C+Dt_i)\Delta y_i}{(A+B/(T-t_i))(t_i-t_{i-1})}\right)^{\eta}\right) \qquad [8.6]$$

From the unit history under investigation and analogous units, the average lifetime is taken to be $1722 \pm 20\%$ hours. As such, the parameters of the initial Weibull delay time distribution $p_0(x_0)$ can be estimated separately using equation [7.10], for $m = 1$, as $\alpha = 0.0005692$ and $\beta = 27.5$ As has already been discussed, the incremental shift in the first principal component between monitoring points is used as the input to the filtering model and the estimation process. Re-scaling the incremental shifts to consider the duration over which the shift occurred, i.e. the time between successive monitoring points presents more information regarding the interpretation of the data when used as input to the stochastic filter. Figure 8.5 illustrates the proportional shifts in the first PC over time for $n = 30$ monitoring points.

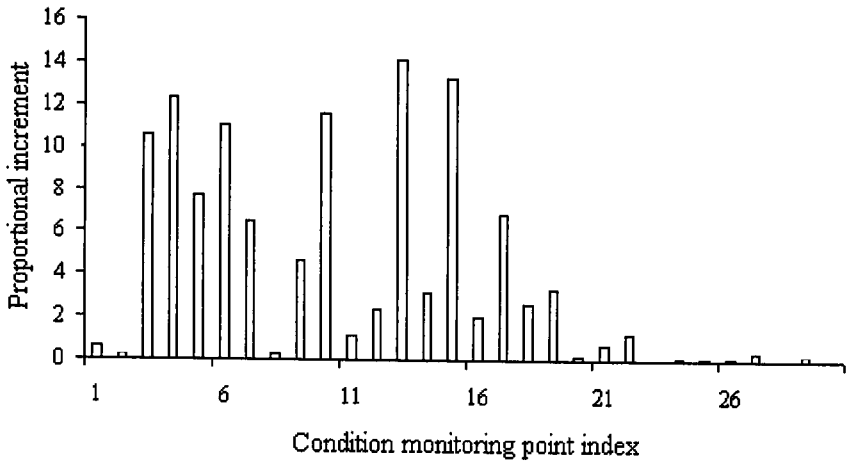

Figure 8.5 - Illustrating shifts in the first PC relative to the time between CM points

Using the Matlab© optimisation algorithm 'fmincon', the parameter estimates obtained for the stochastic filter are given in table 8.3.

Parameter	Estimate	Variance
$\hat{\alpha}$	0.0005692	$3.447*10^{-9}$
$\hat{\beta}$	27.5	$3.063*10^{3}$
$\hat{A}$	2.3487	0.431
$\hat{B}$	1.3358	$1.22*10^{4}$
$\hat{C}$	0.2672	0.072
$\hat{D}$	0.0027	$1.33*10^{-6}$
$\hat{\eta}$	0.5787	$5.564*10^{-3}$

Table 8.3 - The parameter estimates for the stochastic filter

8.3.5 Results

The performance of the two models when applied to the data set is now compared. Figures 8.6 and 8.7 illustrate the conditional failure time densities obtained at each monitoring point for the PHM and the filter respectively.

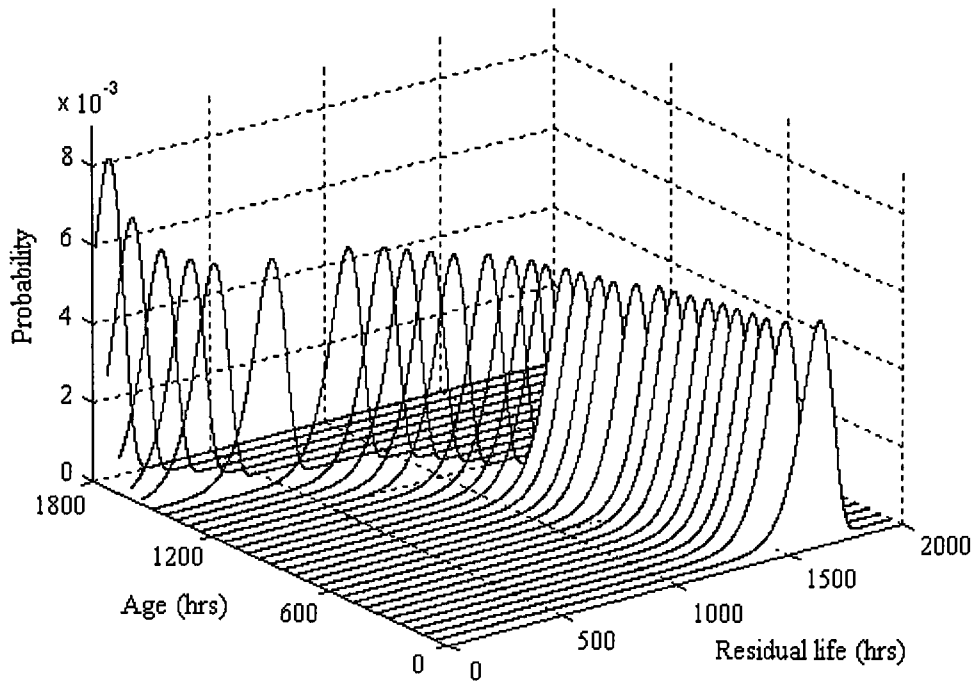

Figure 8.6 - Illustrating the conditional densities obtained at each CM point using the proportional hazards model

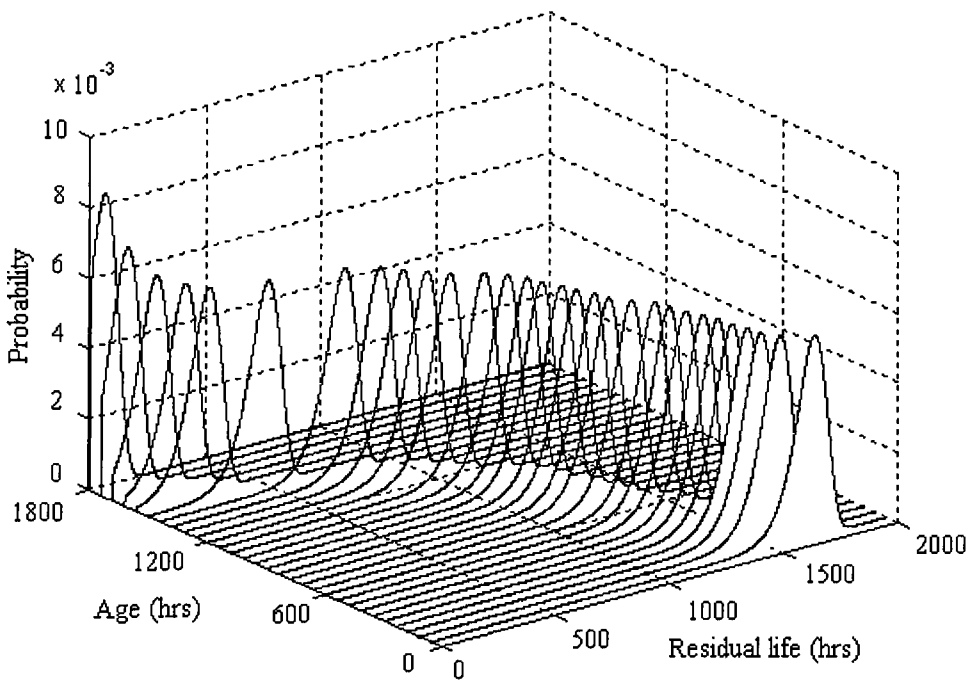

Figure 8.7 - Illustrating the conditional densities obtained at each CM point using the stochastic filter

There is not a significant difference between the proposed PHM and filter for this case, as is evident in figures 8.6 and 8.7. The total MSE over all available monitoring points is 185215.7 in the case of the PHM and the filter fairs marginally better with 175173.5. Similarly, the associated replacement decisions and expected attributable costs are almost identical with the PHM decisions being marginally more conservative with a slightly larger average cost. This is due to the slight increase in variance observed when using the PHM. Applying a standard Weibull survival analysis with parameters $\alpha = 0.0005692$ and $\beta = 27.5$, we obtain an MSE of 175203.9. The results for this case indicate that with a single history for model fitting, it is very difficult to establish a correlation between the CM information and the underlying residual life. This conclusion is supported by the fact that the filter offers only a slight improvement on the Weibull survival analysis and the PHM is in fact worse.

8.3.6 Further

The set-up proposed for the stochastic filter in equation [7.1] is designed to model an incremental mean wear process where a pattern similar to that illustrated in figure 8.8 is expected.

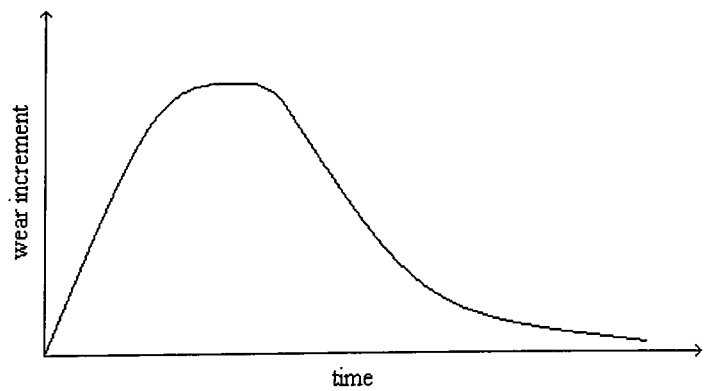

Figure 8.8 - Illustrating the expected incremental wear process over time

However, as is illustrated in figure 8.9, when using the actual incremental data and not the mean process, the observed pattern is less consistent.

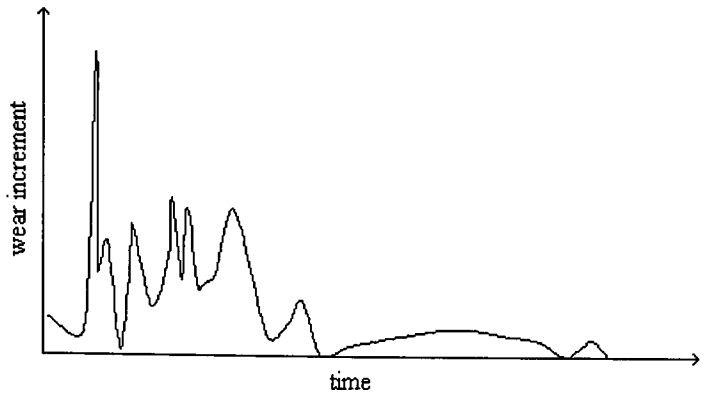

Figure 8.9 - Illustrating the actual incremental wear process over time

Taking these fluctuations into consideration, the following transformation of the condition information is proposed creating a smoothed input to the stochastic filter;

$$\theta_i = \frac{1}{i}\sum_{r=1}^{i}\left(\frac{y_r - y_{r-1}}{t_r - t_{r-1}}\right) \qquad [8.7]$$

where, y_r is the principal component at the rth monitoring point and the history of condition information is denoted by $\underline{\theta}_i = \{\theta_1, \theta_2, ..., \theta_i\}$. θ_i could be described as the moving average rate of wear over an increment however, note that θ_i does not equate to y_i / t_i. The following relationship is proposed between the transformed information and the underlying residual life;

$$\mathbf{E}[\theta_i \mid x_i] \propto t_i^{-1} A e^{-Bx_i/t_i} \qquad [8.8]$$

As with the relationship proposed for the previous filter, the information is incorporated in the filtering process through the specification of $p(\theta_i \mid x_i)$ as

$$p(\theta_i \mid x_i) = \rho_i \eta (\rho_i \theta_i)^{\eta-1} e^{-(\rho_i \theta_i)^{\eta}} \qquad [8.9]$$

where, $\rho_i = t_i / A e^{-Bx_i/t_i}$. Using the filtering equations [6.30] – [6.33], a closed-form expression is established for the ith iteration of the filter as

$$p_i(x_i \mid \underline{\theta}_i) = \frac{(x_i + t_i)^{\beta-1} e^{-(\alpha(x_i+t_i))^{\beta}} \prod_{h=1}^{i} \phi_h(x_i, t_i)}{\int_0^{\infty} (u + t_i)^{\beta-1} e^{-(\alpha(u+t_i))^{\beta}} \prod_{h=1}^{i} \phi_h(u, t_i)\, du} \qquad [8.10]$$

for which we define the function

$$\phi_h(u, t_i) = \frac{e^{-\left(\frac{t_h \theta_h}{Ae^{-B(u+t_i-t_h)/t_h}}\right)^{\eta}}}{(Ae^{-B(u+t_i-t_h)/t_h})^{\eta}} \qquad [8.11]$$

Using equation [6.34] and the same process of iterative reduction and cancellation, the likelihood function for parameter estimation is obtained as

$$\mathcal{L}(\alpha, \beta, A, B, \eta) = \alpha^{\beta} \beta T^{\beta-1} e^{-(\alpha T)^{\beta}} \prod_{i=1}^{n} \eta t_i^{\eta} (Ae^{-B(T-t_i)/t_i})^{-\eta} \theta_i^{\eta-1} e^{-\left(\frac{t_i \theta_i}{Ae^{-B(T-t_i)/t_i}}\right)^{\eta}} \qquad [8.12]$$

The log-likelihood function for this case is

$$l(\alpha, \beta, A, B, \eta) = \log(\alpha^{\beta}\beta) + (\beta - 1)\log(T) - (\alpha T)^{\beta} + n\log(\eta) + \eta \sum_{i=1}^{n} \log\left(\frac{t_i}{Ae^{-B(T-t_i)/t_i}}\right)$$

$$+ (\eta - 1)\sum_{i=1}^{n} \log(\theta_i) - \sum_{i=1}^{n} \left(\frac{t_i \theta_i}{Ae^{-B(T-t_i)/t_i}}\right)^{\eta} \qquad [8.13]$$

The parameter estimates for the stochastic filter using the proposed transformation are obtained using the Matlab© algorithm 'fmincon' from the optimisation toolbox and are given in table 8.4.

Parameter	Estimate	Variance
$\hat{\alpha}$	0.0005692	$3.447*10^{-9}$
$\hat{\beta}$	27.5	$3.063*10^{3}$
$\hat{A}$	5739.3	$1.496*10^{6}$
$\hat{B}$	0.2446	$1.21*10^{-3}$
$\hat{\eta}$	1.0012	0.077

Table 8.4 - The parameter estimates for the stochastic filter using the data transformation

The conditional densities obtained at each monitoring point for this second stochastic filter are given in figure 8.10. Although, figures 8.7 and 8.10 for the two filtering models are very similar, the second model incorporating the transformation of the data as the CM input shows an improved performance when compared with the PHM and the first filtering model.

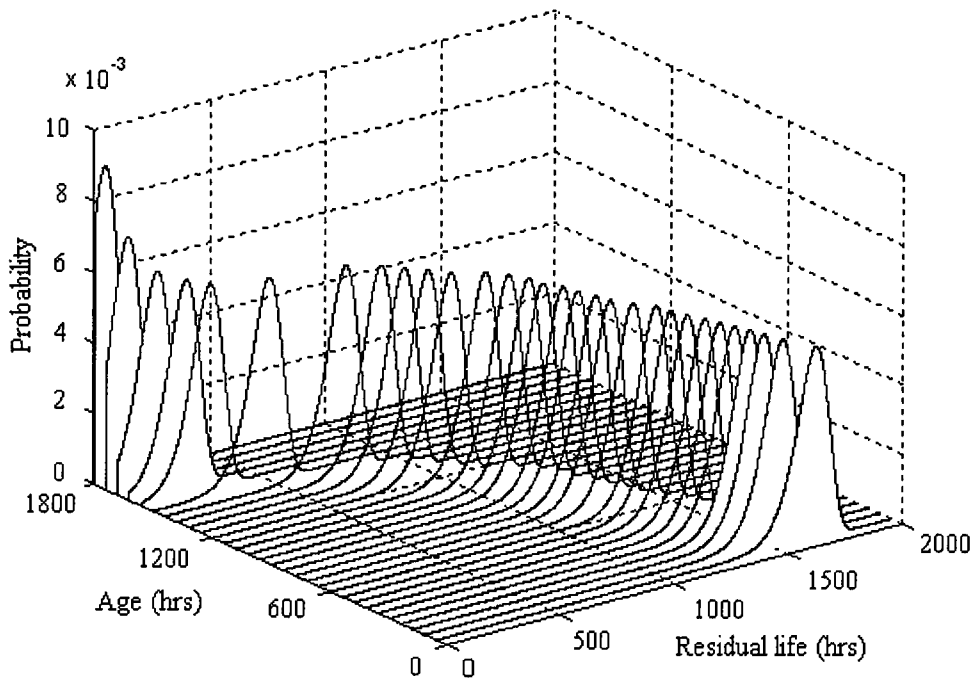

Figure 8.10 - Illustrating the conditional densities obtained at each CM point using the stochastic filter with the data transformation

The MSE of the conditional density observed at the first few CM points is marginally greater for the second filter however confidence in the mean estimate increases over time and the total MSE observed is 164799.8. This increased confidence is reflected in the replacement decisions and associated costs which are slightly less conservative at a reduced cost. However, any differences in the model fitting between the PHM and the two filtering models for this first simple case are negligible.

8.4 Case study 2

For case 1, both the PHM and the two proposed filters naturally provided a decent fit to the data used to establish them as only a single units CM history was used for construction. In this case, we apply the original filter and the PHM to a data set consisting of multiple unit histories and compare the fit of the two models. The second filtering set-up proposed by equation [8.8] and [8.9] is not pursued further as it did not provide a substantial improvement in the first case and initial analysis for this second case did not prove favourable.

8.4.1 The data

The original data set for this second case consists of 25 unit histories. The histories include the CM readings attributable to the concentrations of foreign metal particles in oil samples as the unit ages. However, the actual failure times are not recorded which complicates the estimation process and the model fit comparison. We immediately dismiss one of the available histories due to a lack of CM information. Of the remaining 24 units, m = 12 are selected at random and utilised in establishing the models (units 1, 4, 6, 10, 12-14 and 18-22). The remaining 12 units are used to test the applicability of the established models to new data of the same type. PCA is again employed to remove any collinearity and reduce the dimensionality of the available data for input to the PHM and the filter. A Scree plot is given in figure 8.11 that illustrates the necessity for principle component retention.

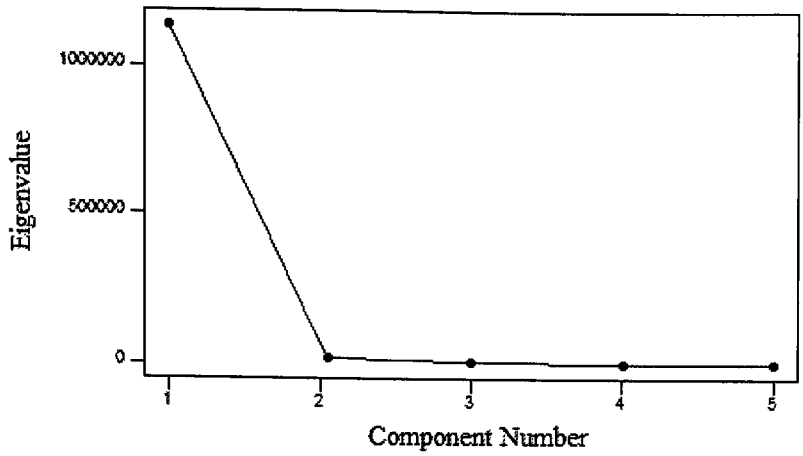

Figure 8.11 – A Scree plot of principle components 1-5

From figure 8.11, it is clear that the first principle component incorporates a substantial amount of the original information. As such, we disregard components 2-5 and retain the first as

$$PC\_1 = 0.953*Fe + 0.247*Cu + 0.136*Al + 0.097*Cr + 0.049*Ni$$

Figure 8.12 illustrates the first principle component for the 12 unit histories that are used to establish the models. Linear regression is used to streamline the estimation process when the data is left-censored.

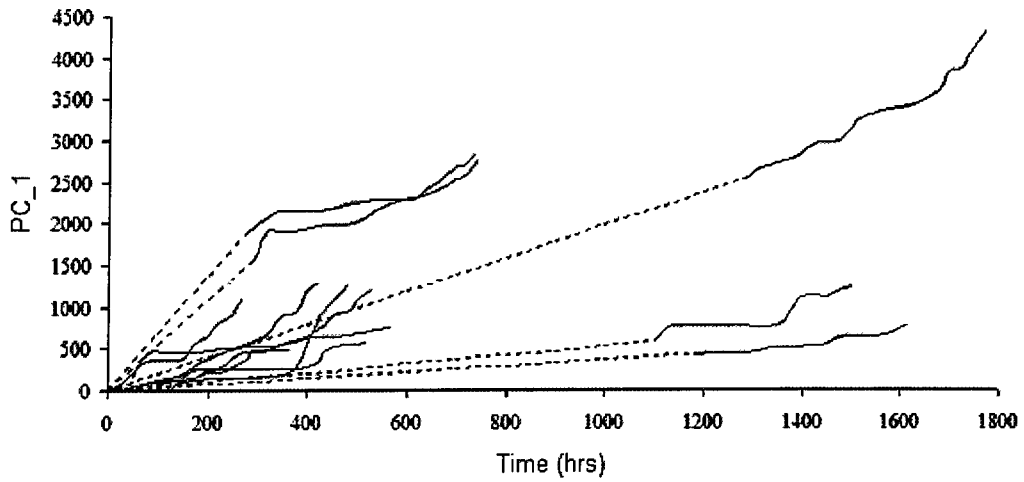

Figure 8.12 - The first principle component for units 1, 4, 6, 10, 12-14, 18-22

8.4.2 Parameter estimation

Using the stochastic filtering and PHM methodologies, we look to establish the conditional density $p_i(x_i \mid \underline{Y}_i)$ or $p_i(x_i \mid y_i)$ respectively, where, x_i is the residual life of the component at time t_i. The functional form of the distributions used for this case are the same as those used in the first case and are given by equations [8.1] and [8.2] for the filter and the hazard for the PHM given by equation [7.1] for a single condition input. However, for reasons that we will now discuss, the parameter estimation process and the analysis of the adequacy of model fit are subject to modification for this case. When the exact failure times of the units used for parameterisation of the models are known, the estimation process is relatively straightforward for both techniques. The PHM uses the failure times, and the final CM readings upon failure, directly in the construction of the likelihood function. See equation [7.3] – [7.5] for the log-likelihood function. In the case of the stochastic filter, knowledge of the actual underlying residual life as $x_i = T - t_i$ enables a convenient reduction in the complexity of the likelihood function for optimisation purposes. See the Appendix for details. However, as discussed in the introduction to this case, the actual failure times of the components are unknown. We only have information on the time of the final CM readings. As a result, we are forced to establish the models on the basis that the data is right-censored and that each component could in fact have continued operating for a substantial amount of time after the final CM point and simply was not monitored because of preventative reconditioning. For reasons that will become apparent, we firstly introduce the estimation of the intitial residual life distribution, as typically used in the filtering process, under the condition that the failure times are unknown. We attempt to establish an approximate range over which the $m = 12$ units

used for estimation purposes are assumed to fail. The lower limit of this range is established as

$$L = \min_j \{t_{jn_j}\} \qquad [8.14]$$

where, t_{jn_j} is the final CM time for the jth unit, for $j = 1, 2, \ldots, m$. It is obvious that all the units considered must fail sometime after L. The upper limit of the range for the unknown failure times is constructed as

$$U = \max_j \{t_{jn_j} + 2\psi_j\} \qquad [8.15]$$

where, ψ_j is established as

$$\psi_j = \sum_{i=1}^{n_j} (t_{ji} - t_{j,i-1})/n_j \qquad [8.16]$$

Now, taking $\underline{\phi}$ to represent the unknown parameters of $p(x_0)$, the expression

$$\int_L^U p(x_0 ; \underline{\phi})\, dx_0 \qquad [8.17]$$

is established for the chosen distributional form and maximised with respect to the unknown parameters and subject to the constraint that 90% of the distributions probability mass lies within the interval (L, U). The constraint is used to represent the uncertainty in the actual failure times. For this case, a Weibull form is assumed for the initial residual life distribution as $x_0 \sim \text{We}(\alpha, \beta)$ and the estimated parameters are given in table 8.5.

Parameter	Estimate	Variance
α	0.0008191	2.836×10^{-7}
β	2.43	9.856

Table 8.5 – The parameters of the residual life distribution $p(x_0)$

Figure 8.13 (below) illustrates the density $p(x_0)$ with 90% of the mass shaded between the two limits L and U.

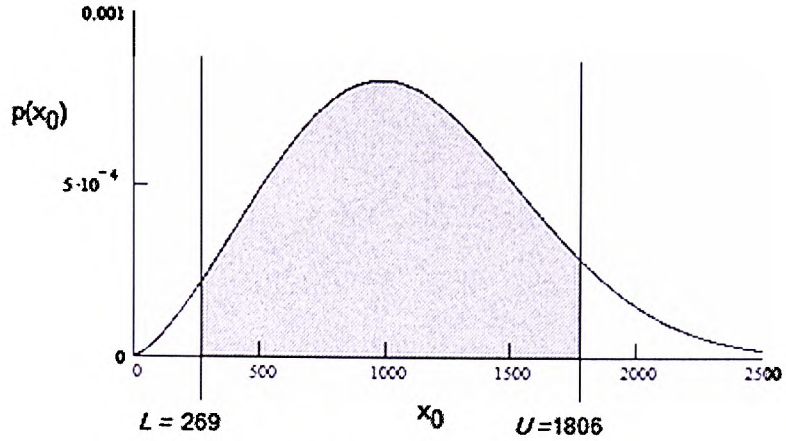

Figure 8.13 – The initial residual life distribution $p(x_0)$

8.4.2.1 The proportional hazards model – parameter estimation

The observable range of the first principle component is divided into discrete covariate states as shown in table 8.6.

State	Range	Mid-range
1	$0 \leq z < 750$	375
2	$750 \leq z < 1500$	1125
3	$1500 \leq z < 2250$	1875
4	$z \geq 2250$	2800

Table 8.6 – The discrete covariate states for the PHM

where, the mid-range is the covariate value adopted for input to the PHM whenever the reading for the first principle component falls within the designated interval. From the 12 unit histories used to parameterise the models, we obtain the discrete state transition rate matrix as shown in table 8.7.

	1	2	3	4
1	-	0.006739	0	0
2	0	-	0.003245	0
3	0	0	-	0.002731
4	0	0	0	-

Table 8.7 – The transition rate matrix for the PHM

Equation [6.12] gives the likelihood function for data sets that incorporate a mixture of complete and right-censored unit histories. When the monitoring process is suspended at some point for all the unit histories and the failure times are not recorded, the likelihood function becomes

$$\mathcal{L} = \prod_{j=1}^{m} R(t_{jn_j}, \underline{Y}_j) = \prod_{j=1}^{m} e^{-U(t_{jn_j}, y_j(t_{jn_j}))} \qquad [8.18]$$

For this particular case, we have

$$\mathcal{L} = \exp\left(-\sum_{j=1}^{m} \sum_{d=1}^{q_j} e^{\gamma' A_{jd}} ((V_{jd}/\eta)^{\beta} - (V_{j,d-1}/\eta)^{\beta}) \right) \qquad [8.19]$$

However, we found that the value of the log-likelihood function increased indefinitely as $\gamma \to 0$ and $\eta \to \infty$. We therefore concluded that more information was required to obtain appropriate parameter estimates. Returning to the initial residual life distribution developed in the previous sub-section, we define

$$\tilde{T} = \bar{x}_0 = \int_0^{\infty} x_0 \, p(x_0; \alpha, \beta) dx_0 \qquad [8.20]$$

as a further input to the PHM estimation process. The likelihood function becomes

$$\mathcal{L} = \prod_{j=1}^{m} h(\tilde{T}, y_j(t_{jn_j})) R(t_{jn_j}, \underline{Y}_j) \qquad [8.21]$$

where,

$$h(\tilde{T}, y_j(t_{jn_j})) = \frac{\beta}{\eta}\left(\frac{\tilde{T}}{\eta}\right)^{\beta-1} e^{\gamma y_j(t_{jn_j})} \qquad [8.22]$$

and $R(t_{jn_j}, \underline{Y}_j)$ is unchanged from the formulation given by equations [6.13] and [6.14]. Note that, upon failure for a given unit, we assume the discretised covariate to be in the same state as at the final monitoring point. The mean parameter estimates and their respective variances are given in table 8.8.

Parameter	Estimate	Variance
β	2.82	0.387
η	1084	3.869×10^{4}
γ	3.286×10^{-5}	9.452×10^{-8}

Table 8.8 – The PHM parameter estimates for case 2 with right-censored histories

A likelihood profile was then undertaken to confirm the parameter estimates by fixing the parameters individually and performing line searches on the remaining parameter spaces.

8.4.2.2 The stochastic filter – parameter estimation

To parameterise the filtering model, the likelihood function for this particular case is

$$\mathcal{L} = \prod_{j=1}^{m}\left(\prod_{i=1}^{n_j} p(\Delta y_{ji} \mid \underline{Y}_{j,i-1}) P_{j,i-1}(x_{j,i-1} > t_{ji} - t_{j,i-1} \mid \underline{Y}_{j,i-1})\right) P_{jn_j}(x_{jn_j} \le U - t_{jn_j} \mid \underline{Y}_{jn_j}) \qquad [8.23]$$

where, the only modification from previous expressions is the final term which is the conditional cdf for the residual life (rather than the pdf) and represents the uncertainty in the actual residual life at the final monitoring point due to the lack of failure time information. The only information available to guide the estimation process is the artificial limit U established in expressions [8.15] and [8.16].

$$\mathcal{L}=\prod_{j=1}^{m}\left(\int_{0}^{U-t_{jn_j}}\alpha^{\beta}\beta(u+t_{jn_j})^{\beta-1}e^{-(\alpha(u+t_{jn_j}))^{\beta}}\prod_{i=1}^{n_j}\eta\left(\frac{C+Dt_{ji}}{(A+B(u+t_{jn_j}-t_{ji})^{-1})(t_{ji}-t_{j,i-1})}\right)^{\eta}\right.$$

$$\left.\ldots\times\ldots\Delta y_{ji}{}^{\eta-1}e^{-\left(\frac{(C+Dt_{ji})\Delta y_{ji}}{(A+B(u+t_{jn_j}-t_{ji})^{-1})(t_{ji}-t_{j,i-1})}\right)^{\eta}}du\right) \qquad [8.24]$$

Maximising equation [8.24] with respect to the unknown parameter set, we obtain the parameter estimates given in table 8.9. The variance about the mean estimate is also stated for each parameter. Note that, the estimates of α and β that parameterise the initial residual life distribution were given previously in table 8.5.

Parameter	Estimate	Variance
A	0.033	7.349×10^{-4}
B	2.441	7.023
C	0.015	1.851×10^{-4}
D	3.531×10^{-6}	2.771×10^{-11}
η	0.739	2.249×10^{-3}

Table 8.9 – The estimated parameters for filter 1

Optimisation of the likelihood function given by equation [8.24] can be quite complex and in some situations solutions may not be attainable. Further alternatives to the problem of parameter estimation for the stochastic filter include the use of the minimum mean square error (MMSE) criterion however, for this particular example the likelihood function given by equation [8.24] proved to be sufficient and the MMSE criterion was not required.

8.4.3 Comparing the models

With the availability of the functional form of the conditional distribution $p_{ji}(x_{ji} \mid \underline{Y}_{ji})$ at the ith CM point for the jth unit and a known failure time, we can analyse and compare the fit of the two models using the MSE criterion of chapters 6

and 7 and case 1 of this chapter. However, as discussed in length in the previous two sub-sections, the failure times are not available for this case and we are required to use the approximate range constructed for all the failure times. We are assuming that the failure time for unit j falls within the range $T_j \in (t_{jn_j}, U)$ where, t_{jn_j} is the final CM point for unit j and the upper limit U is defined by equation [8.15]. The criterion for model comparison employed here, involves the selection of the model that maximises the conditional probability of failure within the interval that starts immediately after the final CM point for the unit in question and ends at the approximate upper limit for all the units, U, as illustrated in figure 8.14.

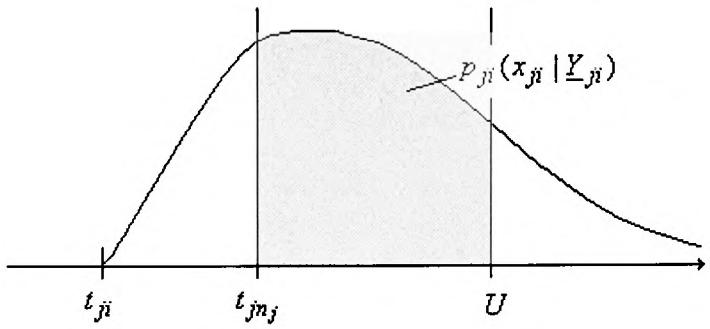

Figure 8.14 – Illustrating the failure interval of the conditional density

At the ith CM point for unit j, the probability of failure within the prescribed interval is given by

$$\xi_{ji} = P_{ji}(t_{jn_j} - t_{ji} < x_{ji} \leq U - t_{ji} \mid \underline{Y}_{ji}) = \int_{t_{jn_j} - t_{ji}}^{U - t_{ji}} p_{ji}(x_{ji} \mid \underline{Y}_{ji})\, dx_{ji} \qquad [8.25]$$

The measure of comparison is simply the summation of the probability mass that is available at each CM point for all the units considered divided by the total number of CM points. For a given model, we have

$$Total_{\xi}(Model) = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n_j}\xi_{ji}}{\sum_{j=1}^{m} n_j} \qquad [8.26]$$

and the most appropriate model for the particular case is then given by

$$\max\{Total_{\xi}(PHM),\ Total_{\xi}(filter)\} \qquad [8.27]$$

The advantage of this criterion is that the output from each model must be in the range $\xi_{ji} \in (0,1)$ at any given monitoring point. This means that the quality of the prediction is directly comparable across both the models that are under scrutiny and the various stages of the CM process for a given history. For most of the CM points considered, both models produced similar curve structures for the conditional residual life distribution. Figures 8.15 – 8.18 illustrate the conditional densities obtained using the filter and the PHM for units 5 and 16 from the new data set.

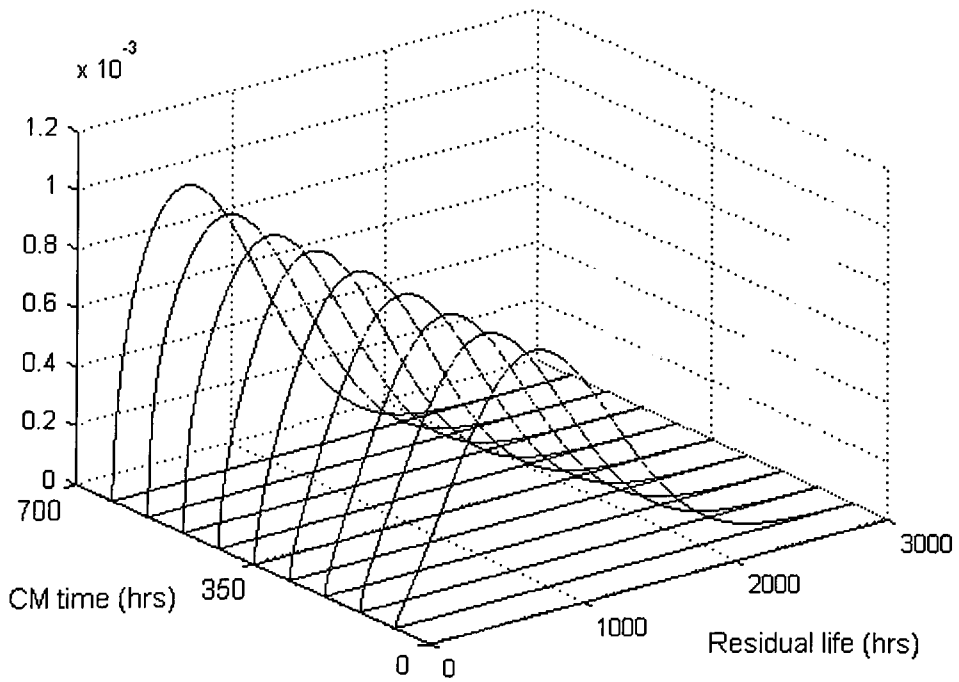

Figure 8.15 - Illustrating the conditional density obtained at each CM point for unit 5 using the stochastic filter

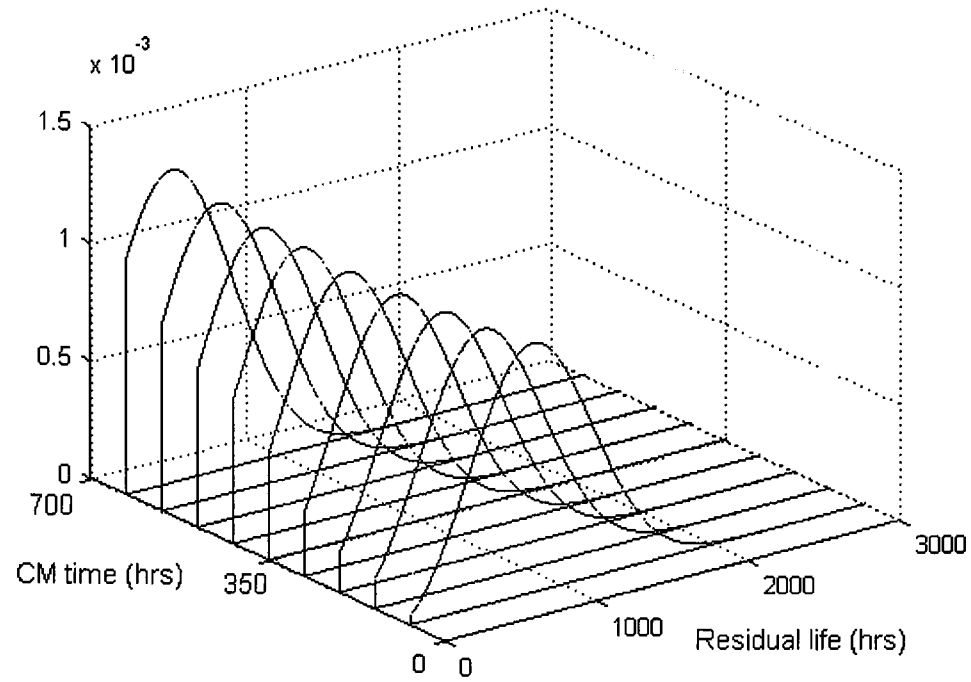

Figure 8.16 - Illustrating the conditional density obtained at each CM point for unit 5 using the PHM

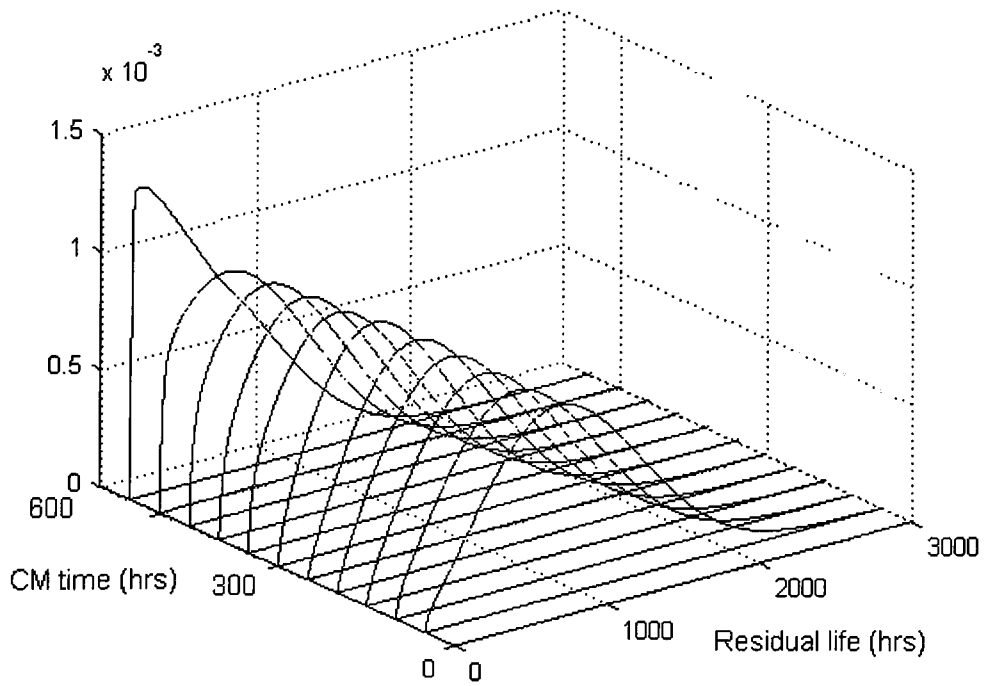

Figure 8.17 - Illustrating the conditional density obtained at each CM point for unit 16 using the stochastic filter

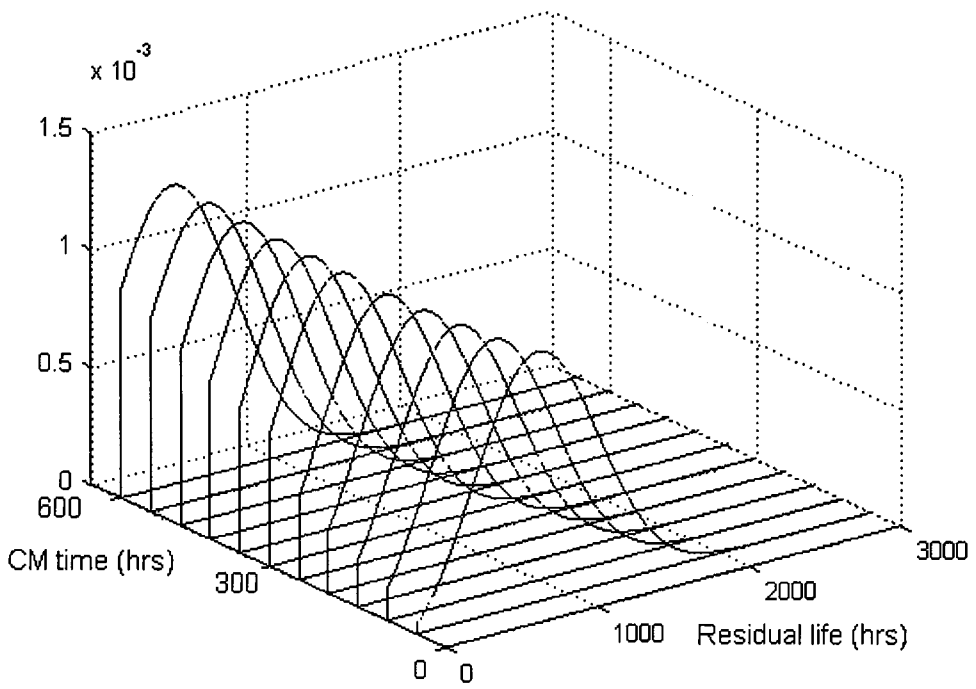

Figure 8.18 - Illustrating the conditional density obtained at each CM point for unit 16 using the stochastic filter

The filter and the PHM give very similar results for this second case in both the analysis of fit and when applying the models to new data. For the conditional residual life distribution developed using the filtering approach, the average probability mass falling within the desired range is 0.572. The PHM fairs slightly better with an average of 0.596. When applying the models to new data, the filter produces an average of 0.705 and the PHM produces 0.742. As with the first oil-based case, any differences between the models are negligible and as such, it is difficult to make any conclusive recommendations regarding model selection for this scenario.

8.5 Discussion

The objective of this chapter has been to demonstrate the use of the PHM and the stochastic filter for oil-based CM scenarios with multiple information parameters.

The first case study illustrates the application of PCA as a technique for reducing the dimension of the CM information making estimation of the model parameters more efficient. In a practical scenario, both models would ideally require a greater number of unit histories for parameterisation purposes however, the case study does indicate how the models might perform in a scenario where limited monitoring information is available. Two different filtering set up's are compared with the PHM and the filtering formulations produced marginally better results for the first case.

In the second case, multiple histories are available however, the data is both right and left-censored for some of the units and this produces problems for both model construction and testing. An alternative criterion is used to compare the models and the PHM produces slightly better results than those obtained using the filter. It is clear that the lack of exact failure time information has a bearing on the fitting of both models to the data with neither model providing a substantial improvement on results obtained using a standard Weibull survival analysis that does not utilise the CM information. Another reason for the lack of fit could be that there is not a sufficient level of correlation between the observed oil-based CM information and the underlying residual life of the components for the case considered.

In terms of further research into the two models discussed in this chapter, it would be useful to develop a means of incorporating more variability in the failure time input to the PHM process to provide more flexibility in the resulting model when limited data is available for construction. With regard to the stochastic filtering approach, it is clear that the model formulation provides flexibility in the construction and as such, an alternative set-up could be considered to represent the relationship between the monitored information and the underlying residual life.

Chapter 9. Further stochastic filtering options for condition-based maintenance applications

9.1 Introduction

In this chapter, we discuss some further stochastic filtering options for a discrete time CM observation process when the underlying system dynamics and relationships are non-linear. The techniques are proposed for situations when alternative approximations or sub-optimal approaches to the state estimation problem could prove useful. The most notable examples of this type of situation are as follows. Firstly, when an excessive number of components are under scrutiny and the necessary computational power is lacking to apply the probabilistic filter of chapters 6, 7 and 8. Secondly, when the relationship between the observed information and the underlying state is not known precisely (possibly due to multiple potential failure modes) or can change over time. Initially, we consider linearisations of the non-linear systems using Taylor expansions of the observation process to develop approximate filters for the residual life of a component. If appropriate distributional forms are selected and the relationship between the state and observation process is described effectively for a particular application, the probabilistic filtering framework used in chapters 6, 7 and 8 generally provides better estimates and is more flexible when compared with the techniques that we discuss in this chapter. This is due to the fact that there is no necessity for approximation when using the general non-linear probabilistic filter in either the treatment of the system and observation expressions or when establishing a closed form expression for the state estimation problem. However, it should be noted that there is a substantial amount of computation involved in recursively establishing the conditional density using the probabilistic filtering approach and numerical approximations are required to obtain

solutions. We then discuss the potential for combining the output given by a number of candidate models when the underlying dynamics change over time or are unknown for the particular component being monitored and as such, a single definitive representation of the system is not available. As with the state estimation techniques explored in the previous two chapters, the state of a component/unit, for the problems discussed in this chapter, is defined as the residual life x_i that remains before the component fails. At the ith CM point at time t_i, the information vector $\underline{y}_i$ becomes available to refine estimates of x_i and $\underline{Y}_i$ represents the CM history $\{\underline{y}_1, \underline{y}_2, ..., \underline{y}_i\}$ available until that point. The update expression for the residual life between subsequent CM points is

$$x_i = x_{i-1} - (t_i - t_{i-1}) \qquad [9.1]$$

for $x_{i-1} > t_i - t_{i-1}$ and is not defined otherwise. The relationship between $\underline{y}_i$ and x_i is given by

$$\underline{y}_i = h(x_i, t_i, \underline{e}_i) \qquad [9.2]$$

or is described using a probabilistic relationship as $p(\underline{y}_i \mid x_i)$ where $\underline{e}_i$ is the noise.

9.2 EKF's for CM applications with limited computational power

As discussed in chapter 2, there are a number of varieties of the extended Kalman filter (EKF) available in the literature on stochastic state estimation techniques. From a computational perspective, the Kaman filter for discrete time systems is a useful approach as it can be parameterised using just the first two moments. Efficient updating and prediction equations are easily established to obtain the parameters of the conditional distribution. Thus, it could potentially be a useful technique in a CBM context when a large number of components are monitored

simultaneously. The Kalman filter can be derived within the framework of the general non-linear filter used in chapters 6, 7 and 8 when the system and observation dynamics evolve linearly and the model errors are assumed to be independent and follow 0-mean Gaussian white noise processes. However, in reality, these assumptions rarely hold. As noted briefly in the introduction, EKF's are designed to enable the application of the standard Kalman filtering methodology to a linearised version of the non-linear system in question. The linearisarion is achieved using Taylor expansions of the state and observation equations where, a typical EKF utilises the first term in the series and (as the name implies) the 2nd-order EKF also uses the second term. In this section, we introduce a semi-deterministic form of the EKF and then apply the principles to a CBM application where, the deterministic element is designed to facilitate the exact relationship between realisations of the actual underlying residual life at different CM points throughout the life of the component. We then illustrate the application of the model with a simulation-based example before discussing the potential for extending the modelling principles by incorporating the 2nd-order terms of the Taylor expansions.

9.2.1 A semi-deterministic extended Kalman filter (EKF)

Initially, we discuss the semi-deterministic EKF for a general deterministic vector state before applying and adapting the technique for a residual life estimation problem using vibration based CM information. The same principles will apply when using oil-based information with the only difference being the specification of the relationship between the current CM reading and the underlying residual life. The evolution of a general state vector $\underline{x}_i$ is described by the non-linear function

$$\underline{x}_{i+1} = f(\underline{x}_i) + \underline{v}_i \qquad [9.3]$$

for a discrete time process where, for a deterministic relationship, $\underline{v}_i$ is a 0- mean process with covariance matrix $\underline{0}$ and is henceforth removed from consideration. The relationship between the observed information vector and the underlying state at the ith discrete time point is described by the non-linear function

$$\underline{y}_i = h(\underline{x}_i) + \underline{e}_i \qquad [9.4]$$

where, the measurement errors are normally distributed as $\underline{e}_i \sim N(\underline{0}, \underline{R}_i)$. Defining $\hat{\underline{x}}_{i|i} = E(\underline{x}_i \mid \underline{Y}_i)$ as the estimate of $\underline{x}_i$ at time t_i and $\hat{\underline{x}}_{i+1|i} = E(\underline{x}_{i+1} \mid \underline{Y}_i)$ as the one-step prediction of $\underline{x}_{i+1}$ at time t_i, the non-linear functions f and h are linearised as

$$f(\underline{x}_i) \approx f(\hat{\underline{x}}_{i|i}) + f'(\hat{\underline{x}}_{i|i})(\underline{x}_i - \hat{\underline{x}}_{i|i}) \qquad [9.5]$$

$$h(\underline{x}_i) \approx h(\hat{\underline{x}}_{i|i-1}) + h'(\hat{\underline{x}}_{i|i-1})(\underline{x}_i - \hat{\underline{x}}_{i|i-1}) \qquad [9.6]$$

Using these approximations, the state transition expression becomes

$$\underline{x}_{i+1} = f'(\hat{\underline{x}}_{i|i})\underline{x}_i + \underline{u}_i \qquad [9.7]$$

where, $\underline{u}_i = f(\hat{\underline{x}}_{i|i}) - f'(\hat{\underline{x}}_{i|i})\hat{\underline{x}}_{i|i}$. Similarly, the relationship between the observed information and the underlying state becomes

$$\underline{y}_i = h'(\hat{\underline{x}}_{i|i-1})\underline{x}_i + \underline{e}_i + \underline{w}_i \qquad [9.8]$$

where, $\underline{w}_i = h(\hat{\underline{x}}_{i|i-1}) - h'(\hat{\underline{x}}_{i|i-1})\hat{\underline{x}}_{i|i-1}$. Applying the Kalman filtering process to the linearised system, the equation for updating the mean estimate of the state at the ith recursion of the filtering process is

$$\begin{aligned} \hat{\underline{x}}_{i|i} &= \hat{\underline{x}}_{i|i-1} + k_i[\underline{y}_i - h'(\hat{\underline{x}}_{i|i-1})\hat{\underline{x}}_{i|i-1} - \underline{w}_i] \\ &= \hat{\underline{x}}_{i|i-1} + k_i[\underline{y}_i - h(\hat{\underline{x}}_{i|i-1})] \end{aligned} \qquad [9.9]$$

where, the gain function is

$$k_i = \underline{P}_{i|i-1}h'(\hat{\underline{x}}_{i|i-1})^T[h'(\hat{\underline{x}}_{i|i-1})\underline{P}_{i|i-1}h'(\hat{\underline{x}}_{i|i-1})^T + \underline{R}_i]^{-1} \qquad [9.10]$$

For a semi-deterministic version of the EKF, the one-step forecast for the mean state vector is simply

$$\underline{\hat{x}}_{i+1|i} = f'(\underline{\hat{x}}_{i|i})\underline{\hat{x}}_{i|i} + \underline{\tilde{u}}_i = f(\underline{\hat{x}}_{i|i}) \qquad [9.11]$$

using the original transition expression given by equation [9.3]. The covariance matrix is updated using

$$\underline{P}_{i|i} = \underline{P}_{i|i-1} - \underline{P}_{i|i-1}h'(\underline{\hat{x}}_{i|i-1})^T[h'(\underline{\hat{x}}_{i|i-1})\underline{P}_{i|i-1}h'(\underline{\hat{x}}_{i|i-1})^T + \underline{R}_i]^{-1}h'(\underline{\hat{x}}_{i|i-1})\underline{P}_{i|i-1} \qquad [9.12]$$

that can be written as

$$\underline{P}_{i|i} = \underline{P}_{i|i-1} - k_i\, h'(\underline{\hat{x}}_{i|i-1})\underline{P}_{i|i-1} \qquad [9.13]$$

using the gain function given by equation [9.10]. Finally, the covariance is predicted using

$$\underline{P}_{i+1|i} = f'(\underline{\hat{x}}_{i|i})\underline{P}_{i|i}f'(\underline{\hat{x}}_{i|i})^T \qquad [9.14]$$

This concludes the description of the semi-deterministic EKF algorithm for general discrete time state-vector and observation-vector processes. We now consider the application of the methodology to CBM applications using vibration information. As noted earlier, the application of the methodology would be identical when considering oil-based CM information with the only difference being the form of the function *h*.

9.2.2 A semi-deterministic EKF for residual life prediction using vibration monitoring information

We consider a single information parameter in the form of the overall vibration level of an individual monitored component. Modelling the transition in the underlying state between two successive monitoring points, from equation [9.1], we have

$$x_{i+1} = x_i - (t_{i+1} - t_i)$$

as the change in the residual life over the duration between the ith and (i+1)th CM point when we have $x_i > t_{i+1} - t_i$. The relationship between the observed vibration parameter and the underlying residual life is described by the expression

$$y_i = a + be^{-cx_i} + e_i \qquad [9.15]$$

at the ith monitoring point where, e_i represents the measurement noise. A one-step prediction of the mean residual life is achieved using the current estimate $\hat{x}_{i|i}$ and the deterministic relationship given by equation [9.1] as

$$\hat{x}_{i+1|i} = \hat{x}_{i|i} - (t_{i+1} - t_i) \qquad [9.16]$$

and due to the lack of uncertainty in equation [9.1], and the fact that $f'(\underline{\hat{x}}_{i|i}) = 1$ for all i, the variance about the mean estimate remains as $P_{i+1|i} = P_{i|i}$ until further information is obtained in the form of y_{i+1}. Upon observing y_i, the mean estimate of the underlying residual life is updated as

$$\hat{x}_{i|i} = \hat{x}_{i|i-1} + k_i\left(y_i - a - be^{-c\hat{x}_{i|i-1}}\right) \qquad [9.17]$$

where, using equation [9.10], the gain function is

$$k_i = \frac{P_{i|i-1}h'(\hat{x}_{i|i-1})}{h'(\hat{x}_{i|i-1})^2 P_{i|i-1} + R_i} \qquad [9.18]$$

for $h'(\hat{x}_{i|i-1}) = -bce^{-c\hat{x}_{i|i-1}}$. The variance is updated using equation [9.13] as

$$P_{i|i} = P_{i|i-1} - k_i h'(\hat{x}_{i|i-1})P_{i|i-1} \qquad [9.19]$$

9.2.3 EKF for residual life prediction using vibration information - example

In this example, we demonstrate the application of a predetermined semi-deterministic EKF to a vibration monitoring scenario using simulated data. Equation [9.15] is parameterised using a = 5, b = 20 and c = 0.012 and we assume that the parameters have been estimated using the CM histories of analogous components.

The error terms are random variables that are distributed as $e_i \sim N(0, 0.016y_i^2)$ and the initial residual life distribution is taken to be $x_0 \sim N(200, 2000)$. Note that, we are only interested in modelling the second stage of component monitoring and the mean value of 200 is the expected residual life upon the initiation of a defect. As with the modelling of vibration information undertaken in chapter 7, we assume that a technique such as the statistical process control approach has been used to determine the origin of the defective operational stage. The form of the variance function for e_i is chosen to reflect the fact that in most vibration monitoring scenarios, the random variation increases as the vibration level increases.

Three 2nd-stage CM histories are simulated using the following steps;

1. The duration of the lifetime of the particular component is simulated first. This is achieved using inversion on the initial density by the solution of

$$u = \int_0^{x_0} p_0(s)ds \qquad [9.20]$$

for x_0, where, u is a uniform random variable on the range (0, 1).

2. At equidistant intervals, we simulate the vibration reading conditioned on the underlying residual life at that CM point. We have

$$y_i = E(y_i \mid x_i = x_0 - t_i) + e_i = h(t_i, x_0 - t_i) + e_i \qquad [9.21]$$

where, e_i is also simulated using inversion on its probability density function.

The simulated histories are illustrated in figure 9.1.

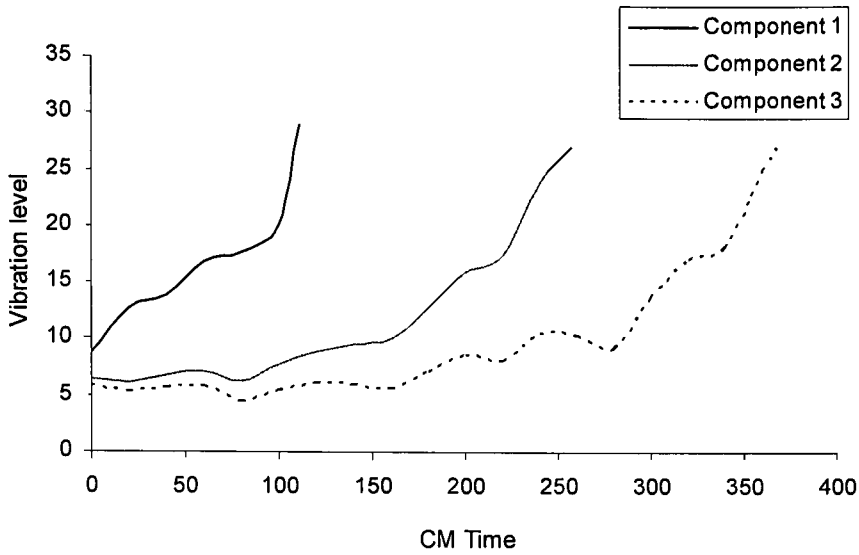

Figure 9.1 – The simulated CM histories

We now demonstrate the ability of the model to track the underlying residual life using the age of the component and the monitoring information for the three cases. The third example is included to illustrate how the model copes with outlying cases. Figures 9.2 – 9.4 illustrate the mean estimate of the residual life at each monitoring point for component histories 1 – 3 respectively.

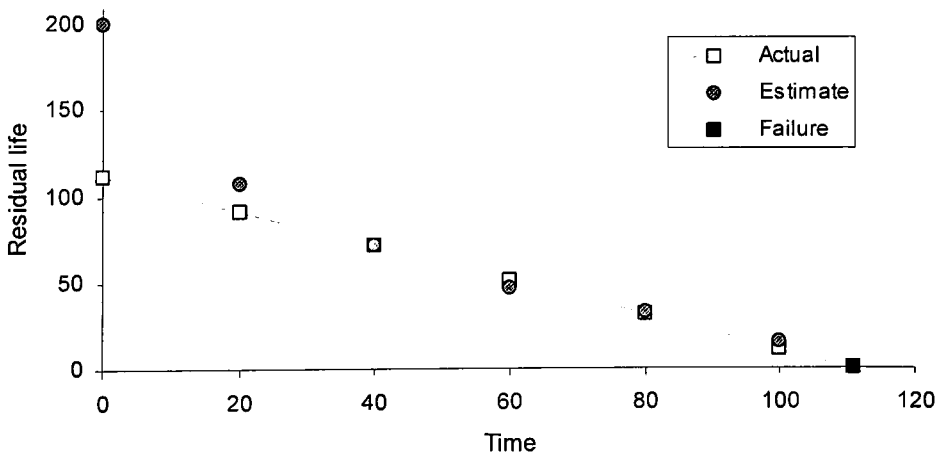

Figure 9.2 – Comparing the mean estimate and the actual underlying residual life at each monitoring point in history 1

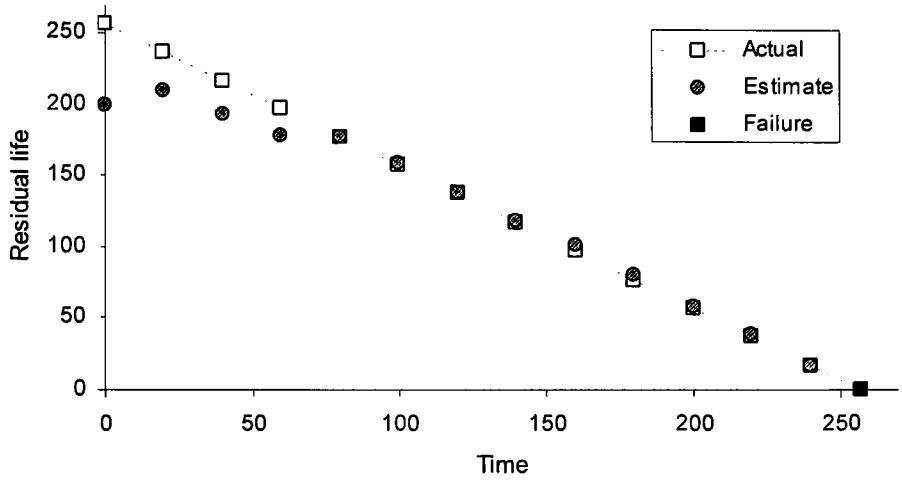

Figure 9.3 – Comparing the mean estimate and the actual underlying residual life at each monitoring point in history 2

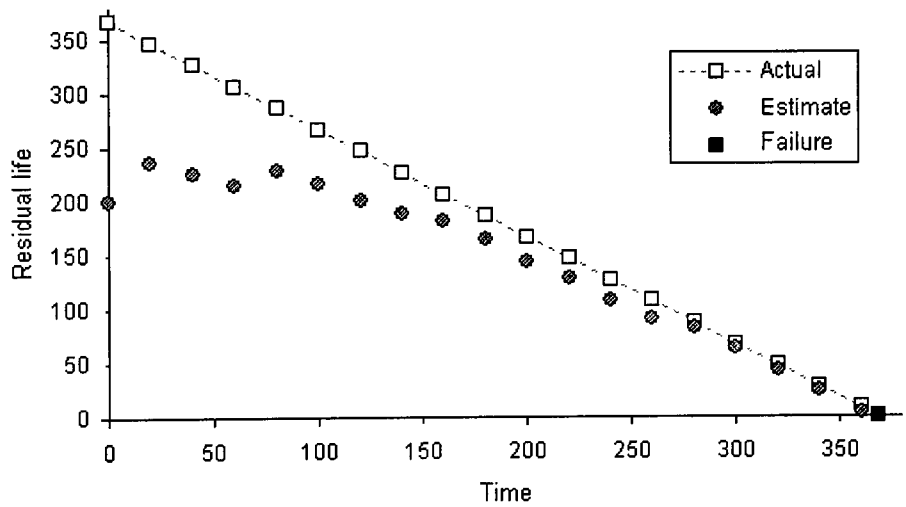

Figure 9.4 – Comparing the mean estimate and the actual underlying residual life at each monitoring point in history 3

As we discussed at length in chapters 6, 7 and 8, a point estimate is not as useful as the definition of the conditional distribution in the construction of reliable decision models. If a suitable model has been defined, we would expect the variance about the mean to decrease over time as more CM information is obtained. Figures 9.5 –

9.7 illustrate the conditional residual life distribution produced at each CM point for vibration histories 1 – 3 respectively.

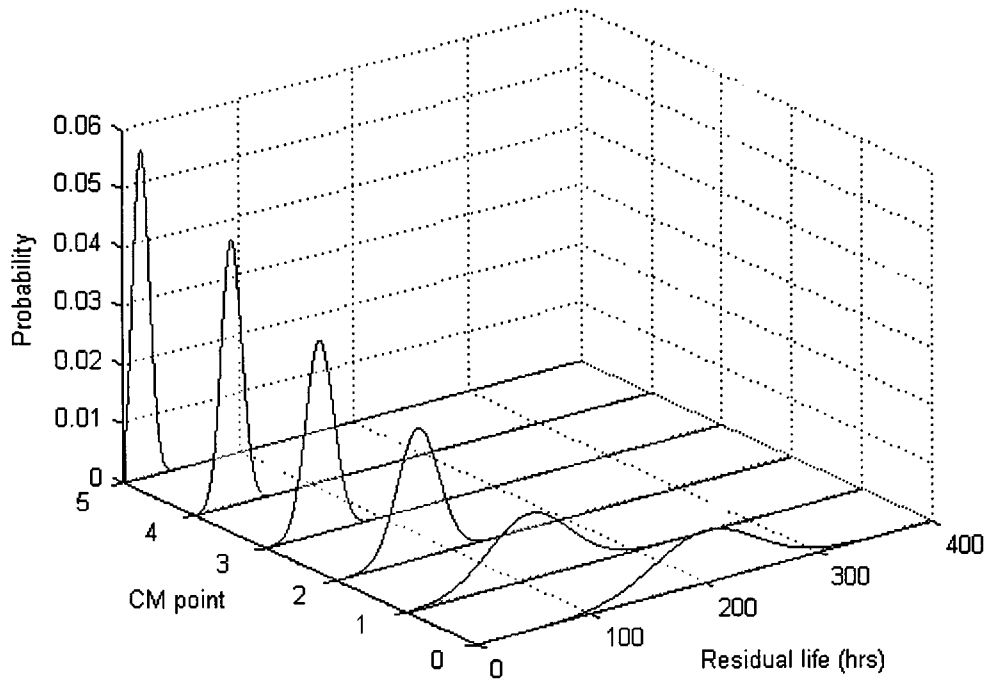

Figure 9.5 – Illustrating the conditional distribution obtained at each CM point for history 1

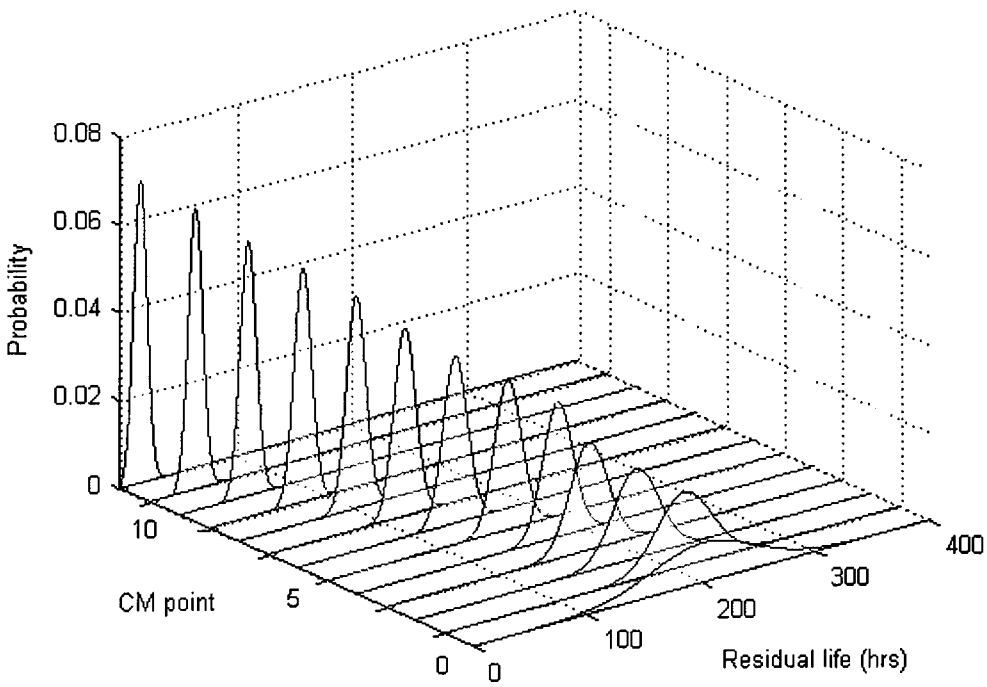

Figure 9.6 – Illustrating the conditional distribution obtained at each CM point for history 2

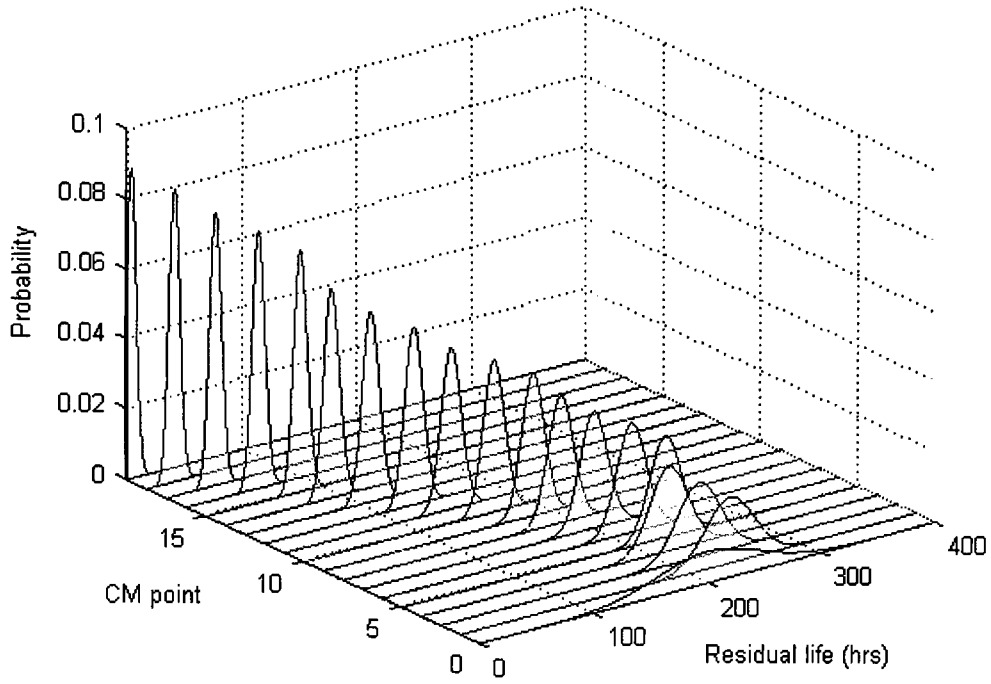

Figure 9.7 – Illustrating the conditional distribution obtained at each CM point for history 3

It is clear from the various figures that the semi-deterministic EKF performs well for this particular example. Figures 9.2 – 9.4 demonstrate that the model tracks the underlying residual life quite rapidly for all the simulated histories considered and figures 9.5 – 9.7 illustrate the reduction in the variance about the mean estimate as more information is received. Naturally, the level of convergence is dependent on the accuracy of the proposed model for the particular case. In this case, the accuracy is reflected in the specification of the variance parameter, R_i, that is selected for this example in accordance with the analogous vibration monitoring histories used in the case studies of chapter 7.

9.2.4 Second-order extended Kalman filtering

In this section, we consider the second order terms in the Taylor series expansion of the non-linear equations that describe the dynamics of the system. In some situations, the extension proposed here will produce greater accuracy when applied

to problems involving the estimation of the residual life of a component. Any improvement in accuracy will be dependent on the manner in which the relationship between the observed information and the underlying residual life is defined. Returning to the definition of a general vector state, the evolution of the state and the relationship between the observed information and the underlying state are given by equations [9.3] and [9.4] respectively. Incorporating the second order terms, the non-linear functions f and h are approximated as

$$f(\underline{x}_i) \approx f(\hat{\underline{x}}_{i|i}) + f'(\hat{\underline{x}}_{i|i})(\underline{x}_i - \hat{\underline{x}}_{i|i}) + \bar{\underline{f}}_i \qquad [9.22]$$

$$h(\underline{x}_i) \approx h(\hat{\underline{x}}_{i|i-1}) + h'(\hat{\underline{x}}_{i|i-1})(\underline{x}_i - \hat{\underline{x}}_{i|i-1}) + \bar{\underline{h}}_i \qquad [9.23]$$

where, $f'(\hat{\underline{x}}_{i|i})$ and $h'(\hat{\underline{x}}_{i|i-1})$ are as before and $\bar{\underline{f}}_i$ is a vector with elements

$$\bar{f}_{ij} = \frac{1}{2}(\underline{x}_i - \hat{\underline{x}}_{i|i})^T \left.\frac{\partial^2 f_j}{\partial \underline{x}^2}\right|_{\underline{x}=\hat{\underline{x}}_{i|i}} (\underline{x}_i - \hat{\underline{x}}_{i|i}) \qquad [9.24]$$

that can be approximated as

$$\bar{f}_{ij} \approx \frac{1}{2} tr\left[\underline{P}_{i|i} \left.\frac{\partial^2 f_j}{\partial \underline{x}^2}\right|_{\underline{x}=\hat{\underline{x}}_{i|i}}\right] \qquad [9.25]$$

Similarly, $\bar{\underline{h}}_i$ is a vector with elements

$$\bar{h}_{ij} = \frac{1}{2}(\underline{x}_i - \hat{\underline{x}}_{i|i-1})^T \left.\frac{\partial^2 h_j}{\partial \underline{x}^2}\right|_{\underline{x}=\hat{\underline{x}}_{i|i-1}} (\underline{x}_i - \hat{\underline{x}}_{i|i-1}) \approx \frac{1}{2} tr\left[\underline{P}_{i|i-1} \left.\frac{\partial^2 h_j}{\partial \underline{x}^2}\right|_{\underline{x}=\hat{\underline{x}}_{i|i-1}}\right] \qquad [9.26]$$

The linearised system is given by equations [9.7] and [9.8] where, the compensators $\underline{u}_i$ and $\underline{w}_i$ become

$$\underline{u}_i = f(\hat{\underline{x}}_{i|i}) - f'(\hat{\underline{x}}_{i|i})\hat{\underline{x}}_{i|i} + \bar{\underline{f}}_i \qquad [9.27]$$

$$\underline{w}_i = h(\hat{\underline{x}}_{i|i-1}) - h'(\hat{\underline{x}}_{i|i-1})\hat{\underline{x}}_{i|i-1} + \bar{\underline{h}}_i \quad [9.28]$$

The filtering update and prediction equations for the general state 2nd- order semi-deterministic EKF are

$$\hat{\underline{x}}_{i|i} = \hat{\underline{x}}_{i|i-1} + k_i[\underline{y}_i - h(\hat{\underline{x}}_{i|i-1}) - \bar{\underline{h}}_i] \quad [9.29]$$

$$\hat{\underline{x}}_{i+1|i} = f(\hat{\underline{x}}_{i|i}) + \bar{\underline{f}}_i \quad [9.30]$$

The remaining equations for the gain function, equation [9.10], and the covariance matrix, equations [9.12] - [9.14], are unchanged from the 1st-order definition. This concludes the description of the algorithm. As noted in the introduction to this section, the necessity for utilising a 2nd-order version of the semi-deterministic EKF is dependent on the particular definition of the relationship between the observed information and the underlying residual life. When considering the vibration monitoring scenario modelled in section 9.2.2, the inclusion of 2nd-order terms is not necessary and the application of the model to the example of section 9.2.3 did not produce a tangible improvement on the results obtained using the standard algorithm. As such, the analysis is omitted here. However, there are many other means of condition monitoring and indeed other ways of modelling the relationship between vibration information and residual life for which, the 2nd-order semi-deterministic EKF could prove to be a useful approach.

9.3 Limited memory filter

If the dynamics of a particular state and observation process over time are not known to a satisfactory degree of precision or are particularly changeable (e.g. multiple potential failure modes), it may be useful to reduce the impact of earlier observations when estimating the underlying state. Each observation refines the estimation or filtering process and if subsequent observations maintain the expected path defined

by previous observations, the variance about the estimate of the residual life decreases (i.e. the variance of the distribution $p_i(x_i \mid \underline{Y}_i)$ is reduced). It is therefore a logical conclusion that, if the dynamics of the process are changeable as the component ages, it may take a number of observations in a new direction for the filtering process to adapt given the fact that earlier observations still carry weight within the filtering formulation. One method of reducing this estimation lag is to remove the impact of these 'earlier' observations. The objective of the limited memory filtering process is to define the residual life as conditioned on a limited number of observations up until the current CM readings. This random variable is defined as $x_i \mid \underline{y}_{k+1},\ldots,\underline{y}_i$, for $k < i$, and any information obtained before t_{k+1} is removed from the estimation process by the definition of the conditional density for the random variable as

$$p(x_i \mid \underline{y}_{k+1},\ldots,\underline{y}_i) = \frac{p_i(x_i \mid \underline{Y}_i)p(\underline{Y}_i)p(x_i)}{p(x_i \mid \underline{Y}_k)p(\underline{Y}_k)p(\underline{y}_{k+1},\ldots,\underline{y}_i)} \qquad [9.31]$$

where, the constituent elements of the reduced memory filter are firstly

$$p(x_i \mid \underline{Y}_k) = \frac{p_k(x_i + t_i - t_k \mid \underline{Y}_k)}{\int_{t_i - t_k}^{\infty} p_k(u \mid \underline{Y}_k)du} \qquad [9.32]$$

Secondly, we have

$$p(\underline{y}_{k+1},\ldots,\underline{y}_i) = \int_0^{\infty} p(\underline{y}_{k+1},\ldots,\underline{y}_i \mid x_i)p(x_i)\,dx_i \qquad [9.33]$$

where, the conditional density is established as

$$p(\underline{y}_{k+1},\ldots,\underline{y}_i \mid x_i) = \frac{p_i(x_i \mid \underline{Y}_i)p(\underline{Y}_i)}{p(x_i \mid \underline{Y}_k)p(\underline{Y}_k)} \qquad [9.34]$$

Finally, we have the probability of observing the history $\underline{Y}_i$ as

$$p(\underline{Y}_i) = \prod_{j=1}^{i} p(\underline{y}_j \mid \underline{Y}_{j-1}) = \prod_{j=1}^{i} \int_0^{\infty} p(\underline{y}_j \mid x_j) p(x_j \mid \underline{Y}_{j-1}) dx_j \qquad [9.35]$$

which is the product of the individual probability of observing each information vector conditioned on the previous observations. Determination of an appropriate lag k would also have to be undertaken in a practical scenario and would prove particularly difficult with a monitoring process undertaken at irregular intervals.

9.4 Model combinations

9.4.1 Introduction

The methodology proposed in this section essentially consists of running a number of probabilistic stochastic filters in parallel and defining an estimate of the residual life of a component as a weighted combination of their respective output. Firstly, we consider a situation where, the underlying dynamics are fixed and we assume that they conform to one of the proposed models for the case. This model is for use when the behaviour can correspond to a number of distinct behavioural types, we are simply unaware which type the current component conforms to. An example considered later in this chapter involves the modelling and estimation of the residual life of a component when the behaviour can correspond to one of two potential failure modes. The behaviour is assumed to manifest itself in the form of failure time clustering as demonstrated in figure 9.8. Separate models are established for each scenario and a recursive procedure is developed to determine, during the life of a component, which model the underlying dynamics conform to using both the age and the available CM history.

We then consider the potential for the dynamics to evolve or fluctuate during the life of a component. We assume that that, at any given stage, the dynamics conform to one of the proposed models and that unknown transitions between models occur over

time. The transition probabilities must be estimated from available data and are modelled using a Markov chain.

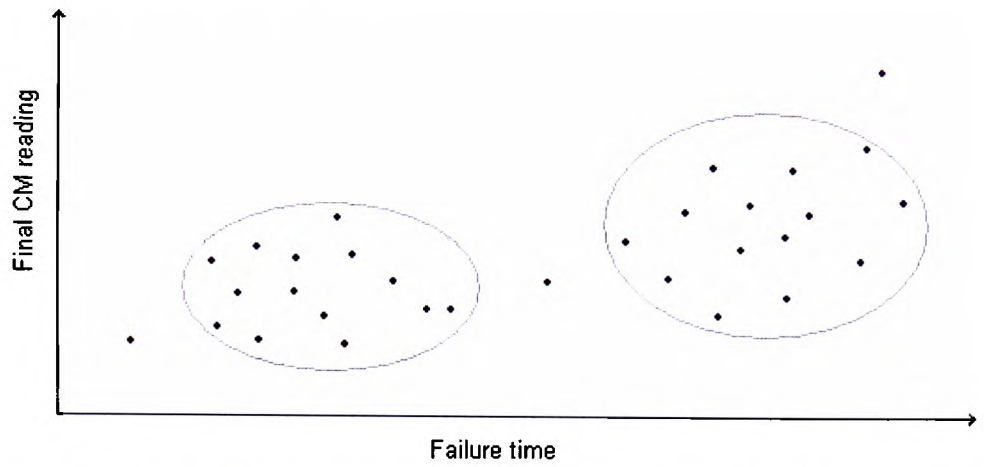

Figure 9.8 - Illustrating the clustering of failure times when two different failure modes exist

9.4.2 Fixed dynamics

We define r different models, each pertaining to an individual and distinct failure type. The notation M_j represents model j (for j = 1, 2, ..., r) and the models are assumed to be parameterised using only available component histories that are relevant to the respective failure modes. The prior probability that the underlying dynamics of the CM process for a given component will correspond to model j is denoted as $p(M_j \mid \underline{Y}_0)$ and is assumed known. Considering multiple indicators of condition obtained simultaneously at each discrete monitoring point (denoted by $\underline{y}_i$) at time t_i, we have

$$p(M_j \mid \underline{Y}_i) = p(M_j \mid \underline{y}_i, \underline{Y}_{i-1}) \qquad [9.36]$$

as the conditional probability that the underlying dynamics of the current CM process correspond to model j given the monitoring history available until that point in time. By the application of Bayes' law we obtain

$$p(M_j \mid \underline{y}_i, \underline{Y}_{i-1}) = \frac{p(\underline{y}_i \mid \underline{Y}_{i-1}, M_j) p(M_j \mid \underline{Y}_{i-1})}{p(\underline{y}_i \mid \underline{Y}_{i-1})} \qquad [9.37]$$

where, the initial probability $p(M_j \mid \underline{Y}_0)$ is assumed to be known and $p(M_j \mid \underline{Y}_{i-1})$ represents the probability that the underlying dynamics conform to model j from the previous recursion of the process. This is the means by which our best judgement regarding the actual dynamics (and hence the underlying residual life of the unit) is updated at each monitoring point. We also have

$$p(\underline{y}_i \mid \underline{Y}_{i-1}, M_j) = \int_0^\infty p(\underline{y}_i \mid x_i, M_j) p(x_i \mid \underline{Y}_{i-1}, M_j) dx_i \qquad [9.38]$$

on the assumption that $p(\underline{y}_i \mid x_i, \underline{Y}_{i-1}, M_j) = p(\underline{y}_i \mid x_i, M_j)$, i.e. $\underline{y}_i$ is controlled by x_i and M_j only. The denominator of equation [9.37] is obtained by enumerating over all the possible scenarios as

$$p(\underline{y}_i \mid \underline{Y}_{i-1}) = \sum_{k=1}^{r} p(\underline{y}_i \mid \underline{Y}_{i-1}, M_k) p(M_k \mid \underline{Y}_{i-1}) \qquad [9.39]$$

In the linear case we are able to establish a conditional density for the residual life that is fully parameterised using the weighted first two moments attributable to each model. In the general non-linear case, we are restricted to obtaining a weighted mean point estimate of the residual life as

$$\hat{x}_i = E(x_i \mid \underline{Y}_i) = \sum_{j=1}^{r} \int_0^\infty x_i p(x_i \mid \underline{Y}_i, M_j) p(M_j \mid \underline{Y}_i) dx_i \qquad [9.40]$$

9.4.3 Evolving dynamics

When the underlying dynamics of a given unit are assumed to vary over time as the unit ages, we introduce a time-invariant Markov chain with transition probabilities

$$a_{kj} = p(M_i = M_j \mid M_{i-1} = M_k) = p(M_{ji} \mid M_{k,i-1}) \qquad [9.41]$$

that corresponds to the conditional probability that the underlying dynamics currently conform to model j at the ith monitoring point given that they conformed to model k at the previous monitoring point. The objective of the combined modelling approach with evolving dynamics is to establish the conditional distribution

$$p_i(x_i \mid \underline{Y}_i) = \sum_{j=1}^{r} p(x_i \mid M_{ji}, \underline{Y}_i) p(M_{ji} \mid \underline{Y}_i) \qquad [9.42]$$

where, M_{ji} is representative of the fact that the underlying dynamics conform to model j at the ith monitoring point. Both terms in equation [9.42] require some explanation. The first is established as

$$p(x_i \mid M_{ji}, \underline{Y}_i) = p(x_i \mid M_{ji}, \underline{y}_i, \underline{Y}_{i-1}) = \frac{p(\underline{y}_i \mid x_i, M_{ji}) p(x_i \mid M_{ji}, \underline{Y}_{i-1})}{p(y_i \mid M_{ji}, \underline{Y}_{i-1})} \qquad [9.43]$$

where, the probabilistic relationship $p(\underline{y}_i \mid x_i, M_{ji})$ is available from the model specification and we have

$$p(x_i \mid M_{ji}, \underline{Y}_{i-1}) = \sum_{k=1}^{r} p(x_i \mid M_{ji}, M_{k,i-1}, \underline{Y}_{i-1}) p(M_{k,i-1} \mid M_{ji}, \underline{Y}_{i-1}) \qquad [9.44]$$

where, in this context, $p(x_i \mid M_{ji}, M_{k,i-1}, \underline{Y}_{i-1}) = p(x_i \mid M_{k,i-1}, \underline{Y}_{i-1})$, as the one step prediction of x_i is available from the previous recursion and is not dependent on the current model given the lack of reliance on $\underline{y}_i$. We also have the reverse transition expression

$$p(M_{k,i-1} \mid M_{ji}, \underline{Y}_{i-1}) = \frac{p(M_{ji} \mid M_{k,i-1}, \underline{Y}_{i-1}) p(M_{k,i-1} \mid \underline{Y}_{i-1})}{p(M_{ji} \mid \underline{Y}_{i-1})} = \frac{a_{kj} p(M_{k,i-1} \mid \underline{Y}_{i-1})}{\sum_{k=1}^{r} a_{kj} p(M_{k,i-1} \mid \underline{Y}_{i-1})} \qquad [9.45]$$

and the denominator of equation [9.43] is established as

$$p(y_i \mid M_{ji}, \underline{Y}_{i-1}) = \int_0^{\infty} p(\underline{y}_i \mid x_i, M_{ji}) p(x_i \mid M_{ji}, \underline{Y}_{i-1}) dx_i \qquad [9.46]$$

Now, we consider the second term of equation [9.42]. Assuming that the initial probability that the underlying dynamics (at the start of the CM process for a new component) correspond to model j, $p(M_{j0} \mid \underline{Y}_0)$, is known, we again employ Bayes' theorem to recursively obtain

$$p(M_{ji} \mid \underline{Y}_i) = p(M_{ji} \mid \underline{y}_i, \underline{Y}_{i-1}) = \frac{p(\underline{y}_i \mid \underline{Y}_{i-1}, M_{ji}) p(M_{ji} \mid \underline{Y}_{i-1})}{p(\underline{y}_i \mid \underline{Y}_{i-1})} \quad [9.47]$$

where, the constituent elements of the numerator are

$$p(\underline{y}_i \mid \underline{Y}_{i-1}, M_{ji}) = \int_0^\infty p(\underline{y}_i \mid x_i, M_{ji}) p(x_i \mid \underline{Y}_{i-1}, M_{ji}) dx_i \quad [9.48]$$

and

$$p(M_{ji} \mid \underline{Y}_{i-1}) = \sum_{k=1}^{r} p(M_{ji} \mid M_{k,i-1}) p(M_{k,i-1} \mid \underline{Y}_{i-1}) = \sum_{k=1}^{r} a_{kj}\, p(M_{k,i-1} \mid \underline{Y}_{i-1}) \quad [9.49]$$

The denominator is given by enumerating over the prediction available from all the potential models as

$$p(\underline{y}_i \mid \underline{Y}_{i-1}) = \sum_{k=1}^{r} p(\underline{y}_i \mid \underline{Y}_{i-1}, M_{ki}) p(M_{ki} \mid \underline{Y}_{i-1}) \quad [9.50]$$

In the following example, we return to the modelling of a component life with fixed, but unknown, underlying dynamics. The combined model with evolving dynamics is only presented here in a theoretical form. An example is not included due to time limitations.

9.4.4 Example – fixed dynamics

In this example, we consider the modelling and estimation of the residual life of a component using vibration information when two potential failure modes are assumed to have been identified from relevant data in a scenario similar to that illustrated in figure 9.8. When the monitoring process commences for a new

component, the underlying dynamics are fixed but unknown. We develop two separate stochastic filters (model's 1 and 2) to represent each potential eventuality. The models are developed using the same functional forms but are parameterised independently using relevant analogous component histories. The models are conducted in parallel and their respective output weighted according to the probability that the underlying dynamics correspond to each model. In this example, we simulate a cycle of data according to each modelling formulation and investigate the ability of the prescribed methodology to track the appropriate underlying model and the residual life of the component. The estimate of the residual life at each monitoring point is compared with a general model (model 3) that is developed and parameterised using all the available monitoring information, i.e. the histories are not classified according to any 'failure type' and are all grouped together for parameter estimation purposes. This is achieved by simulating a large number of cycles of CM data corresponding to each of models 1 and 2 and parameterising a general model using all the simulated output. We then compare the weighted output from models 1 and 2 with the output from model 3 to ascertain the benefit of the combined modelling approach for this particular scenario.

The filtering expression for model j is

$$p_{ji}(x_i \mid \underline{Y}_i, M_j) = \frac{p(y_i \mid x_i, M_j) p_{j,i-1}(x_i + t_i - t_{i-1} \mid \underline{Y}_{i-1}, M_j)}{\int_0^\infty p(y_i \mid x_i, M_j) p_{j,i-1}(x_i + t_i - t_{i-1} \mid \underline{Y}_{i-1}, M_j) dx_i} \qquad [9.51]$$

for $j = 1, 2, 3$. The constituent elements of model j are the initial residual life distribution

$$p_{j0}(x_0) = \frac{\alpha_j (\alpha_j x_0)^{\beta_j - 1} e^{-\alpha_j x_0}}{\Gamma(\beta_j)} \qquad [9.52]$$

which is defined as a Gamma distribution for each model but parameterised independently. Similarly, the distribution governing the conditional relationship between the observed vibration reading and the underlying residual life is taken to be Gaussian for all the models as

$$p(y_i \mid x_i, M_j) = \frac{e^{-\frac{1}{2}\left(\frac{y_i-\mu_{ji}}{\sigma_{ji}}\right)^2}}{\sigma_{ji}\sqrt{2\pi}} \qquad [9.53]$$

where, for model j (M_j), we have $\mu_{ji} = A_j + B_j e^{-C_j x_i}$ as the expected vibration level at the ith monitoring point given a particular realisation of the underlying residual life and, analogous to the example of section 9.2.3, the standard deviation parameter is proportional to the vibration level as $\sigma_{ji} = d_j y_i$. The parameters of models 1 and 2 are specified in table 9.1 where, $\bar{x}_0$ is the average life of a component under each scenario.

Parameter	Model 1	Model 2
A	5	5
B	17.3	21
C	0.025	0.01
d	0.126	0.141
$\bar{x}_0$	200	640
α	0.218	0.115
β	44.205	74.504

Table 9.1 – The parameters of models 1 and 2

The expected CM paths for the average life corresponding to model formulations 1 and 2 are illustrated in figure 9.9. The general model (model 3) is constructed with the same forms as models 1 and 2, given by equations [9.52] and [9.53], and the parameters are estimated using 100 simulated histories. 50 of the histories are generated according to model 1 and 50 according to model 2. The reasoning for this

is that, for simplicity and to demonstrate the methodology, we develop a scenario in which both contingencies are equally likely, i.e. before the monitoring process begins, we have the initial probabilities $p(M_1) = p(M_2) = 0.5$.

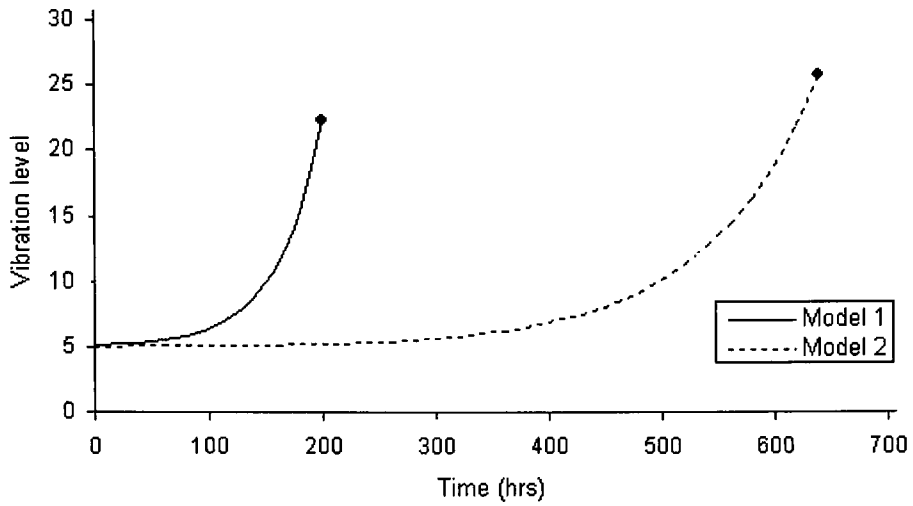

Figure 9.9 – Illustrating the expected CM path for each model

Using the simulated CM histories from models 1 and 2, the estimated parameters of the general model (model 3) are given in table 9.2 below.

Parameter	General Model
A	5.482
B	17.702
C	0.02
d	0.195
α	0.00778
β	3.266

Table 9.2 – The parameters of the general model (model 3)

Using the filtering expression given by equation [9.51] and the constituent elements defined by equations [9.52] and [9.53], the first three recursions of the filtering process are now derived for model j ($j = 1, 2, 3$). At the first CM point, we have

$$p_{j1}(x_1 \mid \underline{Y}_1) = \frac{(x_1 + t_1)^{\beta_j - 1} e^{-\frac{\left(y_1 - A_j - B_j e^{-C_j x_1}\right)^2}{2\sigma_{j1}^2} - \alpha_j (x_1 + t_1)}}{\int_0^{\infty} (u + t_1)^{\beta_j - 1} e^{-\frac{\left(y_1 - A_j - B_j e^{-C_j u}\right)^2}{2\sigma_{j1}^2} - \alpha_j (u + t_1)} du} \qquad [9.54]$$

and at the second CM point, we obtain

$$p_{j2}(x_2 \mid \underline{Y}_2) = \frac{(x_2 + t_2)^{\beta - 1} e^{-\frac{\left(y_2 - A_j - B_j e^{-C_j x_2}\right)^2}{2\sigma_{j2}^2} - \frac{\left(y_1 - A_j - B_j e^{-C_j (x_2 + t_2 - t_1)}\right)^2}{2\sigma_{j1}^2} - \alpha_j (x_2 + t_2)}}{\int_0^{\infty} (u + t_2)^{\beta - 1} e^{-\frac{\left(y_2 - A_j - B_j e^{-C_j u}\right)^2}{2\sigma_{j2}^2} - \frac{\left(y_1 - A_j - B_j e^{-C_j (u + t_2 - t_1)}\right)^2}{2\sigma_{j1}^2} - \alpha_j (u + t_2)} du} \qquad [9.55]$$

and finally, at the third CM point, we have

$$p_{j3}(x_3 \mid \underline{Y}_3) =$$

$$\frac{(x_3 + t_3)^{\beta_j - 1} e^{-\frac{\left(y_3 - A_j - B_j e^{-C_j x_3}\right)^2}{2\sigma_{j3}^2} - \frac{\left(y_2 - A_j - B_j e^{-C_j (x_3 + t_3 - t_2)}\right)^2}{2\sigma_{j2}^2} - \frac{\left(y_1 - A_j - B_j e^{-C_j (x_3 + t_3 - t_1)}\right)^2}{2\sigma_{j1}^2} - \alpha_j (x_3 + t_3)}}{\int_0^{\infty} (u + t_3)^{\beta_j - 1} e^{-\frac{\left(y_3 - A_j - B_j e^{-C_j u}\right)^2}{2\sigma_{j3}^2} - \frac{\left(y_2 - A_j - B_j e^{-C_j (u + t_3 - t_2)}\right)^2}{2\sigma_{j2}^2} - \frac{\left(y_1 - A_j - B_j e^{-C_j (u + t_3 - t_1)}\right)^2}{2\sigma_{j1}^2} - \alpha_j (u + t_3)} du} \qquad [9.56]$$

From the equations developed for recursions 1 – 3, it is evident that a closed form expression is available for a general stage *i* as

$$p_{ji}(x_i \mid \underline{Y}_i, M_j) = \frac{(x_i + t_i)^{\beta_j - 1} e^{-\alpha_j (x_i + t_i) - \sum_{k=1}^{i} \frac{\left(y_k - A_j - B_j e^{-C_j (x_i + t_i - t_k)}\right)^2}{2\sigma_{jk}^2}}}{\int_0^{\infty} (u + t_i)^{\beta_j - 1} e^{-\alpha_j (u + t_i) - \sum_{k=1}^{i} \frac{\left(y_k - A_j - B_j e^{-C_j (u + t_i - t_k)}\right)^2}{2\sigma_{jk}^2}} du} \qquad [9.57]$$

An essential element in both the parameter estimation process and the determination of $p(M_j \mid \underline{Y}_i)$, see equations [9.36] and [9.37], is the distribution $p(y_i \mid \underline{Y}_{i-1}, M_j)$ given by equation [9.38]. For the functional forms used in this example, we have

$$p(y_i \mid \underline{Y}_{i-1}, M_j) = \frac{\int_0^{\infty} (x_i + t_i)^{\beta_j - 1} e^{-\alpha_j (x_i + t_i) - \sum_{k=1}^{i} \frac{\left(y_k - A_j - B_j e^{-C_j (x_i + t_i - t_k)}\right)^2}{2\sigma_{jk}^2}} dx_i}{\sigma_{ji}\sqrt{2\pi} \int_{t_i - t_{i-1}}^{\infty} (u + t_{i-1})^{\beta_j - 1} e^{-\alpha_j (u + t_{i-1}) - \sum_{k=1}^{i-1} \frac{\left(y_k - A_j - B_j e^{-C_j (u + t_{i-1} - t_k)}\right)^2}{2\sigma_{jk}^2}} du} \qquad [9.58]$$

As described for the EKF example earlier in the chapter, the failure times are simulated using inversion on the initial life distribution, $p(x_0)$. The vibration readings are then generated at each CM point using inversion on the conditional density $p(y_i \mid x_i)$. We now simulate a case corresponding to each of the two model formulations and demonstrate the ability of the proposed methodology to track the appropriate model and the underlying residual life. We compare the estimations of residual life and the prediction errors obtained using the combined weighted modelling approach with those obtained using the general model at each CM point in the simulated histories. The prediction errors are obtained at the *i*th CM point as

$$e_i = ((x_i - \mathbf{E}[x_i \mid \underline{Y}_i])^2)^{1/2} \qquad [9.59]$$

As with the case studies in chapters 6, 7 and 8 the mean-square error (MSE) about the simulated failure time is used as a criterion for comparing the combined and general models. Considering the combined model, the MSE attributable to each of the contributing models is weighted according to the probability that each model provides an appropriate representation of the underlying dynamics for the particular component.

9.4.4.1 *Case 1*

For this first case, a cycle of CM data is simulated with the underlying dynamics corresponding to model 1. The failure time for the cycle is 193 hours and figure 9.10 demonstrates the ability of the recursive process to track the appropriate model according to equation [9.37].

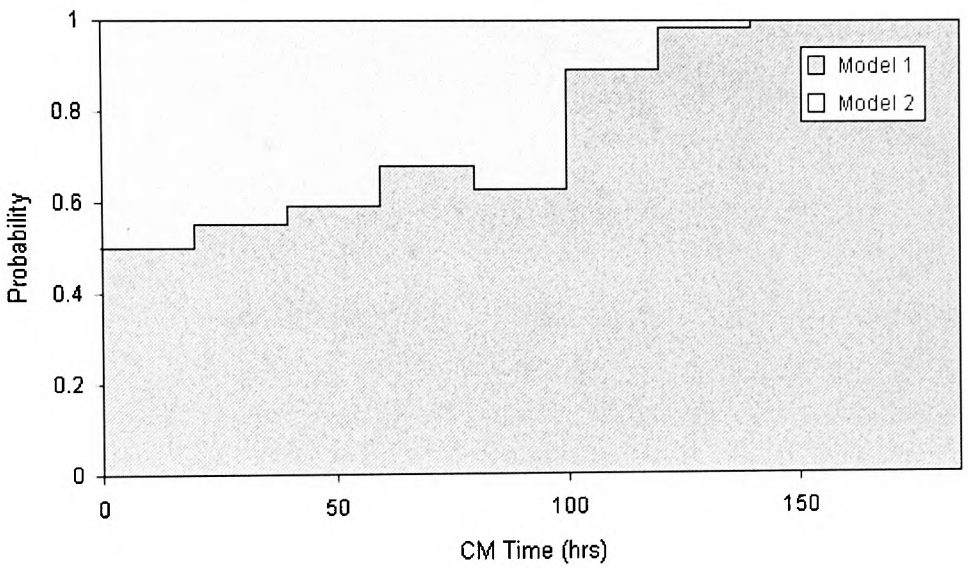

Figure 9.10 – Illustrating the tracking of the appropriate underlying model for case 1

Figure 9.11 illustrates the tracking of the residual life at CM points throughout the life of the component. We compare the estimations of residual life given by the combined weighted modelling approach proposed in this chapter and the general model.

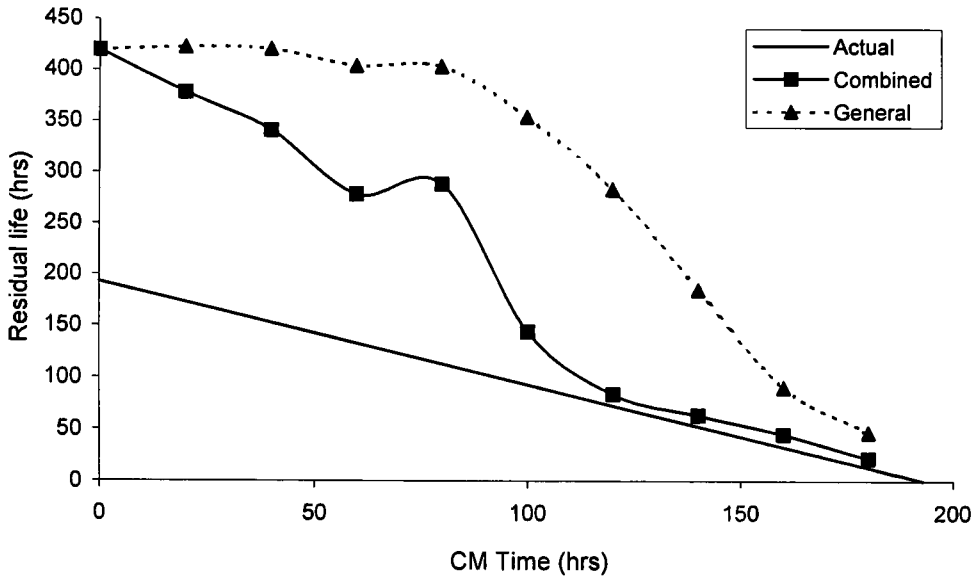

Figure 9.11 – Comparing the residual life predictions obtained using the combined and general models for case 1

Figures 9.10 clearly illustrates that the methodology tracks the appropriate modelling formulation for this particular case and figure 9.11 demonstrates a clear improvement on the residual life prediction capability when compared with the general model. In addition, the sum of squared errors for the combined model is 808.19 compared with 1776.6 for the general model. The superiority of the combined approach is enhanced further by the MSE statistic of 345115 for the combined model and 732541 for the general model.

9.4.4.2 *Case 2*

For this second case, the CM process is simulated according to the formulation for model 2 with a failure time for the component of 651 hours. Figures 9.12 and 9.13 illustrate the tracking of the appropriate model and the residual life respectively.

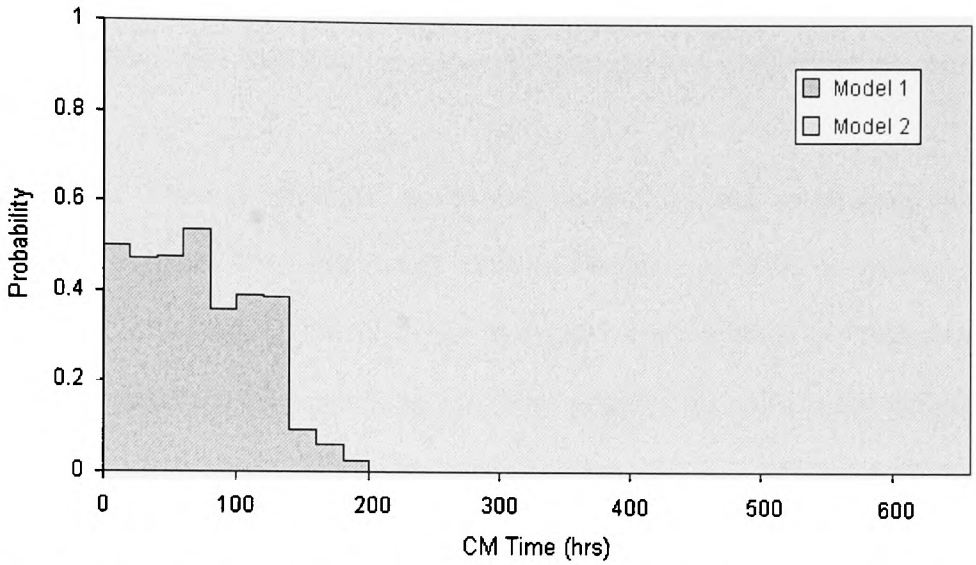

Figure 9.12 – Illustrating the tracking of the appropriate underlying model for case 2

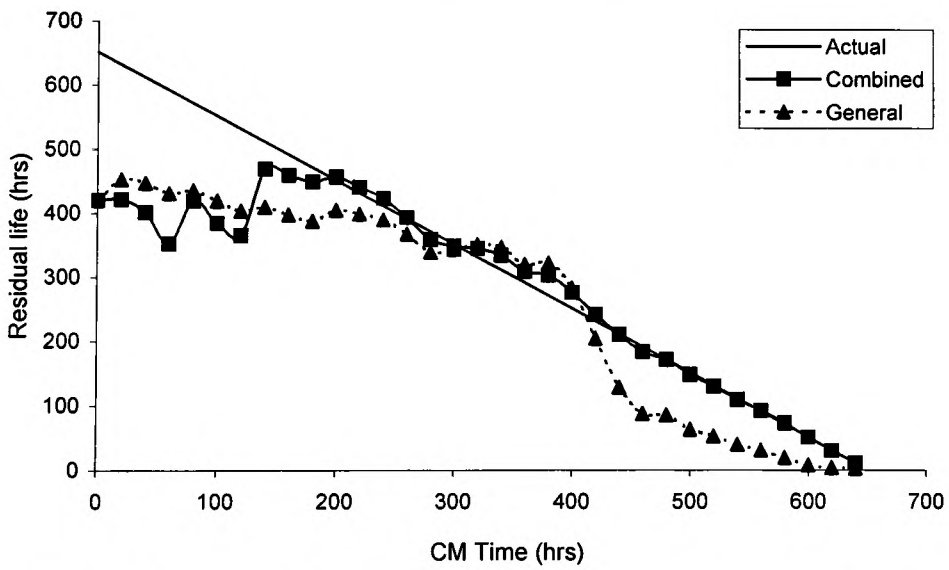

Figure 9.13 – Comparing the residual life predictions obtained using the combined and general models for case 2

As with the first case, it is clear from figures 9.12 and 9.13 that the combined approach tracks the appropriate model quickly for this second case and that the estimates of the residual life are more accurate when compared with those obtained using the general model. This conclusion is again confirmed by the fit statistics; the sum of squared errors is 1422.9 for the combined model and 2234.3 for the general model and the MSE is 585240 for the combined model and 1050250 for the general model.

Cases 1 and 2 have demonstrated that in some situations, it may be advantageous to group the available CM histories and construct a number of filters to represent the specified contingencies. The filters are then applied in parallel to new component CM information and the output from each filter weighted according to the likelihood that the model is the appropriate representation for the current components underlying dynamics. A further consideration could be to model the risk associated with parallel competing failure modes.

9.5 Discussion

In this chapter, we have introduced a number of variations on the general probabilistic filtering approach that are designed to represent particular scenarios or cater for specific modelling or computational needs. Initially, we considered the use of extended Kalman filters when limited computational power is available and/or a large number of components are being monitored in parallel. Then we discussed the potential application of a limited memory stochastic filter designed to facilitate for modelling inaccuracies or fluctuating dynamics over time. Finally, we investigated the potential to represent a number of potential contingencies for a monitoring process using individual stochastic filters and weight the resulting output accordingly. However, the potential usage of stochastic filtering in a CM context is

not limited to the scenarios discussed in this thesis. For instance, in chapter 7 and the EKF example of this chapter, we consider a 2-stage approach to both the vibration monitoring of a component and the inferential process regarding the prediction of residual life. The stochastic filtering process begins once the component is deemed to be in a 'defective' condition (stage 2). The initial phase of monitoring and the fault detection process are tackled using some other means. However, it may be possible to obtain a more consistent probabilistic filtering approach by facilitating for both stages of the CM process. See also Wang (2004b) regarding an alternative solution to this problem. The approach suggested here is to model the relationship $\underline{y}_i \mid x_i$ as

$$p(\underline{y}_i \mid x_i) = p_0(\underline{y}_i \mid x_i)(1-\varepsilon_i) + p_1(\underline{y}_i \mid x_i)\varepsilon_i \qquad [9.60]$$

where,

$$\varepsilon_i = \begin{cases} 0 & ; t_i < T_0 \\ 1 & ; t_i \geq T_0 \end{cases}$$

and T_0 is a random variable representing the start of the defective stage of operation.

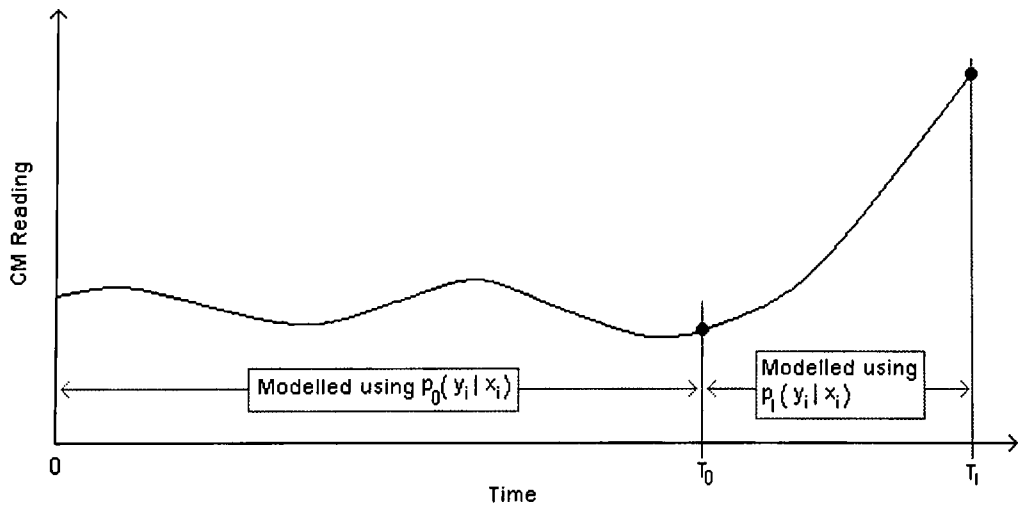

Figure 9.14 - A two-stage CM modelling approach

We are required to estimate

$$\hat{\varepsilon}_{i|i} = P(t_i \geq T_0 \mid \underline{Y}_i) \qquad [9.61]$$

at the ith CM point. The most crucial element of this modelling process is the estimation of T_0 using the distribution developed to represent the relationship $\underline{y}_i \mid x_i$ during the initial 'normal' stage of component life, $p_0(\underline{y}_i \mid x_i)$. One option is to utilise the 'innovations' process that is a available for any predictable and observable process. In this case, we have $\underline{v}_i = \underline{y}_i - \hat{\underline{y}}_i$ where

$$\hat{\underline{y}}_i = \mathbf{E}[\underline{y}_i \mid x_i] = \int_0^\infty \underline{y}_i p_0(\underline{y}_i \mid x_i) d\underline{y}_i \qquad [9.62]$$

At time t_i, $\underline{v}_i$ is a $\underline{0}$-mean vector with covariance matrix $\underline{A}_i$. For an appropriate lag, L, we define the measure

$$D_i = \sum_{k=i-L}^{i} \underline{v}_k' \underline{A}_k \underline{v}_k \qquad [9.63]$$

The probability that, by time t_i, a fault has arisen and the component is currently operating in a defective state is then given by specification of

$$\hat{\varepsilon}_{i|i} = P(t_i \geq T_0 \mid \underline{Y}_i) = P(t_i \geq T_0 \mid \underline{y}_{i-L}, \underline{y}_{i-L+1}, \ldots, \underline{y}_i) = P(t_i \geq T_0 \mid D_i) \qquad [9.64]$$

Alternatively, an initial fault detection rule could be developed as

$$\varepsilon_i = \begin{cases} 0 & ; D_i \leq c \\ 1 & ; D_i > c \end{cases}$$

for some predefined limit c. The potential for constructing a 2-stage filtering process designed for CM applications such as vibration monitoring could be an interesting topic for future research.

In addition, future research on the topics covered in this chapter will require that consideration be given to the influence of preventive maintenance and changes in

other environmental variables. The inclusion of these additional factors will render the state expression

$$x_{i+1} = x_i - (t_{i+1} - t_i)$$

invalid. As a result, the expression will have to be adapted to consider some further deterministic and random elements. Also, tests for the adequacy of model fit must be established. Further considerations for stochastic filtering applications include the use of alternative state definitions, such as the generic term 'wear', as it can be argued that the definition of residual life is subject to the assumption that the value is in some way predetermined, which it is not.

Appendix:

The stochastic filter - parameter estimation
(Chapter 8, case study 1)

Considering a single unit history, the likelihood function for stochastic filtering parameter estimation using the information from one significant principle component y_i is given by equation [6.34] for $m = 1$ as

$$\mathcal{L} = \left(\prod_{i=1}^{n} p(\Delta y_i \mid \underline{Y}_{i-1}) P_{i-1}(x_{i-1} > t_i - t_{i-1} \mid \underline{Y}_{i-1}) \right) p_n(x_n = T - t_n \mid \underline{Y}_n) \qquad [A1]$$

The objective of this appendix is to demonstrate that for the stochastic filter proposed in chapter 8, case study 1, the likelihood function given by equation [A1] reduces to

$$\mathcal{L} = p(x_0 = T) \prod_{i=1}^{n} p(\Delta y_i \mid (x_i = T - t_i)) \qquad [A2]$$

We have $p(\Delta y_i \mid \underline{Y}_{i-1})$ from equation [6.32] with Δy_i replacing $\underline{y}_i$ in the formulation. Now, substituting $p(x_i \mid \underline{Y}_{i-1})$ from equation [6.33] into $p(\Delta y_i \mid \underline{Y}_{i-1})$ produces

$$p(\Delta y_i \mid \underline{Y}_{i-1}) = \frac{\int_0^{\infty} p(\Delta y_i \mid x_i) p_{i-1}(x_i + t_i - t_{i-1} \mid \underline{Y}_{i-1}) dx_i}{\int_{t_i - t_{i-1}}^{\infty} p_{i-1}(u \mid \underline{Y}_{i-1}) du} \qquad [A3]$$

Then substituting equations [A3] into equation [A1], the probabilistic form of the likelihood function reduces to

$$\mathcal{L} = \left(\prod_{i=1}^{n} \int_0^{\infty} p(\Delta y_i \mid x_i) p_{i-1}(x_i + t_i - t_{i-1} \mid \underline{Y}_{i-1}) dx_i \right) p_n(x_n = T - t_n \mid \underline{Y}_n) \qquad [A4]$$

With the chosen distributional forms, the conditional residual delay time distribution at the (i-1)th monitoring point is

$$p_{i-1}(x_{i-1} \mid \underline{Y}_{i-1}) = \frac{(x_{i-1}+t_{i-1})^{\beta-1} e^{-(\alpha(x_{i-1}+t_{i-1}))^{\beta}} \prod_{h=1}^{i-1} \phi_h(x_{i-1}, t_{i-1})}{\int_0^{\infty} (u+t_{i-1})^{\beta-1} e^{-(\alpha(u+t_{i-1}))^{\beta}} \prod_{h=1}^{i-1} \phi_h(u, t_{i-1})\, du} \qquad [A5]$$

where,

$$\phi_h(u, t_{i-1}) = \frac{e^{-\left(\frac{(C+Dt_h)\Delta y_h}{(A+B(u+t_{i-1}-t_h)^{-1})(t_h-t_{h-1})}\right)^{\eta}}}{(A+B(u+t_{i-1}-t_h)^{-1})^{\eta}} \qquad [A6]$$

and $\prod_{h=1}^{i-1} \phi_h(.) = 1$ for $i = 1$. From equation [8.2] we have

$$p(\Delta y_i \mid x_i) = \eta \left(\frac{C+Dt_i}{t_i - t_{i-1}}\right)^{\eta} \Delta y_i^{\eta-1} \frac{e^{-\left(\frac{(C+Dt_i)\Delta y_i}{(A+Bx_i^{-1})(t_i-t_{i-1})}\right)^{\eta}}}{(A+Bx_i^{-1})^{\eta}} \qquad [A7]$$

Inserting equations [A5] and [A7] with the substitution $x_{i-1} = x_i + t_i - t_{i-1}$ into the likelihood function of equation [A4] gives

$$\mathcal{L} = \left(\prod_{i=1}^{n} \int_0^{\infty} \frac{(x_i+t_i)^{\beta-1} e^{-(\alpha(x_i+t_i))^{\beta} - \left(\frac{(C+Dt_i)\Delta y_i}{(A+B/x_i)(t_i-t_{i-1})}\right)^{\eta}} \prod_{h=1}^{i-1} \phi_h(x_i + t_i - t_{i-1}, t_{i-1})}{(A+B/x_i)^{\eta} \int_0^{\infty} (u+t_{i-1})^{\beta-1} e^{-(\alpha(u+t_{i-1}))^{\beta}} \prod_{h=1}^{i-1} \phi_h(u, t_{i-1})\, du} dx_i \right.$$

$$\left. \ldots \times \ldots \eta \left(\frac{C+Dt_i}{t_i - t_{i-1}}\right)^{\eta} \Delta y_i^{\eta-1} \right) p_n(x_n = T - t_n \mid \underline{Y}_n) \qquad [A8]$$

Now, with the result

$$p_n(x_n = T - t_n \mid \underline{Y}_n) = \frac{T^{\beta-1} e^{-(\alpha T)^{\beta}} \prod_{h=1}^{n} \phi_h(T - t_n, t_n)}{\int_0^{\infty} (u+t_n)^{\beta-1} e^{-(\alpha(u+t_n))^{\beta}} \prod_{h=1}^{n} \phi_h(u, t_n)\, du} \qquad [A9]$$

and by recognising that for $h = i$, we have

$$\phi_i(x_i, t_i) = (A + B/x_i)^{-\eta} e^{-\left(\frac{(C+Dt_i)\Delta y_i}{(A+B/x_i)(t_i - t_{i-1})}\right)^{\eta}} \qquad [A10]$$

By equating the functions

$$\phi_h(x_i + t_i - t_{i-1}, t_{i-1}) = \phi_h(x_i, t_i) = \frac{e^{-\left(\frac{(C+Dt_h)\Delta y_h}{(A+B(x_i+t_i-t_h)^{-1})(t_h - t_{h-1})}\right)^{\eta}}}{(A + B(x_i + t_i - t_h)^{-1\eta}} \qquad [A11]$$

the likelihood function becomes

$$\mathcal{L} = \left(\prod_{i=1}^{n} \eta \left(\frac{C + Dt_i}{t_i - t_{i-1}} \right)^{\eta} \Delta y_i^{\eta - 1} \right) \left(\prod_{i=1}^{n} \frac{\int_0^{\infty} (u + t_i)^{\beta-1} e^{-(\alpha(u+t_i))^{\beta}} \prod_{h=1}^{i} \phi_h(u, t_i) du}{\int_0^{\infty} (u + t_{i-1})^{\beta-1} e^{-(\alpha(u+t_{i-1}))^{\beta}} \prod_{h=1}^{i-1} \phi_h(u, t_{i-1}) du} \right)$$

$$\ldots \times \ldots \frac{T^{\beta-1} e^{-(\alpha T)^{\beta}} \prod_{h=1}^{n} \phi_h(T - t_n, t_n)}{\int_0^{\infty} (u + t_n)^{\beta-1} e^{-(\alpha(u+t_n))^{\beta}} \prod_{h=1}^{n} \phi_h(u, t_n) du} \qquad [A12]$$

Finally, by the cancellation of successive terms we establish the relationship

$$\prod_{i=1}^{n} \frac{\int_0^{\infty} (u + t_i)^{\beta-1} e^{-(\alpha(u+t_i))^{\beta}} \prod_{h=1}^{i} \phi_h(u, t_i) du}{\int_0^{\infty} (u + t_{i-1})^{\beta-1} e^{-(\alpha(u+t_{i-1}))^{\beta}} \prod_{h=1}^{i-1} \phi_h(u, t_{i-1}) du}$$

$$= \alpha^{\beta} \beta \int_0^{\infty} (u + t_n)^{\beta-1} e^{-(\alpha(u+t_n))^{\beta}} \prod_{h=1}^{n} \phi_h(u, t_n) du \qquad [A13]$$

and the likelihood function reduces to

$$\mathcal{L} = \alpha^{\beta}\beta T^{\beta-1}e^{-(\alpha T)^{\beta}}\left(\prod_{i=1}^{n}\eta\left(\frac{C+Dt_i}{t_i-t_{i-1}}\right)^{\eta}\Delta y_i^{\eta-1}\right)\prod_{h=1}^{n}\phi_h(T-t_n,t_n)$$

$$= \alpha^{\beta}\beta T^{\beta-1}e^{-(\alpha T)^{\beta}}\prod_{i=1}^{n}\eta\left(\frac{C+Dt_i}{(A+B(T-t_i)^{-1})(t_i-t_{i-1})}\right)^{\eta}\Delta y_i^{\eta-1}e^{-\left(\frac{(C+Dt_i)\Delta y_i}{(A+B(T-t_i)^{-1})(t_i-t_{i-1})}\right)^{\eta}}$$

[A14]

which is

$$\mathcal{L} = p(x_0 = T)\prod_{i=1}^{n} p(\Delta y_i \mid (x_i = T - t_i))$$

thus establishing the desired result.

REFERENCES

Abdel-Hameed M (1975) 'A gamma wear process', *IEEE Transactions on Reliability*, vol.24, No. 2, 152-153

Akaike H (1974) 'A new look at statistical model identification', *IEEE Transactions on Automatic Control*, vol. 19 (6), 716 – 723

Ansell JI, Phillips MJ (1989) 'Practical problems in the statistical analysis of reliability data (with discussion)', *Applied Statistics*, vol.38, No.2, 205-247

Aoki M (1967) '*Optimisation of Stochastic Systems, Topics in Discrete Time Systems*', Academic Press: New York, London

Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) 'A tutorial on particle filters for online non-linear/non-Gaussian Bayesian tracking' *IEEE Transactions on Signal Processing*, vol.50, No.2

Ascher H, Feingold H (1984) '*Repairable Systems Reliability: Modelling, Inference, Misconceptions and Their Causes*', New York: Marcel Dekker.

Austin J, Jackson T, Fletcher M, Jessop M, Cowley P, Lobner P (2003) 'Predictive maintenance: Distributed aircraft engine diagnostics', *The Grid*, Editors: Foster I, Kesselman C, 2nd Edition, Morgan Kaufmann.

Aven T (1996) 'Condition based replacement policies – a counting process approach', *Reliability Engineering and System Safety*, vol. 51, 275-281

Baker RD, Scarf PA, Wang W (1997) 'A delay-time model for repairable machinery; maximum likelihood estimation of optimum inspection intervals' *IMA Journal of Mathematics Applied in Business and Industry*, vol. 8, 83-92

Baker RD, Wang W (1992) 'Estimating the delay-time distribution of faults in repairable machinery from failure data', *IMA Journal of Mathematics Applied in Business and Industry*, No.3, 259-281.

Baker, R D & Wang, W (1993) 'Developing and testing the delay-time model', *Journal of the Operational Research Society*, vol.44, No.4, 361-374

Baker, RD, Christer, AH (1994) "Operational research modelling of engineering aspects of maintenance", *European Journal of Operational Research*, vol.73, 407-422

Banjevic D, Jardine AKS (2004) 'Calculation of reliability function and remaining useful life for a Markov failure time process' *MIMAR Conference Proceedings*

Banjevic D, Jardine AKS, Makis V, Ennis M (2001) 'A control limit policy and software for CBM', *INFOR*, vol.39, 32-50

Barlow RE, Hunter LC (1960) 'Optimum preventive maintenance policies', *Operation Research*, vol.8, 90-100

Barlow RE, Proschan F (1965) '*Mathematical Theory of Reliability*', Wiley, NY

Basawa IV, Rao P (1980) '*Statistical Inference for Stochastic Processes*', Academic Press, London

Bedford T, Cooke R (2003) '*Probabilistic Risk Analysis: Foundations and Methods*', CUP

Bendell A, Walley M, Wightman DW, Wood LM (1986) 'Proportional hazards modelling in reliability analysis – an application to brake discs on high speed trains' *Quality in Reliability and Engineering International*, vol.2, 45-52

Bernardo J, Smith A (2000) '*Bayesian Theory*', John Wiley & Sons: NY. 2nd edition.

Carr MJ, Christer AH (2003) 'Incorporating the potential for human error in maintenance models', *Journal of the Operational Research Society*, vol.54, No.12, 1249-1253

Chatfield C, Collins AJ (1980) '*Introduction to Multivariate Analysis*' London, Chapman & Hall.

Chillcott J, Christer AH (1991) 'Modelling of condition based maintenance at the coal face', *International Journal of Production Economics*, vol. 22, 1-11

Christer AH (1973) 'Innovatory decision making, the role and effectiveness of theories of decision practice', (Editors: White DL, Brown KC), Hodder & Stoughton, London, 369-377

Christer AH (1982) 'Modelling inspection policies for building maintenance', *Journal of the Operational Research Society*, vol.33, 723-732

Christer AH (1999) 'Developments in delay time modelling', *Journal of the Operational Research Society*, vol.50, No.11, 1120-1137

Christer AH, Lee C, Wang W (2000) 'A data deficiency based parameter estimating problem and case study in delay time PM modelling', *International Journal of Production Economics*, vol.67, 63-76

Christer AH, Redmond DF (1992) 'Revising models of maintenance and inspection', *International Journal of Production Economics*, vol. 24, 227-234

Christer AH, Waller W (1984a) 'Delay-time models of industrial inspection maintenance problems', *Journal of the Operational Research Society*, vol.35, No.5, 401-406

Christer AH, Waller W (1984b) 'Reducing production downtime using delay-time analysis', *Journal of the Operational Research Society*, vol.35, No.6, 499-512

Christer AH, Wang W (1992)'A model of condition monitoring inspection of production plant' *International Journal of Production Research*, vol. 30, 2199-2211

Christer AH, Wang W (1995) 'A simple condition monitoring model for a direct monitoring process' *European Journal of Operational Research*, vol.82, 258-269

Christer AH, Wang W, Baker RD, Sharp J (1995a) 'Modelling maintenance practice of production plant using the delay time concept', *IMA Journal of Mathematics Applied in Business and Industry*, vol.6, 67-83

Christer AH, Wang W, Choi K, Sharp JM (1998) 'The delay-time modelling of preventive maintenance of plant given limited PM data and selective repair at PM', *IMA Journal of Mathematics Applied in Medicine and Biology*, vol.15, 355-379

Christer AH, Wang W, Sharp J (1997) 'A state space condition monitoring model for furnace erosion prediction and replacement', *European Journal of Operational Research*, vol.101, 1-14

Christer AH, Whitelaw J (1983) 'An operational research approach to breakdown maintenance: problem recognition', *Journal of the Operational Research Society*, vol.34, 1041-52

Christer, AH, Wang W, Sharp J, Baker RD (1995b) 'Stochastic maintenance modelling of high-tech steel production plant', (Editors: Christer AH, Osaki S, Thomas LC), *Stochastic Modelling in Innovative Manufacturing*, Springer, 196-214

Collacott RA (1977) '*Mechanical Fault Diagnosis and Condition Monitoring*', London: Chapman and Hall

Comon P (1994) 'Independent components analysis – a new concept?', *Signal Processing*, vol.36, 287-314.

Coolen FRA, Dekker R (1995) 'Analysis of a 2-phase model for optimisation of condition monitoring intervals', *IEEE Transactions on Reliability*, vol.44, 505-511.

Cox DR (1972) 'Regression Models and Life Tables (with Discussion)', *Journal of the Royal Statistical Society*, Series B, vol. 34, 187-220

Cox DR (1973) 'Regression models and life tables', *Journal of the Royal Statistical Society*, Series B, vol.34, 187-220

Cox DR, Oakes D (1984) '*Analysis of Survival Data*', Chapman and Hall

Crowder M (2001) '*Classical Competing Risks*', Chapman & Hall/CRC

Dagg R, Newby M (1998) 'Optimal overhaul intervals with imperfect inspection and repair', *IMA Journal of Mathematics Applied in Business and Industry*, vol.9, 381-391

Doob JL (1953) '*Stochastic Processes*', Wiley, New York

Easry DJ, Marshall AW, Proschan F (1973) 'Shock models and wear processes', *Annals of Probability*, vol.1, 627-649

Fisher LD, Lin DY (1999) 'Time dependent covariates in the Cox proportional hazards model', *Annual Review of Public Health*, vol. 20, 145-157

Forney GD (1973) 'The Viterbi algorithm', *Proceedings IEEE*, vol.61, 268-278

Gilgmayr J (1987) 'An age-wear dependent model of failure', *IEEE Transcripts in Reliability*, vol.36, 581-585

Gong C, Tang K (1997) 'Monitoring machine operations using on-line sensors', *European Journal of Operational Research*, vol. 96, 479-492

Grimmett G, Stirzaker D (1982) '*Probability and Random Processes*', Oxford Press.

Harvey AC (1989) '*Forecasting Structural Time-Series Models and the Kalman-Filter*', Cambridge University Press

Heinrich G, Jensen U (1992) 'Optimal replacement rules based upon different information levels', *Naval Research Logistics*, vol.39, 937-955

Jack N (1991) 'Availability measures for components maintained by repair-replacement policies', IMA Journal of Management Mathematics, vol.3, 107-115

Jazwinski AH (1970) '*Stochastic Processes and Filtering Theory*', Academic Press, London

Jia X, Christer AH (2002) 'Case experience comparing the RCM approach to plant maintenance with a modelling approach', (Editors: Murthy P, Blischke W), *Case studies in Reliability and Maintenance*, J Wiley & Son Inc., 477-494

Johnson NI, Kotz S (1972) '*Distributions in Statistics: Continuous Multivariate Distributions*', New York, Wiley

Jolliffe IT (1986) '*Principal Component Analysis*', New York: Springer-Verlag.

Jutten C, Herault J (1991) 'Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture', *Signal Processing*, vol.24, 1-10

Kaio N, Osaki S (1989) 'Comparison of inspection policies', *Journal of the Operational Research Society*, vol.40, 499-503

Kalbfleisch JD, Prentice RL (1973) & (1980) '*The Statistical Analysis of Failure Time Data*', Wiley

Kallianpur G (1980) '*Stochastic Filtering Theory*', Springer: New York.

Kendall M, Stuart A (1979) '*The Advanced Theory of Statistics, Volume 2 – Inference and Relationship*', 4th Edition, Charles Griffin & Company Ltd

Koch G (1986) 'A dynamical approach to reliability theory', (Editors: Serra A, Barlow RE), *Theory of Reliability; Proceedings of the International School of Physics 'Enrico Fermi', Course XCIV*, 215-240, North Holland, Amsterdam

Krishnan V (1984) '*Non-Linear Filtering and Smoothing*', Wiley

Kumar D, Westberg U (1997) 'Maintenance scheduling under age replacement policy using proportional hazard modelling and total-time-on-test plotting', *European Journal of Operational Research*, vol.99, 507-515

Lawless JF (1982) '*Statistical Models and Methods for Lifetime Data*', Wiley

Lin D, Banjevic D, Jardine AKS (2005) 'Using principal components in a proportional hazards model with applications in condition-based maintenance', *Journal of the Operational Research Society*, Published online: August 31, 2005

Lin D, Makis V (2002) 'State of model parameter estimation for transmissions on heavy hauler trucks using oil data', *Proceedings of COMADEM International*

Lin D, Makis V (2004) 'On-line parameter estimation for a failure-prone system subject to condition monitoring', *Journal of Applied Probability*, vol.41, 211–220

Liptser R (1997) 'Lecture notes on Stochastic Processes and Stochastic Control', http://www.eng.tau.ac.il/~liptser/

Liptser R, Shryaev AN (1989) '*Theory of Martingales*', Kluwer, Dordrecht

Ljung L (1999) '*System Identification – Theory for the User*', Prentice Hall

Makis V, Jardine AKS (1991a) 'Computation of optimal policies in replacement models', *IMA Journal of Mathematics Applied in Business and Industry*, vol.3, 169-176

Makis V, Jardine AKS (1991b) 'Optimal replacement policies in the proportional hazards model', *INFOR*, vol.30, 172-183

Makis V, Jardine AKS (1992) 'Optimal replacement in the proportional hazards model', *INFOR*, Vol.30, 1

Makis V, Wu J, Gao Y (2005) 'An application of DPCA to oil data for CBM modelling', *pending publication.*

Man KF, Tang KS, Kwong S (1999) '*Genetic Algorithms*', Springer-Verlag

McKone KE, Weiss EN (1998) 'Total productive maintenance: bridging the gap between practice and research', *Production and Operations Management*, vol.7, No.4, 335-351

Meinhold RJ, Singpurwalla ND (1983) 'Bayesian Analysis of a Commonly Used Model for Describing Software Failures', *The Statistician*, vol.32, No.2, 168-73

Meinhold RJ, Singpurwalla ND (1986) 'A Kalman-filter approach to life testing', (Editors: Serra A, Barlow RE), *Theory of Reliability; Proceedings of the International School of Physics 'Enrico Fermi', Course XCIV*, 215-240, North Holland, Amsterdam

Mercer A (1961) 'Some simple wear dependent renewal processes', *Journal of the Royal Statistical Society*, vol.23, 368-376

Moore DS, McCabe GP (2003) 'Introduction to the Practice of Statistics', 4th edition, WH Freeman, NY

Nairac A, Townsend N, Carr R, King S, Cowley P, Tarassenko L (1999) 'A system for the analysis of jet engine vibration data', *Integrated Computationally Aided Engineering*, vol. 6, 53 – 65

Newby M (1993) 'A critical look at some point-process models for repairable systems', *IMA Journal of Mathematics Applied in Business and Industry*, Vol.4, 375-394.

O'Hancock D, Synovec RE (1989) 'Early detection of C-130 aircraft engine malfunction by principal components analysis of the wear metals in C-130 engine oil', *Applied Spectroscopy*, vol.43, No. 2

Park KS (1988) 'Optimal continuous wear limit replacement under periodic inspections', *IEEE Transcripts in Reliability*, vol.37, 293-294

Parmigiani G (1996) 'Optimal scheduling of fallible inspections', *Operational Research*, vol.44, 360-367

Patankar MS, Taylor JC (2003) 'Posterior probabilities of causal factors leading to unairworthy dispatch after maintenance', *Journal of Quality in Maintenance Engineering*, vol.9, No.1, 38-47

Rabiner LR, Juang BH (1986) 'An introduction to hidden Markov models', *IEEE Acoustics, Speech and Signal Processing* (magazine), January (1986), 4-16

Rao CR (1965) '*Linear Statistical Inference and its Applications*', Wiley

Ross S (1996) '*Stochastic Processes*', 2nd edition, Wiley

Ross SM (1970) '*Applied Probability Models with Optimization Applications*', Holden-Day

Ross SM (1983) '*Stochastic Processes*', New York; Wiley

Russell SJ, Norvig P (2003) '*Artificial Intelligence: A Modern Approach*', Prentice Hall

Scarf PA (1997) 'A survey of papers in condition monitoring', *European Journal of Operational Research*

Sorenson HW (1980) '*Parameter Estimation: Principles and Problems*', Marcel Dekker

Starr G, Rao BKN (2001) '*Condition Monitoring and Diagnostic Engineering Management*', Elsevier Science

Steedman JPSD, Whittaker GD (1973) 'Computers improve plant performance', *Maintenance Engineer*, vol.17, 38-40

Stoer J, Bulirsch R (1980) '*Introduction to Numerical Analysis*', Springer-Verlag

Streit RL, Barrett RF (1990) 'Frequency line tracking using hidden Markov models', *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol.38, 586-598

Tandon N, Choudry A (1999) 'A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings', *Tribology International*, vol.32, 469-480

Therneau TM, Grambsch PM (2000) '*Modeling Survival Data: Extending the Cox Model*', Springer, NY

Vlok PJ, Coetzee JL, Banjevic D, Jardine AKS, Makis V (2002) 'Optimal component replacement decisions using vibration monitoring and the Proportional Hazards Model', *Journal of the Operational Research Society*, vol.53, 193-202

Wang W (1997) 'Subjective estimation of the delay time distribution in maintenance modelling', *European Journal of Operational Research*, vol.99, No.3, 516-529

Wang W (2000) 'A model to determine the critical level and monitoring intervals in condition based maintenance', *International Journal of Production Research*, vol. 38, 1425-1436

Wang W (2002) 'A model to predict the residual life of rolling element bearings given monitored condition information to date' *IMA Journal of Management Mathematics*, vol. 13, 3-16

Wang W (2003) 'Modelling condition monitoring intervals: A hybrid of simulation and analytical approaches', *Journal of the Operational Research Society*, vol.54, 273-282

Wang W (2004 a) 'A wear prediction model and its application to oil-based monitoring', *Submitted for publication*

Wang W (2004 b) 'Modelling the identification of the initial point of a random fault using available condition monitoring information' *proceedings of 5th IMA conference in maintenance and reliability modelling*, Eds, W. Wang, P. Scarf and M. Newby, 5-7 April 2004, University of Salford

Wang W, Christer AH (2000) 'Towards a general condition based maintenance model for a stochastic dynamic system' *Journal of the Operational Research Society* vol. 51, 145-155

Wang W, Scarf PA, Smith MAJ (2000) 'On the application of a model of condition based maintenance', *Journal of the Operational Research Society*, vol.51, 1218-1227

Wang W, Zhang W (2002) 'A model to predict equipment residual life based upon oil analysis data' *Proceedings of COMADEM International Conference*, Editors: RBKN Rao, AK Nandi.