

## Research Article

# Applying Hierarchical Bayesian Neural Network in Failure Time Prediction

Ling-Jing Kao<sup>1</sup> and Hsin-Fen Chen<sup>2</sup>

<sup>1</sup> Department of Business Management, National Taipei University of Technology, 10608, Taiwan

<sup>2</sup> Graduate Institute of Industrial and Business Management, National Taipei University of Technology, 10608, Taiwan

Correspondence should be addressed to Ling-Jing Kao, [lingjingkao@ntut.edu.tw](mailto:lingjingkao@ntut.edu.tw)

Received 31 December 2011; Accepted 21 February 2012

Academic Editor: Jung-Fa Tsai

Copyright © 2012 L.-J. Kao and H.-F. Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid technology development and improvement, the product failure time prediction becomes an even harder task because only few failures in the product life tests are recorded. The classical statistical model relies on the asymptotic theory and cannot guarantee that the estimator has the finite sample property. To solve this problem, we apply the hierarchical Bayesian neural network (HBNN) approach to predict the failure time and utilize the Gibbs sampler of Markov chain Monte Carlo (MCMC) to estimate model parameters. In this proposed method, the hierarchical structure is specified to study the heterogeneity among products. Engineers can use the heterogeneity estimates to identify the causes of the quality differences and further enhance the product quality. In order to demonstrate the effectiveness of the proposed hierarchical Bayesian neural network model, the prediction performance of the proposed model is evaluated using multiple performance measurement criteria. Sensitivity analysis of the proposed model is also conducted using different number of hidden nodes and training sample sizes. The result shows that HBNN can provide not only the predictive distribution but also the heterogeneous parameter estimates for each path.

## 1. Introduction

In this high technology era, the society operations highly depend on various machinery and equipments. Once the machinery or equipment is broken down, enormous trouble and economics cost will be brought to the entire society. To enhance the product reliability, the methodologies to assess product reliability have received much discussion in both academics and industries. Among several mature techniques, degradation testing provides an efficient way for reliability assessment when product quality is associated with a time-varying degradation process. Typically, degradation measures can provide more reliability information, particularly in modeling the failure-causing mechanism, than time-to-failure data in few or no failure situation.

Predicting the remaining lifetime of a product is also an important issue in quality control. For example, knowing the remaining equipment lifetime can help in optimizing the machine maintenance schedule. The equipment lifetime is traditionally studied by fitting a statistical probability distribution, and most of these statistical models are constructed to study various degradation processes of a product. Examples include Lu and Meeker [1], Lu et al. [2], and Meeker and Hamada [3]. The method of stochastic processes is the alternative used to study the degradation data. Examples can be found in Doksum and Hoyland [4].

Most of above methods emphasize on parameter estimations or the process of hypothesis testing. Under the assumption that data follows a certain probability distribution, the statistical inference is made based on the asymptotic theory. The statistical inferences based on the asymptotic theory are valid only if the sample size is large or close to infinite. When the sample information is small or when the discrete data are provided, the finite sample property of the estimation based on the asymptotic theory is not held. Therefore, nonparametric or semiparametric statistics have been proposed to perform the reliability prediction. However, these statistical methods are far from perfect because the overfitting problem usually leads to inaccurate parameter estimates.

Due to the data limitations and the drawbacks of classical statistics approaches, Bayesian approach provides the solution from a different perspective. Unlike these frequentist's approaches which consider the data random and the test statistics or estimators are investigated over imaginary samples  $f(y | \theta)$ , Bayesian approach regards the sampling distribution irrelevant to the statistical inferences because it considers events that have not occurred yet. Bayesian inference is conducted using Bayes theorem in which posterior distribution is defined by the likelihood function which contains sample information times the prior distribution of parameter of the interest. Since Bayesian inference follows the formal rules of probability theory, Bayes estimators are consistent, asymptotically efficient, and admissible under mild conditions. The detail discussion of Bayesian approach can be found in Bernardo and Smith [5], Gelman et al. [6], Robert and Casella [7], and Liu [8].

Lately Bayesian has been applied in the fatigue crack growth prediction in the literature. For example, Zheng and Ellingwood [9] generalize a stochastic fatigue crack growth model by incorporating a time-dependent noise term described by arbitrary marginal distributions and autocorrelations to model the uncertainty in the crack growth under constant amplitude loading. Zhang and Mahadevan [10] propose a Bayesian procedure to quantify the uncertainty in mechanical and statistical model selection and the uncertainty in distribution parameters. The procedure is applied to a fatigue reliability problem, with the combination of two competing crack growth models and considering the uncertainty in the statistical distribution parameters for each model. Akama [11] performs the Bayesian analysis to estimate an appropriate value of the uncertain propagation rate of cracks that can be initiated at the wheel seat of a Shinkansen vehicle axle.

Neural network (NN) is the other popular prediction method. Neural network is a computer-intensive, algorithmic procedure for transforming inputs into desired outputs using highly inter-connected networks of relatively simple processing elements (often termed neurons, units, or nodes; we will use nodes thereafter). Neural networks are modeled following the neural activity in the human brain. The essential features of a neural network are the nodes, the network architecture describing the connections between the nodes, and the training algorithm used to find values of the network parameters (weights) for a particular network. The nodes are connected to one another in the sense that the output from one node can be served as the inputs to other nodes. Each node transforms an input to an output using some specified function that is typically monotone, but otherwise arbitrary. This function

depends on parameters whose values must be determined with a training set of inputs and outputs. Network architecture is the organization of nodes and the types of connections permitted. The nodes are arranged in a series of layers with connections between nodes in different layers, but not between nodes in the same layer.

Several researchers also integrate neural network algorithm with Bayesian theory, which has been known as Bayesian neural network, in prediction. For examples, Neal [12] applied Hybrid Markov chain Monte Carlo (MCMC) numerical integration techniques for the implementation of Bayesian procedures. Müller and Rios Insua [13] proposed a MCMC scheme based on a static or dynamic number of hidden nodes. In their subsequent paper, they have extended their research results by releasing the constraint of number of hidden nodes [13]. Also, Holmes and Mallick [14] used Bayesian neural network modeling in the regression context.

In this paper, we conduct a hierarchical Bayesian neural network analysis with MCMC estimation procedure in the failure time prediction problem. Here, hierarchy means that the coefficients in our constructed HBNN model are specified by random effect distributions. We attempt to use this hierarchical structure to determine if the heterogeneity exists among paths. The advantage of proposed HBNN model cannot only provide a better failure time prediction by incorporating the heterogeneity of components and autocorrelated structure of error term but also provide a predictive distribution for the target value. Different from previous research, the proposed HBNN model can successfully offer the full information of parameter estimation and covariance structure. Engineers can use the heterogeneity estimates to identify the causes of the quality differences and further enhance the product quality.

The data of the fatigue crack growth from Bogdanoff and Kozin [15] is used to illustrate the proposed model. To demonstrate the effectiveness of the proposed model, the prediction performance of the proposed model is evaluated using multiple performance measurement criteria. Sensitivity evaluation of the proposed model is also conducted using different number of hidden nodes and training sample sizes. The result shows that HBNN can provide not only the predictive distribution but also the heterogeneous parameter estimates for each path.

The rest of this paper is organized as follows: Section 2 introduces the proposed HBNN model for failure time prediction. In Section 3, the fatigue crack growth data from Bogdanoff and Kozin [15] is illustrated, and the model estimation procedure is provided. Failure time prediction and sensitivity analysis are demonstrated in Section 4. Concluding remarks are offered in Section 5.

## 2. HBNN Model for Failure Time Prediction

To model failure time, we adapted the growth-curve equation used by Liski and Nummi [16] as follows:

$$t_{i,j} = \beta_{i,m=0} + \sum_{m=1}^M \beta_{i,m} \Gamma_{i,m,j} + \varepsilon_{i,j}, \quad (2.1)$$

$$\Gamma_{i,m,j} = \psi(\varpi_{i,k=0,m} + \varpi_{i,k=1,m} y_{i,j} + \varpi_{i,k=2,m} \ln(y_{i,j})),$$

where  $y_{i,j}$  is the  $j$ th crack length of the  $i$ th path and  $t_{i,j}$  is the observed cycle time of the  $i$ th path, where  $i = 1, 2, \dots, N$  and  $j = 1, 2, \dots, n_i$ . In addition,  $\beta_{i,m}$  are the weights of the  $i$ th path

attached to the hidden nodes  $m$  ( $m = 1, 2, \dots, M$ ),  $M$  is the number of hidden nodes,  $\Gamma_{i,m,j}$  is the output of the  $m$ th hidden node when the  $j$ th crack length of the  $i$ th path is presented,  $\varpi_{i,k=1,m}$  are the weights from the first input,  $y_{i,j}$ , to the hidden node  $m$ , and  $\varphi$  is the activation function. Typically, the choice of  $M$  depends upon the problem under consideration. The testing results of neural network with combinations of different numbers of hidden nodes have been investigated. In the present case, we have set the number of hidden nodes  $M$  equal to 3 because it gives the best predicting result.

According to the above equation, we know that there are totally  $N$  paths from a given population, and  $n_i$  observations are available for path  $i$  at fixed crack lengths  $y_{i,1}, y_{i,2}, \dots, y_{i,n_i}$  (i.e., The observations at length  $y_{i,1}, y_{i,2}, \dots, y_{i,n_i}$  are  $t_{i,1}, t_{i,2}, \dots, t_{i,n_i}$ , resp.). Herein, we assume that the conditional distribution of  $t_{i,j}$  given  $y_{i,j}$  is normally distributed as  $f(t_{i,j} | y_{i,j}) \sim N(\beta_{i,0} + \sum_{m=1}^M \beta_{i,m} \Gamma_{i,m,j}, \sigma^2)$ . It means that each value of  $y_{i,j}$  produces a random value of  $t_{i,j}$  from a normal distribution with a mean of  $\beta_{i,0} + \sum_{m=1}^M \beta_{i,m} \Gamma_{i,m,j}$  and a variance  $\sigma^2$ . Moreover, from literature [17, 18], we understand that degradation signals are usually autocorrelated in nature. We also noticed that the values of the first-order autocorrelation  $r_1$  of the residuals in Lu and Meeker [1] are not exactly equal to 2.0. Therefore, we suspected that the error term might be characterized as a first-order autoregressive process. Based on this finding, we proposed a new parametric crack growth model with autocorrelated errors as the following equations:

$$t_{i,j} = \beta_{i,0} + \sum_{m=1}^M \beta_{i,m} \Gamma_{i,m,j} + \varepsilon_{i,j}, \quad (2.2)$$

$$\Gamma_{i,m,j} = \varphi(\varpi_{i,0,m} + \varpi_{i,1,m} y_{i,j} + \varpi_{i,2,m} \ln(y_{i,j})), \quad (2.3)$$

$$\varepsilon_{i,j} = \rho_i \varepsilon_{i,j-1} + Z_{i,j}, \quad (2.4)$$

where  $\rho_i$  is the autocorrelation coefficient and  $Z_{i,j}$  is a normal distributed error with  $N(0, \sigma^2)$  form. Note that the elements  $t_{i,1}, t_{i,2}, \dots, t_{i,n_i}$  in (2.2) are independent given  $\beta_{i,m}, \varpi_{i,k,m}, \sigma^2, \rho_i$  and  $y_{i,j}$ , where  $k$  is the number of inputs. The function  $\varphi(\cdot)$  is referred to as an activation function in a neural network. Typically, the activation function is nonlinear. Some of the most common choices of  $\varphi(\cdot)$  are the logistic and the hyperbolic tangent functions. In order to describe the heterogeneity varying from path to path, we characterized  $\beta_i$  by a 4-variate normal distribution with mean vector  $\bar{\beta}$  and covariance matrix  $V_\beta$ ,  $\beta_i | \bar{\beta}, V_\beta \sim N_4(\bar{\beta}, V_\beta)$ ,  $i = 1, 2, \dots, N$ , and  $\varpi_{i,m}$  is characterized by a 3-variate normal distribution with mean vector  $\bar{\varpi}_m$ , and covariance matrix  $V_{\varpi_m}$  for  $m = 1, 2$ , and 3. Equations (2.2)–(2.4) specify a general model for studying when observed cycle time sensitivity to crack length may increase. The heterogeneity among paths is captured by parameters  $\beta_i, \varpi_{i,k,m}$ , and the specification of covariates  $V_\beta$  and  $V_{\varpi_m}$ .

According to the above setting, the likelihood function for the data can be written as

$$l(\beta_i, \varpi_{i,m}, \sigma^2, \rho_i | \{t_{i,j}\}) \propto \prod_{i=1}^N \prod_{j=1}^{n_i} [t_{i,j} | Y_{i,j}, \beta_i, \varpi_{i,m}, \sigma^2, \rho_i]. \quad (2.5)$$

To reduce the computational burden of posterior calculation and exploration, we define  $\bar{\beta} \sim N_4(\alpha, V_\alpha)$ ,  $V_\beta \sim \text{Inv.Wishart}(f_0, G_0)$ ,  $\bar{\varpi}_m \sim N_3(p, V_p)$ ,  $V_{\varpi_m} \sim \text{Inv.Wishart}(f_1, G_1)$ ,

$\sigma^2 \sim \text{Inv.Gamma}(r_0/2, S_0/2)$ ,  $\rho_i \sim \text{Uniform}(-1, 1)$  as the conjugate priors on the parameters  $\beta_i, \bar{\beta}, V_{\beta}, \bar{\omega}_{i,m}, \bar{\omega}_m, V_{\bar{\omega}_m}, \sigma^2$ , and  $\rho_i$ , respectively. Typically, the selection of priors is problem-specific. Some have even criticized Bayesian approach as relying on “subjective” prior information. However, we should also notice that the basis of prior information could be “objective” or data-based. The power prior developed by Ibrahim et al. [19] is an example of it. However, in most empirical cases, the utilization of diffuse prior for parameters is a reasonable default choice.

By using the Bayes theorem with the sample information and prior distribution of each parameter, the posterior distribution of each parameter can be derived. The posterior distributions and the details of the estimation procedure can be referred to Carlin and Louis [20]. The posterior distributions of estimated parameters can be summarized as the full conditional probability formulas shown in the Appendix.

In addition to the posterior distribution for the estimated parameter, the predictive distribution of the unobserved cycle time,  $t_{\text{pred}}$ , given the observed cycle time,  $t$ , is one of our main objectives. The predictive distribution is analytically intractable because of the requirement of highly dimensional numerical integration. However, the Markov chain Monte Carlo (MCMC) method provides an alternative, whereby we sample from the posterior directly and obtain sample estimates of the quantities of interest, thereby performing the integration implicitly [21]. In other words, Bayesian analysis of hierarchical models has been made feasible by the development of MCMC methods for generating samples from the full conditionals of the parameters given the remaining parameters and the data.

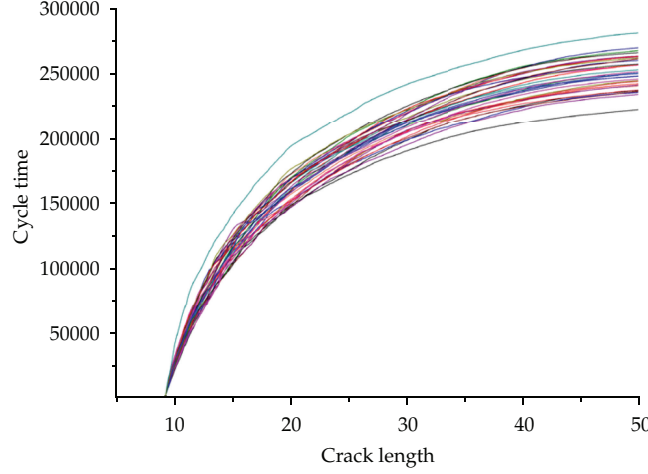
Among these MCMC methods, Gibbs sampling algorithm is one of the best known estimation procedures that uses simulation as its basis [22] and will be used herein to estimate our parameters. It has been shown that, under the mild condition, the Markov chain will converge to a stationary distribution [23]. Beginning with the conditional probability distributions in (5), the Gibbs and Metropolis-Hasting sampling procedure uses recursive simulation to generate random draws. Details of the conditional distributions for the full information model are available upon request. The values of these random draws are then used as the conditional values in each conditional probability distribution, and according to the procedure, generated random draws are carried out again in the next iteration. After numerous reiterated simulations are performed in this way, the convergent results yield random draws that are the best estimates of the parameters.

### 3. Illustrative Example and Model Estimation

#### 3.1. Illustrative Example

We use the fatigue crack growth data from Bogdanoff and Kozin [15] as an illustrative example to demonstrate the modeling procedure and effectiveness of the proposed Hierarchical Bayesian neural network approach. Figure 1 is a plot of the total 30 sample degradation paths. It is obvious that variability amongst paths does exist. There are several possible factors, such as different operating conditions and different material properties, which could cause the variability. Therefore, it is a big challenge to construct a model to capture the statistical properties of degradation paths and to predict failure time.

In this data set, there are 30 sample paths in total and each sample path has 164 paired observations, cycle time, and crack length. The cycle time is observed at some fixed crack lengths. We predefined the path as “failure” as soon as its crack length reaches a particular



**Figure 1:** Thirty paths of fatigue crack growth data from Bogdanoff and Kozin [15].

critical level of degradation (i.e.,  $D_f = 49$  mm) and assumed the experiment was terminated at 40 mm. In other words, based on the measurements of degradation from 9 mm to 40 mm, we would like to model the degradation process and use the proposed model to predict the failure time  $t_{i,j} = D_f$  to the assumed critical level for the degradation path (i.e., crack length = 49 mm). As mentioned, because the fatigue experiment was conducted on paths with fixed crack length, we are interested in the predicted failure time for the path when a specific crack length (i.e., 49 mm) is reached.

### 3.2. Model Estimation

Because the coefficients  $\beta_{i,m}$  and  $\varpi_{i,k,m}$  used to depict the degradation process are high dimensional, it is difficult to integrate out these parameters to obtain the distribution of failure time, especially when complex interactions among random parameters are present. To solve this problem, estimation was carried out using the Markov chain Monte Carlo methods using *R* language. The chain ran for 20,000 iterations, and the last 10,000 iterations are used to obtain parameter estimates. Convergence was assessed by starting the chain from multiple points and inspecting time-series plots of model parameters. Posterior draws from the full conditional are used to compute means and standard deviations of the parameter estimates.

Table 1 reports the posterior mean and standard deviation of the parameters for the proposed model. It shows that the values of  $\Gamma_{m=1,j}$ ,  $\Gamma_{m=2,j}$ , and  $\Gamma_{m=3,j}$  become steady when  $\ln(y_{i,j})$  becomes a large number. The covariance matrix of the heterogeneity distribution is reported in Table 2. It shows that the posterior mean of the diagonal elements of matrix  $V_\beta$  are ranged from 0.0002 to 0.0004. Compared to the outputs of hidden nodes ( $\Gamma_{m,j}$  ranged from 0 to 1), all these diagonal elements are not really small. It represents that the heterogeneity across paths does exist. According to above findings, we can conclude that the proposed HBNN model can successfully determine the heterogeneity across various paths even though, in this particular data set, we were unable to provide explanation to the cause of heterogeneities because of the limited information in data.



**Table 1:** Estimated mean and STD for posterior parameters.

	Estimated parameter					$\bar{\rho}_i$
	$\bar{\beta}$	$\bar{\omega}_{m=1}$	$\bar{\omega}_{m=2}$	$\bar{\omega}_{m=3}$	$\sigma^2$	
Posterior Mean	$\begin{bmatrix} -0.0013 \\ 0.3218 \\ 0.2925 \\ 0.2749 \end{bmatrix}$	$\begin{bmatrix} -8.0899 \\ 0.1952 \\ 0.8385 \end{bmatrix}$	$\begin{bmatrix} -38.3625 \\ -0.6606 \\ 19.3739 \end{bmatrix}$	$\begin{bmatrix} -26.610 \\ -0.1878 \\ 9.8855 \end{bmatrix}$	0.001521	0.7806321
Posterior STD	$\begin{bmatrix} 0.0153 \\ 0.0119 \\ 0.0178 \\ 0.0152 \end{bmatrix}$	$\begin{bmatrix} 0.0075 \\ 0.0003 \\ 0.0012 \end{bmatrix}$	$\begin{bmatrix} 0.0093 \\ 0.0008 \\ 0.0093 \end{bmatrix}$	$\begin{bmatrix} 0.0019 \\ 0.0070 \\ 0.0084 \end{bmatrix}$	0.000313	0.073125

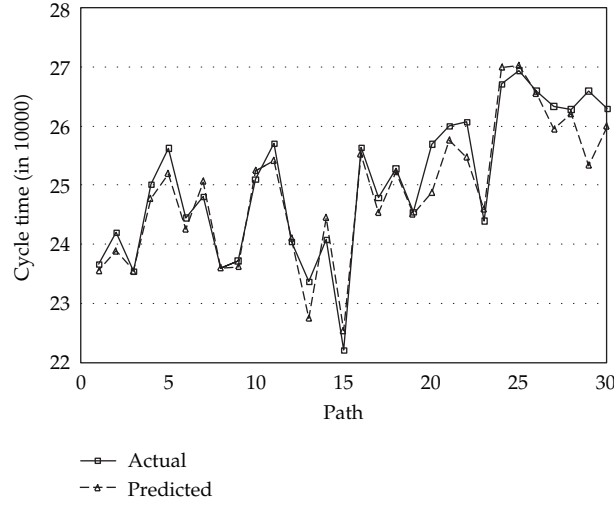
**Table 2:** Covariance matrix of the distribution of heterogeneity.

Estimated parameter	Posterior Mean	Posterior STD
$V_{\beta}$	$\begin{bmatrix} 0.043 & 0.016 & 0.023 & 0.021 \\ 0.023 & 0.001 & 0.009 & 0.044 \\ 0.029 & 0.050 & 0.024 & 0.020 \\ 0.023 & 0.018 & 0.006 & 0.041 \end{bmatrix}$	$\begin{bmatrix} 0.0043 & 0.0024 & 0.0044 & 0.0006 \\ 0.0051 & 0.0022 & 0.0044 & 0.0051 \\ 0.0030 & 0.0009 & 0.0040 & 0.0030 \\ 0.0039 & 0.0003 & 0.0001 & 0.0033 \end{bmatrix}$
$V_{\bar{\omega}_{m=1}}$	$\begin{bmatrix} 0.062 & 0.022 & 0.044 \\ 0.074 & 0.053 & 0.016 \\ 0.033 & 0.046 & 0.063 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.013 & 0.038 \\ 0.055 & 0.039 & 0.068 \\ 0.051 & 0.006 & 0.018 \end{bmatrix}$
$V_{\bar{\omega}_{m=2}}$	$\begin{bmatrix} 0.144 & 0.051 & 0.020 \\ 0.026 & 0.003 & 0.024 \\ 0.081 & 0.043 & 0.099 \end{bmatrix}$	$\begin{bmatrix} 0.033 & 0.021 & 0.028 \\ 0.006 & 0.011 & 0.057 \\ 0.066 & 0.012 & 0.069 \end{bmatrix}$
$V_{\bar{\omega}_{m=3}}$	$\begin{bmatrix} 0.075 & 0.005 & 0.033 \\ 0.036 & 0.076 & 0.001 \\ 0.020 & 0.045 & 0.073 \end{bmatrix}$	$\begin{bmatrix} 0.071 & 0.016 & 0.034 \\ 0.029 & 0.001 & 0.057 \\ 0.062 & 0.002 & 0.078 \end{bmatrix}$

## 4. Failure Time Prediction and Sensitivity Analysis

### 4.1. Failure Time Prediction

The model estimation shown in Section 3 allows us to predict failure time  $t_{i,j}$  to the assumed critical level of degradation (i.e.,  $D_f = 49$  mm) based on the measurements of degradation from 9 mm to 40 mm. In order to demonstrate the effectiveness of the proposed hierarchical Bayesian neural network model, the prediction performance is evaluated using the following performance measures: the root mean square error (RMSE), mean absolute difference (MAD), mean absolute percentage error (MAPE), and root mean square percentage error (RMSPE). The definitions of these criteria were summarized in Table 3. RMSE, MAD, MAPE, and RMSPE are measures of the deviation between actual and predicted failure times. The smaller the deviation, the better the accuracy. The failure time prediction results using the proposed hierarchical Bayesian neural network model are computed and summarized in Figure 2 and Table 4. Table 4 shows that RMSE, MAD, MAPE, and RMSPE of the HBNN model are 0.37340, 0.27121, 1.058%, and 1.440%, respectively. It can be observed that these values are very small. It indicates that there is a smaller deviation between the actual and predicted failure times obtained by the proposed model. Moreover, the proposed HBNN can provide not only posterior estimates of the spatial covariance but also a natural way to incorporate the model uncertainty in statistical inference.



**Figure 2:** Prediction of failure time at 49 mm (when data collection is stopped at 40 mm).

**Table 3:** Performance measures and their definitions.

Metrics	Calculation
RMSE	$RMSE = \sqrt{\sum_{i=1}^{30} (T_i - P_i)^2 / 30}$
MAD	$MAD = \sum_{i=1}^{30}  T_i - P_i  / 30$
MAPE	$MAPE = \sum_{i=1}^{30}  (T_i - P_i) / T_i  / 30$
RMSPE	$RMSPE = \sqrt{\sum_{i=1}^{30} ((T_i - P_i) / T_i)^2 / 30}$

\*Note that  $T$  and  $P$  represent the actual and predicted failure time, respectively.

**Table 4:** Summary of failure time prediction results by HBNN model.

Models	RMSE	MAD	MAPE	RMSPE
HBNN	0.37340	0.27121	1.058%	1.440%

## 4.2. Sensitivity Analysis

To evaluate the sensitivity of the proposed method, the performance of the HBNN model was tested using different number of hidden nodes and training sample sizes. In this section, we set the number of hidden nodes as 3, 4, 5, and 6. And three different sizes of the training dataset (observations collected from 9 (mm) to 30 (mm), 9 (mm) to 35 (mm), and 9 (mm) to 40 (mm) resp.) were considered. The prediction results made by the HBNN model are summarized in Table 5 in terms of RMSE, MAD, MAPE, and RMSPE.

According to the table, the HBNN model has a lower RMSE, MAD, MAPE, and RMSPE with observations collected from 9 (mm) to 40 (mm) than with observations collected from 9 (mm) to 30 (mm). This is because the sample size of the 9 (mm) to 30 (mm) dataset is smaller than the sample size of the 9 (mm) to 40 (mm) dataset. However, the RMSE, MAD, MAPE, and RMSPE are almost the same for the cases of hidden nodes = 3, 4, 5, or 6. This result suggests that there is no difference for the predictions when the number of hidden nodes varies.



**Table 5:** Sensitivity analysis.

# of hidden nodes	Training dataset	RMSE	MAD	MAPE	RMSPE
3	from 9 (mm) to 30 (mm)	0.40576	0.29048	1.108%	1.543%
	from 9 (mm) to 35 (mm)	0.38022	0.27297	1.081%	1.486%
	from 9 (mm) to 40 (mm)	0.37340	0.27121	1.058%	1.440%
4	from 9 (mm) to 30 (mm)	0.40568	0.29060	1.089%	1.558%
	from 9 (mm) to 35 (mm)	0.38070	0.27326	1.057%	1.487%
	from 9 (mm) to 40 (mm)	0.37366	0.27088	1.089%	1.467%
5	from 9 (mm) to 30 (mm)	0.40592	0.29036	1.082%	1.520%
	from 9 (mm) to 35 (mm)	0.38076	0.27350	1.056%	1.456%
	from 9 (mm) to 40 (mm)	0.37391	0.27065	1.062%	1.435%
6	from 9 (mm) to 30 (mm)	0.40606	0.29047	1.104%	1.536%
	from 9 (mm) to 35 (mm)	0.38051	0.27393	1.098%	1.489%
	from 9 (mm) to 40 (mm)	0.37357	0.27068	1.076%	1.461%

## 5. Conclusion

In this paper, we applied the HBNN approach to model the degradation process and to make the failure time prediction. In the process of developing the HBNN model, the MCMC was utilized to estimate the parameters. Since the prediction of failure time made by HBNN model can sufficiently represent the actual data, the time-to-failure distribution can also be obtained successfully. In order to demonstrate the effectiveness of the proposed hierarchical Bayesian neural network model, the prediction performance of the proposed model is evaluated using multiple performance measurement criteria. Sensitivity evaluation of the proposed model is also conducted using different number of hidden nodes and training sample sizes. As the results reveal, using HBNN can provide not only the predictive distribution but also accurate parameter estimate. By specifying the random effects on the coefficients  $\beta_i$  and  $\varpi_{i,m}$  in the HBNN model, the heterogeneity varying across individual products can be studied. Based on these heterogeneities, the engineers will be able to conduct a further investigation in the manufacturing process and then to find out the causes of differences.

For the future research, statistical inferences of failure time based on degradation measurement, such as failure rate and tolerance limits, can be further evaluated given the predicted failure time. In addition, for some highly reliable products, it is not easy to obtain the failure data even under the elevated stresses. In such case, accelerated degradation testing (ADT) can be an alternative that provides an efficient channel for failure time prediction. The proposed HBNN approach can also be applied to depict the stress-related degradation process by including those stress factors as covariates in the model.

## Appendix

### The Full Conditional Probability of Estimated Parameters

$$[\beta_i | \text{rest}] \propto [t_i | Y_i, \beta_i, \varpi_{i,m}, \sigma^2, \rho_i] \cdot [\beta_i | \bar{\beta}, V_\beta],$$

$$[\bar{\beta} | \text{rest}] \propto \prod_{i=1}^N [\beta_i | \bar{\beta}, V_\beta] \cdot [\bar{\beta} | \mu_{\bar{\beta}}, V_{\bar{\beta}}],$$

$$\begin{aligned}
[V_\beta | \text{rest}] &\propto \prod_{i=1}^N [\beta_i | \bar{\beta}, V_\beta] \cdot [V_\beta | f_0, G_0], \\
[\varpi_{i,m} | \text{rest}] &\propto \prod_{j=1}^{n_i} [t_i | \Gamma_i, \beta_i, \varpi_{i,m}, \sigma^2, \rho_i] \cdot [\varpi_{i,m} | \bar{\varpi}_m, V_{\varpi_m}], \\
[\bar{\varpi}_m | \text{rest}] &\propto \prod_{i=1}^N [\varpi_{i,m} | \bar{\varpi}_m, V_{\varpi_m}] \cdot [\bar{\varpi}_m | p, V_p], \\
[V_{\varpi_m} | \text{rest}] &\propto \prod_{i=1}^N [\varpi_{i,m} | \bar{\varpi}_m, V_{\varpi_m}] \cdot [V_{\varpi_m} | f_1, G_1], \\
[\sigma^2 | \text{rest}] &\propto \prod_{i=1}^N [t_i | Y_i, \beta_i, \varpi_{i,m}, \sigma^2, \rho_i] \cdot [\sigma^2 | r_0, s_0], \\
[\rho_i | \text{rest}] &\propto [t_i | Y_i, \beta_i, \varpi_{i,m}, \sigma^2, \rho_i] \cdot [\rho_i].
\end{aligned} \tag{A.1}$$

## References

- [1] C. J. Lu and W. Q. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, no. 2, pp. 161–174, 1993.
- [2] C. J. Lu, W. Q. Meeker, and L. A. Escobar, "A comparison of degradation and failure-time analysis methods for estimating a time-to-failure distribution," *Statistica Sinica*, vol. 6, no. 3, pp. 531–546, 1996.
- [3] W. Q. Meeker and M. Hamada, "Statistical tools for the rapid development & evaluation of high-reliability products," *IEEE Transactions on Reliability*, vol. 44, no. 2, pp. 187–198, 1995.
- [4] K. A. Doksum and A. Hoyland, "Models for variable-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution," *Technometrics*, vol. 34, no. 1, pp. 74–82, 1992.
- [5] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, John Wiley & Sons, 1994.
- [6] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, London, UK, 1995.
- [7] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, New York, NY, USA, 1999.
- [8] S. I. Liu, "Bayesian model determination for binary-time-series data with applications," *Computational Statistics & Data Analysis*, vol. 36, no. 4, pp. 461–473, 2001.
- [9] R. Zheng and B. R. Ellingwood, "Stochastic fatigue crack growth in steel structures subjected to random loading," *Structural Safety*, vol. 20, no. 4, pp. 305–324, 1998.
- [10] R. Zhang and S. Mahadevan, "Model uncertainty and Bayesian updating in reliability-based inspection," *Structural Safety*, vol. 22, no. 2, pp. 145–160, 2000.
- [11] M. Akama, "Bayesian analysis for the results of fatigue test using full-scale models to obtain the accurate failure probabilities of the Shinkansen vehicle axle," *Reliability Engineering and System Safety*, vol. 75, no. 3, pp. 321–332, 2002.
- [12] R. M. Neal, *Bayesian Learning for Neural Networks*, Springer, New York, NY, USA, 1996.
- [13] P. Müller and D. Rios Insua, "Issues in bayesian analysis of neural network models," *Neural Computation*, vol. 10, pp. 571–592, 1998.
- [14] C. C. Holmes and B. K. Mallick, "Bayesian wavelet networks for nonparametric regression," *IEEE Transactions on Neural Networks*, vol. 11, no. 1, pp. 27–35, 2000.
- [15] J. L. Bogdanoff and F. Kozin, *Probabilistic Models of Cumulative Damage*, John Wiley, New York, NY, USA, 1984.

- [16] E. P. Liski and T. Nummi, "Prediction in repeated-measures models with engineering applications," *Technometrics*, vol. 38, no. 1, pp. 25–36, 1996.
- [17] R. B. Chinnam, "On-line reliability estimation for individual components using statistical degradation signal models," *Quality and Reliability Engineering International*, vol. 18, no. 1, pp. 53–73, 2002.
- [18] R. B. Chinnam and P. Baruah, "A neuro-fuzzy approach for on-line reliability estimation and condition based-maintenance using degradation signals," *International Journal of Materials and Product Technology*, vol. 20, no. 1–3, pp. 166–179, 2004.
- [19] J. G. Ibrahim, L. M. Ryan, and M.-H. Chen, "Using historical controls to adjust for covariates in trend tests for binary data," *Journal of the American Statistical Association*, vol. 93, no. 444, pp. 1282–1293, 1998.
- [20] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, CRC Press, London, UK, 2000.
- [21] S. P. Brooks, "Markov chain Monte Carlo method and its application," *Journal of the Royal Statistical Society D*, vol. 47, part 1, pp. 69–100, 1998.
- [22] A. E. Gelfand and A. F. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.
- [23] H. J. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, vol. 24, Springer, New York, NY, USA, 2001.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

