Bayesian Modeling of Pitting Corrosion in Steam Generators

by

Dan Mao

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering

Waterloo, Ontario, Canada, 2007

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Steam generators in nuclear power plants experienced varying degrees of under-deposit pitting corrosion. A probabilistic model to accurately predict pitting corrosion is necessary for effective life-cycle management of steam generators.

This thesis presents an advanced probabilistic model of pitting corrosion characterizing the inherent randomness of the pitting process and measurement uncertainties of the in-service inspection (ISI) data obtained from eddy current (EC) inspections. A Bayesian method is developed for estimating the model parameters. The proposed model is able to estimate the number of actual pits, the actual pit depth as well as the maximum pit depth, which is the main interest of the pitting corrosion model.

A MATLAB program of the Markov chain Monte Carlo technique is developed to perform the Bayesian estimations. Simulation experiments are performed to check the behavior of the Bayesian method. Results show that the MCMC algorithm is an effective way to estimate the model parameters. Also, the effectiveness and efficiency of Bayesian modeling are validated.

A comprehensive case study is also presented on the in-service inspection data of pitting corrosion in a steam generator unit. The Weibull distribution is found to be an appropriate probability distribution for modeling the actual pit depth in steam generators.

Acknowledgements

I would like to express my greatest gratitude to my supervisor, Professor Mahesh D. Pandey, for his patient guide and financial support. Without his help, this thesis could not have been completed.

I would also like to thank my co-supervisor, Professor S. T. Ariaratnam, for his kind encouragement and generous final support. The teaching assistantship for CivE 303 was a very pleasant and rewarding experience of working under him.

The raw pitting corrosion data were preprocessed by Dr. Mikko Jyrkama, to whom I am very grateful. Without his excellent work, the data would have never been in useful form.

My special thanks are given to my husband, parents and sister for their love, understanding, support and sacrifice.

Finally, I would like to thank many friends at the university: Ying An, Wensheng Bu, Xiaoguang Chen, Tianjin Cheng, Suresh V. Detla, Qinghua Huang, Hongtao Liu, Yuxin Liu, Guru Prakash, Anup Sahoo, Papia Sultana, and Jinyu Zhu.

To

My Parents

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Steam Generators in Nuclear Power Plants

Canadian Deuterium Uranium (CANDU™) reactors, designed by and a trademark of Atomic Energy of Canada Limited (AECL), are used in commercial nuclear power plants in Canada and many other countries. The initial choice of natural uranium (0.7% U-235) as fuel and heavy water (deuterium) as both a moderator and a coolant made the CANDU design unique from other types of reactors in the world (Tapping et al., 2000).

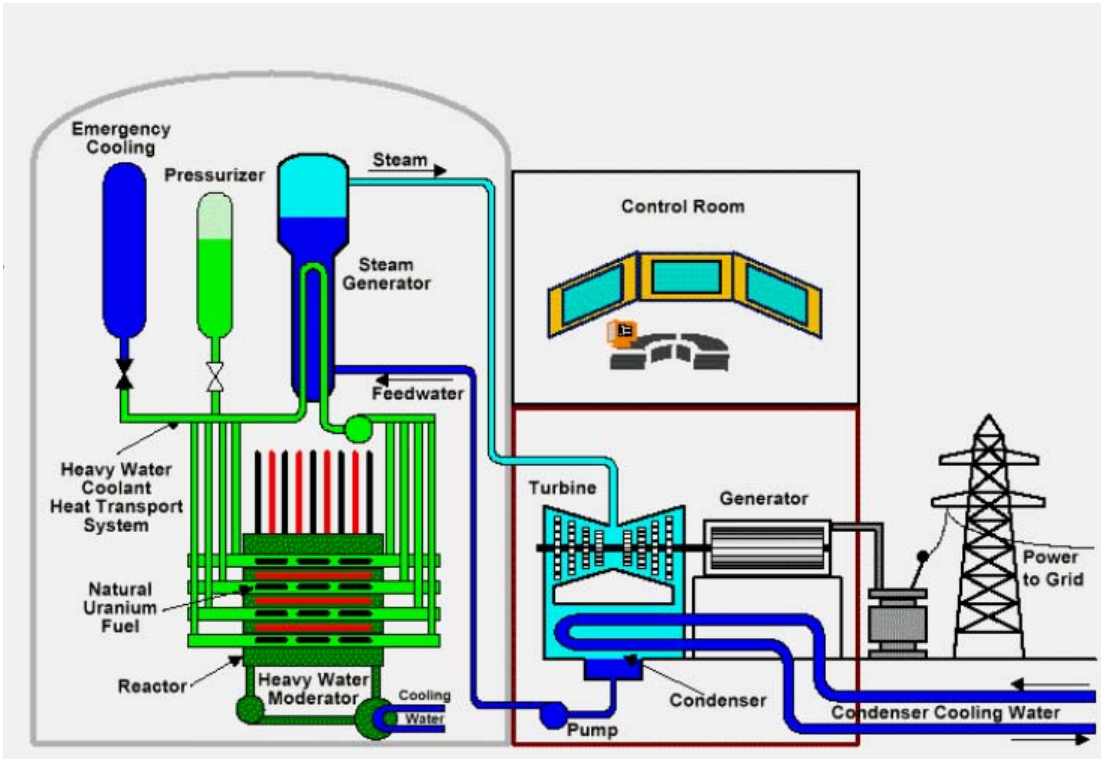

Figure 1.1: Schematic layout of the CANDU nuclear power plant system

Figure 1.1 shows a schematic layout of a typical CANDU nuclear power plant (NPP). Among the many systems, structures and components in the NPP, the steam generator is the subject of the thesis.

Currently operating CANDU steam generators are vertical re-circulating steam generators manufactured. A typical CANDU steam generator is shown in Figure 1.2. The primary side of the steam generator mainly consists of the primary side of the tubesheet and several thousand small diameter alloy tubes that are bent into an inverted U-shape. The alloy U-tubes are welded to the primary side of the alloy clad carbon steel tubesheet and rolled into the tubesheet. The secondary side of the steam generator consists of a steam separating equipment, a tube bundle shroud, the secondary side of the tubesheet, the secondary side of the tube bundle, pre-heater baffles and tube support plates. Appendix A lists major design specifications of CANDU steam generator in the current five Canada's nuclear power generation stations (Tapping et al., 2000).

Steam generator is one of the critical components in CANDU NPP. It has two main functions: safety barrier between the radioactive primary coolant loop to the non-radioactive secondary coolant loop, and heat exchange or production of steam for the turbine generator (Nickerson and Maruska, 1998).

In the primary coolant loop, the heavy water inside the fuel channel (Figure 1.1) is heated by the energy released from the nuclear fission and kept under high pressure to raise its boiling point and avoid significant steam formation in the reactor core. The hot pressurized heavy water then enters the steam generators through the inlet. It passes through the "hot leg" of the tube bundle, in the shape of an inverted U, giving up heat to the feedwater outside the tube bundle. After the heat exchange, the heavy water goes

down along the cold leg of the tube bundle and return via the outlet back to the reactor core.

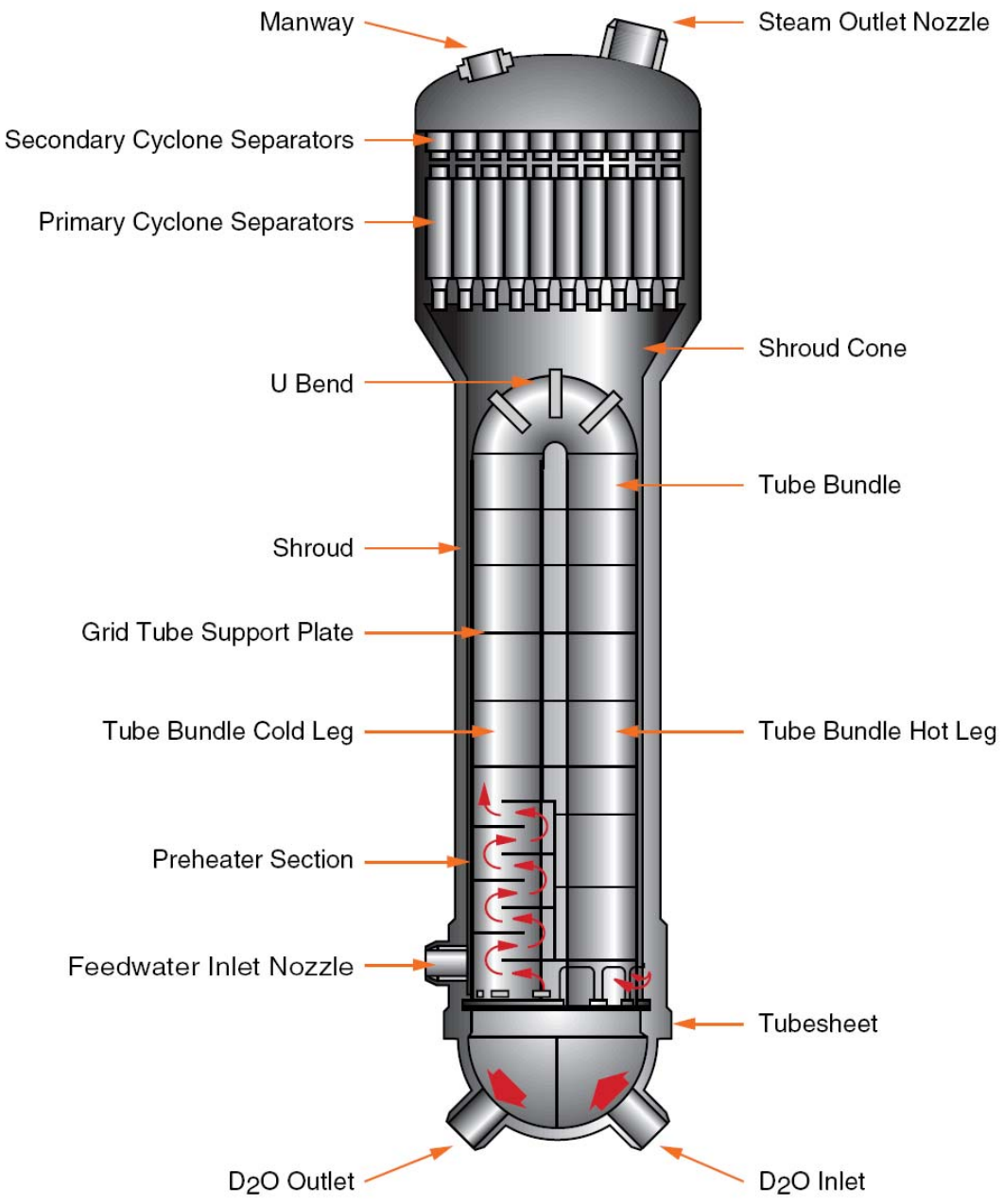

Figure 1.2: The schematic diagram of a CANDU steam generator (from Snook, 2001)

In the secondary side, the less-pressurized feedwater enters the pre-heater section and is forced to flow over the heavy water outlet and the U-bend bundle. The feedwater turns to the steam-water mixture with the re-circulating saturated water and rise to the top of the U-tube bundle. The steam-water mixture eventually turns to steam leaving the steam separators at a high speed through the steam separating equipment to a conventional turbine, with a generator attached to it to generate electricity. In addition, the saturated water from the steam separating equipment re-circulates to the bottom of the tube bundle. After acquiring enough heat of vaporization, the saturated water leaves the boiler as saturated steam.

Since the primary heavy water contains radioactive particles and is isolated from the feedwater only by the U-tube walls, the alloy U-tubes form part of the primary boundary for isolating these radioactive particles. It is, therefore, important that the U-tubes be maintained defect-free so that no breaks will occur in the U-tubes to contaminate the feedwater. However, to date steam generators in nuclear power plants have experienced varying degrees of under-deposit pitting corrosion (Millett, 1998) which is a highly undesirable event. Therefore, pitting corrosion is our main concern in this thesis and will be discussed next.

1.2 Motivation

Pitting corrosion is a form of localized corrosion and typified by formation of cavities resulting from local metal dissolution within a passivated surface area (Galvele, 1983). Once initiated, the pitting corrosion can lead to rapid penetration at small discrete areas

and cause failure due to perforation, although the total corrosion, measured by weight loss, might be very negligible.

Eddy current tests of the tubes have indicated that the pitting corrosion occurs on tubes near the tubesheet at levels corresponding to the levels of sludge that accumulates on the tubesheet (Maruska, 2002). The sludge is mainly from oxides and copper compounds along with traces of other metals that have settled out of the feedwater onto the tubesheet. The correlation between sludge levels and the pit location strongly suggests that the sludge deposits provide a site for concentration of a phosphate solution or other corrosive agents at the tube wall that result in pitting corrosion (Jones, 1996).

Therefore, an effective life-cycle management of steam generators, including both effective intervention methods and accurate prediction of pitting damages, is in great necessity for nuclear power plants to manage the pitting corrosion problems. Current intervention methods include water lancing and chemical cleaning, i.e., removal of the deposits in the steam generators (Maruska, 2002).

To predict the pitting damages, we need either a mechanistic model or an empirical model. Unfortunately, the under-deposit pitting corrosion is a complex process of which the physical-chemical mechanism has not yet been well understood. No mechanistic model is available. Instead, empirical models may be built based on on-site observational data (the number of pits and their sizes) during different inspection outages. But this is also a very challenging problem. First of all, both the generation and growth of pits are of random nature. Moreover, as a localized damage, the pits are usually very difficult to detect. Therefore, the probability of detection (POD) should be considered. When a pit is detected, the measurement of its size suffers from measurement errors. The measurement

uncertainties become even worse for in-service inspections, due to the limited access, tight time budget and the existence of sludge deposit.

To accurately quantify risks associated with the pitting corrosion and facilitate a successful risk-based life cycle management of steam generators, the need for an advanced probabilistic model that takes into account the various uncertainties of the pitting corrosion from in-service inspection is therefore compelling.

1.3 Objectives and Scope

The thesis mainly focuses on pitting corrosion and is aimed at developing a probabilistic model of pitting corrosion for life-cycle management of steam generators. The inherent randomness of pitting corrosion and POD and measurement errors of the in-service inspection tools are to be integrated in the proposed model. In the thesis a statistical approach is developed to estimate the parameters of the proposed model. The approach deals with the pit generation and pit growth in a systematic way and its parameters are estimated using a Bayesian methodology. In particular, a Markov chain Monte Carlo simulation technique is developed to estimate the model parameters. The model is used to estimate the number and size of pits for steam generators. It is also used to estimate the distribution of maximum pit depth, which is one of the major decision-making parameters in life-cycle management program of steam generators.

1.4 Organization

This thesis is divided into six chapters including this introductory one. In Chapter 2, background information of the physical-chemical mechanism of corrosion and pitting corrosion is introduced. Existing probabilistic models and statistical methodologies of pitting corrosion are also reviewed. In Chapter 3 a probabilistic model of pitting corrosion that integrates all ingredients of associated uncertainties is formulated. Then the necessity of the Bayesian methodology is discussed. Chapter 4 introduces basic concepts and algorithms of the Markov Chain Monte Carlo simulation. Simulation experiments are performed to verify the behavior of the method and to validate the MATLAB code developed by the author. In Chapter 5, a case study of in-service inspection data from a steam generator is undertaken. Major findings of the thesis are concluded in Chapter 6, with future research topics recommended.

.

Chapter 2

Literature Review

This chapter reviews basic concepts of pitting corrosion and existing literature on probabilistic modeling of pitting corrosion. Since most pitting corrosion processes are electrochemical reactions that involve transfer of electronic charges (Jones, 1996), it is necessary to begin with a brief discussion on the electrochemical basis of corrosion.

2.1 Principle of Pitting Corrosion

2.1.1 Corrosion as an Electrochemical Process

Consider an example of a metal, denoted by M, corroding in hydrochloric acid solution (Figure 2.1). In ionic form the reaction is expressed as

$$M + nH^+ + nCl^- \rightarrow M^{n+} + nCl^- + (n/2)H_2 \quad (2.1)$$

Eliminating Cl^- from both sides, (2.1) gives

$$M + nH^+ \rightarrow M^{n+} + (n/2)\, H_2 \quad (2.2)$$

This reaction can be further separated as two electrode, or half-cell reactions as

Anodic reaction: $M \rightarrow M^{n+} + ne^-$ (2.3a)

Cathodic reaction: $nH^+ + ne^- \rightarrow (n/2)\, H_2$ (2.3b)

Reaction (2.3a), known as the anodic reaction, is an oxidation in which the valence of the metal M increases from 0 to $+n$ by liberating electrons. The other reaction (2.3b), known

as the cathodic reaction, is a reduction in which the oxidation state of hydrogen decreases from +1 to 0, consuming electrons.

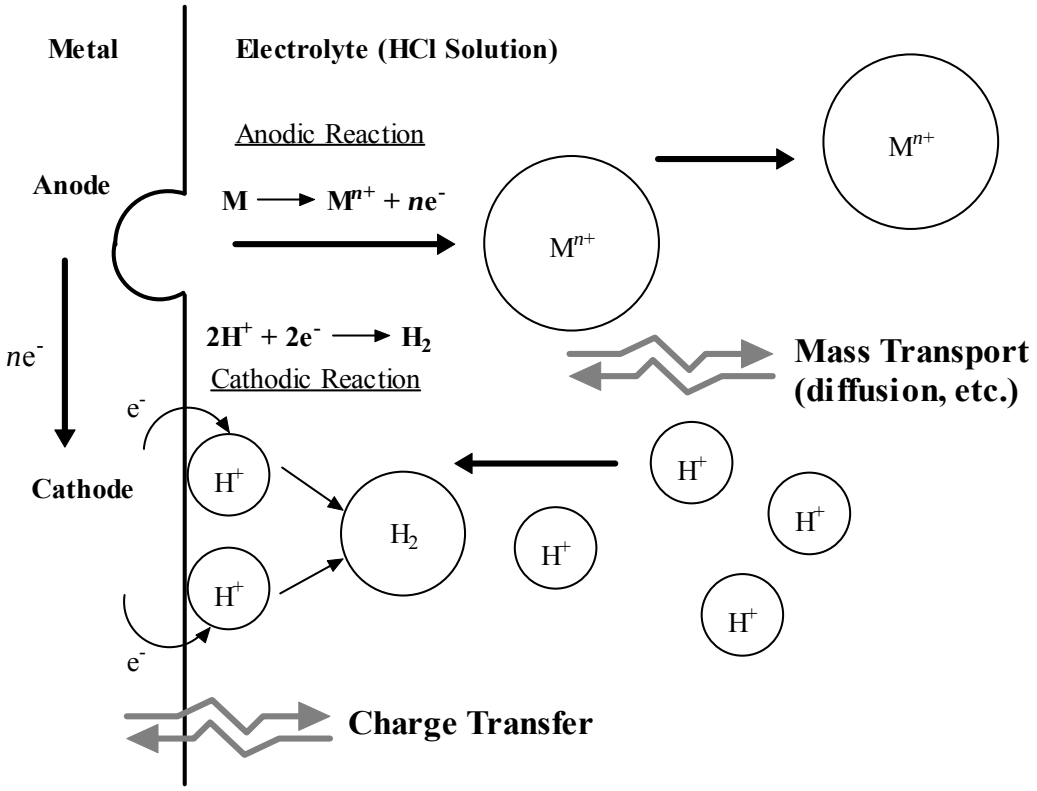

Figure 2.1: Corrosion of metal *M* in hydrochloric acid solution

Thermodynamics and electrochemical kinetics are the foundations of corrosion study. While the former gives an understanding of the energy changes involved in an electrochemical reaction and hence determines the spontaneous direction for the reaction, the latter provides basis of predicting the corrosion rate, which is often our major concern.

According to thermodynamics, for any electrochemical reaction such as equation (2.1), there is always a change of free energy associated with it. The fundamental relationship between the free-energy change, ΔG, and the electrochemical potential, E, is expressed as (Jones, 1996):

$$\Delta G = -nFE \tag{2.4}$$

where n is the number of electrons that is exchanged in the reaction, and F is Faraday's constant, 96,500 coulombs per mole. The potential E is the summation of the two half-cell potentials, e_a and e_c, for the corresponding anodic and cathodic reactions. The half-cell potentials can be determined only by experiments with a common reference electrode (e.g., the standard hydrogen electrode) in a solution of standard concentration. These potentials are called standard electrode potentials, which should be adjusted for actual concentration through the so-called Nernst equation (Jones 1996). The negative sign in equation (2.4) is used to conform to the convention that a positive potential results in a negative free-energy change for a spontaneous reaction. The spontaneous reaction direction is the direction of which the ΔG is negative. The more negative the value of ΔG, the greater the tendency for the reaction to go.

A large negative ΔG does not necessarily mean a high corrosion rate, however. To estimate the corrosion rate, we need to study the reaction kinetics. Similar to the standard half-cell potentials in thermodynamics, a fundamental quantity for reaction kinetics is the exchange current density, the rate of oxidation or reduction at an *equilibrium* electrode expressed in terms of current per unit of area. The exchange current density can also be determined only by experiments. Note when the electrodes are in equilibrium state, no appreciable corrosion will occur. When a metal is corroding in an aqueous environment,

the potential of the metal will no longer be at an equilibrium value. This deviation from equilibrium half-cell electrode potential e_a or e_c is called *polarization*. It is the polarization that causes a net surface reaction rate.

There are two types of electrochemical polarization, activation and concentration, in any electrochemical cell. The effect of the activation polarization is expressed by the Butler-Volmer equation as

$$\eta_{act} = \beta \log\left(i / i_0\right) \tag{2.5}$$

where η_{act} is the polarization or potential change, i_0 the exchange current density in equilibrium, i the actual current density, and β known as the Tafel constant. The activation polarization can occur at both the anode and cathode.

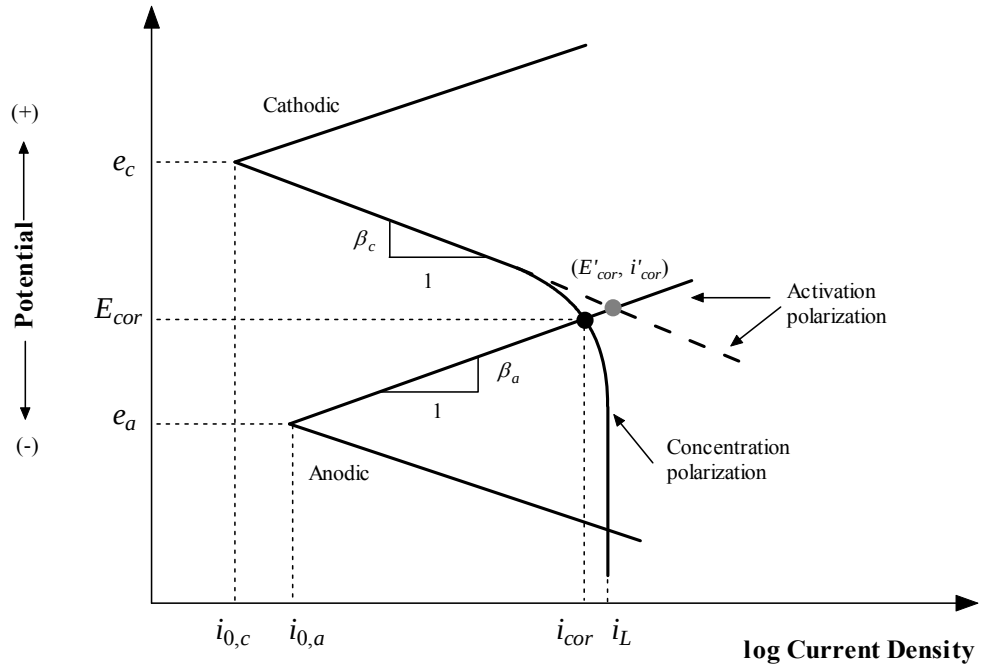

Figure 2.2: Potential-current density relationship in electrochemical kinetics

When the cathodic reagent at the corroding surface is in short supply, a concentration polarization will appear according to the Nernst equation. As shown in Figure 2.1, the hydrogen ions are depleted at the surface, causing a decrease of cathodic potential. The concentration polarization is a function of current density as

$$\eta_{conc} = k \log\left(1 - i / i_L\right) \tag{2.5}$$

where η_{conc} is the concentration polarization, k a constant, and i_L the limiting current density, depending upon the diffusivity of the reacting species.

Corrosion rates can be determined by plotting the polarization curves of the two half-cell electrodes, as shown in Figure 2.2. The cross point of the anodic and cathodic curves correspond to the corrosion potential E_{cor} and corrosion current density i_{cor}. Figure 2.2 also shows the effect of the concentration polarization. Without the concentration polarization in the cathode, the cross point would be the grey one, giving higher corrosion potential E'_{cor} and current density i'_{cor}.

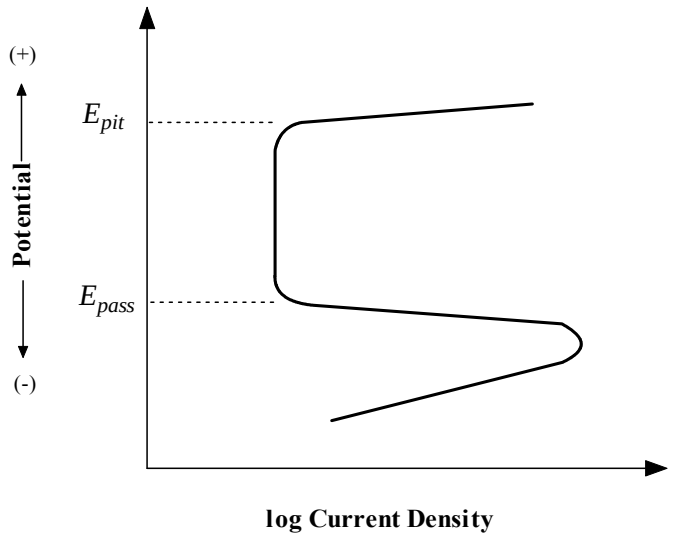

Figure 2.3: Passivity of metal corrosion

It seems from Figure 2.2 that the corrosion rate could increase as long as the corrosion potential and anodic polarization increase. But this is not true. As a matter of fact, in many metals including iron, nickel, chromium and titanium, the corrosion rate decreases above some critical potential as shown in Figure 2.3. This critical potential is called passivity potential, E_{pass}, while the corrosion resistance above E_{pass} despite a high anodic polarization is defined as passivity. Passivity is caused by formation of a thin, protective, corrosion-product surface film that acts as a barrier to the anodic dissolution reaction. At higher potential, the passive film breaks down and some localized corrosion may occur, one type of which is the pitting corrosion that is to be discussed next.

2.1.2 Stages of Pitting Corrosion

Pitting corrosion as a localized attack of metal surface is a complex process with a sequence of stages that includes: (1) pitting initiation or nucleation, (2) meta-stable pitting, (3) stable pitting and (4) repassivation or stable pit death (Strehblow, 2002).

The different stages of pitting corrosion can be best explained in connection with its potential dependence, as shown in Figure 2.4. This potential current curve is measured by electrochemical methods which are conducted by applying potentials stepwise (potentiostatically) and recording the resulting current. Starting with a protective passive film on the metal surface, with the electrochemical potential increasing, the passive film breaks down and pitting corrosion is initiated. At lower potentials, pit initiation is followed by rapid repassivation. This stage is usually referred to as metastable pitting. The resulting current transients differ widely with respect to the peak current height as well as the lifetime. Increasing the potential generally leads to larger current transients

with higher peak currents and longer lifetimes, indicating an extended pit growth period, as shown in the right panels of Figure 2.4. Above a certain potential or potential range, a transition to stable pit growth occurs. Even above the pitting potential, repassivation may still occur, showing the stochastic character of localized corrosion processes.

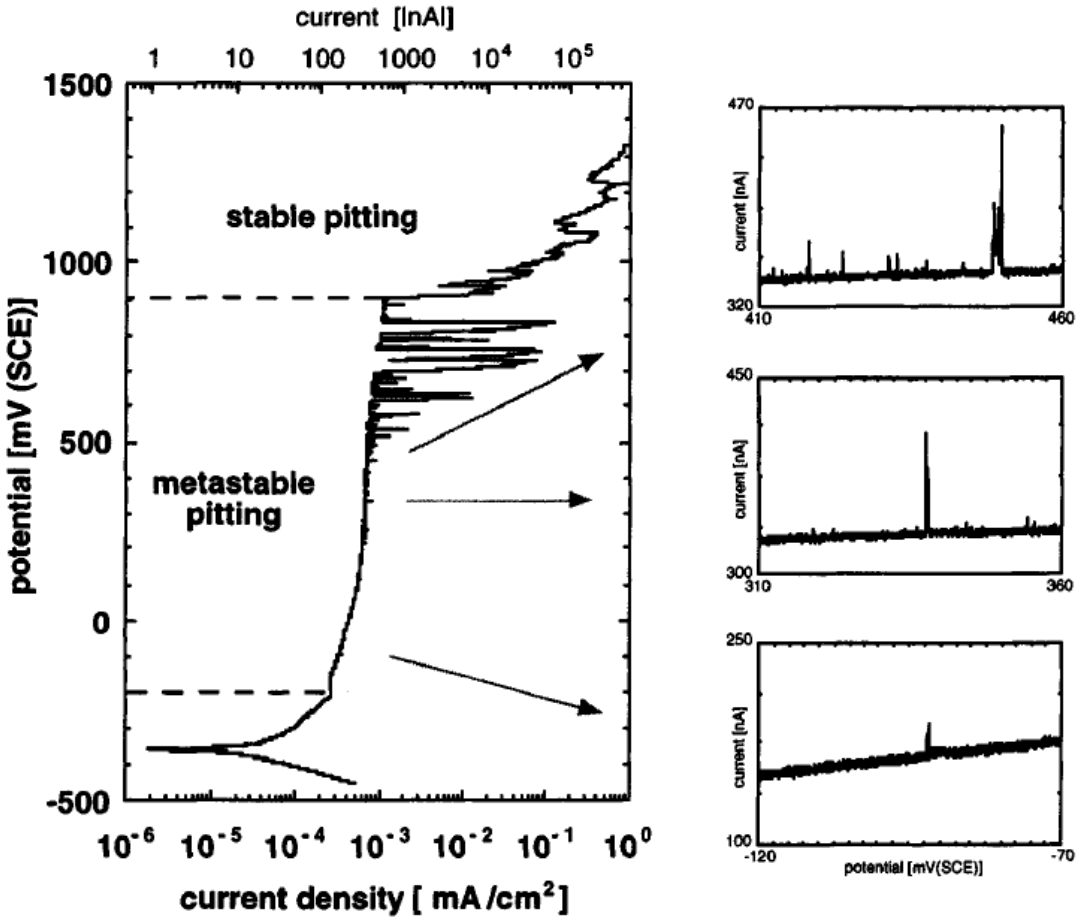

Figure 2.4: Typical potential curve showing the different stages of pitting corrosion (from Bohni, 2000)

2.1.3 Under-deposit Pitting in Steam Generators

To date the mechanism of pitting initiation is not well understood. Despite this, some reasonable explanations can still be made for the pitting corrosion in steam generators tubed with different alloys. This pitting is more accurately described as *outside surface under-deposit corrosion* (Jones, 1996), because most of the pits have been found at the tubesheet and alloy tubes under sludge pile accumulated at the secondary side of the steam generators.

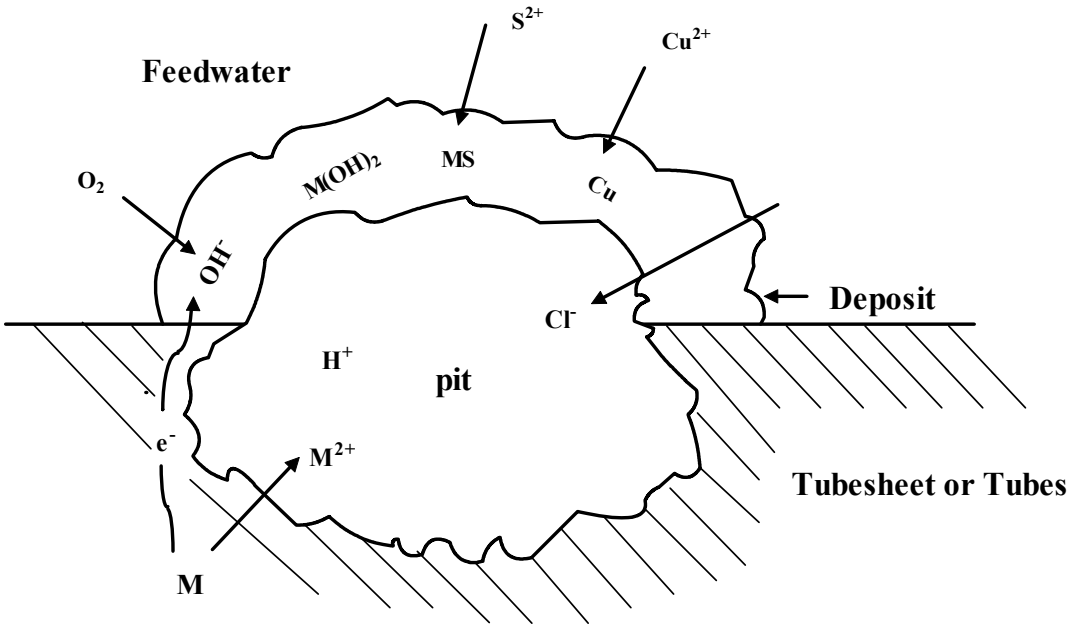

Figure 2.5: Schematic mechanism of the occluded cell in under-deposit pitting

The mechanism of under-deposit pitting initiation is schematically shown in Figure 2.5 (Jones, 1996). The deposits on the outside surface include oxide, hydroxide, and sulfide compounds of several metals such as copper, nickel, chromium and iron. An

important common feature of these deposits is the electrical conductivity. Because of this, the outside surface of the deposits can serve as the cathode for reduction of dissolved oxygen and other dissolved oxidizers from the bulk solutions. The cathodic reactions in the steam generator units include a deferential aeration cell, defined by

$$O_2 + 2H_2O + 4e^- \rightarrow 4OH^- \quad (2.6)$$

as well as the reduction of copper ions from leakage of the condenser tubes:

$$Cu^{2+} + 2e^- \rightarrow Cu \quad (2.7)$$

Moreover, local boiling on the tubesheet or tube surfaces concentrates aggressive species, such as chlorides and sulphides due to condenser leakage and water treatment programs, at sludge pile. The cell electrolyte within the pit is acidified by hydrolysis of chloride or sulfide as

$$M^{2+} + 2H_2O + 2Cl^- \rightarrow M(OH)_2 + 2HCl \quad (2.8)$$

which produce local pH reductions in favor of pit initiation (Bardal, 2004).

So far, we have qualitatively described the physical process of pitting corrosion based mainly on the electrochemical theory. But to quantify the pitting process, we need to consider its stochastic nature. For example, the pitting potential E_{pit} of a metal is not a fixed value --- actually large scatter in data has been found common even under well-control laboratory conditions (Shibata, 2000). Because of the existence of metastable stage as well as the repassivation in both metastable and stable stages, the initiation and growth of pits are of stochastic nature. Therefore, stochastic models are preferred in modeling the process of pitting corrosion. Next we are to review the existing literature on probabilistic modeling of pitting corrosion.

2.2 Probabilistic Modeling of Pitting Corrosion

Statistical and stochastic aspects of pitting corrosion were first studied in the late 1970s (Shibata, 1976; Shibata and Takeyama, 1977). In large-scale engineering structures, the measurement of pit depth and frequency is time consuming and expensive and hence, only the deepest pits is studied since they cause the failure of the system. Thus, the extreme value statistics developed by Gumbel (1958) are widely used in the application of maximum pit depth distribution. Shibata (1991, 1996) reviewed the historical background of the extreme value statistics and the development in the theory and application to corrosion.

Extreme value distributions include three types of asymptotic distributions for an infinite number of samples:

Type I: Gumbel distribution $F(x) \sim \exp\left(-\exp\left[-x\right]\right)$

Type II: Cauchy distribution $F(x) \sim \exp\left(-x^{-k}\right)$

Type III: Weibull distribution $F(x) \sim \exp\left(-\exp\left[\omega - x\right]^{k}\right)$

where x is a random variable representing the maximum pit depth, and k and ω are constants.

When corrosion data for the maximum value are collected, one must decide which type of extreme value distribution should be fitted to the observed data. For this purpose, several methods for a goodness-of-fit (GOF) test, including chi-square test, Komologorov-Simirnov test, correlation coefficient test, and others, can be used. A more systematic way for the GOF test, proposed by Shibata (1996), is to use the shape parameter of a generalized extreme value (GEV) distribution defined by:

$$F(x) = \exp\left(-1 - k\frac{x-\mu}{\sigma}\right)^{1/k}, \qquad \text{for} \quad kx \le \sigma + \mu k \ \text{ and } \ k \ne 0 \tag{2.9}$$

where μ and σ are the location and scale parameters, respectively, and k is the shape parameter. The shape parameter is an indicator of the distribution type by the sign and the absolute value: when $k = 0$, Type I; $k < 0$, Type II and $k > 0$, Type III.

However, the extreme value statistical theory suffers from some limitations. It is static as the time variable is not involved. Furthermore, the assumption for extreme value theory is that the maximum pit depth is independent, which may not be realistic.

Recently, Melchers (2004) studied the pitting corrosion of mild steel in marine immersion environment and presented new field data. A multiphase phenomenological model was proposed for general pit depth as a function of period of exposure. Later he found a bi-modal probability distribution may fit the data better. This finding casts doubts on the conventional use of extreme value theory in representing the uncertainty associated with maximum pit depth (Melchers, 2005). He suggested high dependence of maximum pit depth should be the major reason for such a contradiction.

Shibata (1996) also reviewed the model for pitting generation, for which a homogeneous Poisson process was discussed. Associated with the Poisson distribution of the pit number, an exponential distribution was derived for induction time between two successive occurrences of new pits. Further experimental data on the distribution of the induction time for pit generation revealed that more complicated stochastic models should be presented to fit the data. Thus, Shibata proposed another two types of models as follows:

- Pure birth stochastic models, only considering pit generation events;

- Birth and death stochastic models, assuming stochastic pit generation and subsequent pit repassivation processes.

The latter models seem to be fitted more satisfactorily to the various cases studied experimentally.

Provan and Rodriguez (1989) applied a nonhomogeneous, continuous-time Markov chain to describe the growth of the maximum pit depth in pitting corrosion systems. They also incorporated this Markov model into a failure control system designed to enhance the reliability of structures where pitting corrosion occurs (Rodriguez and Provan, 1989). Let $H(t)$ denote the maximum pit depth at time t. Suppose that when $H(t) \geq h$, the component fails due to the pitting corrosion. Assume that the range $[0, h]$ is divided into n (usually equal-distance) intervals, and each interval is labeled as a discrete state of the pit depth. Then the growth of pitting corrosion follows a Markov Chain if the following equation holds:

$$\Pr\left(H_{t+1} = j \middle| H_t = i, H_{t-1} = k, \cdots, H_0 = i_0\right) = \Pr\left(H_{t+1} = j \middle| H_t = i\right) \tag{2.11}$$

That is, given the present state $H_t = i$, the future $H_{t+1} = j$ is conditionally independent of the past $H_{t-1} = k, \cdots, H_0 = i_0$. A continuous-time Markov chain is fully characterized by its transition intensity matrix

$$Q^{(t)} = \begin{bmatrix} q_{11}^{(t)} & q_{12}^{(t)} & q_{13}^{(t)} & \cdots & q_{1n}^{(t)} \\ 0 & q_{22}^{(t)} & q_{23}^{(t)} & \cdots & q_{2n}^{(t)} \\ 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & q_{n-1n-1}^{(t)} & q_{n-1n}^{(t)} \\ 0 & 0 & 0 & \cdots & q_{nn}^{(t)} = 0 \end{bmatrix} \tag{2.12}$$

where $q_{ij}^{(t)} = \Pr\left(H_{t+1} = j \middle| H_t = i\right)$ for $i < j$ and $q_{ii}^{(t)} = -\sum_{j=i+1}^{n} q_{ij}^{(t)}$, $i = 1, \cdots, n-1$. Note that $q_{nn}^{(t)} = 0$ as state n is the absorbing state, i.e. once the pitting depth reaches the state "n", the component fails and it can never revert to the early states automatically. According to Kolmogorov-Chapman equation of Markov chains, the conditional probability distribution satisfies the following differential equations:

$$\frac{\partial P(\tau, t)}{\partial t} = -Q(t)P(\tau, t) \tag{2.13}$$

Sheikh et al. (1990) treated pitting corrosion as a time-dependent damage process characterized by an exponential or logarithmic pit growth and demonstrated that the time-to-first-leak for the pipeline is distributed according to a Weibull extreme value reliability model.

Mola et al. (1990) proposed a stochastic model applied to a process where pits initiate at reacting inclusions in the metal yielding the development of quasi-hemispherical pits. The predictions of the model were compared to experimental data obtained from 361SS specimens in a borate-phosphate buffer containing sodium chloride.

2.3 Summary

Corrosion is an electrochemical reaction in nature. Pitting corrosion is a localized form of corrosion and consists of two major steps: pitting generation processes and pitting growth processes. Both steps are stochastic in nature due to the existence of meta-stable pitting, uncertain pitting potential and varying repassivation of stable pits. Several probabilistic

and statistical models have been proposed in the literature for modeling the pitting corrosion. However, a model validated by field data is still lacking.

Chapter 3

Bayesian Modeling of Pitting Corrosion

In this chapter a probabilistic model of pitting corrosion is presented to characterize the inherent randomness of the pitting process and measurement uncertainties of the in-service inspection (ISI) data. A Bayesian method is proposed for estimating the model parameters. When the data is limited and contaminated by measurement uncertainties, Bayesian methodology is a good alternative for calibrating and updating the model. The proposed model is used to estimate the distribution of maximum pit depth, which is the main interest of the pitting corrosion model.

The formulation of the problem and model specifications are discussed in Section 3.1 and 3.2. Likelihood function of model parameters for the ISI data is derived in Section 3.3, which is followed by discussions of the Bayesian method and prior distributions in Section 3.4. The chapter is concluded with a few remarks on Bayesian computation, which is the major topic of Chapter 4.

3.1 Formulation of Problem

For the alloy tubes suffering from pitting corrosion, actual pits occur randomly on the tube surface with a finite number and different sizes. In this thesis, a pit is treated as a one-dimensional damage, i.e., the pit depth. The modeling of the pits includes both the pit number and the pit size. An ideal model would relate the two quantities to the dynamics of pitting generation and pitting growth. But such a model needs more, reliable ISI data to calibrate, which is lacking in our study. Rather, we pragmatically take a static viewpoint.

That is, the number of new pits between two successive inspection campaigns is assumed to be a random variable, and once the pits occur, they do not grow so that they can be modeled by a random variable. Once these two random variables are well characterized, the distribution of the maximum pit size can be easily derived.

But the problem is confounded by the measurement uncertainties of the ISI tools for steam generators. The measurement uncertainties consist of the uncertainty of detection and measurement errors of pit sizes when detected (Figure 3.1). Because of limited detection capability of the inspection tools, some pits especially those with small size may not be detected. For those having been detected, the actual readings of their depth from the tool suffer from measurement errors.

The uncertainty of detection is often characterized probability of detection (POD). Let H denote the pit depth and use a binary random variable D to indicate whether or not a pit is detected, i.e.,

$$D = \begin{cases} 1, & \text{if the pit is detected;} \\ 0, & \text{otherwise.} \end{cases} \tag{3.1}$$

The POD is then defined as a conditional probability depending on the pit depth:

$$\text{POD}(h) = P(D = 1 \,|\, H = h) \tag{3.2}$$

Note that the POD function is not a cumulative distribution function, although for most modern ISI tools, POD(h) is 0 when h is 0 and it is 1 when h is big enough. In Figure 3.1 a step function is plotted for illustration purpose only. It is shown that the pits cannot be detected when the size is less than a certain threshold value; greater than that, the pits can be 100% detected. More discussions of the POD function have been presented in Section 3.2.4.

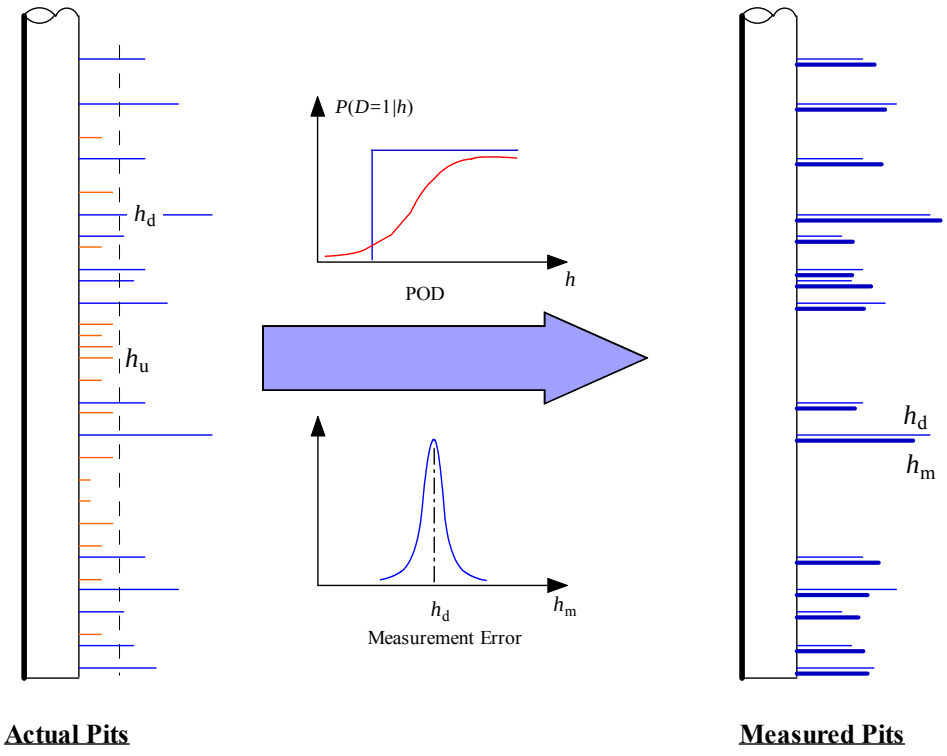

Figure 3.1: Measurement Uncertainties of ISI Data

After the detection with uncertainty, the actual pits are divided into two groups: the detected pits, denoted by $\mathbf{h_d} = \left(h_{d1}, h_{d2}, \cdots, h_{d,n_d}\right)$, and the undetected pits, denoted by $\mathbf{h_u} = \left(h_{u1}, h_{u2}, \cdots, h_{u,n_u}\right)$, where n_d and n_u denote the number of detected and undetected pits, respectively. The actual total number of pits is denoted by n and clearly $n = n_d + n_u$.

For the detected pits, the measured pit depth, denoted by $\mathbf{h_m} = \left(h_{m1}, h_{m2}, \cdots, h_{m,n_d}\right)$, differs from their actual depth by an additive random measurement error **e**, i.e.,

$$\mathbf{h_m} = \mathbf{h_d} + \mathbf{e} \tag{3.3}$$

where $\mathbf{e} = (e_1, e_2, \ldots e_{nd})$. In Figure 3.1, the measurement error with a normal distribution is illustrated as an example.

To summarize the problem, from the ISI, we obtain the number and depth measurements of detected pits n_d and $\mathbf{h}_m$. Based on these data, we want to estimate the actual number of pits n and the actual pit depth $\mathbf{h}$. The POD and measurement errors may be available as background information from the inspection tools. Our final goal is to use these estimations to estimate the maximum pit depth while eliminating the effects of POD and measurement errors.

3.2 Model Specifications

3.2.1 Assumptions

Several assumptions are made as the following:

(i) The actual pit numbers N for an inspection campaign is a random variable and follows a Poisson distributions with mean λ, i.e.,

$$\Pr(N = n) = \frac{\lambda^n}{n!} e^{-\lambda} \tag{3.4}$$

for $n = 0, 1, 2, \ldots$.

(ii) The actual sizes of pits $H_1, H_2, \cdots, H_N$ are independent random variables and they follow a Weibull distribution.

(iii) The pit number N_i and pit size H_i are independent.

(iv) A logistic function with threshold is assumed for the POD.

(v) The measurement errors, e_1, e_2, …, are assumed to be independent and identically Gaussian distributed with zero mean and known variance.

Next we discuss the assumed distributions and functions in detail.

3.2.2 Weibull Distribution

A Weibull distribution is defined with the probability density function (PDF) as follows:

$$f(h) = \gamma\beta h^{\beta-1} \exp(-\gamma h^{\beta}), \quad h > 0 \tag{3.5}$$

in which $\gamma > 0$ is the scale parameter and $\beta > 0$ is the shape parameter that controls the shape of the PDF.

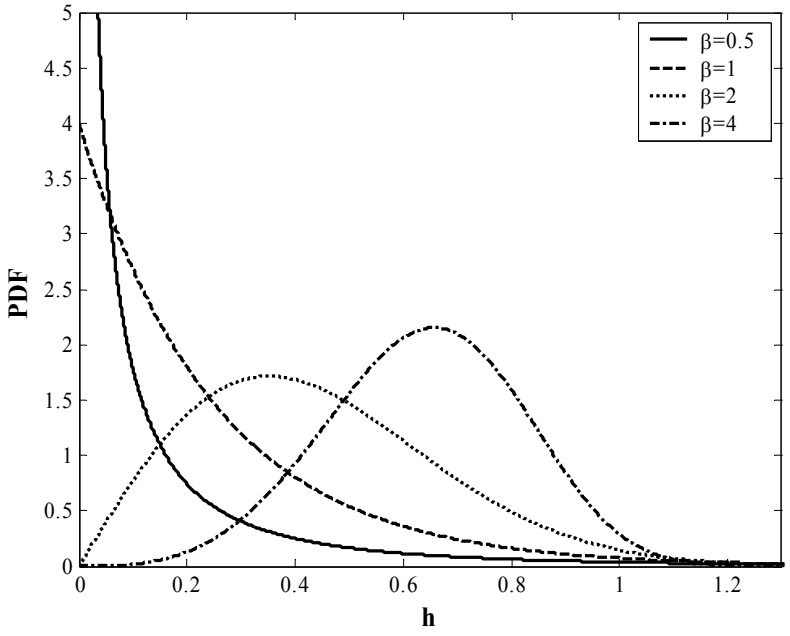

Figure 3.2: PDF of the Weibull distribution with different shape parameters

The reason for the Weibull distribution as the distribution of pit depth is because of its flexibility. It can mimic the behaviour of other statistical distributions such as the normal and exponential distributions. Figure 3.2 shows the effect of the shape parameter on the shape of the Weibull PDF, where $\gamma = 4$ fixed. Depending on the value of β, the shape of the PDF takes a variety of forms. When $\beta = 1$, the Weibull distribution reduces to an exponential distribution whereas for $\beta > 1$, the Weibull distribution is of bell shape and approximates normal and log-normal distributions.

3.2.3 Probability of Detection

A logistic function with threshold is used for POD:

$$\mathrm{POD}(h) = \begin{cases} 1 - \dfrac{1 + \exp(-qh^*)}{1 + \exp\left[q(h - s - h^*) \right]}, & \text{if } h > s \\ 0, & \text{otherwise} \end{cases} \tag{3.6}$$

where s is the threshold of detection, introduced previously.

Figure 3.3 displays the POD curves with different values of h^* where $s = 0.05$ and $q = 20$. It is shown that when $h = h^*$, the POD at three cases are all about 0.2. Hence, h^* can be considered as a location parameter indicating the starting point of "good" detection.

Figure 3.4 displays the POD curves with different values of q where $s = 0.05$ and $h^* = 0.2$. Clearly, q is an index measuring the quality of detection. Bigger q implies better detectability.

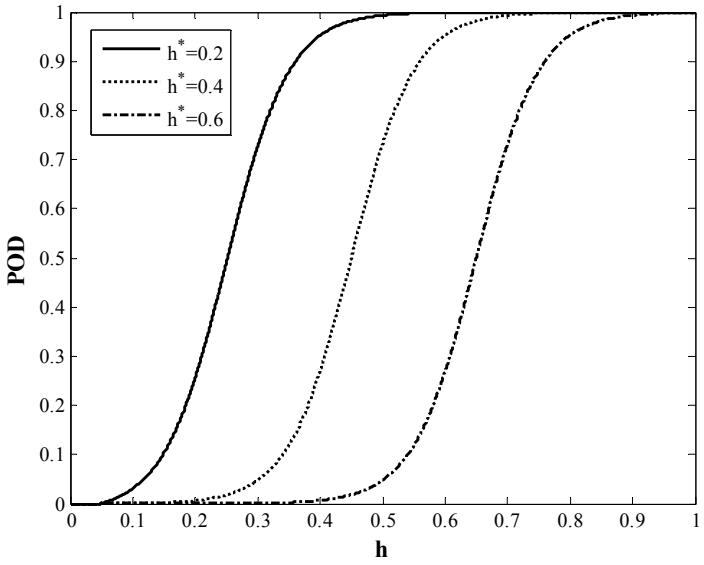

Figure 3.3: POD curves with different values of h^*

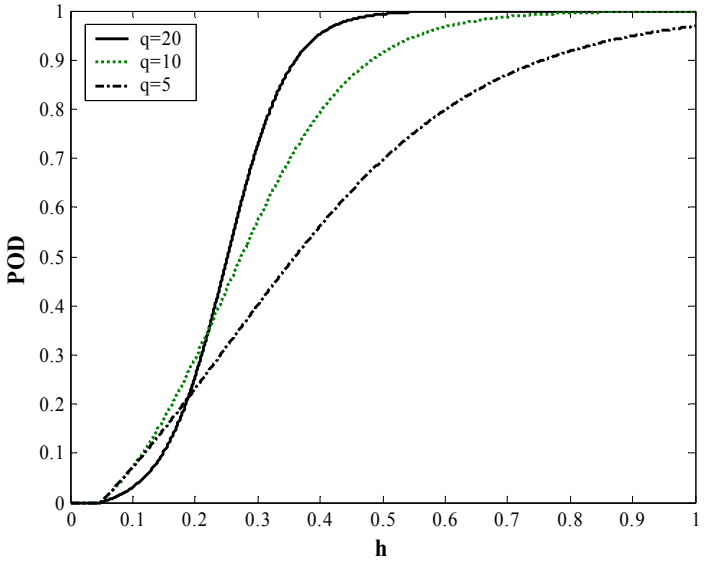

Figure 3.4: POD curves with different values of q

3.2.4 Model of Measurement Errors

A normal distribution of the measure error E is presented as following:

$$f_E(e) = \frac{1}{\sigma_E \sqrt{2\pi}} \exp\left(-\frac{e^2}{2\sigma_E^2} \right) \tag{3.7}$$

where mean is zero and variance is σ_E^2.

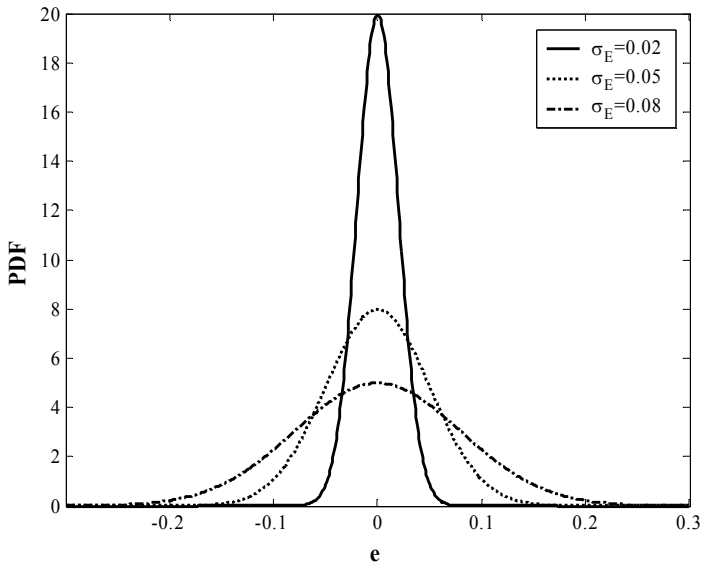

Figure 3.5: Measurement error curves with different values of σ_E

Figure 3.5: displays the measurement error curves with different values of σ_E. The standard deviation σ_E describes the width of the bell shape of the normal distribution. There is no doubt that a smaller standard deviation means more accurate measurement.

3.2.5 Distribution of Maximum Pit Depth

Let us denote the maximum pit depth by $H_{\max}$. Based on the assumptions made above, its cumulative distribution function can be derived as

$$\begin{aligned} F_{H_{\max}}(h) &= \Pr\left(H_{\max} \le h\right) = \Pr\left[\bigcap_{i=1}^{N}\left(H_i \le h\right)\right] = \sum_{n=0}^{+\infty}\left[\Pr\left(H_i \le h\right)\right]^n \Pr\left(N = n\right) \\ &= \sum_{n=0}^{+\infty}\left[F_H(h)\right]^n \frac{\lambda^n}{n!} e^{-\lambda} = \exp\left\{-\lambda\left[1 - F_H(h)\right]\right\} = \exp\left\{-\lambda \exp\left(-\gamma h^{\beta}\right)\right\} \end{aligned} \tag{3.8}$$

Equation (3.8) shows that the CDF of the maximum pit depth is a function of the three parameters λ, γ and β.

3.3 Likelihood function

Maximum likelihood method is considered to be a classic way to estimate the unknown model parameters based on the observed or measured data. In our case, the parameters include: the scale parameter γ and shape parameter β of $f_H(h)$, the intensity parameter λ of the Poisson distribution for the number of pits. Although we are not going to explore the possibility of ML method, we derive the likelihood function in this section as it is one of the two components of the Bayesian inference, which will be discussed in the next section.

Let us denote the model parameters by a vector $\boldsymbol{\xi}=\left[\boldsymbol{\theta}=\left(\gamma, \beta\right), \lambda\right]$. Given a sample of data $\mathbf{h}_m = h_{m1}, h_{m2}, \cdots, h_{m,n_d}$, of the measured pits from an inspection campaign, the likelihood function of the model parameters is defined as the joint probability density function of $\mathbf{H}_m$ at the value of $\mathbf{h}_m$ and the probability of $N_d = n_d$, i.e.,

$$L(\boldsymbol{\xi}\,|\,\mathbf{h}_m, n_d) = f_{\mathbf{H}_m}\left(h_{m1}, h_{m2}, \cdots, h_{m,n_d}\right) \times \Pr\left(N_d = n_d\right) \tag{3.9}$$

The product of the two terms results from assumption (iii) in Section 3.2.1. Since the actual pit sizes H_i (i = 1, …, n_d) are independent and so are the measurement errors, the measured pit sizes H_{mi} are also independent. Therefore, we have

$$f_{\mathbf{H}_m}\left(h_{m1}, h_{m2}, \cdots, h_{m,n_d}\right) = \prod_{i=1}^{n_d} f_{H_m}\left(h_{mi}\right) \tag{3.10}$$

Next we derive the PDF of measured pit size $f_{H_m}(h)$ and probability of observed number of pits $\Pr\left(N_d = n_d\right)$.

According to (3.3), $H_m = H_d + e$, the PDF of H_m is the convolution of the PDF of H_d and the measurement error e, i.e.,

$$f_{H_m}(h_m) = \int_0^\infty f_{H_d}(y)\, f_E(h_m - y)\mathbf{d}y \tag{3.11}$$

in which $f_E(.)$ can be found in (3.7). But the PDF of the detected pit depth H_d is a conditional probability, which is expressed as

$$f_{H_d}(h) = f_H(h\,|\,D = 1) = \frac{f\left(\mathrm{D}{=}1\,|\,H = h\right) f_H(h)}{\Pr\left(\mathrm{D}{=}1\right)} \tag{3.12}$$

The first term of the numerator is just the POD; c.f. (3.2). The denominator is the unconditional pit detection probability, which depends on the chosen actual pit depth density $f_H(h)$ and is the function of $\boldsymbol{\theta}$ as

$$p_f(\boldsymbol{\theta}) = \Pr(D = 1) = \int_0^\infty \mathrm{POD}(h)\, f_H(h)\mathrm{d}h \tag{3.13}$$

Then the Equation (3.12) turns to

$$f_{H_d}(h) = \frac{1}{p_f(\boldsymbol{\theta})} \mathrm{POD}(h) f_H(h) \tag{3.14}$$

in which $f_H(h)$ and POD(h) can be found in (3.5) and (3.6), respectively. Substituting Equation (3.14) into Equation (3.11), we have

$$f_{H_m}(h_m) = \frac{1}{p_f(\boldsymbol{\theta})} \int_0^\infty POD(y) f_H(y) f_E(h_m - y) \mathbf{d}y = \frac{\int_0^\infty POD(y) f_H(y) f_E(h_m - y) \mathbf{d}y}{\int_0^\infty POD(y) f_H(y) \mathbf{d}y} \tag{3.15}$$

For the probability of the number of detected pits, we first notice that the number of total pits follows a Poisson distribution. Given the number of total pits $N = n$, the number of detected pits is just a Binomial distribution with "success probability" (here it is the detection probability) $p_f(\boldsymbol{\theta})$, denoted by $Bi\left(. \middle| n,\, p_f(\boldsymbol{\theta})\right)$. Therefore, we have

$$\begin{aligned} \Pr(N_d = n_d) &= \sum_{n=0}^{\infty} \Pr(N = n)\, Bi\left(n_d \middle| n,\, p_f(\boldsymbol{\theta})\right) \\ &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \binom{n}{n_d} \left[p_f(\boldsymbol{\theta})\right]^{n_d} \left[1 - p_f(\boldsymbol{\theta})\right]^{n-n_d} \end{aligned} \tag{3.16}$$

Substituting both (3.15) and (3.16) into (3.9), we could finally get the likelihood function of the parameters γ, β and λ.

Now the difficulty of using the maximum likelihood method is clear. First of all, the likelihood function is very complicated. Although the pit number and pit sizes are assumed independent, the parameters for these two random quantities are intertwined by the POD in terms of $p_f(\boldsymbol{\theta})$ and hence λ cannot be separately estimated from the estimation of γ and β. Secondly, the evaluation of the likelihood function involves two integrations and infinite summations. Although numerical quadrature and truncated summations can be used for approximation, the maximization procedure could be easily

ruined by the round-off errors and hung up without convergence. Thirdly, the ISI data is often limited. Even in the case that a maximum likelihood estimate of the parameters can be obtained, their confidence intervals could be so wide that a point estimate is of very little practical significance.

In contrast, Bayesian method is a good alternative for this situation. As we can see next, the integration can be avoided in Bayesian method by a data augmentation technique. Furthermore, the information from the limited data is enhanced by prior information expressed by prior distributions of the parameters. Also the presentation of the estimation as a posterior distribution seems more reasonable for decision-making in the life-cycle management.

3.4 Bayesian Choices

3.4.1 Basic Concepts

From the Bayesian perspective, the estimation of unknown model parameters $\boldsymbol{\xi} = \left(\xi_1, \xi_2, \ldots, \xi_q\right)$ is an update of its prior distribution based on information inferred from the given data through the assumed probabilistic model (Robert 2001). The prior distribution of the parameters is denoted by $\pi(\boldsymbol{\xi})$ and the information acquired from the data **D** is often expressed by the likelihood function $L\left(\boldsymbol{\xi}\middle|\mathbf{D}\right)$. The Bayes rule yields a new distribution for the parameters as

$$p(\boldsymbol{\xi}\middle|\mathbf{D}) = \frac{\pi(\boldsymbol{\xi})L\left(\boldsymbol{\xi}\middle|\mathbf{D}\right)}{\int \pi(\boldsymbol{\xi})L\left(\boldsymbol{\xi}\middle|\mathbf{D}\right)d\boldsymbol{\xi}} \tag{3.17}$$

And the new distribution is called the posterior distribution of the parameters. Note the denominator in (3.17) does not depend on $\boldsymbol{\xi}$ and, with fixed data, is a normalization constant. Hence quite often, it is sufficient to express the posterior as the following non-normalized form:

$$p(\boldsymbol{\xi}|\mathbf{D}) \propto \pi(\boldsymbol{\xi})L\left(\boldsymbol{\xi}|\mathbf{D}\right) \tag{3.18}$$

Once we have the posterior distribution, the inference about the parameters can be made from it. For example, the posterior mean of a function of $\boldsymbol{\xi}$, $g(\boldsymbol{\xi})$, is

$$E\left(g(\boldsymbol{\xi})|\mathbf{D}\right) = \frac{\int g(\boldsymbol{\xi})\pi(\boldsymbol{\xi})L\left(\boldsymbol{\xi}|\mathbf{D}\right)d\boldsymbol{\xi}}{\int \pi(\boldsymbol{\xi})L\left(\boldsymbol{\xi}|\mathbf{D}\right)d\boldsymbol{\xi}} \tag{3.19}$$

Also the marginal distribution of a component of the parameter vector, say ξ_1, can be evaluated as

$$p\left(\xi_1|\mathbf{D}\right) = \frac{\int \pi(\xi_1)L\left(\xi_1|\mathbf{D}\right)d\xi_2\cdots d\xi_q}{\int \pi(\xi_1)L\left(\xi_1|\mathbf{D}\right)d\xi_1 d\xi_2\cdots d\xi_q} \tag{3.20}$$

For the pitting corrosion problem, we have already derived the likelihood function for the parameters in the proposed model. In order to use the Bayesian method, we need to specify the prior distributions, which are to be discussed next.

3.4.2 Prior Distributions

There are many ways to specify the prior distributions, including elicitation of expert opinion, non-informative priors, conjugate priors, etc. Among them, conjugate prior distributions are often preferred for both conceptual and computational reasons. We shall

not discuss the choice of priors in detail; rather, we simply list the prior distributions we choose for the analysis. For more discussions of this issue, refer to, for example, Aven (2003), Robert (2001), Gelman et al. (2004).

For the intensity parameter λ for the number of pits, a Gamma distribution with parameters a and b is selected, i.e.,

$$\pi(\lambda) = \frac{b^a \lambda^{a-1}}{\Gamma(a)} e^{-\lambda b} \tag{3.21}$$

This prior is considered conjugate with the data, which are assumed a Poisson distribution. If the actual number of pits were observed as n, then the posterior distribution of λ would be

$$p(\lambda) \propto \pi(\lambda) \Pr(N = n) \propto \lambda^{a+n-1} e^{-(b+1)\lambda} \tag{3.22}$$

which is another Gamma distribution with parameters $(a + n)$ and $(b + 1)$.

The parameters of the Weibull distribution for the actual pit depth include the scale parameter γ and shape parameter β. For γ, the following Gamma distribution with parameters A and B is used as the prior distribution:

$$\pi(\gamma) = \frac{B^A \gamma^{A-1}}{\Gamma(A)} e^{-\gamma B} \tag{3.23}$$

For β, we consider a Beta distribution as the prior distribution with the parameters r, t, β_l and β_r:

$$\pi(\beta) = \frac{\Gamma(r+t)}{\Gamma(r)\Gamma(r)(\beta_r - \beta_l)} \left(\frac{\beta - \beta_l}{\beta_r - \beta_l} \right)^{r-1} \left(\frac{\beta_r - \beta}{\beta_r - \beta_l} \right)^{t-1} \tag{3.24}$$

for $\beta_l < \beta < \beta_r$. It can be shown that the posterior distribution of γ given β is again a gamma distribution and thus the gamma prior is conjugate, whereas the prior of β is not conjugate.

3.5 Remarks

Equations (3.19) and (3.20) involve high-dimensional integrations. Although conjugate priors lead to analytical solutions for the posterior distribution, this is only for rare applications. More often than not, analytic evaluations of the integrations in Equation (3.19) and (3.20) are just impossible, which may be the major disadvantage of a Bayesian method. Take the proposed pitting corrosion model for example, even we intentionally choose conjugate prior distributions as many as possible, the complicated likelihood function makes the posteriors no more conjugate. Fortunately, a newly developed simulation technique, named Markov Chain Monte Carlo, can effectively address the problems of Bayesian computation. This technique will be presented in Chapter 4.

Chapter 4

Markov Chain Monte Carlo Method

This chapter briefly reviews basic concepts of Markov chain Monte Carlo (MCMC) method at first. Then the iterative sampling procedure for the proposed model, presented in Chapter 3, is described. Simulation experiments are performed to illustrate the behavior of the Bayesian method.

4.1 Introduction to MCMC

In Bayesian statistics, high-dimensional integration is a challenging task, as discussed in Chapter 3. Monte Carlo (MC) simulation is an effective technique for such a task (Rubinstein, 1981). By drawing random samples from prescribed distributions, the integral can be estimated from sample statistics. More specifically, Monte Carlo simulation evaluates the following integration

$$E\left[g(x)\right] = \int g(x)\,p(x)dx \tag{4.1}$$

by drawing N_{sim} samples $\left\{X_t, t = 1, \cdots, N_{sim}\right\}$ from the distribution $p(x)$ and then approximating it by

$$\hat{g} \approx \frac{1}{N_{sim}} \sum_{t=1}^{N_{sim}} g(X_t) \tag{4.2}$$

Laws of large numbers ensure that the approximation can be made as accurate as desired by increasing the simulated sample size N_{sim}. When the samples $\left\{X_t\right\}$ are independent, the variance of the estimate is

$$Var(\hat{g}) \approx \frac{1}{N_{sim}} Var[g(X)] \tag{4.3}$$

The square root of the variance is just the standard error of the estimate and is used for assessing the accuracy of the estimation.

In general, drawing samples $\{X_t\}$ independently from the distribution *p*(.) is not feasible, since *p*(.) can be quite non-standard. For example, the posterior distribution in equation (3.18) is only partially known. To completely portray the distribution, its normalization constant, which itself is an integral, needs to be evaluated. In such cases, the conventional sampling methods such as inverse function methods are not applicable. However, to evaluate $E[X_t]$ the $\{X_t\}$ need not be independent. A stationary dependent sequence is as good as the independent one as long as the marginal distribution follows the prescribed distribution. One way of doing this is through a Markov chain having the distribution *p*(.) as its stationary distribution. This method is called MCMC.

MCMC is a general simulation technique based on drawing samples iteratively from proposed distributions and then correcting those draws in each step of the process to better approximate the target posterior distribution when this target distribution cannot be directly sampled (Gilks, 1996; Robert, 2001). The key to MCMC is to construct a Markov chain of which the stationary distribution is the specified posterior distribution and to run the simulation long enough so that the distribution of the current draws is close enough to this stationary distribution.

A random sequence, $\{X_0, X_1, X_2, \cdots\}$, is Markovian when given X_t, the next state X_{t+1} is conditionally independent of the history of the chain $\{X_0, X_1, X_2, \cdots X_{t-1}\}$, i.e.,

$$\Pr\left(X_{t+1}\middle|X_t, X_{t-1}, \cdots, X_1, X_0\right) = \Pr\left(X_{t+1}\middle|X_t\right) \tag{4.4}$$

The probability $\Pr\left(X_{t+1}\middle|X_t\right)$ is called transition probability and is abbreviated as P_{ij} for $\Pr\left(X_{t+1} = j\middle|X_t = i\right)$. To generate a sample sequence, we only need to specify the initial state and its transition probability. An important distribution of the Markov chain is its stationary distribution. A stationary distribution π satisfies

$$\sum_i \pi(i) P_{ij} = \pi(j) \tag{4.5}$$

in which the summation takes for all the state space *i*. There is a strict mathematical theory for the conditions of the transition probability for existence and uniqueness of the stationary distribution. For details, refer to Roberts (1996).

The key reason that MCMC works is that, after some simulation steps, the Markov chain will gradually forget its initial state and eventually converge to its stationary distribution, which is the target distribution we want to sample. The proof of this convergence was presented by Gelman (2004) and will not be discussed here.

Thus, after a sufficiently long burn-in of say N_b iterations, points $\left\{X_t; t = N_b + 1, \cdots, N_{sim}\right\}$ will be stationary with the marginal distribution being approximately the target distribution. Discarding the Burn-in samples, we have

$$\hat{g} \approx \frac{1}{N_{sim} - N_b} \sum_{t=N_b+1}^{N_{sim}} g(X_t) \tag{4.6}$$

We have to pay some price for the easier sampling, however. In addition to the discarded burn-in samples, the estimate (4.6) is less efficient than that in the independent case. Indeed, the variance of the estimate is

$$
\begin{aligned}
Var\left(\hat{g}\right) &\approx \frac{1}{\left(N_{sim}-N_b\right)^2}\sum_{s=N_b+1}^{N_{sim}}\sum_{t=N_b+1}^{N_{sim}} Cov\left[g(X_s),g(X_t)\right] \\
&= \frac{1}{\left(N_{sim}-N_b\right)} Var\left[g\left(X\right)\right]\left\{1+2\sum_{t=N_b+1}^{N_{sim}}\left(1-\frac{t}{N_{sim}-N_b}\right)\rho_t\right\}
\end{aligned}
\tag{4.7}
$$

where ρ_t is the coefficient of correlation between $g\left(X_s\right)$ and $g\left(X_{s+t}\right)$. Hence, when $\rho_t > 0$ the variance of the estimate is larger than that in the independence case in which $\rho_t = 0$.

It is now clear that MCMC should be an effective simulation technique. Next we discuss how to construct the transition probability so that its stationary distribution is the target distribution. Two algorithms --- Gibbs Sampler and Metropolis-Hasting algorithm --- are discussed mainly through examples. Theoretical discussions of the algorithms are beyond the scope of the thesis and can be referred to, for example, Gilks et al. (1996), Roberts (2001) and Gelman (2004).

4.2 Gibbs Sampler

4.2.1 Description of Algorithm

Gibbs sampler is also called alternating conditional sampling and it is suitable for cases where the conditional distribution can be easily sampled. Suppose the parameter vector θ includes *d* components or can be grouped into *d* sub-vectors, i.e., $\theta = (\theta_1, \cdots \theta_d)$. At each iteration *t*, rather than taking samples once for all components, the Gibbs sampler draws samples of each θ_j^t in turn from the conditional distribution given all the other components of θ, $p(\theta_j \left| \theta_{-j}^{t-1} \right.)$, where θ_{-j}^{t-1} represents all the components of θ, except for

θ_j, at their current values, i.e., $\theta_{-j}^{t-1} = (\theta_1^t, \cdots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, \cdots, \theta_d^{t-1})$. This algorithm is very effective when the parameters or some of the parameters are conditionally conjugate. For a simple example, suppose the prior distributions for the mean and variance of a Normal random variable are Normal and inverse Gamma, respectively. Then given the mean, the conditional posterior distribution for the variance is an inverse Gamma, whereas given the variance, the conditional posterior of the mean is a Normal distribution. Since the two conditional posterior distributions are both known, we could, if we want, draw samples using the Gibbs sampler.

For many practical problems, however, not all but only a part of the conditional posterior distributions of the parameters are known and easy to sample directly from them. In those cases the Gibbs sampler is often combined with the Metropolis-Hastings algorithm, which will be discussed in the next section.

4.2.2 Example: A Bivariate Normal Distribution

Consider a bivariate normal random vector

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \tag{4.8}$$

To take a sample of it, we can draw two independent random variates from the standard normal distribution and then take a linear transformation of the two variates. Instead of doing so, we here use a Gibbs sampler to take samples of the bivariate normal distribution.

Given one component of the random vector, the other component is again a normal distribution. That is,

$$X_1 \left| (X_2 = x_2) \sim N\left(\rho x_2, 1-\rho^2\right) \right.$$
$$X_2 \left| (X_1 = x_1) \sim N\left(\rho x_1, 1-\rho^2\right) \right. \tag{4.9}$$

Using the Gibbs sampler, we start with a given points (x_1^0, x_2^0) and then use the above univariate, conditional normal distributions to take samples for x_1 and x_2 alternately. The effects of starting points will be eliminated along the increase of iteration steps, as illustrated in Figure 4.1 where the first 10 iterations of four different sample paths are plotted. Although they start at much dispersed points, the trend of convergence to the center area is common.

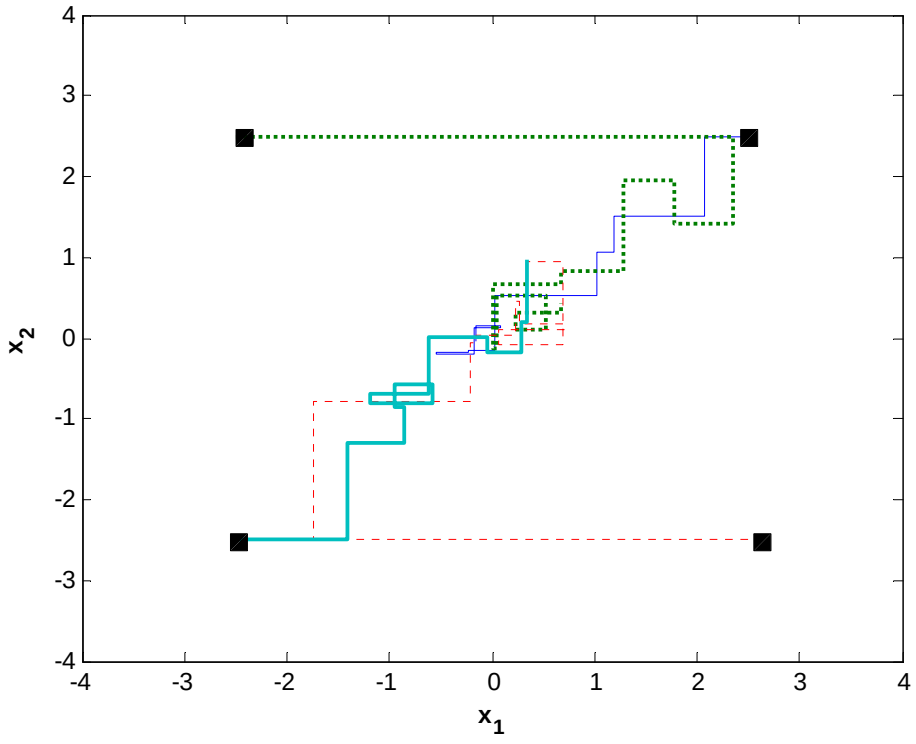

Figure 4.1: First 10 steps of the Gibbs sampler with four starting points

The first 500 steps of the four sample paths are plotted in Figure 4.2. It is clear that the sequences reach approximate convergence. Figure 4.3 shows the samples after 500 steps and the contour of the bivariate normal PDF. The samples match the contour very well, meaning the stationary distribution obtained by the Gibbs sampler is the target distribution---bivariate normal.

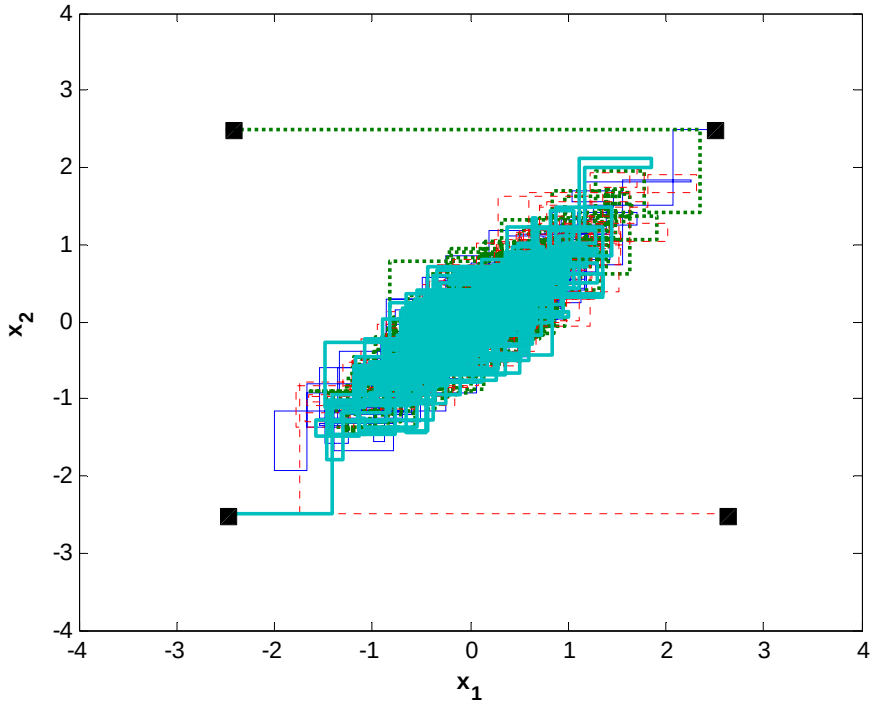

Figure 4.2: First 500 steps of the Gibbs sampler with the four starting points

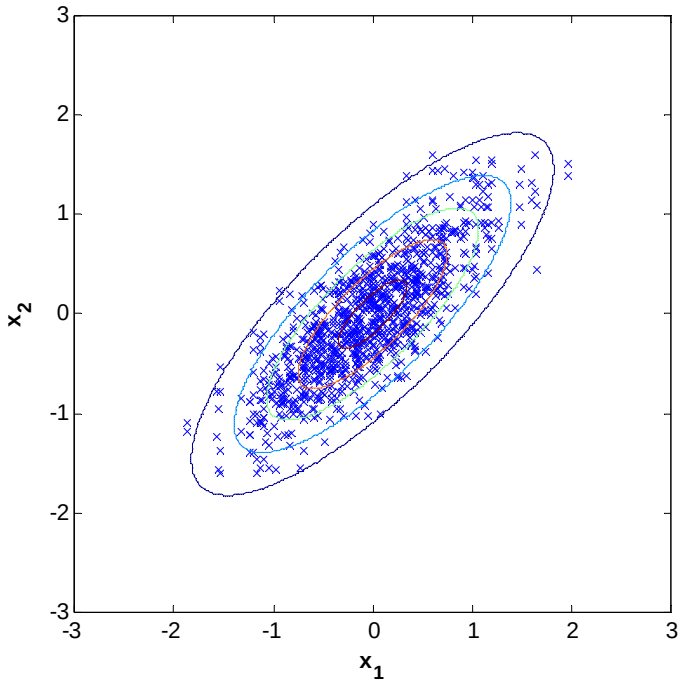

Figure 4.3: Samples from Gibbs sampler after 500 iterations

4.3 Metropolis-Hastings Algorithm

4.3.1 Description of Algorithm

The Metropolis-Hastings (M-H) algorithm is a general method for constructing a Markov chain that draws samples from transition distributions for arbitrary posterior distributions.

Consider generating a random draw θ from the target posterior distribution $p(\theta)$. The M-H algorithm proceeds as follows:

(i) Draw a starting point θ^0, for which $p(\theta^0) > 0$.

(ii) Given θ^{t-1}, sample a proposal θ^* from a jumping distribution (or proposal distribution) at time t, $J_t(\theta^* \left| \theta^{t-1} \right.)$.

(iii) Calculate the ratio of the densities as

$$r = \frac{p(\theta^*) / J_t(\theta^* \left| \theta^{t-1} \right.)}{p(\theta^{t-1}) / J_t(\theta^{t-1} \left| \theta^* \right.)} \tag{4.10}$$

(iv) Generate a uniform random draw u from the interval between 0 and 1.

(v) Set

$$\theta^t = \begin{cases} \theta^*, & \text{if } u \le \min(r,1) \\ \theta^{t-1}, & \text{otherwise} \end{cases}.$$

(vi) Repeat (ii) to (v) for $t = 1, 2, \cdots$.

Unlike the Gibbs sampler, the M-H algorithm take a new value of θ with probability $r \le 1$. Another essential feature of the M-H algorithm is that the $p(.)$ is not necessarily a proper probability density function. Furthermore, the proposal distribution has to satisfy the following conditions: (1) we can quickly and easily generate random draws from the proposal distribution and (2) the proposal distribution should be distributed with the same range as the target distribution.

This is very important as we have already seen that in Bayesian inferences the posterior distribution is known only up to a normalized constant. The unknown constant is cancelled out in equation (4.10). Because of this feature, the M-H algorithm becomes very attractive in Bayesian inference.

4.3.2 A Gaussian Example

To illustrate the effectiveness of the M-H algorithm, a simple Gaussian example is presented. Let θ be a standard Normal random variable, i.e., $\theta \sim N(0, 1)$. Instead of using the traditional independent sampling technique, here we use the M-H algorithm for

sampling. For illustration purpose, we choose the jumping distribution as $\theta^* \sim N\left(\theta_{t-1}, 0.5^2\right)$.

The sample sequence from the M-H algorithm is shown in Figure 4.4. After 500 steps, the sequence can be considered stationary as neither the mean nor variance changes with the increase of the number of iterations.

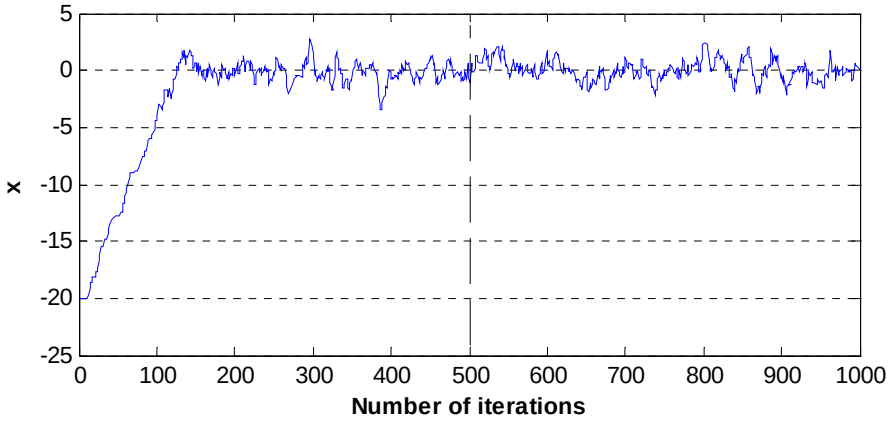

Figure 4.4: Sample sequence of standard normal distribution from M-H algorithm

Figure 4.5 compares the stationary sequence from the M-H sampling with the independent sequence sampled directly from the normal distribution. Although the sequence from the Markov chain looks very different from the independent sequence, the fitted PDF from the M-H sequence is as close as the independent sampling fits to the theoretical PDF.

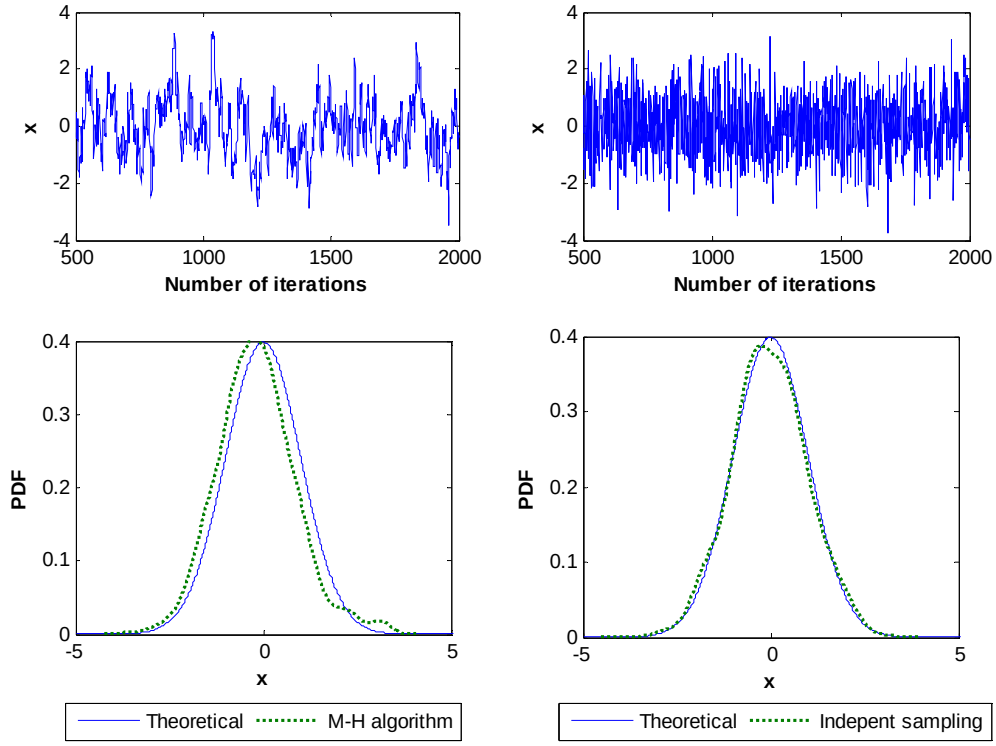

Figure 4.5: Comparison of the M-H sequence with an independent sequence

4.3.3 A Weibull Example: Hybrid Algorithm

In this section we attempt to illustrate the idea of combining the Gibbs sampler with the M-H algorithm to sample a multi-dimensional posterior distribution through an example of a Weibull distribution. We call it a hybrid algorithm.

Suppose we have a sample of data with size n, $(x_1, x_2, \ldots, x_n)$, from a Weibull distribution. Let the prior distribution of the scale parameter γ be a Gamma distribution with the hyperparameters $A = 2$ and $B = 0.1$, as defined in Equation (3.23). Also let the prior distribution of the shape parameter β be a Beta distribution with the

hyperparameters $r = 2,\ s = 3$ and defined on the interval $[\beta_l = 0, \beta_r = 5]$, as shown in Equation (3.24). Then it can be shown that the conditional posterior distribution of γ given β is still a gamma distribution, i.e.

$$\left(\gamma \middle| \beta, x_i\right) \sim Ga\left(A + n, B + \sum_{i=1}^{n} x_i^{\beta}\right) \tag{4.11}$$

Similarly, the conditional posterior distribution of β given γ is found to be as

$$p\left(\beta \middle| \gamma, x_i\right) \propto \beta^n \left(\beta - \beta_l\right)^{r-1} \left(\beta_r - \beta\right)^{s-1} \exp\left(-\gamma \sum_{i=1}^{n} x_i^{\beta}\right) \prod_{i=1}^{n} x_i^{\beta-1} \tag{4.12}$$

which is not a fully specified PDF. Clearly the beta distribution is not a conjugate prior.

Now that the conditional posterior of γ can be easily sampled but that of β cannot. We can thus combine the Gibbs sampler with the M-H algorithm to draw samples from the posterior distributions. That is, given a sample of vector $\left(\gamma^{t-1}, \beta^{t-1}\right)$ at the (t-1)$^{\text{st}}$ step, γ^t is sampled from (4.11) with $\beta = \beta^{t-1}$, following the fashion of a Gibbs sampler, whereas β^t should be sampled from (4.12) using a M-H algorithm. The iteration is initialized at an arbitrary starting point $\left(\gamma^0, \beta^0\right)$. Since the prior distribution of the shape parameter β is a bounded distribution at $[\beta_l, \beta_r]$, a simple uniform distribution over the same interval can be used as the proposal distribution for the M-H algorithm.

To illustrate, a random sample of size 20 from the Weibull distribution with $\gamma = 20$ and $\beta = 2$ is simulated. Figure 4.6 plots the sequences of the Markov chain for β and γ using the hybrid algorithm, respectively. Unlike like the Gibbs sampler by which the chain of γ moves at each step, the chain of β simulated by the M-H algorithm is more like stepwise random paths. The sample draws are not jumping even for more than 30

successive iterations in some cases. Thus, the number of possible implementations of the M-H algorithm is larger than the Gibbs sampler.

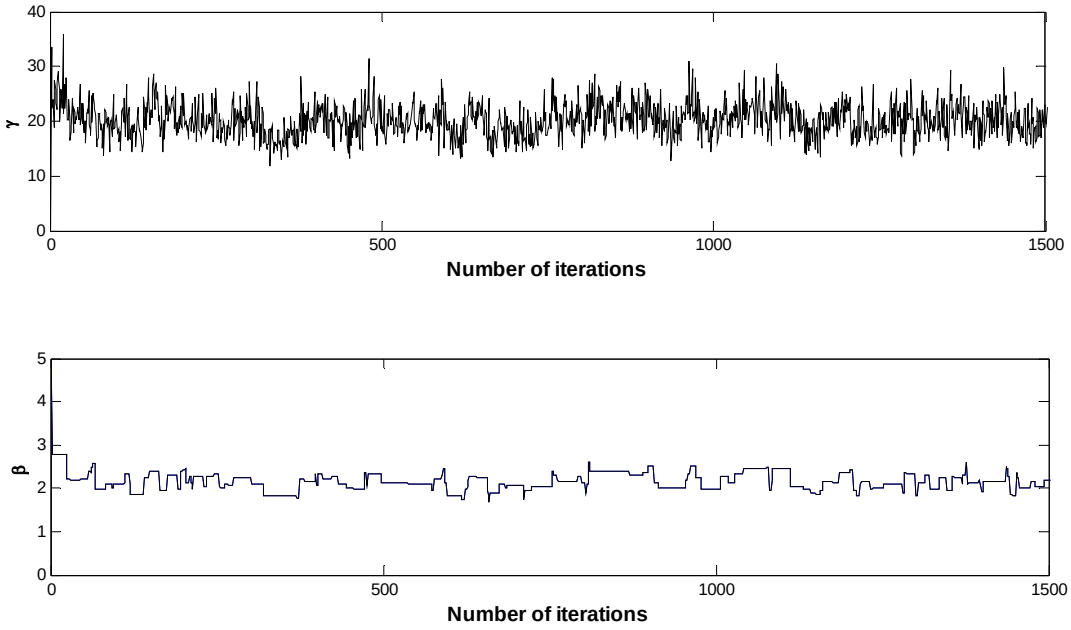

Figure 4.6: Sample sequences from the Hybrid algorithm for the Weibull distribution

We use a burn-in period $N_b = 5000$ iterations and total number of iterations $N_{sim} = 20000$. Figure 4.8 and 4.8 show the prior and posterior distributions of γ and β, respectively. The posterior PDF is plotted using the kernel smoothing estimation from the simulated data (Mathworks, 2007). The posterior mean values of γ and β are 19.612 and 1.8856, respectively, which are very close to their true values, 20 and 2.

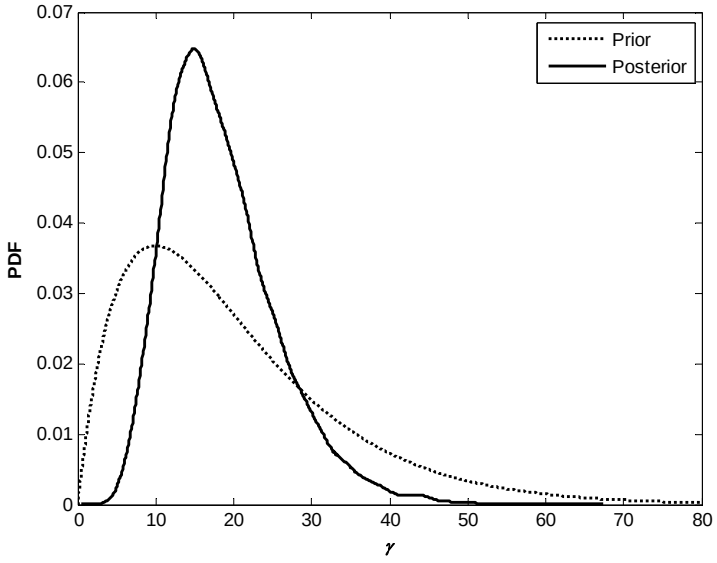

Figure 4.7: Prior and posterior distributions of γ

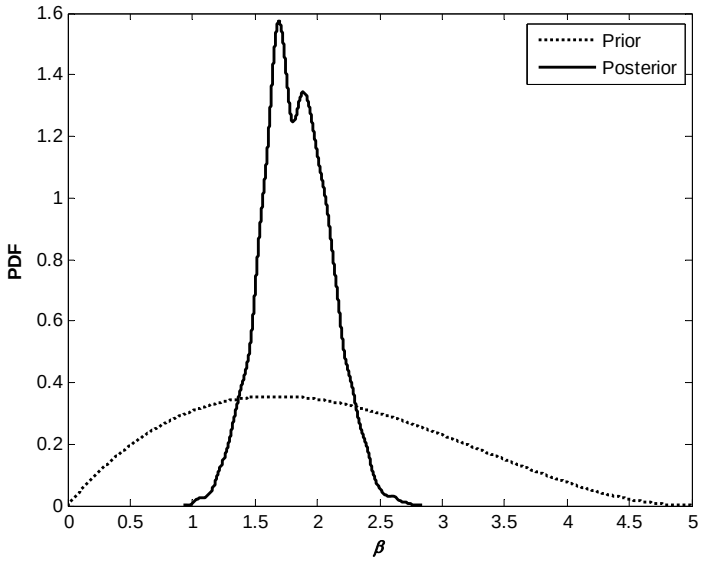

Figure 4.8: Prior and posterior distributions of β

4.4 MCMC for the Proposed Pitting Corrosion Model

4.4.1 Simplification of the Likelihood Function

In Section 3.3 the likelihood function for the proposed pitting corrosion model is derived, which involves in an infinite summation and several multi-dimensional integrations. When the likelihood is multiplied by the proposed prior distributions in Section 3.4, the posterior distribution will become extremely complicated. In such a case, the M-H algorithm might be the only applicable sampling method but it should be definitely inefficient.

To avoid the difficulty of evaluating the likelihood function, a data augmentation technique is used. As shown in Figure 4.9, there are several missing intermediate data between the parameters and the observations. If the missing data were known, the likelihood of the parameters would be much more simplified. For example, when the actual pit sizes (h_d or h_u) and the pit number (n) known, the likelihoods for β and γ and for λ are separable, the former being the product of the Weibull PDF as follow

$$L\left(\boldsymbol{\theta}\middle|\mathbf{h_d},\mathbf{h_u}\right)=\left(\gamma\beta\right)^{n_d+n_u}\left(\prod_{i=1}^{n_d}h_{d,i}^{\beta-1}\prod_{i=1}^{n_u}h_{u,i}^{\beta-1}\right)\times\exp\left\{-\gamma\left(\sum_{i=1}^{n_d}h_{d,i}^{\beta}+\sum_{i=1}^{n_u}h_{u,i}^{\beta}\right)\right\}$$

and the latter a Poisson distribution function, $L\left(\lambda\middle|n\right)=\frac{\lambda^n}{n!}e^{-\lambda}$. Of course, we are not able to directly obtain the missing data. But in the framework of MCMC simulation, at each sampling step, the simulated parameters can be used for generating the missing data by using the assumed probabilistic model. For example, once λ is obtained, we can generate a sample of n from the assumed Poisson distribution. After the actual pit numbers

obtained, the parameters can be updated using the Bayesian rule. The detail of the iteration algorithm is described in the next subsection.

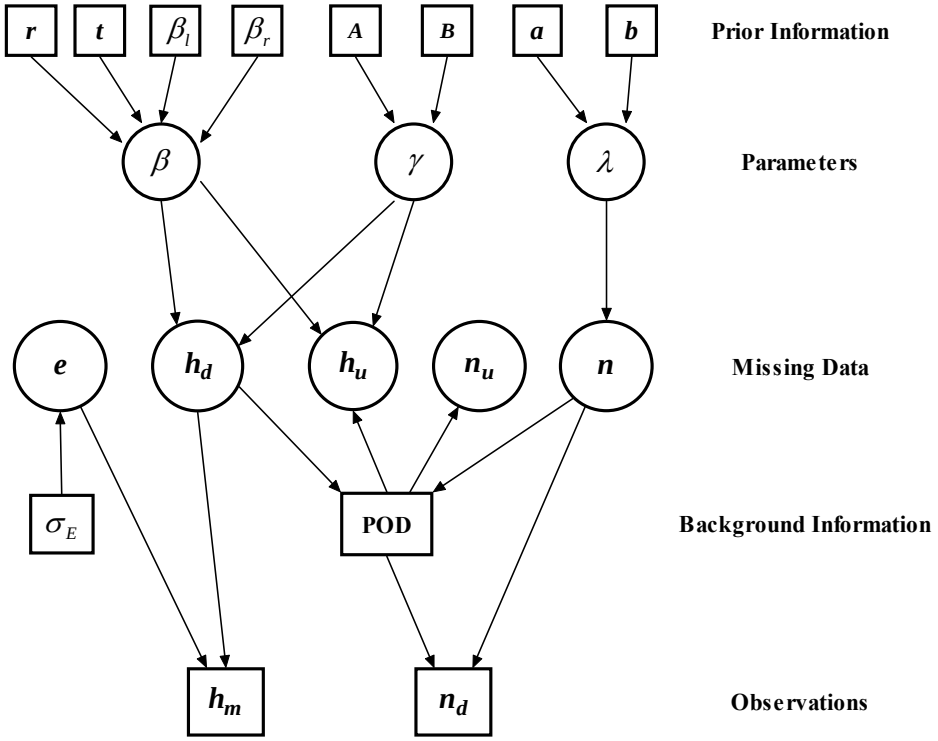

Figure 4.9: Structure of the Bayesian pitting corrosion model

4.4.2 Algorithm

For this complicated high-dimensional model, the combined algorithm discussed in 4.3.3 should be used. The detailed derivations of the posterior distributions for all the parameters are listed in Appendix C.

Now, we describe the $(j+1)^{st}$ iteration of the sampling iteration where j denotes the iteration index. Also, the implementation consists of two iterated steps: a) computation of

the conditional distribution of λ knowing $\boldsymbol{\theta}$ and b) computation of the conditional distribution of $\boldsymbol{\theta}$ knowing λ. Assume that after j^{th} iterations we have λ^j, γ^j and β^j. We will detail these two steps next.

Step a

1) Sample λ^{j+1} from $Ga\left(a+n^j, b+1\right)$.

2) Calculate $p_f(\theta^j)$. Here, according to Equation (3.13), $p_f(\theta^j)$ is approximated through an inner simulation

$$p_f(\theta^j) = \frac{1}{1000}\sum_{i=1}^{1000} \text{POD}(y_i^j) \tag{4.13}$$

where $y_1^j, y_2^j, \cdots, y_{1000}^j$ are independent selections form the current pit depth distribution $f\left(.\middle|\theta^j\right)$.

3) Sample n_u^{j+1} from a $Pois\left[\lambda^{j+1}\left(1-p_f(\theta^j)\right)\right]$.

4) Calculate the actual pit number $n^{j+1} = n_u^{j+1} + n_d$.

Step b

5) Simulated a number n_u^{j+1} of "undetected" flaw size $\mathbf{h}_u^{j+1}$ from the current flaw size distribution $f(h|\theta^j)$. Those simulations are performed as follows. Let $\tilde{z}$ be a random drawing from $f(h|\theta^j)$. If $\tilde{z} < s$ where s is the detection threshold, accept this value; otherwise accept it with the probability $1 - POD(\tilde{z})$. This process is continued until n_u^{j+1} pit depths are simulated.

6) Generate $\mathbf{e}^{j+1} = (e_1^{j+1}, \cdots, e_{n_d}^{j+1})$, a vector of n_d independent errors from the distribution $f_E(e)$.

7) Calculate a vector of the detected pit depth $\mathbf{h}_d^{j+1} = \mathbf{h}_m - \mathbf{e}^{j+1}$.

8) Sample γ^{j+1} from $Ga\left(A + n^{j+1}, B + \sum_{i=1}^{n_d} \left(h_{d,i}^{j+1} \right)^{\beta} + \sum_{i=1}^{n_u^{j+1}} \left(h_{u,i}^{j+1} \right)^{\beta} \right)$

9) Since there is no conjugate prior distribution for the parameter β of the Weibull distribution, the simulation of β^{j+1} from its conditional posterior distribution is expressed as

$$p\left(\beta \middle| \gamma^{j+1}, \mathbf{h}_d^{j+1}, \mathbf{h}_u^{j+1}\right) \propto \beta^n \left(\beta - \beta_l\right)^{r-1} \left(\beta_r - \beta\right)^{s-1} \left(\prod_{i=1}^{n_d} \left(h_{d,i}^{j+1} \right)^{\beta-1} \prod_{i=1}^{n_u^{j+1}} \left(h_{u,i}^{j+1} \right)^{\beta-1} \right) \times \exp\left\{ -\gamma \left(\sum_{i=1}^{n_d} \left(h_{u,i}^{j+1} \right)^{\beta} + \sum_{i=1}^{n_u^{j+1}} \left(h_{u,i}^{j+1} \right)^{\beta} \right) \right\} \tag{4.14}$$

We make use of the Hasting-Metropolis algorithm in Section 4.3.3 to get a sample of β.

4.4.3 Simulation Results

A MATLAB program was coded to implement the MCMC simulation described in the previous section. To validate the codes as well as illustrate the behavior of the Bayesian method, a numerical experiment is performed using this MATLAB program.

A random sample of 20 pits from the Weibull distribution with $\gamma = 20$ and $\beta = 2$ is simulated. The Weibull distribution has a theoretical mean of 0.1982, or 19.82 per cent of the through-wall depth. The maximum likelihood estimates for the data is $\hat{\gamma} = 16.5654$

and $\hat{\beta} = 1.7774$. The 20 pits are then filtered with a logistic POD function defined by Equation (3.6) where $h^* = 0.2$ and $s = 0.05$. The parameter q is computed by solving the equation $POD(0.5) = 0.85$ which gives $q = 8$. The 5 pits left after the filtering, or the detected pits are then contaminated with independent normal random errors with $\sigma_E = 0.05$. Finally we obtain the "measured" pit depth $\mathbf{h_m}$ = [0.2490, 0.4172, 0.4416, 0.1961, 0.1621]. Next, we are going to use these measured data to make statistical inferences about the parameters by the MCMC algorithm presented in the previous subsection. For the inference, we use the same random error distribution and the same logistic POD function when implementing the MCMC algorithm.

Recall that the prior distribution of the Poisson parameter λ is a Gamma distribution with hyperparameters a and b. Here we choose a = 20 and b = 1. The prior distributions of β and γ are the same as in Section 4.3.3.

To monitor the convergence of the sample sequences, the procedure proposed by Celux et al. (1999) is used. At each iteration j, we compute the mean value of the simulated pit depth including both detected and undetected ones by $\bar{h}^j = \left(\mathbf{h}_{\mathrm{d}}^j + \mathbf{h}_{\mathrm{u}}^j\right)/n^j$. We also compute the empirical mean value (*EMV*) as

$$EMV = \frac{1}{N_{sim} - N_b} \sum_{j=m+1}^{N_{sim}} \bar{h}^j \tag{4.15}$$

We compare this value with the theoretical mean value (TMV) of the Weibull distribution with the estimated parameters $\hat{\gamma}$ and $\hat{\beta}$, i.e.,

$$TMV = \gamma^{-1/\beta}\Gamma\left(1 + 1/\beta\right) \tag{4.16}$$

where $\Gamma(\cdot)$ is a gamma function.

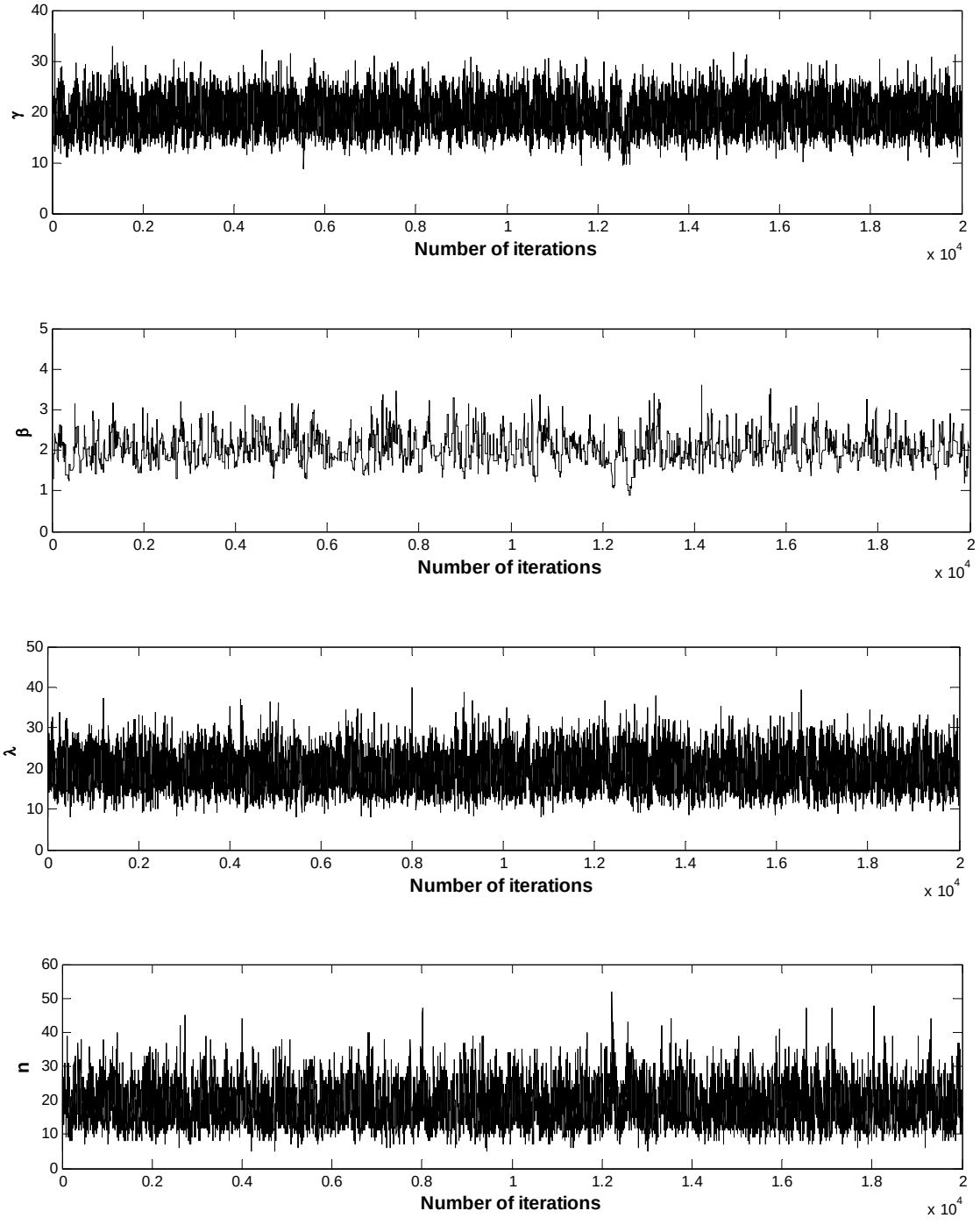

Figure 4.10: MCMC chains of model parameters

When N_{sim} is large enough to ensure a reasonable approximation of the posterior distribution $p\left(\boldsymbol{\theta}\middle|\mathbf{h_m}, n_m\right)$, the values EMV and TMV are expected to be close. Thus, if EMV and TMV are notably different, it can be guessed that N_{sim} is too small to ensure the MCMC convergence for the chosen prior distributions.

Figure 4.10 display the chains simulated by MCMC for the model parameters. All the chains reach their stationary state after 2000 iterations. Therefore a burn-in period $N_b = 2000$ and a total number of iterations $N_{sim} = 20000$ are used.

Figure 4.11 to 4.13 plots the prior and posterior distributions for the parameters γ and β of the Weibull distribution and for the parameter λ of the Poisson distribution.

Table 4.1 lists the posterior means of the $\hat{n}$, $\hat{\lambda}$, $\hat{\gamma}$ and $\hat{\beta}$ as well as the EMV and TMV for two different values of hyperparameter *a*, which represents the prior mean of the number of pits. The estimated EMV and TMV values are very close at both cases, meaning the numbers of burn-in and total simulations are adequate. The Bayesian estimates are quite reasonable when $a = 20$ compared with $a = 50$ according to the small sample size $n_d = 5$. Thus, the prior assumption concerning the actual pit number seems to be significant on the Bayesian inference. Moreover, we notice that the mean pit depth, EMV or TMV, decreases with increasing prior mean number of pits. The reason may be that during the MCMC sampling process the number of the simulated smaller 'unobserved' pits $\mathbf{h_u}$ will increase, to correspond to the assumed large prior mean number of pits with fixed number of detected pits.

Table 4.1: Bayesian estimates of the Weibull distribution

a	EMV	ETMV	$\hat{n}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\beta}$
20	0.2041	0.2042	18	19	19.3729	2.0192
50	0.1373	0.1349	44	47	18.8478	1.5474

The theoretical and estimated Weibull distributions by the Bayesian method are graphically compared in Figure 4.14. Also plotted in this figure is the result of the maximum likelihood method based on the 20 uncontaminated data. The two results are both close to the theoretical curve. Note for the Bayesian analysis, only 5 contaminated measurements were used. Clearly, the influence of measurement uncertainties due to measurement error and POD has been successfully eliminated through the extra prior information.

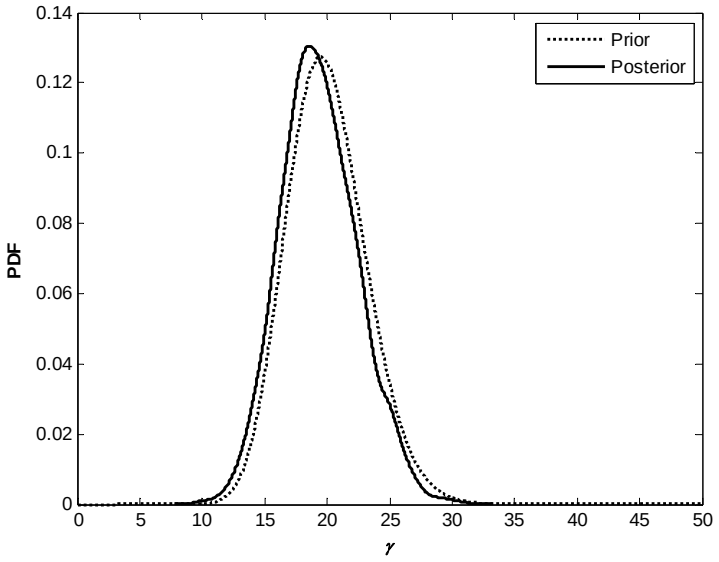

Figure 4.11: Prior and posterior distributions of γ

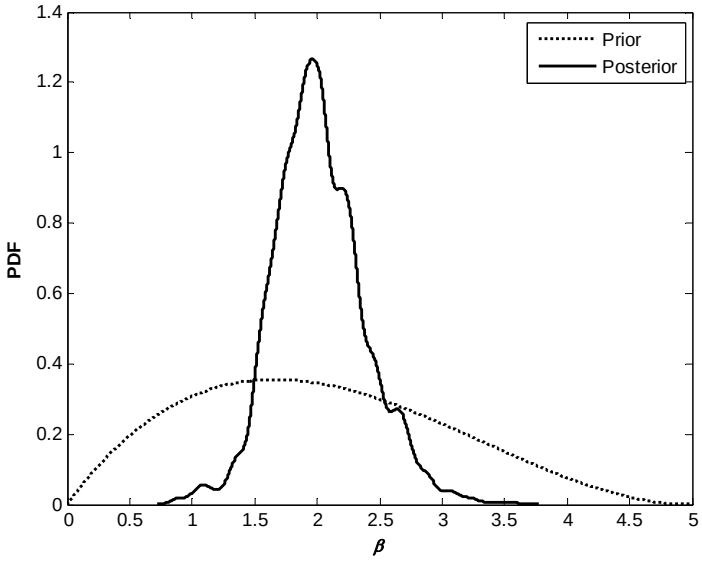

Figure 4.12: Prior and posterior distributions of β

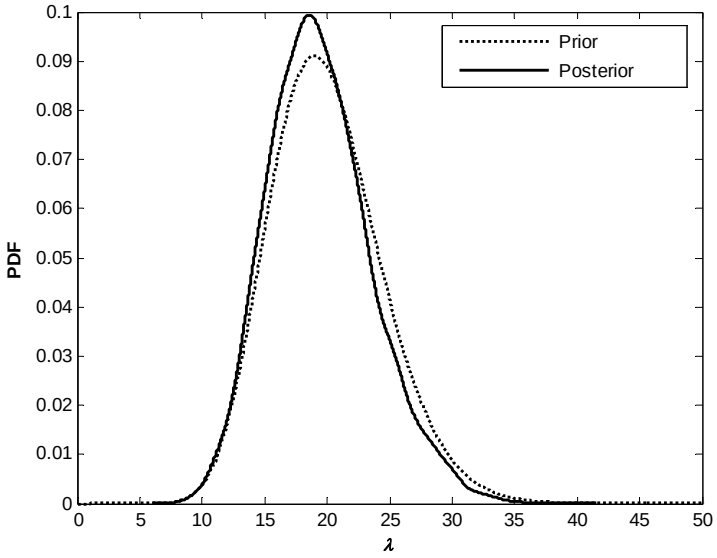

Figure 4.13: Prior and posterior distributions of λ

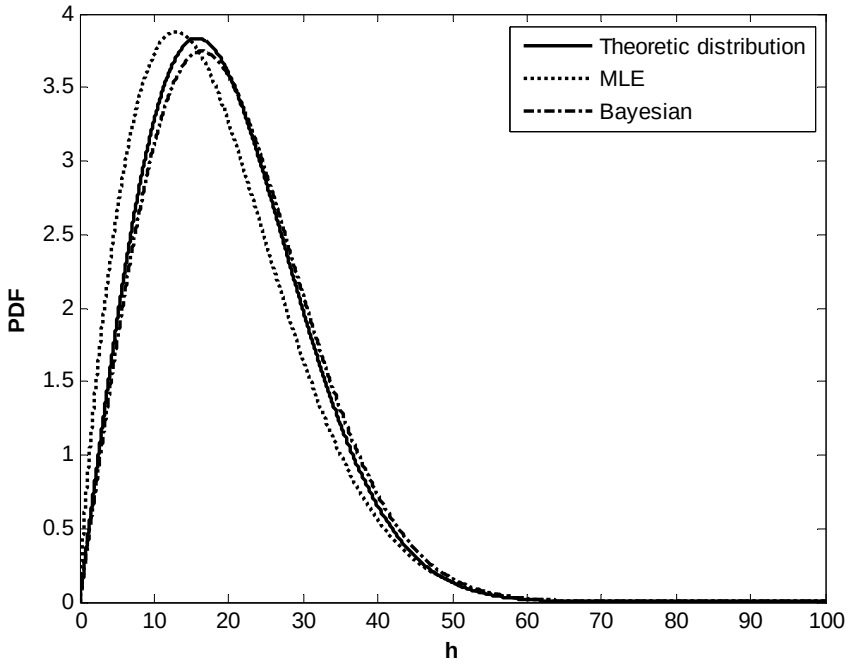

Figure 4.14: Comparison of posterior Weibull PDF with the MLE result

4.5 Summary

The basic concepts and procedures of MCMC methods for Bayesian inference of the proposed pitting corrosion model are discussed in this chapter. In order to simplify the likelihood function, a data augmentation method for missing intermediate data is proposed. A hybrid sampling technique based upon Gibbs sampler and the Metropolis-Hastings algorithm is used. Both simple examples and a simulation experiment are used to verify the proposed methods and to illustrate the good performance of the Bayesian method.

Chapter 5 Case Study

5.1 Overview of the ISI Data

A case study is presented in this chapter on pitting corrosion data collected during nine in-service inspection (ISI) outages of a steam generator. The pitting data of the unit was obtained using eddy current probes.

The detected numbers of new pits are plotted in Figure 5.1. Over the years, two maintenance campaigns, including the water lancing (WL) and chemical cleaning (CC), were performed right after the 2nd and 8th inspection outages. The effects of the maintenance campaigns are obvious: after the two maintenance campaigns, the number of new pit has decreased substantially.

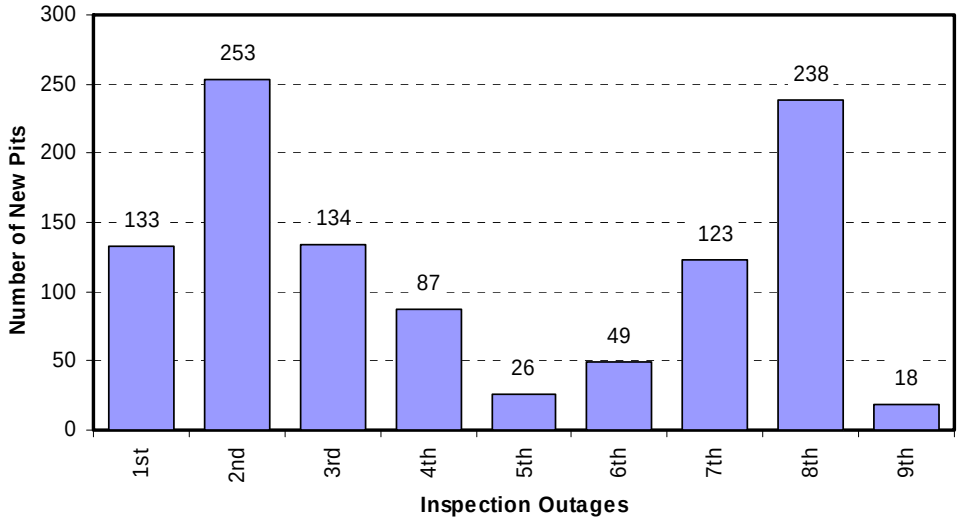

Figure 5.1: Numbers of new measured pits at each inspection outage

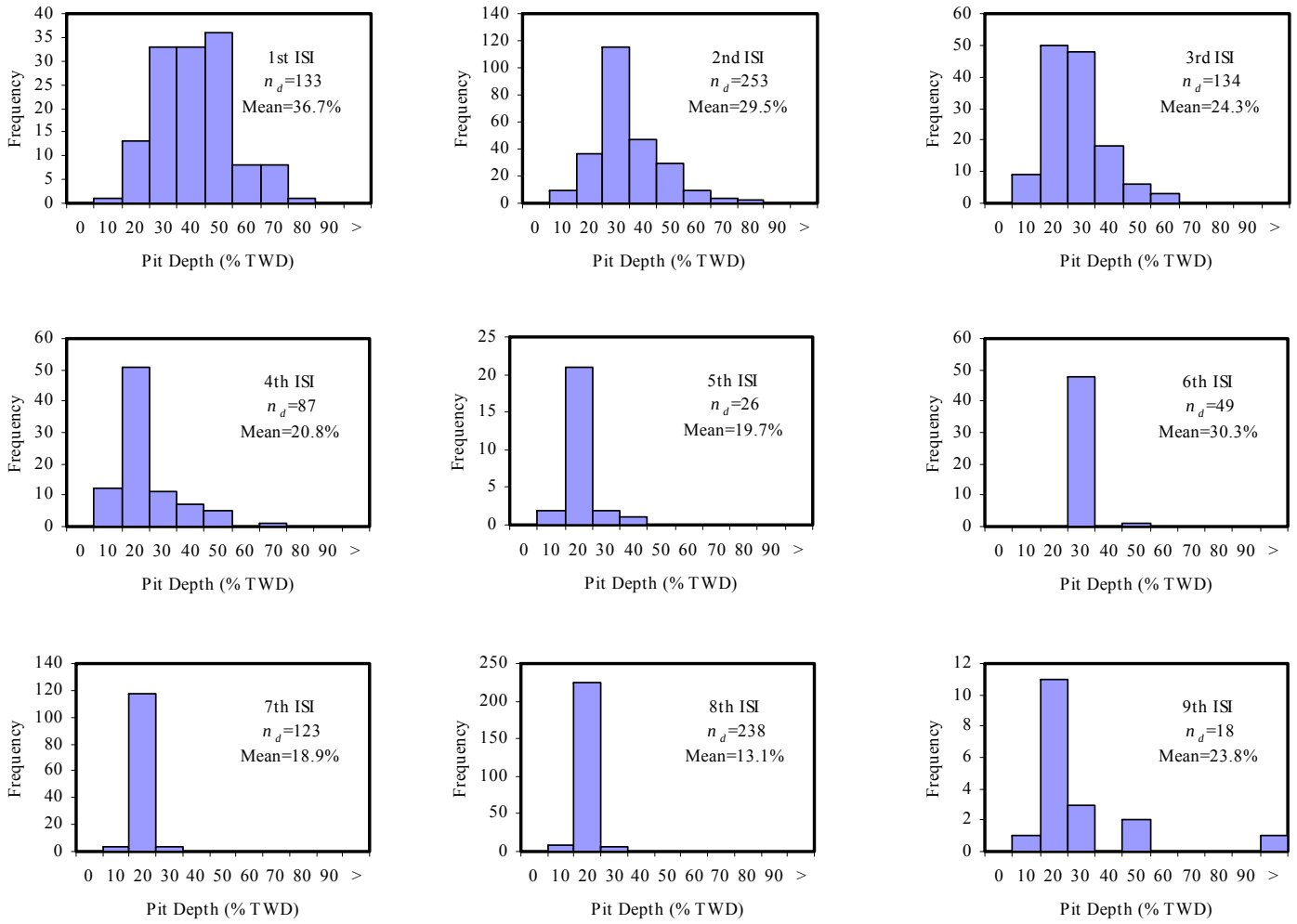

Figure 5.2: Histograms of pit depth for new pits at each ISI outages

The histograms of the new measured pit depth are shown in Figure 5.2, in which the pit depth is expressed at percentages of through-wall depth (TWD). So the measured pit depth should be in the range of 0 to 100%. The pit depth data from the first four inspection campaigns are more spread than the remaining ones. In contrast, the measurements of pit depth at the 6^{th} ISI locate almost exclusively at 30% TWD. Similarly, the data from the 7^{th} and 8^{th} campaigns concentrate at 20% TWD, although a relatively large number of new pits have been observed.

5.2 Bayesian Modeling

5.2.1 Prior Distributions, POD and Measurement Errors

An important component in Bayesian modeling is the prior distributions of the parameters. As we have seen in Section 4.4.3, the prior mean of λ has significant impact on the posterior distributions. Hence, in this case study the Jeffrey's non-informative prior distribution (Robert, 2001) is used for λ. It turns out to be proportional to $\lambda^{-1/2}$, which is equivalent to the Gamma prior in equation (3.21) when $a = 1/2$ and $b = 0$.

For choosing the hyperparameters concerning the actual pit depth, we use guessed values of the mean value and variance of the actual pit depth. Here, based on the histogram shown in Figure 5.2, we roughly assume the actual pit depth distribution with the mean 0.2 (20%) and variance 0.1^2, which correspond to $\gamma = 20$ and $\beta = 2$ for a Weibull model. Thus, we choose $A = 20$, $B = 1$, $[\beta_l, \beta_r] = [0, 5]$, $r = 2$ and $t = 3$.

Due to the design characteristics of inspection tools as well as operational conditions, we use the same POD function and measurement errors as in chapter 4 for this case study.

5.2.2 Results

For the sake of clarity of presentation, we show in Figure 5.3 to 5.5 the sample chains of the model parameters and the corresponding marginal posterior distributions for the 1st, 6th and 9th ISI data only. The three cases correspond respectively to the three representative situations shown in Figure 5.2: spread data (1st to 4th), concentrated data (6th to 8th) and limited data (9th). For all the cases 100,000 iterations are run and the first half of them are taken as burn-ins. Table 5.1 lists the values of EMV and TMV, which are very close for all the 9 cases, indicating adequate numbers of burn-in and total iterations.

Also listed in the table are the posterior estimates of the model parameters. The estimated posterior mean of the shape parameter β for the Weibull distribution varies from 1.2 to 2.6, whereas the posterior mean of γ varies with a wider range from 10 to 33.

Figure 5.6 compares the numbers of measured pits with the estimated actual pit numbers for each set of inspection data. The estimated actual pit numbers do not depends only on the measured pit numbers but also on the measured pit depth. For instance, the measured pit numbers at the 1st and 3rd outages are 133 and 134 respectively. But the estimated actual pit numbers are 304 and 768. The former is much less than the latter even though their measured numbers are almost the same. Recall that the average of the measured pit depth at the 1st inspection is 36.7%, which is greater than the average of 24.2% at the 3rd inspection. This implies that the number of undetected pits (which are usually small in size) is less at the 1st inspection than at the 3rd inspection.

For most cases, estimated pit numbers are reasonably greater than the measured pit numbers, except the 8th outage, for which the actual pit number is estimated as 7748, very

far from the fact that only 238 new pits were observed. This result should be understood as artificial. As a matter of fact, the sample sequences do not converge for the 8th ISI data when β_L is set less than 1. In order to make them convergent, β_L was arbitrarily set greater than 1, which means the Weibull distribution is forced to be of a bell shape. This may be against the reality, as the 8th ISI has the least average value among all the ISI data.

Table 5.1: Posterior mean of the model parameters

Campaigns	EMV	TMV	$\hat{n}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\beta}$
1st	0.2789	0.2768	304	304	10.7849	2.0441
2nd	0.1798	0.1788	1016	1017	11.4888	1.5474
3rd	0.1416	0.1411	768	769	16.3508	1.5058
4th	0.1208	0.1213	637	637	19.9271	1.4898
5th	0.1265	0.1275	183	183	25.6416	1.6665
6th	0.246	0.2453	127	128	28.4364	2.6014
7th	0.1166	0.1164	991	992	31.3599	1.6916
8th †	0.0546	0.0546	7748	7749	33.3929	1.2356
9th	0.1455	0.1458	101	101	16.1863	1.5292

† The sample sequences for this case do not converge when the lower limit β_L of the prior distribution of β is set less than 1. The results shown here are based on lower limit set greater than 1. More explanation can be found in the text.

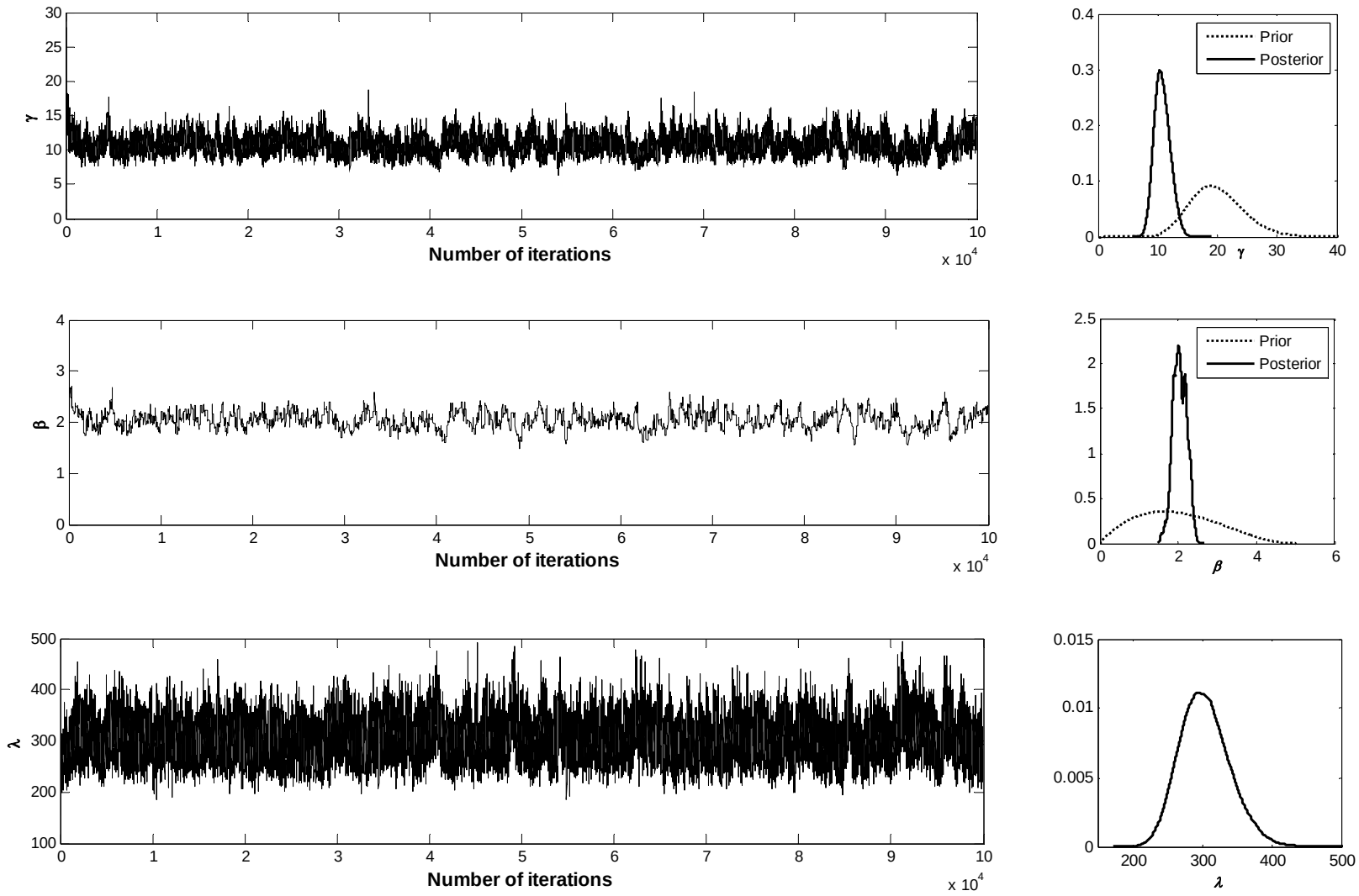

Figure 5.3: MCMC chains and marginal distributions of the model parameters for the 1st ISI data

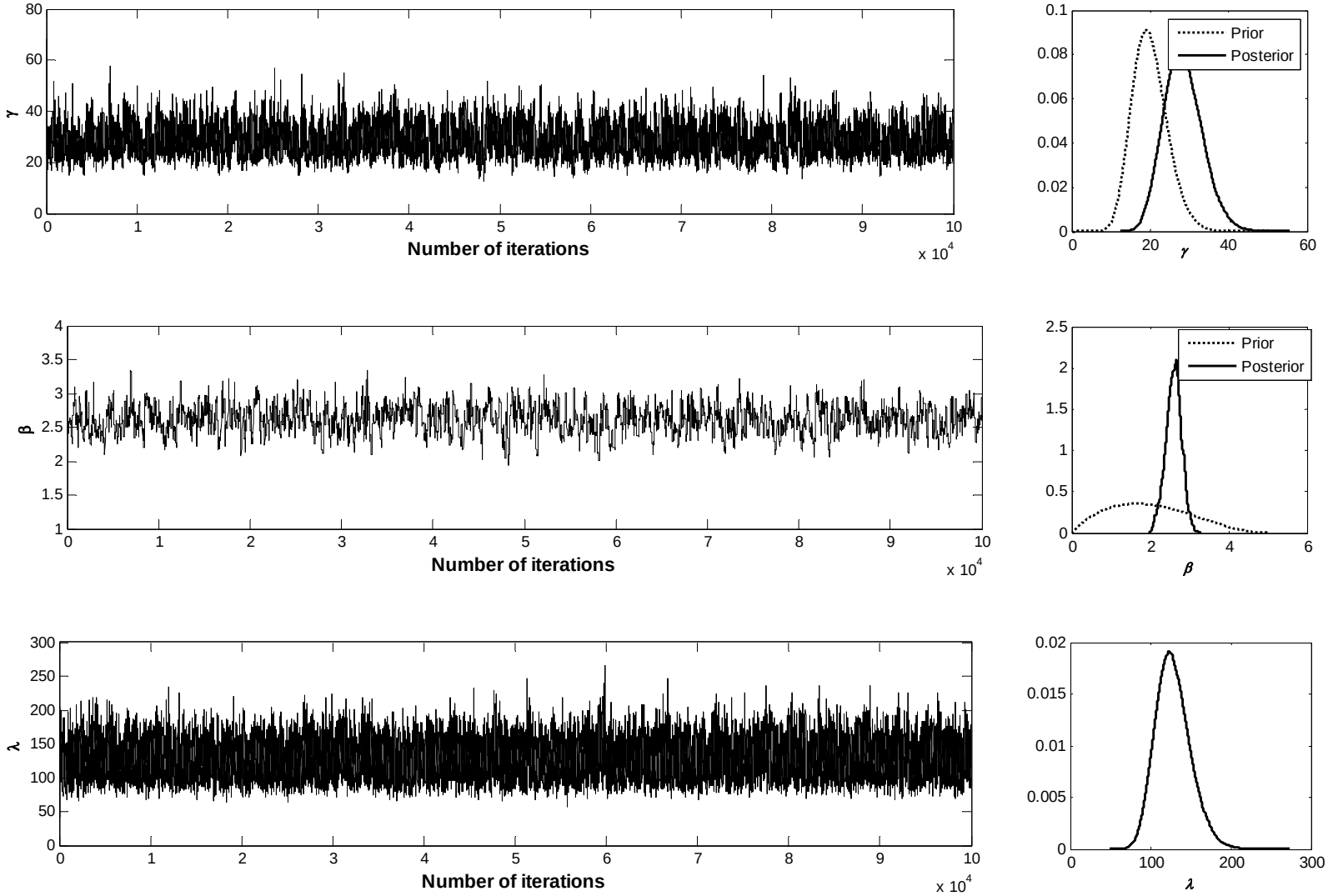

Figure 5.4: MCMC chains and marginal distributions of the model parameters for the 6th ISI Data

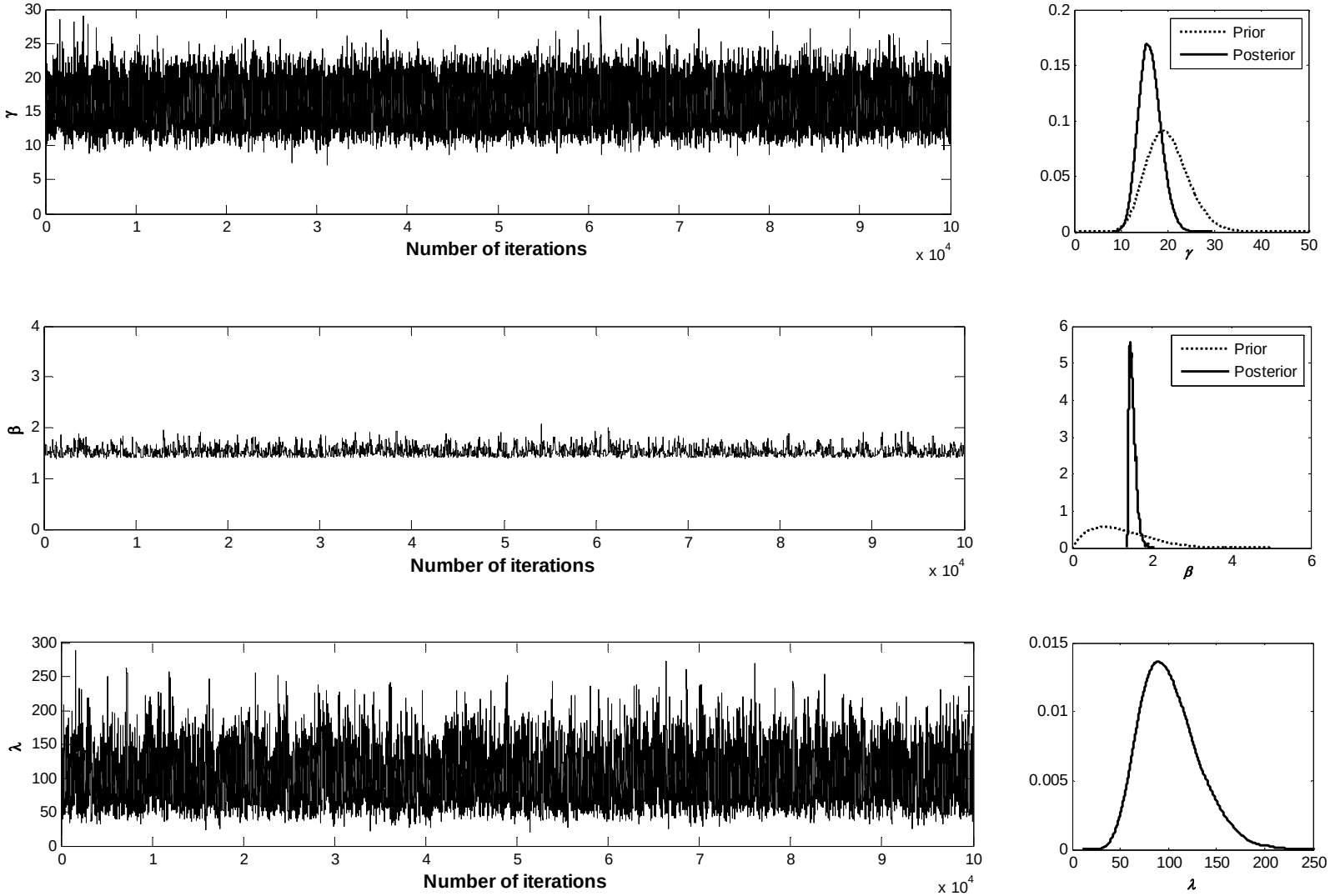

Figure 5.5: MCMC chains and marginal distributions of the model parameters for the 9th ISI data

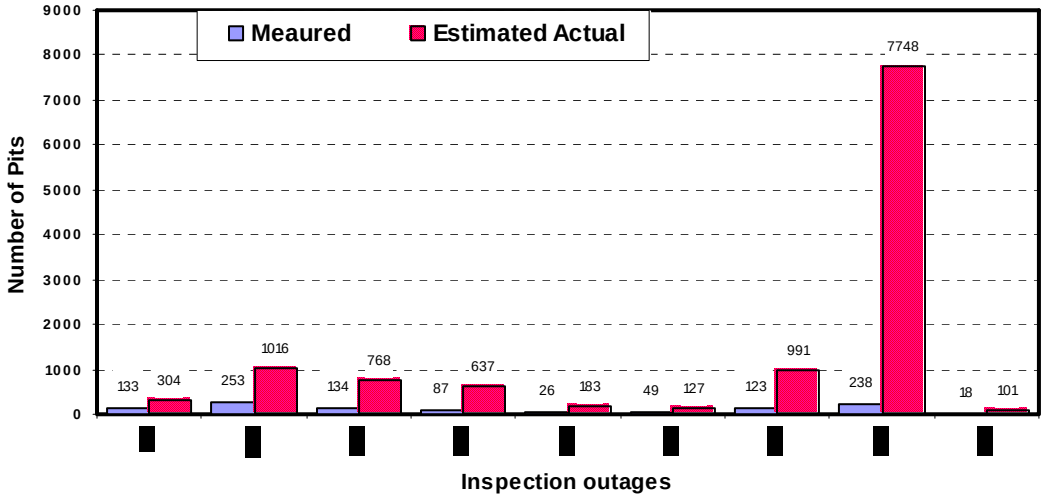

Figure 5.6: Comparison of actual pit numbers with measured pit numbers

5.2.3 Distributions of Maximum Pit Depth

Once the parameters β, γ and λ are estimated, the distribution of maximum pit depth for each inspection campaign can be evaluated using equation (3.8). Figure 5.7 shows the estimated maximum pit depth distribution with comparison of the estimated actual pit depth as well as the measured pit depth for the 1st ISI data. Graphs for the remaining data are shown in Figure 5.8. Among the 9 inspections, the first two predict maximum pit depth greater than 100%, or through-wall corrosion might have been expected. Although there is no data to validate this observation, the fact that water lancing and chemical cleaning was done right after the 2nd inspection suggests that actual through-wall leakage might have been found in reality at that time.

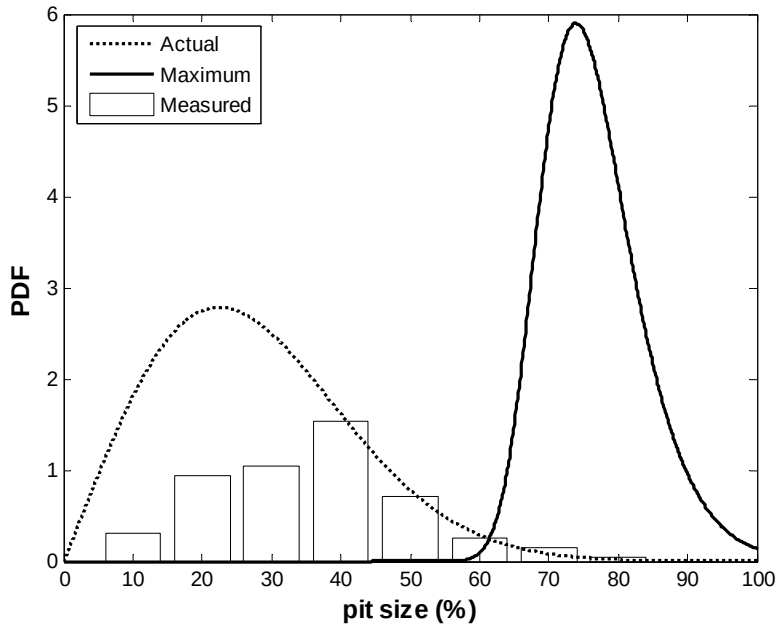

Figure 5.7: Distribution of maximum pit depth for the 1st ISI data

5.3 Summary

This chapter presents a comprehensive case study of pitting corrosion data collected during nine in-service inspection (ISI) outages over years at a steam generator of a nuclear power station. The proposed Bayesian pitting model is fitted separately for each set of data by using the MCMC simulation technique described in Chapter 4. The posterior means of the parameters are estimated and the distributions of maximum pit depth are predicted. The results are considered to agree to the reality.

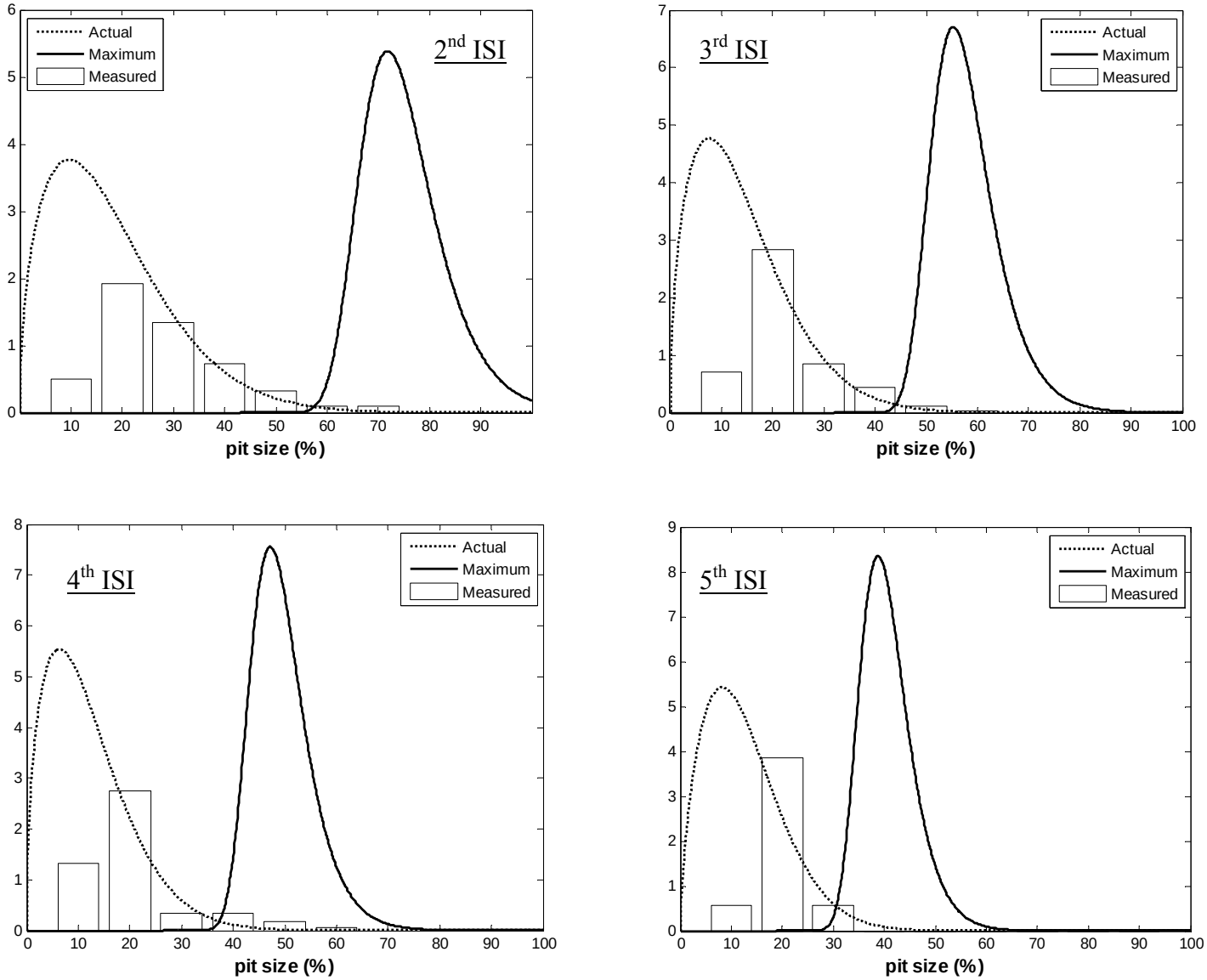

Figure 5.8: Distributions of maximum pit depth

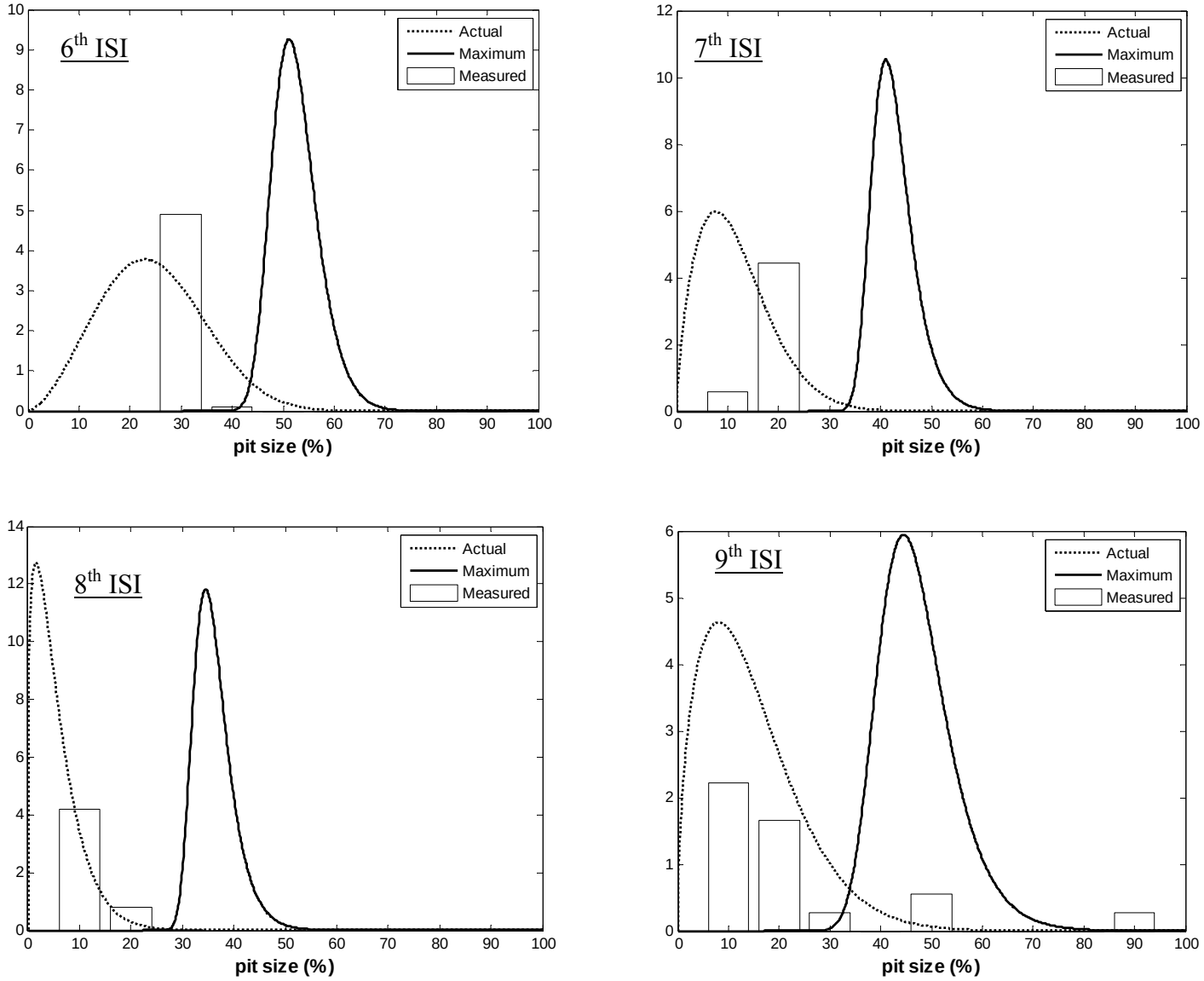

Figure 5.8 (Cont'd): Distribution of maximum pit depth

Chapter 6

Conclusions and Recommendations

6.1 Conclusions

This thesis presents an advanced probabilistic model of pitting corrosion for in-service inspection data to aid life-cycle management of steam generators. This model considers the uncertainties of pitting corrosion data and it is able to capture the influence of measurement errors and POD of inspection tools and predict the actual pit number, the actual pit depth and the maximum pit depth. Because of the lack of in-service data, we propose a Bayesian approach to estimating the model parameters.

The MCMC technique is shown to be effective for Bayesian computation and inference. Simulation experiments are performed to check the behavior of the Bayesian method. The effectiveness and efficiency of Bayesian modeling are verified. Simulation results also shows that choosing appropriate hyper-parameters of the prior distribution is important to get the reliable estimates.

A comprehensive case study is also presented by applying this Bayesian method to the analysis of pitting data obtained in a steam generator unit. The results are thought to agree to the reality indicating that the versatile Weibull distribution is appropriate for modeling the actual pit depth in steam generators.

6.2 Recommendations

This Bayesian model can be extended to a more advanced model. The actual pit number could follow a non-homogeneous Poisson process assuming the pit generation rate is

$$\lambda(t) = ab(t - t_c)^{b-1} \tag{6.1}$$

where t_c is the time when cleaning is performed most recently. Then the maximum pit depth distribution turns to a time-dependent variable as follow:

$$F_{\max}(h;t) = \exp\left\{-a(t-t_c)^b\left[1-F(h)\right]\right\} = \exp\left\{-a(t-t_c)^b \gamma h^{\beta}\right\} \tag{6.2}$$

In this case, the parameters involved in this advanced model include: the scale parameter γ and shape parameter β of $f_H(h)$, a and b of the Poisson process for the number of pits. Therefore, this advanced Bayesian model is able to predict the new pit number over the time.

The choice of the prior distributions is based on conjugate prior strategies in this thesis. Model checking, such as goodness-of-fit tests, is needed to be performed. Also, an efficient rule for the tuning of the prior hyper-parameters should be also developed.

In-service pitting data is suffering from a variety of uncertainty. Each inspection data may be obtained by several inspection tools. Thus, unlike in Chapter 5 that we use the same POD and measurement errors for all the inspection cases, data should be grouped under the same inspection conditions to minimize the random effects of measurement uncertainty. Moreover, measurement errors may also depend on the pit size indicating that random errors will be reduced with larger pit sizes. Size-dependent measurement errors should be considered.

Based on the proposed Bayesian model, optimal maintenance and inspection strategies is needed to be developed for the life-cycle management of steam generators to manage the risk due to pitting corrosion.

Appendix A

Summaries of Design Parameters for CANDU Steam Generators

Station	Pickering A	Pickering B	Bruce A	Bruce B	Darlington	Point Leprea	Gentilly-**2**
Commercial operation	1971-73	1983-86	1977-79	1984-87	1990-93	1983	1983
Reactor Rating (MWe)	4×540	4×540	4×904	4×915	4×935	1×680	1×685
Number of Steam Generators	4×12	4×12	4×8	4×8	4×4	1×4	1×4
Number of Tubes	4×2600	4×2600	4×4200	4×4200	4×4663	4×3550	4×3550
Tube Size (OD) (mm)	12.7	12.7	12.7	12.7	15.9	15.9	15.9
Tube thickness (mm)	1.13~1.2	1.13~1.2	1.13~1.2	1.13~1.2	1.13~1.2	1.13~1.2	1.13~1.2
Tube Material	Monel-400	Monel-400	Inconel-600 HTMA Stress relieved	Inconel-600 HTMA Stress relieved	Inconel-800	Inconel-800	Inconel-800
T_{hot} / T_{cold} (ºC)	293/249	293/249	304/265	304/265	309/265	310/266	310/266
Tube Supports	C-steel lattice bars	C-steel trefoil broached TSP	C-steel trefoil broached TSP	C-steel trefoil broached TSP	401S SS lattice bars	401S SS trefoil broached TSP	401S SS trefoil broached TSP
U-bend Supports	C-steel lattice bars	Staggered C-steel scallop bars	Stacked C-steel scallop bars	Staggered C-steel scallop bars	Flat 410S AVBs	410S SS staggered scallop bars	410S SS staggered scallop bars
Tube Expansion	Hard roll, near secondary face						

Appendix B

MATLAB Codes of MCMC

```
%%%Case study for actual pit depth distribution and actual pit number
%%%Weibull distribution as the actual pit depth distribution

clc;
clear all;
close all;

load 'pitdata.mat';

%%%POD&MEs%%%%%%%%%%%%%%%%%%
sigma=5/100;
q=8;            %%measure the quality of detection
s=0.05;          %%threshold of detection
hstar=0.2;

%%% parameters for prior distributions
A = 20;
B = 1;

a=1/2;
b=0;
BL = 0;
BR = 5;
r = 2;
t = 3;

Nsim = 40000;
Npod=1000;
n0=length(hm);

n = zeros(Nsim,1);
lambda = n;
gam = n;
bet = n;
meanh = n;

%%%set initial values
n(1)=200;
lambda(1)=200;
```

```
gam(1) = 1;
bet(1) = 4;

for i = 2:Nsim
   lambda(i)=gamrnd(a+n(i-1),1/(b+1));

   %%%Pf%%%%%%%%%%%%%%%%
   hpod=weibrnd(gam(i-1),bet(i-1),Npod,1);
   pod = 1-min((1+exp(-q*hstar))./(1+exp(q*(hpod-s-hstar))),1);
   pf=sum(pod)/Npod;

   %%%n%%%%%%%%%%%%%%%%%%
   nu = poissrnd((1-pf)*lambda(i));
   n(i) = nu+n0;

   %%%undetected pit depth%%%%%%%%%%%%
   hu=zeros(nu,1);
   k = 0;
   while k<nu
      htemp=weibrnd(gam(i-1),bet(i-1));
      %%%htemp = wblrnd(gam(i-1)^(-1/bet(i-1)),bet(i-1));
      u = rand;
      upod = min((1+exp(-q*hstar))/(1+exp(q*(htemp-s-hstar))),1);
      if u <= upod
         k = k+1;
         hu(k) = htemp;
      end
   end

   %%%detected Pit Depth%%%%%%%%%%
   k = 1;
   while k<=n0
      e = sigma*randn;
      if e < hm(k) & e > hm(k)-1
         hd(k) = hm(k)-e;
         k = k+1;
      end
   end

   x = [hd';hu];
   meanh(i) = sum(x)/n(i);
```

```
    %%%Gibbs Sampler%%%%
    bet(i) = bet(i-1);
    gam(i) = gamrnd(A+n(i),1/(B+sum(x.^bet(i-1))));
    bt = rand*(BR-BL)+BL;
    w1 = MH_wt(bt,BL,BR,r,t,x,gam(i));
    w0 = MH_wt(bet(i-1),BL,BR,r,t,x,gam(i));
    u = rand;
    if u < min(1,exp(w1-w0))
       bet(i) = bt;
    end
end


%%%Bayesian estimates%%%%%%%%%
mn_bay = mean(n(Nsim/2:Nsim))
ml_bay = mean(lambda(Nsim/2:Nsim))
mg_bay = mean(gam(Nsim/2:Nsim))
mb_bay = mean(bet(Nsim/2:Nsim))
ETMV=mg_bay^(-1/mb_bay)*gamma(1+1/mb_bay)
EMV=mean(meanh(Nsim/2:Nsim))


%%%%%%%%%%%%%%%%%%%%%%%%%%%
function w = MH_wt(u,BL,BR,r,t,x,g)
n = length(x);
w = n*log(u)+(r-1)*log(u-BL)+(t-1)*log(BR-u)+(u-1)*sum(log(x))-g*sum(x.^u);
```

Appendix C

Derivations of Posterior Distributions

Step a

1) The posterior distribution of the parameter λ, given the actual pit number n, is:

$$\begin{aligned} p(\lambda|n) &\propto f(n|\lambda)\times\pi(\lambda) \\ &= \frac{\lambda^n}{n!}e^{-\lambda}\times\frac{\lambda^{a-1}b^a}{\Gamma(a)}e^{-\lambda b} \\ &\propto \lambda^{a+n-1}e^{-\lambda(1+b)} \end{aligned}$$

which is the kernel of Gamma density function with shape (a+n+1) and scale (b+1). Therefore, $p(\lambda|n) = Gamma(a+n, b+1)$.

2) The posterior distribution of the actual pit number n, given the observed pit number n_d and the model parameter vector θ, is

$$p(n|n_d,\theta) \propto f(n_m,\theta|n)\times\pi(n) \propto \binom{n}{n_m}\left(p_f(\theta)\right)^{n_m}\left(1-p_f(\theta)\right)^{n-n_m}\times\frac{\lambda^n}{n!}e^{-\lambda}$$

Let $n_u = n - n_d$. We have

$$\begin{aligned} p(n_u|n_d,\theta) &\propto \frac{\lambda^{(n_u+n_m)}}{(n_u+n_m)!}e^{-\lambda}\bullet\binom{n_u+n_m}{n_m}\left(p_f(\theta)\right)^{n_m}\left(1-p_f(\theta)\right)^{n_u} \\ &\propto \left[\lambda\left(1-p_f(\theta)\right)\right]^{n_u} \end{aligned}$$

which means $p(n_u \mid n_m, \lambda, \gamma) \propto Poisson\left[\lambda\left(1-p_f(\theta)\right)\right]$ and we have the actual pit number $n^{j+1} = n_u^{j+1} + n_m$.

Step b

3) Weibull Distribution:

The conditional posterior distribution of γ, given β, $\mathbf{h}_d$ and $\mathbf{h}_u$, is

$$\begin{aligned} p(\gamma \mid \beta, \mathbf{h}_d, \mathbf{h}_u) &\propto f(\mathbf{h}_d, \mathbf{h}_u \mid \beta, \gamma) \times \pi(\gamma) \\ &= \prod_{i=1}^{n_m}\left[\gamma\beta\left(h_i^d\right)^{\beta-1}\exp\left(-\gamma\left(h_i^d\right)^{\beta}\right)\right]\times \\ &\quad \prod_{i=1}^{n_d}\left[\gamma\beta\left(h_i^u\right)^{\beta-1}\exp\left(-\gamma\left(h_i^u\right)^{\beta}\right)\right]\times\frac{\gamma^{A-1}B^A}{\Gamma(A)}e^{-\gamma B} \\ &\propto \gamma^{A+n-1}\exp\left[-\gamma\left(B+\sum_{i=1}^{n_m}\left(h_i^d\right)^{\beta}+\sum_{i=1}^{n_u}\left(h_i^u\right)^{\beta}\right)\right] \end{aligned}$$

So, we have $p(\gamma \mid \beta, \mathbf{h}_d, \mathbf{h}_u)$ is a gamma distribution

$$p(\gamma \mid \beta, \mathbf{h}_d, \mathbf{h}_u) \sim Ga\left(A+n, B+\sum_{i=1}^{n_m}\left(h_i^d\right)^{\beta}+\sum_{i=1}^{n_u}\left(h_i^u\right)^{\beta}\right)$$

Similarly, we can derived the conditional posterior distribution of β, given γ, $\mathbf{h}_d$ and $\mathbf{h}_u$, as

$$
\begin{aligned}
p(\beta|\gamma,\mathbf{h}_d,\mathbf{h}_u) &\propto L(\mathbf{h}_d,\mathbf{h}_u|\gamma,\beta)\times\pi(\beta) \\
&= \prod_{i=1}^{n_m}\left[\gamma\beta\left(h_i^d\right)^{\beta-1}\exp\left(-\gamma\left(h_i^d\right)^{\beta}\right)\right]\times \\
&\quad \prod_{i=1}^{n_u}\left[\gamma\beta\left(h_i^u\right)^{\beta-1}\exp\left(-\gamma\left(h_i^u\right)^{\beta}\right)\right]\times \\
&\quad \frac{\Gamma(r+t)}{\Gamma(r)\Gamma(r)\left(\beta_r-\beta_l\right)}\left(\frac{\beta-\beta_l}{\beta_r-\beta_l}\right)^{r-1}\left(\frac{\beta_r-\beta}{\beta_r-\beta_l}\right)^{t-1} \\
&\propto \beta^n\left(\beta-\beta_l\right)^{r-1}\left(\beta_r-\beta\right)^{t-1}\times\left(\prod_{i=1}^{n_m}\left(h_i^d\right)^{\beta-1}\prod_{i=1}^{n_u}\left(h_i^u\right)^{\beta-1}\right)\times \\
&\quad \exp\left(-\gamma\left(\sum_{i=1}^{n_m}\left(h_i^d\right)^{\beta}+\sum_{i=1}^{n_u}\left(h_i^u\right)^{\beta}\right)\right)
\end{aligned}
$$

which is not a density function of any known distribution and therefore a Metropolis-Hastings algorithm shall be used for sampling.

References

Aven, T. (2003) Foundations of Risk Analysis. John Wiley & Sons, London.

Bardal, E. (2004) *Corrosion and Protection*. Springer, London.

Bohni, H. (2000) Localized corrosion of passive metals. In *Uhlig's Corrosion Handbook*, R. W. Revie (ed.), 173-190. John Wiley & Sons, New York.

Celux, G., Persoz, M., Wandji, J. N. and Perrot, F. (1999) Using Markov Chain Monte Carlo Methods to Solve Full Bayesian Modeling of PWR Vessel Flaw Distributions. *Reliability Engineering & System Safety* **66**, 243-252

Galvele, J. R. (1983) Pitting corrosion. In *Treatise on Materials Science and Technology*, J. C. Scully (ed.), 1-57. Academic Press, London.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004) *Bayesian Data Analysis*. Chapman & Hall/CRC, London.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996) Introducing Markov chain Monte Carlo. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Speigelhalter (eds.), 1-19. Chapman & Hall, London.

Gumbel, E. J. (1958) *Statistics of Extremes*. Columbia Univ. Press., New York.

Jones, D. A. (1996) *Principles and Prevention of Corrosion*. Prentice Hall, Upper Saddle River.

Maruska, C. C. (2002) Steam generator life cycle management Ontario Power Generation (OPG) experience. *4th CNS International Steam Generator Conferences*.

Mathworks, Inc. (2007) *MATLAB® help documentation*. http://www.mathworks.com /support/

Melchers, R. E. (2004) Pitting corrosion of mild steel in marine immersion environment-part 1: maximum pit depth. *Corrosion* **60**, 824-836.

Melchers, R. E. (2004) Pitting corrosion of mild steel in marine immersion environment-part 2: variability of maximum pit depth. *Corrosion* **60**, 937-944.

Melchers, R. E. (2005) Statistical Characterization of pitting corrosion-part 1: data analysis. *Corrosion* **61**, 655-663.

Melchers, R. E. (2005) Statistical Characterization of pitting corrosion-part 2: probabilistic modeling for maximum pit depth. *Corrosion* **61**, 766-776.

Melchers, R. E. (2005) Representation of uncertainty in maximum depth of marine corrosion pits. *Structural Safety* **27**, 322-334.

Mola, E. E. and Mellein, B. et al. (1990) Stochastic approach for pitting corrosion modeling I. the case of quasi-hemispherical pits. *J. Electrochem. Soc.* **137**, 1384-1390.

Nickerson, J. and Maruska, C. C. (1998) Steam generator management at Ontario Hydro Nuclear Stations. *3rd CNS International Steam Generator Conferences*, 57-66.

Provan, J. W. and Rodriguez III, E. S. (1989) Part I: development of a Markov description of pitting corrosion. *Corrosion Science* **45**, 178-192.

Roberts, G. O. (1996) Markov chain concepts related to sampling algorithms. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Speigelhalter (eds.), 115-130. Chapman & Hall, London.

Robert, C. P. (2001) *The Bayesian Choice*. Springer, New York.

Rodriguez III, E. S. and Provan, J. W. (1989) Part II: development of a general failure control system for estimating the reliability of deteriorating structures. *Corrosion Science* **45**, 193-206.

Rubinstein, R. Y. (1981). Simulation and Monte Carlo Methods. John Wiley & Sons, New York.

Sheikh, A. K., Boah, J. K. and Hansen, D. A. (1990) Statistical modeling of pitting corrosion and pipeline reliability. *Corrosion* **46**, 190-197.

Shibata, T. and Takeyama, T. (1976) Pitting corrosion as a stochastic process. *Nature* **260**, 315-316.

Shibata, T. and Takeyama, T. (1977) Stochastic theory of pitting corrosion. *Corrosion* **33**, 243-251.

Shibata, T. (1991) Evaluation of corrosion failure by extreme value statistics. *ISIJ International* **31**, 115-121.

Shibata, T. (1996) 1996 Whitney Award Lecture: Statistical and stochastic approaches to localized corrosion. *Corrosion Science* **52**, 813-830.

Snook, W. (2001) *http://canteach.candu.org/imagelib/33000-HTS/Stm_Gen.pdf*

Strehblow, H. (2002) Mechanisms of pitting corrosion. In *Corrosion Mechanisms in Theory and Practice*, P. Marcus (ed.), 243-285. Marcel Dekker, New York.

Tapping, R.L., Nickerson, J., Spekkens, P. and Maruska, C (2000) CANDU steam generator life management. *Nuclear Engineering and Design* **197**, 213-223.