STOCHASTIC MODELING OF INSPECTION UNCERTAINTIES AND APPLICATIONS TO PITTING FLAWS IN STEAM GENERATOR TUBES

D. Mao[1], X.-X. Yuan, M. D. Pandey

[1] AMEC NSS, 4th Floor, 700 University Avenue, Toronto, Ontario
[2] Department of Civil Engineering, Ryerson University, Toronto, Ontario
[3] Department of Civil Engineering, University of Waterloo, Waterloo, Ontario

Abstract

Steam generators (SG) are a major pressure retaining component of great safety significance in nuclear power plants. Due to various manufacturing, operation and maintenance activities, as well as material interaction with the surrounding chemical environment, the SG tubes have been subject to a number of degradation modes. Among them, the under-deposit pitting corrosion at outside surfaces of the SG tubes just on top of the tubesheet support plates has had a serious impact on the integrity of the SG tubes. This paper 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 based on Markov Chain Monte Carlo (MCMC) simulation is developed for estimating the model parameters. The proposed model is able to predict the actual pit number, the actual pit depth as well as the maximum pit depth, which is the main interest of the pitting corrosion model.

1. Introduction

Steam generators in nuclear power plants have experienced varying degrees of under-deposit pitting corrosion. 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; Shibata, 1996; Strehblow, 2002). 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 (Tapping et al., 2000; 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. One of the key issues in developing the life-cycle management system is to accurately quantify risks associated with the pitting corrosion. For that purpose, an advanced probabilistic model shall be developed that takes into account the various uncertainties of the pitting corrosion from in-service inspection. The uncertainties include both the inherent randomness in the generation and growth of pits and inspection uncertainties. A localized damage, the pit is usually very difficult to be detected. Therefore, the probability of detection (POD) should be considered. Even if 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.

The paper, following the idea of Celeux et al. (1999) developed for cracking flaws in pressure tubes, 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 paper 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 predict the number and size of pits for steam generators. It is also used to predict the distribution of maximum pit depth, which is one of the major decision-making parameters in life-cycle management program of steam generators.

2. Probabilistic Pitting Corrosion Model

A pit is usually treated as a one-dimensional damage, i.e., the pit depth. A pit model should specify the dynamics of pitting generation and pitting growth. With consideration of the availability of ISI data for model calibration, the following assumptions are made in the paper:

(i) The actual sizes of pits $H_1, H_2, \ldots, H_N$ are independent random variables and they follow a Weibull distribution whose probability density function (PDF) is expressed as

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

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

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

$$\Pr\left(N = n\right) = \frac{\lambda^n}{n!}e^{-\lambda} \tag{2}$$

for n = 0, 1, 2, ….

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

This model involves three parameters: λ, γ and β. Given these parameters, the distribution of the maximum pit depth can be derived. Denote the maximum pit depth by H_{max}. Based on the assumptions made above, its cumulative distribution function can be derived as

$$F_{H_{max}}(h) = \sum_{n=0}^{+\infty} \left[\Pr\left(H_i \le h\right) \right]^n \Pr\left(N = n\right) = \exp\left\{-\lambda \exp\left(-\gamma h^{\beta}\right)\right\} \quad (3)$$

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

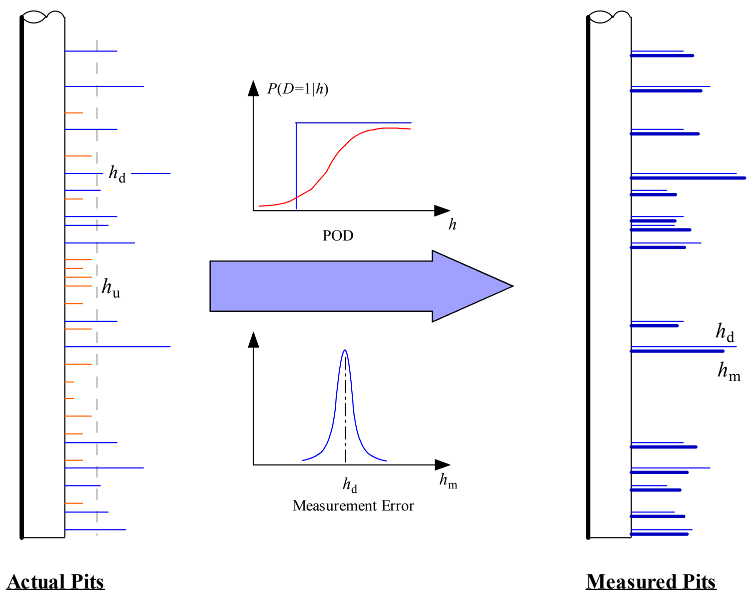

Figure 1: Measurement Uncertainties of ISI Data

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 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), which is defined as a conditional probability depending on the pit depth

$$\text{POD}(h) = P(D = 1 | H = h) \quad (4)$$

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.

A logistic function with threshold is assumed for the POD

$$\mathrm{POD}(h) = \begin{cases} 1 - \dfrac{1+\exp(-qh^*)}{1+\exp\left[q(h-s-h^*)\right]}, & \text{if } h > s \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where s is the threshold of detection, introduced previously; h^* can be considered as a location parameter indicating the starting point of "good" detection; and q is an index measuring the quality of detection. Bigger q implies better detectability.

After the detection with uncertainty, the actual pits are divided into two groups: the detected pits, denoted by $\mathbf{h_d} = \left(h_{d1}, h_{d2}, \ldots, h_{d,n_d}\right)$, and the undetected pits, denoted by $\mathbf{h_u} = \left(h_{u1}, h_{u2}, \ldots, h_{u,n_u}\right)$, 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 $\mathbf{h_m} = \left(h_{m1}, h_{m2}, \ldots, h_{m,n_d}\right)$, differs from their actual depth by an additive random measurement error $\mathbf{e}$, i.e.,

$$\mathbf{h_m} = \mathbf{h_d} + \mathbf{e} \tag{6}$$

where $\mathbf{e} = (e_1, e_2, \ldots e_{nd})$. In this paper, the measurement errors are assumed to be independent and identically Gaussian distributed with zero mean and known variance. The PDF of the measurement errors is expressed as

$$f_E(e) = \frac{1}{\sigma_E\sqrt{2\pi}}\exp\left(-\frac{e^2}{2\sigma_E^2}\right) \tag{7}$$

where σ_E^2 is the variance.

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 predict the maximum pit depth while eliminating the effects of POD and measurement errors.

3. Bayesian Statistical Methods for Estimating the Proposed Model

3.1. Bayesian Approach

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. It can be shown that the likelihood function of the three parameters based on the ISI data involves an infinite summation and several multi-dimensional integrations (Mao, 2007). In this case, Bayesian method is a good alternative to the maximum likelihood method.

For a Bayesian analysis, prior distributions for the parameters should be also specified. In this study, a Gamma distribution is used for the priors of γ and λ, whereas a Beta distribution used for β. Specifically, The prior distribution of γ is a gamma distribution with two hyper-parameters A and B, denoted by $\gamma \sim Ga(A,B)$, and its PDF is expressed as

$$\pi(\gamma) = \frac{B^A \gamma^{A-1}}{\Gamma(A)} e^{-\gamma B} \tag{8}$$

Similarly, we assume $\lambda \sim Ga(a,b)$. The Beta prior for β, however, has four hyper-parameters. Denoted by $\beta \sim Beta(R,T,L,U)$, its PDF is expressed as

$$\pi(\beta) = \frac{\Gamma(r+t)}{\Gamma(r)\Gamma(t)(\beta_r - \beta_l)} \left(\frac{\beta - \beta_l}{\beta_r - \beta_l} \right)^{r-1} \left(\frac{\beta_r - \beta}{\beta_r - \beta_l} \right)^{t-1} \tag{9}$$

in which β_l and β_r are the lower and upper bounds of β. It can be shown that the posterior distribution for γ (or λ) given β (or δ) is again a Gamma distribution and thus the Gamma prior is conditionally conjugate, whereas the Beta prior for β is not conjugate.

With the prior distributions specified, the full Bayesian pitting corrosion model is defined. As illustrated in Figure 2, this model includes five levels. On the bottom of the model are the observations that include the number of detected pit flaws and the measured pit depths at each inspection campaign. On the top are the hyper-parameters for the prior distributions. The observations and hyper-parameters, altogether with the knowledge of measurement errors (σ_E) and the POD, are the known inputs (symbolized by squares in the figure) that are used to derive the posterior distribution of the parameters $\boldsymbol{\xi} = (\gamma, \beta, \lambda)$ through Bayesian updating. Another level is added in the middle to represent the missing data, which includes the number of undetected pits (hence the total number of pits), measurement errors, and the actual depths of both detected and undetected pits. With the aid of this additional level of missing data, the process of Bayesian updating becomes easier, as explained next

To avoid the difficulty of evaluating the likelihood function, a data augmentation technique is used. As shown in Figure 2, 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. In fact, if 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

$$L\left(\boldsymbol{\theta} \middle| \mathbf{h_d}, \mathbf{h_u}\right) = (\gamma\beta)^{n_d + n_u} \left(\prod_{i=1}^{n_d} h_{d,i}^{\beta-1} \prod_{i=1}^{n_u} h_{u,i}^{\beta-1} \right) \times \exp\left\{ -\gamma \left(\sum_{i=1}^{n_d} h_{d,i}^{\beta} + \sum_{i=1}^{n_u} h_{u,i}^{\beta} \right) \right\} \tag{10}$$

and the latter a Poisson distribution function, $L\left(\lambda \middle| n\right) = \frac{\lambda^n}{n!} e^{-\lambda}$.

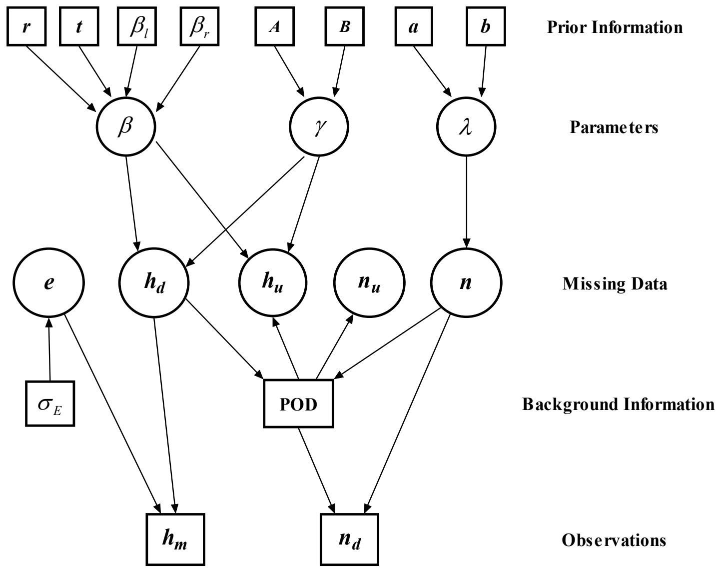

Figure 2: Bayesian pitting corrosion model

3.2. Markov Chain Monte Carlo (MCMC) Simulation

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. Detailed discussions of the method are beyond the scope of the paper and they can be referred to, e.g., Gilks et al. (1996), Roberts (2001), and Gelman et al. (2004). The key of MCMC is to construct a Markov chain of which the stationary distribution is the target posterior distribution and to run the simulation long enough so that the distribution of the current draws is close enough to this stationary distribution. Two common algorithms of MCMC are Gibbs sampler and metropolis-Hasting algorithm (Gilks et al., 1996).

As mentioned in the previous subsection, the data augmentation technique can be used to simplify the likelihood function with the aid of missing data. Of course, we are not able to directly obtain the missing data. In the framework of MCMC simulation, the missing data can be simulated iteratively from the assumed probabilistic model using the simulated parameters. For example, once λ is obtained, we can generate a sample of n from the assumed Poisson distribution. After the actual pit number is obtained, the pit number is then used to update the parameter λ using the Bayesian rule and to generate the actual number of pit sizes. The detail of the iteration algorithm is described next.

The algorithm consists of two iterated steps: a) computation of the conditional distribution of λ knowing $\theta = (\gamma, \beta)$ and b) computation of the conditional distribution of θ knowing λ. Assume that after j^{th} iterations we have λ^j, γ^j and β^j.

Step a

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

2) Calculate $p_f(\theta^j)$ from

$$p_f(\theta^j) = \frac{1}{1000}\sum_{i=1}^{1000} \mathrm{POD}(y_i^j) \tag{11}$$

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 $Pois\left[\lambda^{j+1}\left(1-p_f(\theta^j)\right)\right]$, where $Pois(\alpha)$ denotes a Poisson distribution with mean α.

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 2⁄4 be a random drawing from $f(h|\theta^j)$. If 2⁄4 $< s$ where s is the detection threshold, accept this value; otherwise accept it with the probability $1-POD($2⁄4$)$. 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 $Ga\left(A + n^{j+1}, B + \sum_{i=1}^{n}\left(h_i^{j+1}\right)^{\beta}\right)$, where h^{j+1} include both detected and undetected pits.

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

$$p\left(\beta\middle|\gamma^{j+1}, \mathbf{h}^{j+1}\right) \propto \beta^n \left(\beta-\beta_l\right)^{r-1}\left(\beta_r-\beta\right)^{s-1}\left(\prod_{i=1}^{n_d}\left(h_i^{j+1}\right)^{\beta-1}\right)\exp\left\{-\gamma\sum_{i=1}^{n}\left(h_i^{j+1}\right)^{\beta}\right\} \tag{12}$$

The Metropolis- Hasting algorithm is used to draw a sample of β from the above conditional posterior distribution.

4. Case Study

4.1. Overview of ISI Data

A case study is presented to illustrate the proposed methodology. The pitting corrosion data were collected during an in-service inspection (ISI) outage of a steam generator using eddy current probes.

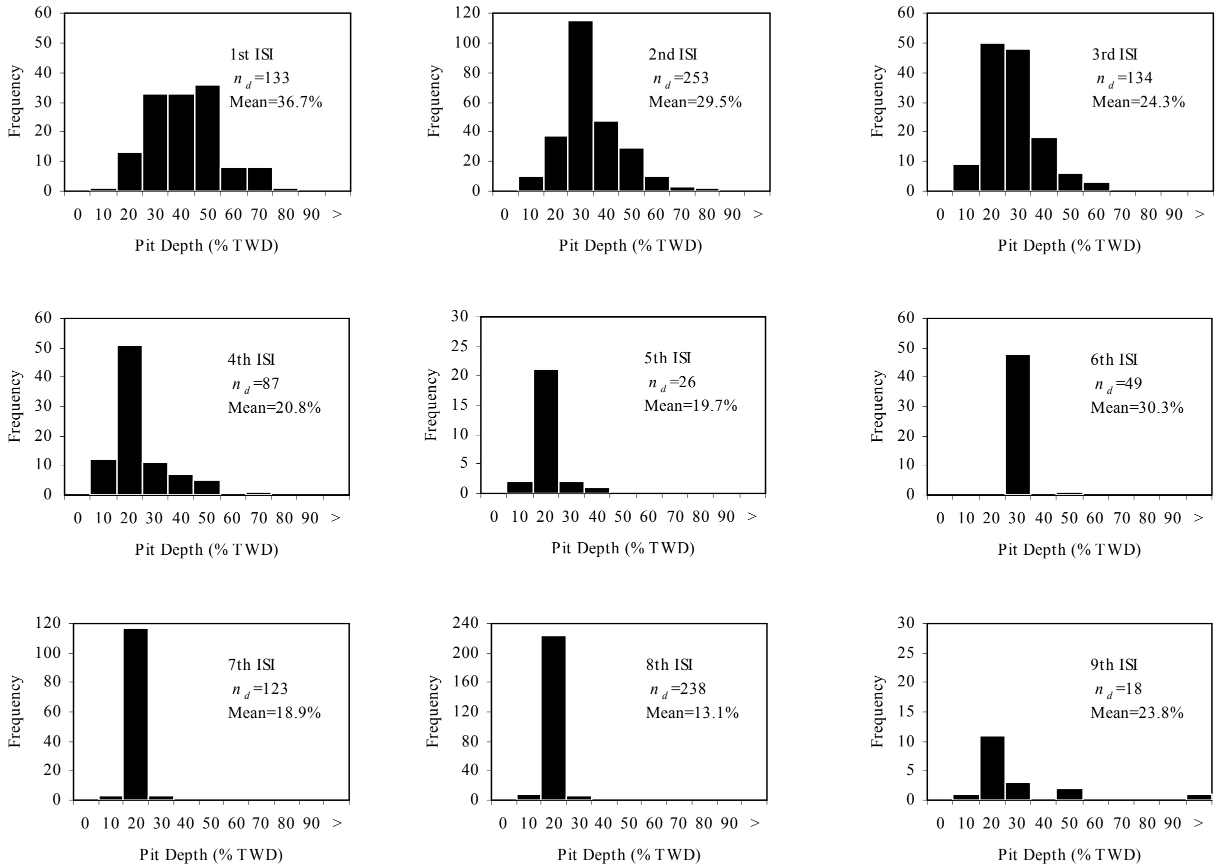

Figure 3: Histograms of pit depth for new pits

The histograms of the new measured pit depth are shown in Figure 3, 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 6th ISI locate almost exclusively at 30% TWD. Similarly, the data from the 7th and 8th campaigns concentrate at 20% TWD, although a relatively large number of new pits have been observed.

In the Bayesian analysis, the prior mean of λ has significant impact on the posterior distributions. 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 3, we 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 $t = 3$.

Based on the design and operational characteristics of the eddy current probes used, we use the POD function as Eq.(5) with q = 20, h^* = 0.10 and s = 0.05. The standard deviation of the measurement error σ_E equals 0.05.

4.2. Results

In the MCMC simulation, 100,000 iterations are run. For the sake of clarity of presentation, we show the sample chains of the model parameters and the corresponding marginal posterior distributions for only the 9th ISI data in Figure 4. It is clear that the samples converge to their corresponding stationary distribution after 20,000 iterations. Therefore we discard the first 20,000 samples as burn-ins and use the remaining to perform statistical analysis.

Figure 5 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.

Figure 6 shows the estimated maximum pit depth distribution with comparison of the estimated actual pit depth as well as the measured pit depth for the selected ISI data. For complete set of the comparisons, refer to Mao (2007). Comparing to the measured pit sizes, the distributions of actual pit sizes have smaller mean values, whereas the distributions of the maximum pit size shifts towards to greater mean values. It is also interesting that the first ISI data predict a significant probability that the maximum pit depth is greater than 100% TWD, i.e., the SG tubes would leak. In fact, after that ISI campaign, a broken SG tube was found indeed. After that event, preventive maintenance measures such as chemical cleaning and water lancing were taken and thereafter the probability of leak has been reduced significant.

5. Conclusions

Under-deposit pitting corrosion at outside surfaces of the steam generator tubes just on top of the tubesheet support plates has had a serious impact on the integrity of the SG tubes. This paper presents an advanced probabilistic model of pitting corrosion characterizing the inherent randomness of the pitting process and measurement uncertainties of the ISI data obtained from eddy current probes. A MCMC-based Bayesian statistical method is developed for estimating the model parameters. The MCMC technique is shown to be effective for Bayesian computation and inference. The proposed model is able to predict the actual pit number, the actual pit depth as well as the maximum pit depth, which is the main interest of the pitting corrosion model.

6. References

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.

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

Mao, D. (2007). Bayesian Modeling of Pitting Corrosion in Steam Generators. Master thesis. University of Waterloo, Waterloo.

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

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

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

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.

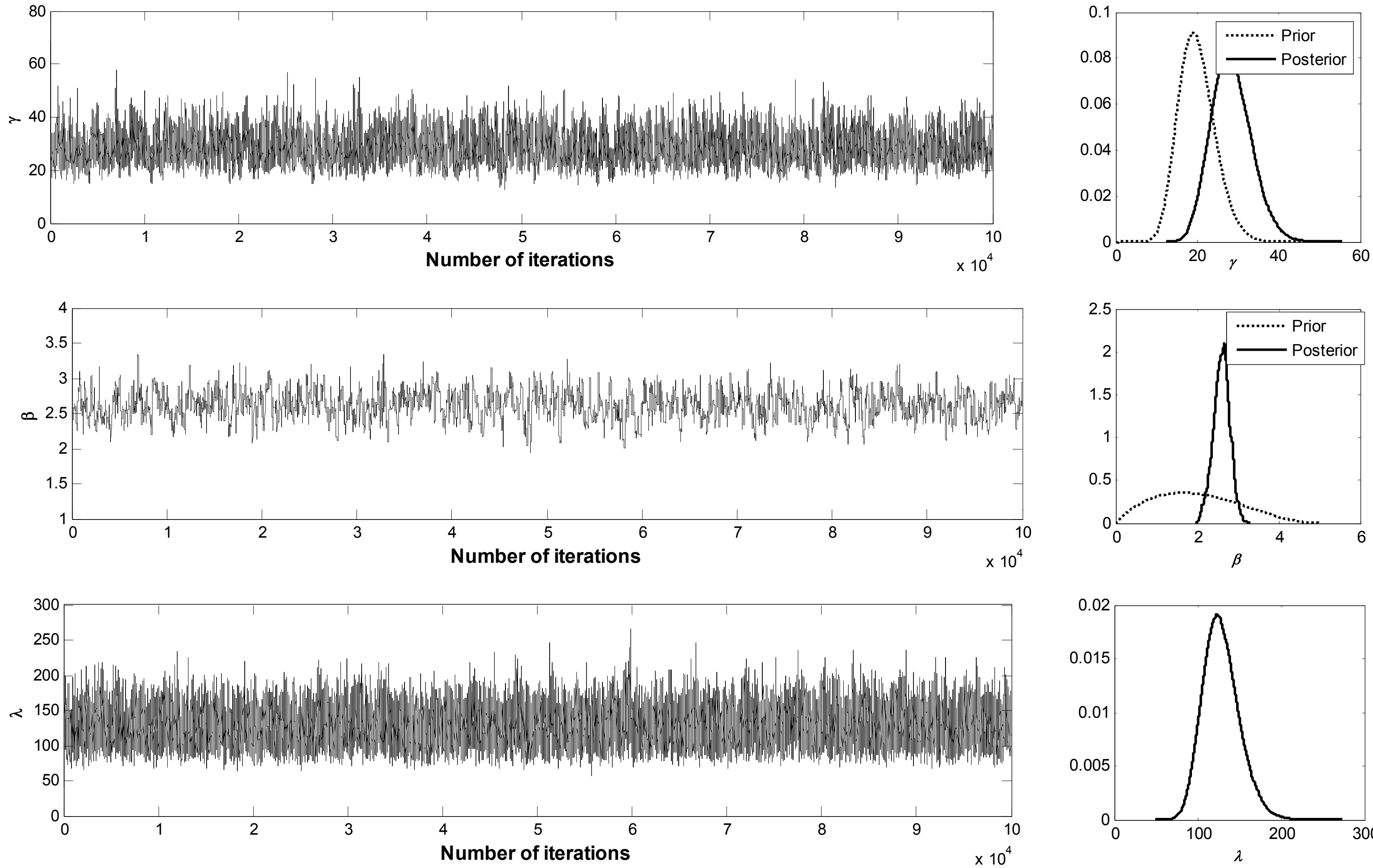

Figure 4: MCMC chains and marginal distributions of the model parameters for the 6^{th} ISI Data

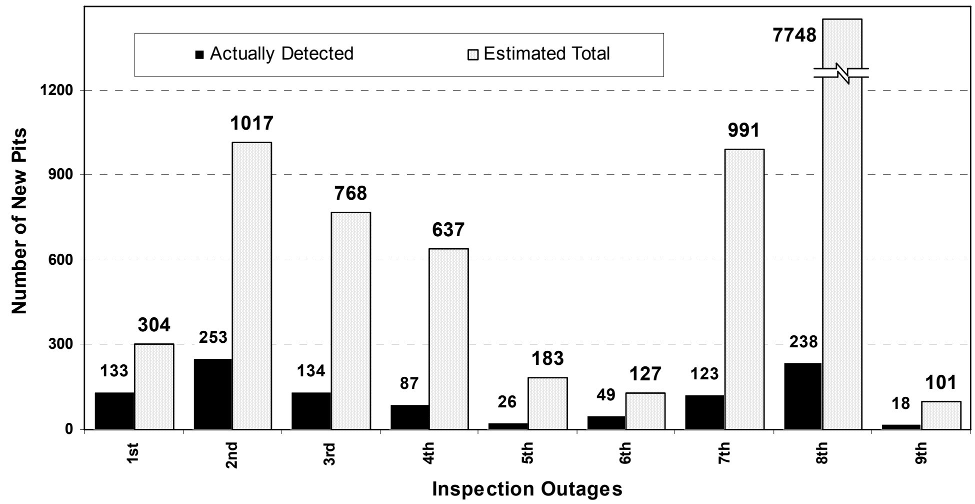

Figure 5: Comparison of actual pit numbers with measured pit numbers

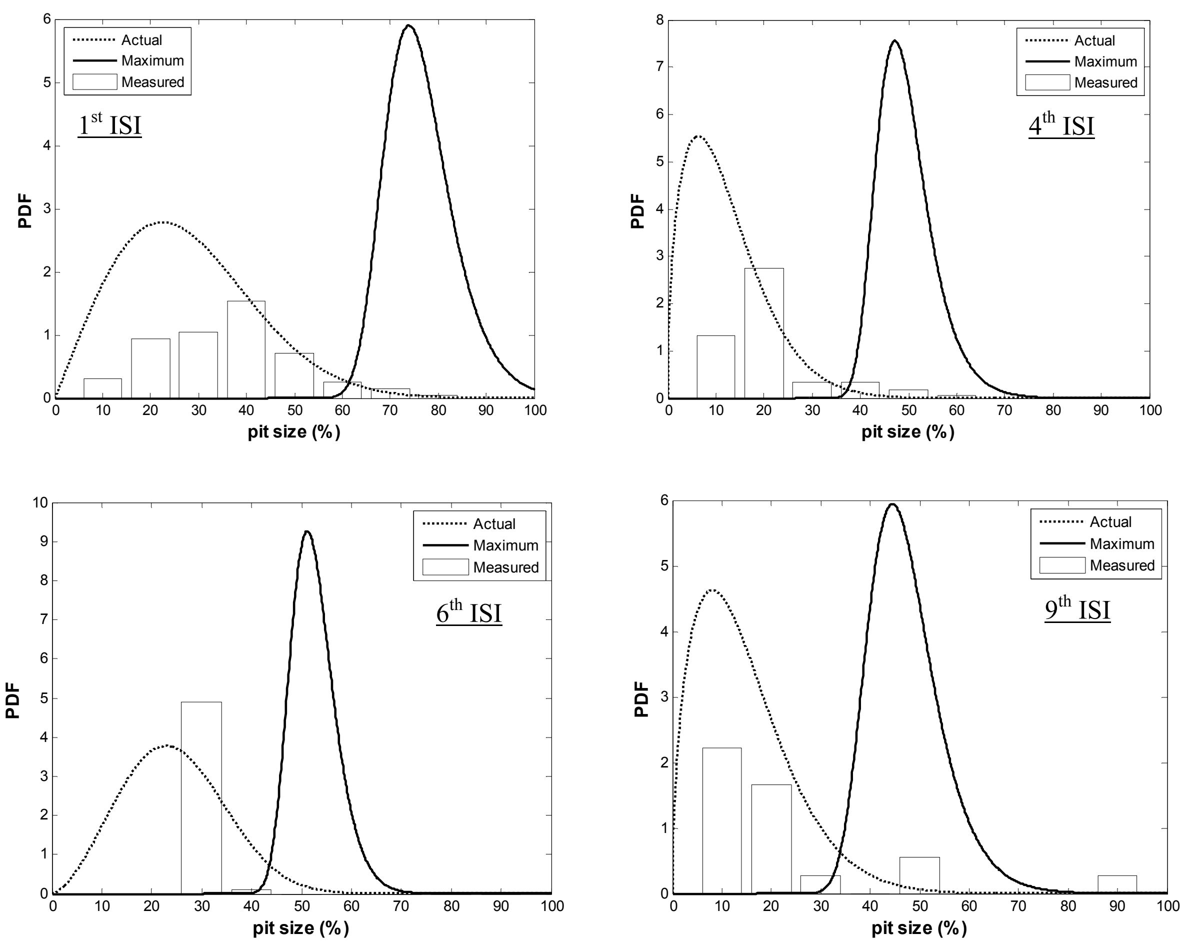

Figure 6: Estimated distributions of actual pit depth and distributions of maximum pit depth