

NONPARAMETRIC METHOD AND HIERARCHICAL  
BAYESIAN APPROACH FOR PARAMETER  
ESTIMATION AND PREDICTION

by

Jing Cai

B.Eng, South China University of Technology, 1997

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE  
in the  
Department of Statistics and Actuarial Science  
Faculty of Science

© Jing Cai 2011

SIMON FRASER UNIVERSITY

Summer 2011

*All rights reserved.*

*However, in accordance with the Copyright Act of  
Canada, this work may be reproduced without authorization under the  
conditions for Fair Dealing. Therefore, limited reproduction of this  
work for the purposes of private study, research, criticism, review and  
news reporting is likely to be in accordance with the law, particularly  
if cited appropriately.*

## APPROVAL

**Name:** Jing Cai  
**Degree:** Master of Science  
**Title of project:** Nonparametric Method and Hierarchical Bayesian Approach for Parameter Estimation and Prediction

**Examining Committee:** Dr. Yi Lu  
Chair

---

Dr. Jiguo Cao, Senior Supervisor  
Simon Fraser University

---

Dr. David Alexander Campbell, Internal Examiner  
Simon Fraser University

---

Dr. Carl James Schwarz, External Examiner  
Simon Fraser University

**Date Approved:** August 19, 2011

---

## Partial Copyright Licence



The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website ([www.lib.sfu.ca](http://www.lib.sfu.ca)) at <http://summit/sfu.ca> and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library  
Burnaby, British Columbia, Canada

# Abstract

Obtaining accurate estimates or prediction from available data is one of the important goals in statistical research. In this thesis, we propose two new statistical methods, with examples of application and simulation studies, to achieve this goal. The parametric penalized spline smoothing procedure is a flexible algorithm that requires no restricted parametric assumption and is proved to obtain more accurate estimates of curves and derivatives than available methods. In the second part of thesis, we propose a hierarchical Bayesian approach to estimate dynamic engineering model parameters and their mixed effects. This approach has the benefits of solving the identifiability problem of model parameters and accurately estimating these parameters from right censored data. It is further investigated with simulated data to perform predictions. Predicting quality with this method is proved to be better than that from procedures without considering censoring situation.

*To Eric, Carolyn and Karis!*

# Acknowledgments

Pursuing a career in the field that analyzes and interprets data was always my dream ever since I studied my first course in statistics after immigrating to Canada. I am feeling fortunate, grateful and thankful that with the professional guidance from my supervisor, the kindness advices and helps from departmental professors and graduate fellows, and the endless supports from my dearest family, my dream finally comes true. I would like to express my deep debt of gratitude to all that have given me help and support.

I am deeply indebted to my supervisor, Dr. Jiguo Cao. Without your patience, guidance and encouragement, I will never be able to navigate the path this far. I would like to thank you for your thoughtful insights in pointing me the correct direction in my research projects.

I am very grateful to have Dr. David Campbell and Dr. Carl Schwarz in my examining committee, and would like to give my appreciation for their valuable time and suggestions that have made my thesis better. I also want to express my gratitude to the departmental professors who have provided me professional training in statistics. In particular, I would like to thank Dr. Joan Hu for her upmost kindness, inspiration and advice which strengthen my confidence in pursuing my dream. I want to thank all my fellow graduate students, especially, Alisha Albert-Green, Crystal Li, Harsha Perera, and Hua Zheng, for their assistance and friendship that have highlighted my two years graduate study.

I know, as a wife and a mom, I have owed a lot to my husband and my two sweetest daughters, who bring me joy, love and hope every day. I also want to thank my parents for their everlasting love and support.

# Contents

<b>Approval</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Dedication</b>	<b>iv</b>
<b>Acknowledgments</b>	<b>v</b>
<b>Contents</b>	<b>vi</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Parametric Penalized Spline Smoothing</b>	<b>3</b>
2.1 Introduction . . . . .	3
2.2 Method . . . . .	5
2.3 Simulation Studies . . . . .	9
2.3.1 Simulation 1 . . . . .	9
2.3.2 Simulation 2 . . . . .	17
2.4 Applications . . . . .	22
2.5 Conclusions . . . . .	24

<b>3</b>	<b>Simultaneous Hierarchical Bayesian Method</b>	<b>25</b>
3.1	Introduction . . . . .	25
3.2	The Dynamic Duration of Load Model . . . . .	28
3.3	Statistical Approach Set-up . . . . .	29
3.3.1	Experimental Design in Duration of Load Study . . . . .	29
3.3.2	Hierarchical Bayesian Approach . . . . .	30
3.3.3	Bayesian approach for constant-load breakage time prediction	35
3.4	Simulation Studies . . . . .	36
3.4.1	Simulation 1 . . . . .	36
3.4.2	Simulation 2 . . . . .	42
3.5	Conclusions . . . . .	45
	<b>Bibliography</b>	<b>47</b>



# List of Tables

2.1	The average point-wise root mean squared errors (RMSEs) for the estimates of the height function and the first and second derivatives using four methods, when the data are simulated based on the PB function as the true function. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4), and Wahba's standard smoothing spline method (SS). The PB function is defined in (2.5). . . . .	13
2.2	The average point-wise root mean squared errors (RMSEs) for the estimates of the height function and the first and second derivatives using four methods, when the data are simulated based on the JPA function as the true function. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4), and Wahba's standard smoothing spline method (SS). The JPA function is defined in (2.4). . . . .	15

2.3	The average point-wise bias, standard deviations (SDs), and root mean squared errors (RMSEs) for the estimates of the curve and the first and second derivatives using three methods, when the data are simulated based on $f(t) = \exp(2t)/\cos(3t)$ as the true function with noise standard deviation 0.05. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4).	19
2.4	The average point-wise bias, standard deviations (SDs), and root mean squared errors (RMSEs) for the estimates of the curve and the first and second derivatives using three methods, when the data are simulated based on $f(t) = \exp(2t)/\cos(3t)$ as the true function with noise standard deviation 0.01. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4).	20
2.5	The average point-wise absolute value of bias, standard deviations (SDs), and root mean squared errors (RMSEs) for the estimates of the curve and the first and second derivatives using three methods, when the data are simulated based on $f(t) = \exp(2t)/\cos(3t)$ as the true function with noise standard deviation 0.005. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4).	21
3.1	The averaged bias(Bias), standard deviations (SD), and root mean squared errors (RMSE) of the estimates of six parameters using data set with 10% right censoring on the constant-load breakage times. "Cen" denotes estimation considering right censoring and "NoCen" indicates not consider censoring.	40

3.2	The averaged bias(Bias), standard deviations (SD), and root mean squared errors (RMSE) of the estimates of six parameters using data set with 20% right censoring on the constant-load breakage times. "Cen" denotes estimation considering right censoring and "NoCen" indicates not consider censoring. . . . .	41
3.3	Table of descriptive statistics of MSPEs, with readings of 2.5% and 97.5% percentiles, median and mean. "Cen" denotes using parameter estimates considering right censoring and "NoCen" indicates using parameter estimates not consider censoring. . . . .	44

# List of Figures

- 2.1 The JPA function 2.4 and the PB function 2.5, and their first and second derivatives using the true parameter values. The dashed lines are the estimates for the curve,  $x(t)$ , the first derivative,  $x'(t)$ , and the second derivative,  $x''(t)$ , using the parametric penalized spline smoothing method (PPSS) when the PB function is the true function used to simulate data. The PPSS estimates are almost identical to the true ones. 10
- 2.2 The point-wise root mean squared errors (RMSE) for the estimates of the curve,  $x(t)$ , the first derivative,  $x'(t)$ , and the second derivative,  $x''(t)$ , using the parametric penalized spline smoothing (PPSS) method (solid lines) and the penalized spline smoothing (PSS4) method (dashed lines) when the PB function is the true function used to simulate data. 14
- 2.3 The point-wise root mean squared errors (RMSE) for the estimates of the curve,  $x(t)$ , the first derivative,  $x'(t)$ , and the second derivative,  $x''(t)$ , using the parametric penalized spline smoothing (PPSS) method (solid lines) and the penalized spline smoothing (PSS4) method (dashed lines) when the JPA function is the true function used to simulate data. 16
- 2.4 The estimates for the curve  $x(t)$ , the first derivative  $x'(t)$ , and the second derivative  $x''(t)$  with the parametric penalized spline smoothing method (PPSS). The solid lines are the true curve,  $f(t) = \exp(2t)/\cos(3t)$ , and the true derivatives. The dash lines are the estimates using PPSS model. The dotted lines are the estimate using parametric nonlinear regression (PNR) method. The circles are the simulation data, which have white noises with the variance  $\sigma^2 = 0.01^2$ . . . . . 18

2.5	The estimate for the height function $x(t)$ , the first derivative $x'(t)$ , and the second derivative $x''(t)$ with the parametric penalized spline smoothing method. The circles are the real height measurements of one girl in a Berkeley growth study. The shaded areas are the corresponding 95% point-wise confidence intervals. . . . .	22
2.6	K-fold corss-validation (KCV) calculated based on 2.3 by setting $K=10$ . It is minimized when $\log_{10}(\lambda_1) = -0.9$ and $\log_{10}(\lambda_2) = -3.9$ using the grid search method. . . . .	23
3.1	Example of stress history in duration of load . . . . .	27
3.2	Histograms of one set of short-term ( $T_r$ ) and long-term ( $T_c$ ) breakage times simulated from the EDRM model. . . . .	38
3.3	Histograms of one set of six parameters' posterior distributions. Red dotted line in each histogram indicates the location of parameter's true value we have assigned. . . . .	39
3.4	Plots of predicted values against true values. In prediction I, data used to estimate parameters had 10% right censored observations. In prediction II, data used to estimate parameters had 20% right censored observations. Dots represent predicted values that used parameter estimates obtained by considering the censoring condition. Crosses ("+") present predicted values that used parameter estimates obtained by without considering the censoring condition . . . . .	43

# Chapter 1

## Introduction

In inferential statistics, it is important to obtain accurate estimates. An optimal estimate provides the "best possible" approximation drawn from the available data, which in turn can improve the quality of prediction. Many estimators with their corresponding estimation process have been proposed for achieving this goal. For example, some commonly used estimators are maximum likelihood, method of moments, minimum variance unbiased estimator and Markov chain Monte Carlo. Estimation process of these estimators may require parametric or nonparametric assumptions. In addition to these existing estimation process, we propose two new methods developed from my master research projects, which use functional data analysis techniques or hierarchical Bayesian approach.

The parametric penalized spline smoothing (PPSS) method, introduced in Chapter 2, estimates time-dependent smooth curves and their derivatives through some combination of B-spline basis functions. It is a flexible procedure because it performs estimation nonparametrically. However, unlike other nonparametric algorithms that completely ignore expert opinion about the underlying function, PPSS penalizes the distance between the nonparametric estimate and the parametric function which is proposed base on expert opinion or prior knowledge of the underlying function. It further controls the roughness of the nonparametric estimate by assigning optimal values to the smooth parameters. From the simulation studies, this method is proved to obtain more accurate estimates than the existing procedures. It is further applied

to the longitudinal growth measurements of one girl, which are provided in a Berkeley growth study (Tuddenham & Snyder, 1954). We illustrate the quality of the estimation and statistical inferences, such as the 95% point-wise confidence interval of the growth curve, first and second derivatives, through tables and graphs in Chapter 2.

Hierarchical Bayesian approach is often employed to investigate nonlinear mixed effects model. We propose a simultaneous hierarchical model, in Chapter 3, to estimate dynamic parameters and their mixed effects for duration-of-load problem in wood science. Our proposed hierarchical approach considers the uniqueness of wood specimens by introducing random effects to the parameters in the existing engineering model. In addition, our proposed Bayesian model can conduct more accurate estimation from censored observations than procedures that do not take into account the censoring situation. It is further proved, through simulation studies, that our proposed method can perform better prediction on unobserved measurements than procedures without considering censoring. Tables and graphs illustrating the quality of estimation and prediction are presented in Chapter 3.

# Chapter 2

## Parametric Penalized Spline Smoothing

### 2.1 Introduction

Intensive research have been conducted to model human growth over time, known as the human height function. Several parametric models have been widely adopted in research and thought to be sensible. For example, Preece and Baines proposed PB-1 model in 1978 (Preece & Baines, 1978); Bock and Thissen suggested a triple logistic model in 1980 (Bock & Thissen, 1980); Kanefuji and Shohoji derived two models in 1990 (Kanefuji & Shohoji, 1990); and Jolicoeur and his co-workers proposed three different models from 1988 to 1992 (Jolicoeur et al., 1992, 1988). However, these parametric models may not be completely valid for the height function, because they cannot properly model the entire growth curve or the rate of change (Jolicoeur et al., 1992).

To relax the parametric assumption, it is also popular to apply nonparametric smoothing methods to estimate the human growth curve and its derivatives (Gasser et al., 1985, 1984; Ramsay et al., 1994; Ramsay & Silverman, 2005). These nonparametric smoothing methods estimate the growth curve completely from the data without making any parametric assumptions. A shortcoming of these methods is that they completely ignore expert opinion or prior knowledge of the growth curve as



reflected in the parametric models.

Motivated by the above dilemma, a general parametric penalized spline smoothing method is proposed in order to combine both information from the data and expert opinion. The parametric penalized spline smoothing method still estimates the underlying function nonparametrically. The nonparametric function is evaluated with three terms: the first term measures the fit of the nonparametric function to the data, the second term measures the distance of the nonparametric function to the parametric model proposed based on expert opinion, and the third term controls the roughness of the nonparametric function. These three terms are added together to evaluate the overall performance of the nonparametric function. The smoothing parameters control the trade-off between fitting to data, fidelity to the parametric model, and the roughness of the nonparametric function.

The rest of the paper is organized as follows. The parametric penalized spline smoothing method is introduced in Section 2. This method has been evaluated by simulation study in Section 3. It is also compared with a parametric nonlinear regression model and the penalized spline smoothing algorithms. The parametric penalized spline smoothing method is demonstrated by estimating the human height function and its derivatives from the real data in Section 4. Finally, conclusions are given in Section 5.

## 2.2 Method

Suppose a function  $x(t)$  is measured at some discrete points  $t_i$ ,  $i = 1, \dots, n$ . These measurements are denoted as  $y_i$  and have a mean  $x(t_i)$ . Our objective is to estimate the function  $x(t)$  from the noisy measurements  $y_i$ .

If we have prior knowledge of  $x(t)$ , a parametric form may be assumed for this underlying function, denoted  $h(t|\boldsymbol{\theta})$ . The parameter  $\boldsymbol{\theta}$  can be estimated by the regression method proposed in Bates & Watts (1988). However, the parametric function may not be completely valid for capturing the true underlying function. Therefore, several nonparametric methods are proposed to estimate  $x(t)$  without making any parametric assumption on  $x(t)$ .

One popular nonparametric method is the spline smoothing method proposed in Ramsay & Silverman (2005). If we let  $\phi_g(t)$ ,  $g = 1, \dots, G$ , be some basis functions, then any smooth function,  $x(t)$ , can be approximated with a linear combination of these basis functions

$$x(t) = \sum_{g=1}^G c_g \phi_g(t) = \boldsymbol{\phi}(t)^T \mathbf{c}, \quad (2.1)$$

where  $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_G(t))^T$  is a vector of basis functions, and  $\mathbf{c} = (c_1, \dots, c_G)^T$  is a vector of basis coefficients. The basis coefficients can be estimated by minimizing the error sum of squares;

$$\text{SSE}(\mathbf{c}) = \sum_{i=1}^n (y_i - x(t_i))^2.$$

Basis functions are often chosen as B-spline basis functions de Boor (2001) because they are only non-zero in local intervals. This property is called *compact support*, which is essential for efficient computation. The B-spline basis functions are determined by the number of knots and their locations. The number of basis functions,  $G$ , equals sum of the implied degree of the B-spline basis functions and the number of interior knots plus one. It is not easy to choose the optimal number of knots and their locations, which is an infinite dimensional optimization problem. Methods for knot selection are discussed in Eubank (1988), Friedman & Silverman (1989), Wahba

(1990), Friedman (1991) and Stone et al. (1997). These methods select knots from a set of candidate knots using a technique similar to stepwise regression.

An alternative method is penalized spline smoothing. It uses a saturated number of basis functions. For example, one knot is put at each location with an observation. To prevent from over-fitting the data, a roughness penalty term is added to control the smoothness of the fitted curve. This is often defined by the derivative of the fitted curve. The basis coefficients are then estimated by minimizing the penalized error sum of squares:

$$\text{PENSSE}(\mathbf{c}) = \sum_{i=1}^n (y_i - x(t_i))^2 + \lambda \int_{t_1}^{t_n} \left[ \frac{d^m x(t)}{dt^m} \right]^2 dt.$$

If our main interest is to estimate the nonparametric function,  $x(t)$ , then  $m$  is usually set to 2; otherwise,  $m$  is chosen as  $m = j + 2$  if the derivative  $d^j x(t)/dt^j$ ,  $j = 1, 2, \dots$ , is required to be estimated.

The penalized spline smoothing method completely relies on the data and ignores any expert opinion of the underlying function. In order to make up for this shortcoming, the parametric penalized spline smoothing method is proposed to combine both information from the data and expert opinion. Suppose some parametric function, denoted  $h(t|\boldsymbol{\theta})$ , is proposed to model the underlying function,  $x(t)$ , based on some expert opinion, which may not be completely valid for the underlying function. The parametric penalized spline smoothing method estimates  $x(t)$  as a linear combination of basis functions as defined in (2.1). A saturated number of basis functions are chosen here. The basis coefficients are estimated by minimizing

$$J(\mathbf{c}|\boldsymbol{\theta}) = \sum_{i=1}^n (y_i - x(t_i))^2 + \lambda_1 \int_{t_1}^{t_n} [x(t) - h(t|\boldsymbol{\theta})]^2 dt + \lambda_2 \int_{t_1}^{t_n} \left[ \frac{d^m x(t)}{dt^m} \right]^2 dt, \quad (2.2)$$

where the first term measures the fit to the data, the second term measures the fidelity to the parametric model, and the last term controls the roughness of the fitted function. The smoothing parameters,  $\lambda_1$  and  $\lambda_2$ , control the trade-off among these three terms.

Given any value of  $\boldsymbol{\theta}$ , the estimate for the basis coefficients can be derived by

minimizing  $J(\mathbf{c}|\boldsymbol{\theta})$ :

$$\hat{\mathbf{c}}(\boldsymbol{\theta}) = [\boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mathbf{Q} + \mathbf{R}]^{-1} [\boldsymbol{\Phi}^T \mathbf{y} + \boldsymbol{\eta}(\boldsymbol{\theta})],$$

where  $\boldsymbol{\Phi}$  is a  $n \times G$  basis matrix with the  $(i, g)$ -entry as  $\phi_g(t_i)$ , the data vector  $\mathbf{y} = (y_1, \dots, y_n)^T$ , and

$$\begin{aligned} \mathbf{Q} &= \lambda_1 \int_{t_1}^{t_n} \boldsymbol{\phi}(t) \boldsymbol{\phi}(t)^T dt, \\ \mathbf{R} &= \lambda_2 \int_{t_1}^{t_n} D^m \boldsymbol{\phi}(t) D^m \boldsymbol{\phi}(t)^T dt, \\ \boldsymbol{\eta}(\boldsymbol{\theta}) &= \lambda_1 \int_{t_1}^{t_n} \boldsymbol{\phi}(t) h(\boldsymbol{\theta}, t) dt. \end{aligned}$$

The conditional estimate  $\hat{\mathbf{c}}(\boldsymbol{\theta})$  is a function of  $\boldsymbol{\theta}$ . The parameter  $\boldsymbol{\theta}$  can be estimated by minimizing the error sum of squares

$$H(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - \hat{x}(t_i)]^2 = \sum_{i=1}^n [y_i - \hat{\mathbf{c}}(\boldsymbol{\theta})^T \boldsymbol{\phi}(t_i)]^2.$$

To summarize, the basis coefficients  $\mathbf{c}$  and  $\boldsymbol{\theta}$  are estimated at two nested levels of optimization. In the inner optimization level,  $\mathbf{c}$  is estimated conditional on  $\boldsymbol{\theta}$ . In the outer optimization level,  $\mathbf{c}$  is removed from the parameter space as a function of  $\boldsymbol{\theta}$ , and  $\boldsymbol{\theta}$  is then estimated. This estimation procedure is called *parameter cascading*, which has been applied to estimate differential equations (Ramsay et al., 2007) and to estimate linear mixed-effects models (Cao & Ramsay, 2010).

One could tune the trade-off between the smoothness of the fitted function and its distance to the parametric function by varying the values of  $\lambda_1$  and  $\lambda_2$  in equation (2.2). This flexibility is one advantage of the parametric penalized spline smoothing method. The values for  $\lambda_1$  and  $\lambda_2$  may be chosen subjectively based on expert opinion of the underlying function. We also propose K-fold cross-validation as the objective criterion to choose the optimal values for  $\lambda_1$  and  $\lambda_2$ .

In the K-fold cross-validation, the original data are randomly partitioned into K subsets. Of the K subsets, K-1 subsets of data are used to estimate the nonparametric function  $\hat{x}^{(-k)}(t)$ , and the remaining subset of data is retained as the validation

data set for testing the fit of the nonparametric function  $\hat{x}^{(-k)}(t)$ . The above cross-validation process is repeated  $K$  times, with each of the  $K$  subsets used exactly once as the validation data set. The  $K$  results are then averaged. The  $K$ -fold cross-validation can be expressed as

$$\text{KCV}(\lambda_1, \lambda_2) = \frac{1}{K} \sum_{k=1}^K \sum_{i \in \pi_k} [y_i - \hat{x}^{(-k)}(t_i)]^2, \quad (2.3)$$

where  $\pi_1, \dots, \pi_K$  is the partition of the set  $\{1, 2, \dots, n\}$ , and  $\hat{x}^{(-k)}(t)$  is the fitted curve from the data excluding those indexed from the set  $\pi_k$ . Ten-fold cross-validation is used in the simulation and application of this paper.

## 2.3 Simulation Studies

Some simulation studies are implemented to evaluate the performance of the parametric penalized spline smoothing method, and to compare it with the parametric nonlinear regression method and the penalized spline smoothing method.

### 2.3.1 Simulation 1

Many reasonable parametric height functions have been proposed. Among them, the JPA function (Jolicoeur et al., 1992),  $h(t|\boldsymbol{\theta})$ , and the PB function (Preece & Baines, 1978),  $f(t|\boldsymbol{\theta})$ , are chosen as the true functions in our simulation studies. These two parametric functions are expressed as

$$\text{JPA} : h(t|\boldsymbol{\theta}_1) = A \frac{[B_1(t+E)]^{C_1} + [B_2(t+E)]^{C_2} + [B_3(t+E)]^{C_3}}{1 + [B_1(t+E)]^{C_1} + [B_2(t+E)]^{C_2} + [B_3(t+E)]^{C_3}} \quad (2.4)$$

$$\text{PB} : f(t|\boldsymbol{\theta}_2) = A - \frac{2(A-B)}{\exp[(t-E)/D_1] + \exp[(t-E)/D_2]}, \quad (2.5)$$

where  $\boldsymbol{\theta}_1 = (A, B_1, B_2, B_3, C_1, C_2, C_3, E)^T$  is a vector of parameters in the JPA function  $h(t|\boldsymbol{\theta}_1)$ , and  $\boldsymbol{\theta}_2 = (A, B, D_1, D_2, E)^T$  is a vector of parameters in the PB function  $f(t|\boldsymbol{\theta}_2)$ . The true parameter values for  $\boldsymbol{\theta}_1$  and  $\boldsymbol{\theta}_2$  are set as the nonlinear regression estimates from real height measurements of one girl in a Berkeley growth study (Tuddenham & Snyder, 1954), which are  $\boldsymbol{\theta}_1 = (166.91, 0.56, 0.13, 0.08, 0.53, 3.41, 23.84, 0.0001)^T$  and  $\boldsymbol{\theta}_2 = (167.291, 158.293, 7.886, 0.804, 14.018)^T$ . Figure 2.1 displays the JPA and PB functions, and their first and second derivatives using the true parameter values. It shows that these two functions are almost identical to each other using the true parameter values, but their first and second derivatives are very different. Each true function is evaluated at 31 points from age 1 to age 18, with spacing quarterly while the child is one year old, annually from two to eight years, and biannually thereafter. These time points are also consistent with the Berkeley growth data. The simulated data are then generated by adding white noise with a variance of  $\sigma^2$  to the true values. We use 3, 7, 10, 20, 30, 40 millimeters as our values of  $\sigma$  and evaluate the three estimation methods under different scales of measurement error. The simulation is implemented with 500 replicates.

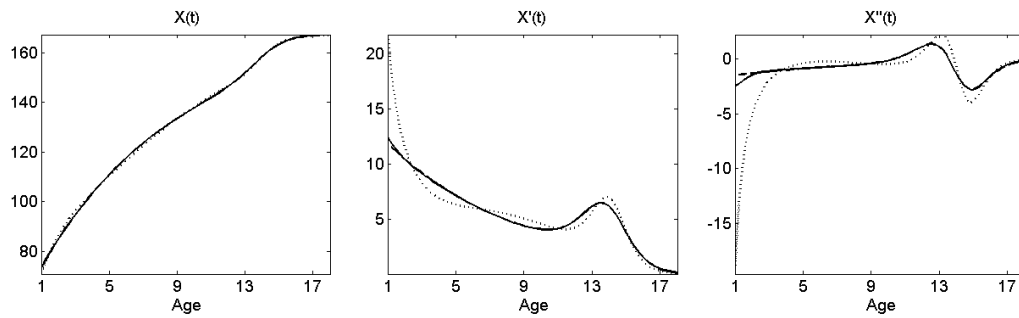


Figure 2.1: The JPA function 2.4 and the PB function 2.5, and their first and second derivatives using the true parameter values. The dashed lines are the estimates for the curve,  $x(t)$ , the first derivative,  $x'(t)$ , and the second derivative,  $x''(t)$ , using the parametric penalized spline smoothing method (PPSS) when the PB function is the true function used to simulate data. The PPSS estimates are almost identical to the true ones.

Four methods are compared based on how accurately they estimate the height function and the first and second derivatives from noisy data. The first method is the parametric nonlinear regression (PNR), where the JPA function is chosen as the parametric function. The second method is the penalized spline smoothing (PSS) method, in which order 6 B-spline basis functions are chosen with one knot put in each measurement location. The roughness penalty term is defined with the second or fourth derivatives of the nonparametric function, which are called PSS2 and PSS4, respectively. The third method is the parametric penalized spline smoothing (PPSS) method. It uses the same basis functions as the penalized spline smoothing method, but the roughness penalty term is defined with the fourth derivatives of the nonparametric function. The parametric function  $h(t|\boldsymbol{\theta}_1)$  in equation (2.2) is chosen as the JPA function. The fourth method is Wahba's standard smoothing splines method, which is implemented using the `gss` package in R.

These four methods are evaluated by comparing their average point-wise root mean squared errors (RMSEs) for the estimates of the height function and the first and second derivatives, which are defined as

$$\begin{aligned}\text{RMSE}(\hat{x}(t)) &= \frac{1}{n} \sum_{i=1}^n \sqrt{\frac{1}{500} \sum_{j=1}^{500} [\hat{x}_j(t_i) - x(t_i)]^2}, \\ \text{RMSE}(\hat{x}'(t)) &= \frac{1}{n} \sum_{i=1}^n \sqrt{\frac{1}{500} \sum_{j=1}^{500} [\hat{x}'_j(t_i) - x'(t_i)]^2}, \\ \text{RMSE}(\hat{x}''(t)) &= \frac{1}{n} \sum_{i=1}^n \sqrt{\frac{1}{500} \sum_{j=1}^{500} [\hat{x}''_j(t_i) - x''(t_i)]^2},\end{aligned}$$

where  $\hat{x}_j(t_i)$ ,  $\hat{x}'_j(t_i)$ ,  $\hat{x}''_j(t_i)$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, 500$ , are the estimated height function and the first and second derivatives in the  $j$ -th simulation at time  $t_i$ , and  $x(t_i)$ ,  $x'(t_i)$  and  $x''(t_i)$ , are the true height function and the first and second derivatives at time  $t_i$ .

When the PB function is the true function used to simulate data, Table 2.1 displays the average point-wise RMSEs for the estimates of the height function and the first and second derivatives using the four methods. The PPSS method is always better



than the other three methods. The PNR method is sensitive to the variance of the noise, performing much better with a small variance. Both the PPSS method and PSS method are relatively robust to the variance of the noise. Because the gss package is only able to estimate  $x(t)$ , the first and second derivatives,  $x'(t)$  and  $x''(t)$ , are estimated using the finite-difference method from the estimated  $\hat{x}(t)$ . The smoothing spline method has smaller RMSEs for  $\hat{x}(t)$  and  $\hat{x}'(t)$  than the PSS method, but the RMSE for  $\hat{x}''(t)$  is very large, which may be caused by the error from the finite-difference method.

In terms of estimating the growth function,  $x(t)$ , the PPSS method is slightly better than the PNR method, and reduced the RMSE by about 15% more than the PSS4 method. When estimating the first derivative,  $x'(t)$ , the PPSS method only reduces the RMSE by 1% more than the PNR method when  $\sigma = 3$ , but this improvement increased to 10% when  $\sigma = 10$ . The RMSEs of the estimated  $\hat{x}'(t)$  with the PPSS method are reduced 32% more than those with the PSS4 method for all values of the standard deviations. Similar results are found when estimating the second derivative,  $x''(t)$ . The RMSE of the estimated second derivatives using the PPSS method are only 55% of that using the PNR method when  $\sigma = 10$ . The PPSS method also reduced the RMSE of the estimated second derivatives by 40% more than the PSS4 method for any scales of noise.

Figure 2.2 displays the point-wise RMSE for the estimates of the curve,  $x(t)$ , the first derivative,  $x'(t)$ , and the second derivative,  $x''(t)$ , using the PPSS and PSS4 methods when the PB function is the true function used to simulate data. The PPSS method has smaller point-wise RMSEs over almost the entire interval than the PSS4 method, although the PSS4 method has slightly smaller RMSEs in [13,16] where the growth curve has some sharp changes.

When the JPA function is the true function used to simulate data, the PNR method will be favoured since it has the correct parametric model for the data. Table 2.2 shows the average point-wise RMSEs for the estimated  $\hat{x}(t)$ ,  $\hat{x}'(t)$ , and  $\hat{x}''(t)$  using the four methods in this scenario. When the noise standard deviation is set to 3, 7 or 10, the PPSS method has the same  $\text{RMSE}(\hat{x}(t))$  with the PNR method, but  $\text{RMSE}(\hat{x}'(t))$  and  $\text{RMSE}(\hat{x}''(t))$  with the PPSS method is larger than the PNR method. The PPSS

Table 2.1: The average point-wise root mean squared errors (RMSEs) for the estimates of the height function and the first and second derivatives using four methods, when the data are simulated based on the PB function as the true function. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4), and Wahba's standard smoothing spline method (SS). The PB function is defined in (2.5).

Noise	Method	RMSE( $\hat{x}(t)$ )	RMSE( $\hat{x}'(t)$ )	RMSE( $\hat{x}''(t)$ )
$\sigma = 3$	PPSS	0.10	0.09	0.11
	PNR	0.10	0.09	0.12
	PSS2	0.13	0.20	0.45
	PSS4	0.12	0.16	0.26
	SS	0.13	0.17	5.20
$\sigma = 7$	PPSS	0.23	0.20	0.27
	PNR	0.23	0.21	0.31
	PSS2	0.29	0.44	0.96
	PSS4	0.26	0.30	0.45
	SS	0.26	0.29	4.83
$\sigma = 10$	PPSS	0.33	0.27	0.34
	PNR	0.33	0.31	0.62
	PSS2	0.42	0.63	1.37
	PSS4	0.38	0.43	0.62
	SS	0.37	0.38	5.70

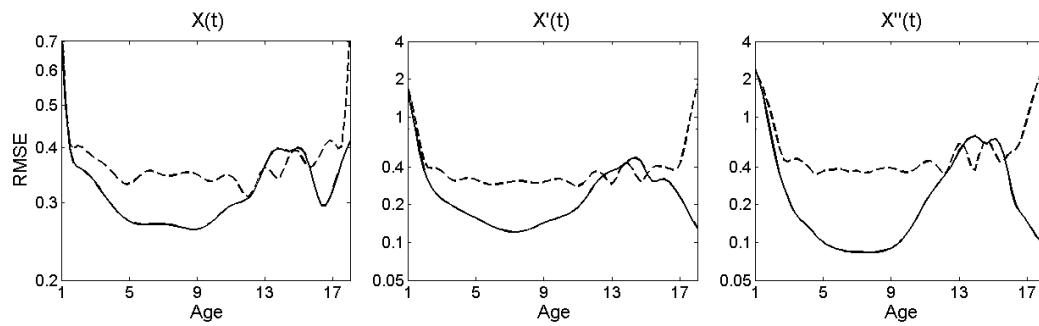


Figure 2.2: The point-wise root mean squared errors (RMSE) for the estimates of the curve,  $x(t)$ , the first derivative,  $x'(t)$ , and the second derivative,  $x''(t)$ , using the parametric penalized spline smoothing (PPSS) method (solid lines) and the penalized spline smoothing (PSS4) method (dashed lines) when the PB function is the true function used to simulate data.

method always has smaller values for  $\text{RMSE}(\hat{x}(t))$ ,  $\text{RMSE}(\hat{x}'(t))$ , and  $\text{RMSE}(\hat{x}''(t))$  than the PSS method. The smoothing spline method has comparable RMSEs for  $\hat{x}(t)$  and  $\hat{x}'(t)$  with the PSS2 method, but the RMSE for  $\hat{x}''(t)$  is very large, which may be caused by the error from the finite-difference method.

Table 2.2: The average point-wise root mean squared errors (RMSEs) for the estimates of the height function and the first and second derivatives using four methods, when the data are simulated based on the JPA function as the true function. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4), and Wahba's standard smoothing spline method (SS). The JPA function is defined in (2.4).

Noise	Method	$\text{RMSE}(\hat{x}(t))$	$\text{RMSE}(\hat{x}'(t))$	$\text{RMSE}(\hat{x}''(t))$
$\sigma = 3$	PPSS	0.09	0.9	0.19
	PNR	0.09	0.08	0.12
	PSS2	0.15	0.33	0.96
	PSS4	0.16	0.28	0.69
	SS	0.16	0.35	23.40
$\sigma = 7$	PPSS	0.22	0.19	0.32
	PNR	0.22	0.18	0.27
	PSS2	0.31	0.54	1.43
	PSS4	0.28	0.39	0.80
	SS	0.31	0.53	14.92
$\sigma = 10$	PPSS	0.32	0.27	0.43
	PNR	0.32	0.27	0.38
	PSS2	0.43	0.72	1.82
	PSS4	0.39	0.5	0.94
	SS	0.42	0.65	15.63

Figure 2.3 shows the point-wise RMSE for the estimates of the curve,  $x(t)$ , the first derivative,  $x'(t)$ , and the second derivative,  $x''(t)$ , using the PPSS and PSS4 methods when the JPA function is the true function used to simulate data. The PPSS method has smaller point-wise RMSEs over most of the interval than the PSS4 method.

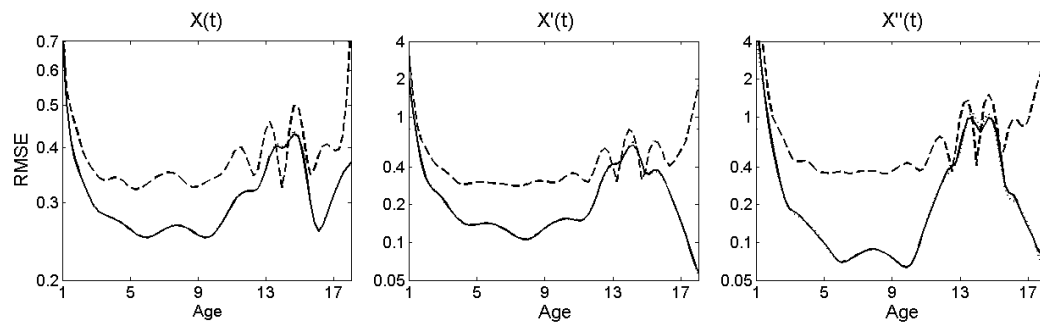


Figure 2.3: The point-wise root mean squared errors (RMSE) for the estimates of the curve,  $x(t)$ , the first derivative,  $x'(t)$ , and the second derivative,  $x''(t)$ , using the parametric penalized spline smoothing (PPSS) method (solid lines) and the penalized spline smoothing (PSS4) method (dashed lines) when the JPA function is the true function used to simulate data.

### 2.3.2 Simulation 2

We choose two parametric functions:

$$h(t|\boldsymbol{\theta}_1) = a + b \cdot t + c \cdot t^2, \quad (2.6)$$

$$f(t|\boldsymbol{\theta}_2) = \frac{\exp(\alpha t)}{\cos(\beta t)}, \quad (2.7)$$

where  $\boldsymbol{\theta}_1 = (a, b, c)^T$  is a vector of parameters in the parametric function  $h(t|\boldsymbol{\theta}_1)$ , and  $\boldsymbol{\theta}_2 = (\alpha, \beta)^T$  is a vector of parameters in the parametric function  $f(t|\boldsymbol{\theta}_2)$ . The function  $f(t|\boldsymbol{\theta}_2)$  is treated as the true parametric function with the true parameter value  $\boldsymbol{\theta}_2 = (2, 3)^T$ . The true function is evaluated at 21 equally-spaced points in  $[-0.1, 0.1]$ . The white noise with the variance,  $\sigma^2$ , are added to the true values to generate the simulated data. The value of  $\sigma$  is chosen as 0.005, 0.01, and 0.05 in three separate simulation studies. Each simulation study is implemented with 500 replicates. The above two parametric functions are chosen because they have almost the same values, but their first and second derivatives are different, as shown in Figure 2.4.

The curve and the first and second derivatives are estimated with the three methods. The first method is the parametric nonlinear regression (PNR) method which uses  $h(t|\boldsymbol{\theta}_1)$  as the parametric function. The second method is the penalized spline smoothing (PSS) method, in which order 6 B-spline basis functions are chosen with one knot put in each measurement location. The roughness penalty term is defined with the second or fourth derivatives of the nonparametric function, called PSS2 and PSS4, respectively. The third method is the parametric penalized spline smoothing (PPSS) method. It uses the same basis functions as the penalized spline smoothing method, and the roughness penalty term defined by the fourth derivative of the nonparametric function. The parametric function  $h(t|\boldsymbol{\theta})$  in equation (2.2) is chosen as the polynomial function  $h(t|\boldsymbol{\theta}_1)$ .

The summary of the estimates of the simulated data, with large noise standard deviation ( $\sigma = 0.05$ ) is displayed in Table 2.3. The PPSS method has the smallest RMSE of the estimates for the curve and the first and second derivatives. The PNR method suffers with the large bias of the estimates, while the standard deviation of

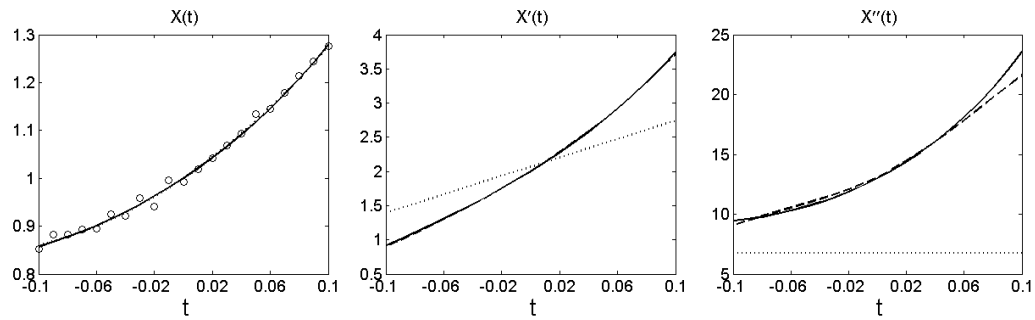


Figure 2.4: The estimates for the curve  $x(t)$ , the first derivative  $x'(t)$ , and the second derivative  $x''(t)$  with the parametric penalized spline smoothing method (PPSS). The solid lines are the true curve,  $f(t) = \exp(2t)/\cos(3t)$ , and the true derivatives. The dash lines are the estimates using PPSS model. The dotted lines are the estimate using parametric nonlinear regression (PNR) method. The circles are the simulation data, which have white noises with the variance  $\sigma^2 = 0.01^2$ .

the estimates using PSS2 and PSS4 methods are large.

Table 2.3: The average point-wise bias, standard deviations (SDs), and root mean squared errors (RMSEs) for the estimates of the curve and the first and second derivatives using three methods, when the data are simulated based on  $f(t) = \exp(2t)/\cos(3t)$  as the true function with noise standard deviation 0.05. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4).

<b>Estimator</b>	<b>Method</b>	$ \text{Bias}  \times 10^3$	$\text{SD} \times 10^3$	$\text{RMSE} \times 10^3$
$\hat{x}(t)$	PPSS	1.85	17.75	17.88
	PNR	1.86	17.76	17.89
	PSS2	1.41	21.30	21.36
	PSS4	0.81	20.47	20.50
<b>Estimator</b>	<b>Method</b>	$ \text{Bias}  * 10^2$	$\text{SD} * 10^2$	$\text{RMSE} * 10^2$
$\hat{x}'(t)$	PPSS	8.92	40.08	41.47
	PNR	37.54	25.87	46.80
	PSS2	8.25	77.69	78.48
	PSS4	3.25	67.17	67.25
<b>Estimator</b>	<b>Method</b>	$ \text{Bias} $	$\text{SD}$	$\text{RMSE}$
$\hat{x}''(t)$	PPSS	3.45	6.50	7.68
	PNR	7.70	3.25	8.52
	PSS2	5.25	32.91	34.79
	PSS4	1.14	20.65	20.69

When the data are simulated with the median scale of noise ( $\sigma = 0.01$ ), the summary for the estimates are displayed in Table 2.4. The estimates for the first and second derivatives with the PPSS method have the smallest RMSE. The PNR method has slightly smaller RMSE of the curve estimates than the PPSS method, but the bias of curve estimates with the PNR method is tenfold of that obtained using the PPSS method. The PNR method also obtains very large biased estimates of the first and



second derivatives.

Table 2.4: The average point-wise bias, standard deviations (SDs), and root mean squared errors (RMSEs) for the estimates of the curve and the first and second derivatives using three methods, when the data are simulated based on  $f(t) = \exp(2t)/\cos(3t)$  as the true function with noise standard deviation 0.01. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4).

Estimator	Method	$ \text{Bias}  \times 10^3$	$\text{SD} \times 10^3$	$\text{RMSE} \times 10^3$
$\hat{x}(t)$	PPSS	0.16	4.48	4.49
	PNR	1.60	3.70	4.11
	PSS2	0.24	5.57	5.57
	PSS4	0.15	4.56	4.56
Estimator	Method	$ \text{Bias}  * 10^2$	$\text{SD} * 10^2$	$\text{RMSE} * 10^2$
$\hat{x}'(t)$	PPSS	0.80	17.30	17.33
	PNR	36.35	5.26	36.96
	PSS2	2.95	34.24	34.53
	PSS4	0.51	18.05	18.06
Estimator	Method	$ \text{Bias} $	SD	RMSE
$\hat{x}''(t)$	PPSS	0.35	7.26	7.28
	PNR	7.51	0.66	7.55
	PSS2	3.47	28.45	30.01
	PSS4	0.29	7.61	7.62

When the data are simulated with the small scale of noise ( $\sigma = 0.005$ ), the summary for the estimates are displayed in Table 2.5. The PPSS method is slightly better than PSS4 method, while the PNR and PSS2 method perform poorly. The PNR method obtains very large biased estimates, because wrong parametric function is assumed. The estimates with the PSS2 method have the extremely large standard deviations.

Table 2.5: The average point-wise absolute value of bias, standard deviations (SDs), and root mean squared errors (RMSEs) for the estimates of the curve and the first and second derivatives using three methods, when the data are simulated based on  $f(t) = \exp(2t)/\cos(3t)$  as the true function with noise standard deviation 0.005. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4).

<b>Estimator</b>	<b>Method</b>	$ \text{Bias}  \times 10^3$	$\text{SD} \times 10^3$	$\text{RMSE} \times 10^3$
$\hat{x}(t)$	PPSS	0.09	2.21	2.21
	PNR	1.60	1.79	2.51
	PSS2	0.13	3.53	3.53
	PSS4	0.09	2.21	2.22
<b>Estimator</b>	<b>Method</b>	$ \text{Bias}  * 10^2$	$\text{SD} * 10^2$	$\text{RMSE} * 10^2$
$\hat{x}'(t)$	PPSS	0.76	8.70	8.75
	PNR	36.73	2.61	36.91
	PSS2	1.55	36.05	36.12
	PSS4	0.77	8.73	8.78
<b>Estimator</b>	<b>Method</b>	$ \text{Bias} $	$\text{SD}$	$\text{RMSE}$
$\hat{x}''(t)$	PPSS	0.46	3.65	3.69
	PNR	7.58	0.33	7.59
	PSS2	3.70	60.08	61.26
	PSS4	0.46	3.66	3.70

## 2.4 Applications

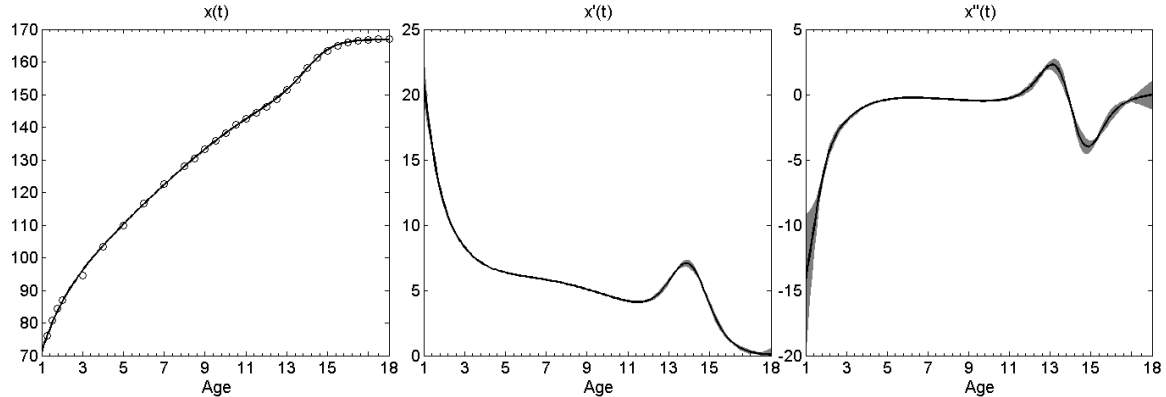


Figure 2.5: The estimate for the height function  $x(t)$ , the first derivative  $x'(t)$ , and the second derivative  $x''(t)$  with the parametric penalized spline smoothing method. The circles are the real height measurements of one girl in a Berkeley growth study. The shaded areas are the corresponding 95% point-wise confidence intervals.

Figure 2.5 displays some real height measurements of one girl in a Berkeley growth study discussed in Tuddenham & Snyder (1954). The girl is measured at 31 non-equally-spaced time points, with four measurements when she is one year old, annual measurements from two to eight years and biannual measurements from eight to eighteen years. The growth dynamic process of this girl can be studied by estimating the height function  $x(t)$ , and the first and second derivatives  $x'(t)$ ,  $x''(t)$  using the parametric penalized spline smoothing method. The JPA function in equation (2.4) is used as the parametric function  $h(t)$  defined in equation (2.2). Ten-fold cross-validation is employed to choose the optimal values for the smoothing parameters. It is displayed in Figure 2.6, which is minimized when  $\log_{10}(\lambda_1) = -0.9$  and  $\log_{10}(\lambda_2) = -3.9$  using the grid search method. The parametric penalized spline smoothing method estimates the parameters in the JPA function as  $A = 166.90$ ,  $B_1 = 0.56$ ,  $B_2 = 0.13$ ,  $B_3 =$

$0.08, C_1 = 0.53, C_2 = 3.42, C_3 = 24.00, E = 0.0001$ .

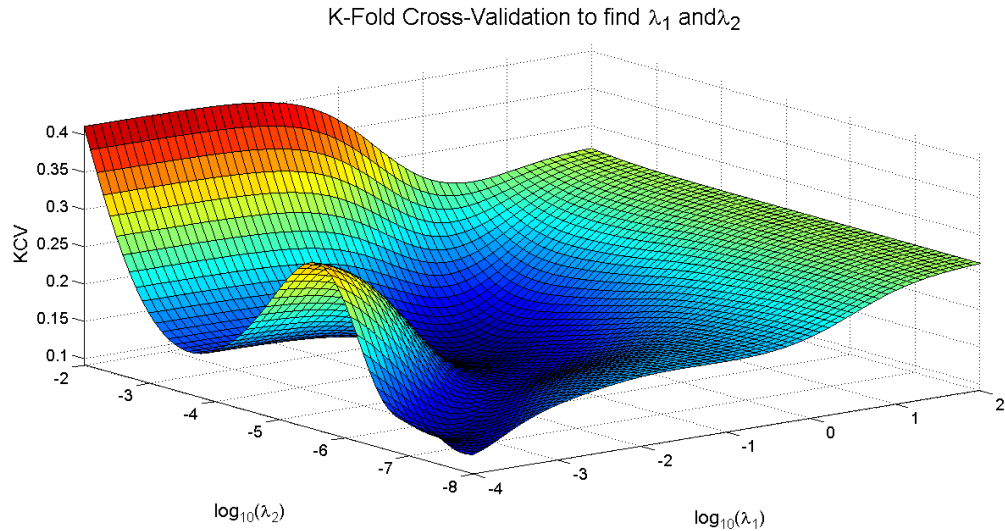


Figure 2.6: K-fold cross-validation (KCV) calculated based on 2.3 by setting  $K=10$ . It is minimized when  $\log_{10}(\lambda_1) = -0.9$  and  $\log_{10}(\lambda_2) = -3.9$  using the grid search method.

The estimates for the height function,  $x(t)$ , and the first and second derivatives,  $x'(t)$  and  $x''(t)$ , are displayed in Figure 2.5. The height function is strictly increasing, and becomes flat around age 16. The growth rate function,  $x'(t)$ , decreases sharply until age 3, and decreases slowly until age 13. It then bumps up at age 13, peaks at age 14, and decreases quickly to zero at age 18. The acceleration function,  $x''(t)$ , quickly increases until age 3, becomes flat from age 3 to age 12, and then has a peak around age 13 and a valley around age 15.

The 95% point-wise confidence intervals for the height function,  $x(t)$ , and the first and second derivatives,  $x'(t)$  and  $x''(t)$ , are obtained with the parametric bootstrap method (Efron & Tibshirani, 1993). The parametric bootstrap method is implemented as follows. The simulated data are generated by adding white noise with the estimated variance  $\hat{\sigma}^2$  to the estimated height function  $\hat{x}(t)$ . The height function,  $x(t)$ , and the first and second derivatives,  $x'(t)$  and  $x''(t)$ , are then estimated from the simulated data with the parametric penalized spline smoothing method. The above process are

done using 500 replications of the simulation. The 95% point-wise confidence intervals are obtained from the 500 estimates of the height function and the first and second derivatives.

## 2.5 Conclusions

It is of great interest to estimate a smooth function and its derivatives accurately. Some parametric models are often proposed based on prior knowledge of the function, but the parametric assumption may not be accurate. The nonparametric smoothing methods are very flexible, but do not consider any expert opinion on the function.

The parametric penalized spline smoothing method combines the advantages of the parametric models and the nonparametric smoothing method. It uses a linear combination of basis function to estimate the underlying function nonparametrically. Expert opinion of the function is adopted by adding an additional penalty term using the squared difference between the fitted function and the parametric function. A roughness penalty term is also defined by the fourth derivative of the fitted function to prevent overfitting where a saturated number of basis functions are employed. The parameter penalized spline method may be considered an extension of the partial spline models introduced in Wahba (1990).

The introduction of the parametric model  $h(t|\theta)$  can be thought of as a "nuisance parameter (model)", and the goal of the simulations is as to test whether this nuisance model is significant or not. The simulation studies show that the parametric penalized spline smoothing method can obtain more accurate estimates of the function and its derivatives than the parametric regression method and penalized spline smoothing method when the parametric model is not correct. In addition, even when the parametric model is correct, the parametric penalized spline smoothing method is comparable to the parametric regression method.

# Chapter 3

## Simultaneous Hierarchical Bayesian Method

### 3.1 Introduction

British Columbia (BC) has a very rich forest resource, where two-thirds of its total landmass (60 million hectares) is covered with forests (BCf, 2011). Because of this unique characteristic, forest industry is a crucial part and accounts for at least 15% of BC's economy. For example, forest industry activity in 2006 accounted for 7.4% of the total provincial GDP and 29% of BC's good producing industry GDP (BCf, 2011). It is also an important employment source that affects almost one-third of all BC rural communities (Dufour, 2002). In addition, forest products are the most important export commodity and are representing more than half of the total international goods exports in BC (BCf, 2011). Hence, studies, such as those related to resource monitoring and wood products' properties modeling, are intensive.

In North American construction industry, such as in house building, wood-based products have been playing a key role. It is common that a structural building is usually subjected to different types of loads. For example, it can be affected by weights from the building structures, such as roofing and walls, weights from occupancies, such as temporary furniture and objects placed inside, or even loads from wind, earthquake and snow. Modeling and predicting the long-term breakage time of wood products

under predetermined stress level is always one main objective to wood engineers. Study on duration-of-load problem, hence, becomes an important research theme.

The load-duration behavior in wood and wood-based products, also called creep-rupture phenomenon, refers to the effect that duration of time or strength is influenced by the applied stress that acts on a wood member. In wood construction, this creep-rupture phenomenon is one distinctive characteristic that needs to be taken into account when designing allowable stress values. Extensive research studies on testing and modeling the time-dependent strength behavior have been conducted since early 1970s (Barrett, 1996). Most of the models have adopted a damage state variable to assess the damage accumulation that is subjected to the loading history. Thus, these models are usually referred as cumulative damage models or damage accumulation models in wood science. Examples of cumulative damage models include the exponential cumulative damage model proposed by Gerhards (1979), the models based on viscoelastic fracture mechanics introduced by Nelson (1985a,b), the models involving a threshold stress ratio proposed by Barrett & Foschi (1978) and Foschi & Barrett (1982) respectively, and their extended models published by Foschi & Yao (1986) and Yao & Foschi (1992). These cumulative damage models can be used to predict the breakage time of wood products under any predetermined stress level.

All the cumulative damage models are presented in a form of ordinary differential equations (ODE) as shown below:

$$\frac{d\alpha(t)}{dt} = f(\tau(t) | \boldsymbol{\theta}), \quad (3.1)$$

where,  $\alpha(t)$  is the variable that measures the accumulative damage and takes values from 0 to 1, with  $\alpha(t) = 0$  implying no damage and  $\alpha(t) = 1$  indicating failure.  $\tau(t)$  is the applied stress history,  $f(\tau(t)|\boldsymbol{\theta})$  is a parametric function of the stress history, and  $\boldsymbol{\theta}$  is the vector of parameters. The breakage time,  $T$ , is defined as the time when  $\alpha(t) = 1$ . Let  $\alpha(t|\tau(t))$  denote the solution of ODE (3.1) under the applied stress history,  $\tau(t)$ , then we have  $\alpha(T|\tau(t)) = 1$ . Currently, both short-term and long-term mechanical tests, also called ramp-load and constant-load tests, on small and clear wood specimens are conducted in order to derive the load-duration adjustment factors. In the ramp-load test, continuous stress is applied with a constant loading rate until

breakage occurs on test specimen. The loading rate must be carefully controlled such that failure is achieved in approximately 1 minute (ASTM-D6815-02, 2002). In the constant-load test, however, continuous stress is applied with the constant loading rate until pre-determined constant stress level. The test specimen is under this constant load until breakage occurs or the end of the test. Thus, the applied stress  $\tau(t)$  in the load-duration tests can be defined as follows:

$$\begin{aligned} \text{Ramp-load test:} \quad & \tau_r(t) = k \times t, \quad 0 \leq t \leq T_r, \\ \text{Constant-load test:} \quad & \tau_c(t) = \begin{cases} k \times t, & 0 \leq t \leq T_0, \\ \tau_c, & T_0 < t \leq T_c, \end{cases} \end{aligned} \quad (3.2)$$

where  $k$  is a constant loading rate,  $T_r$  is the breakage time from the ramp-load test,  $T_0$  is the time that test member starts to be subjected to a constant-load,  $\tau_c$  is the preselected constant load, and  $T_c$  is the breakage time from constant-load test. Based on the definition of the breakage time, we have  $\alpha(T_r|\tau_r(t)) = 1$ , and  $\alpha(T_c|\tau_c(t)) = 1$ . Figure 3.1 below gives an example of the stress history function described in (3.2).

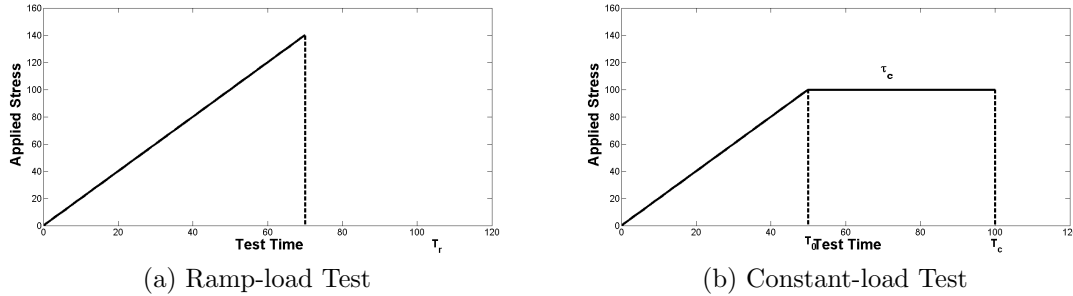


Figure 3.1: Example of stress history in duration of load

Although the form of cumulative damage models are provided based on the expert knowledge, the values of the parameters in the cumulative damage models are usually unknown. Their values have to be estimated from the collected experimental data. Estimating parameters in the cumulative damage models becomes a challenge in wood



engineering, because some cumulative damage models may not have closed-form solutions. In Bayesian statistics, the hierarchical Bayesian nonlinear mixed effect model can perform parameter estimation through the Markov chain Monte Carlo (MCMC) algorithms (Bennett et al., 1996), where closed-form solution to differential equation is not required. Thus, estimation becomes computationally convenient. In our study, we set up a hierarchical Bayesian model and employed Gibbs sampling algorithms to estimate load-duration model parameters and their mixed effects. In addition, since, in practice, not all test specimens' breakage time can be obtained at the end of the constant-load test, censored or missing observations may exist. Current calibration method adopted in estimating load-duration model parameters does not take censoring into account. Our Bayesian approach, however, can make prediction on breakage time for the censored or missing observations.

The rest of this paper is organized as the following. Section 2 will give an introduction on one popular cumulative load-duration damage model. Our proposal on creep-rupture experimental design and implementation using a hierarchical Bayesian model for parameter estimation and constant-load breakage time prediction are presented in section 3. Section 4 gives details of the two simulation studies on our selected cumulative damage model. Finally, our conclusion with discussions are listed in section 5.

## 3.2 The Dynamic Duration of Load Model

Exponential Damage Rate Model (EDRM) proposed by Gerhards & Link (1987) is usually adopted in engineering wood design and thus considered in our study. The dynamic engineering form of this model is given as the following:

$$\frac{d\alpha_i(t)}{dt} = A_i \left( \frac{\tau(t)}{\tau_{si}} - \sigma_0 \right)^{B_i} \quad (3.3)$$

where  $\alpha_i(t)$  is the variable measuring the accumulative damage for the  $i^{th}$  wood specimen, for  $i = 1, 2, \dots, n$ .  $\alpha_i(t) = 1$  indicates failure and  $\alpha_i(t) = 0$  represents no damage.  $A_i$  and  $B_i$  are model parameters that have random effects.  $\tau(t)$  is the

applied stress history that has the form of equation (3.2). Finally,  $\tau_{si}$  is the short-term maximum strength measured in ramp-load test, and it can be expressed as,  $\tau_{si} = k \times T_{ri}$ , the product of loading rate and breakage time from ramp-load test. It is clear that the EDRM model, which has equation (3.3), only depends on the stress level.

### 3.3 Statistical Approach Set-up

#### 3.3.1 Experimental Design in Duration of Load Study

In the EDRM model, parameters  $A_i$  and  $B_i$  are considered having random effects. In order to illustrate that this consideration is reasonable, let us study the mathematical solution for short-term breakage time  $T_r$  in the EDRM model. Assuming stress is applied with a linear loading rate, one can integrate both sides of the equation (3.3) from starting time  $t_0 = 0$  to failure time  $t = T_{ri}$  and obtain the expression of  $T_{ri}$  as,

$$T_{ri} = \frac{B_i \exp(A_i)}{\exp(B_i) - 1} \quad (3.4)$$

From equation (3.4), it is sensible that the short-term failure time for each individual test specimen should be different because of its diversity in wood properties. Similarly, long-term failure time shall be random too. Hence, considering the mixed effect for dynamic model parameters can help to effectively and accurately predict breakage time in wood science. Other parameters, such as  $T_0$ ,  $\tau_c$ , and  $\tau_{si}$  can be obtained through either the expert opinion or calculated from the test measurements.

From equation (3.4), in addition, each short-term breakage time  $T_{ri}$  is a function of  $A_i$  and  $B_i$ . If we re-parameterized them, such that  $A_i^* = \exp(A_i)$  and  $B_i^* = \frac{B_i}{\exp(B_i) - 1}$ , the equation (3.4) becomes,

$$T_{ri} = A_i^* \times B_i^*,$$

It is obvious that  $A_i^*$  and  $B_i^*$  are not identifiable with just one realization of  $T_{ri}$ . However, in reality, one piece of wood specimen can not be broken twice, and

consequently, only one short-term observation can be obtained. This dilemma brings difficulties to estimate and identify each pair of  $A_i$  and  $B_i$ . In practice, wood scientists usually solve this problem by conducting a calibration using simulation data from the assumed failure time distribution, or employing equal-rank assumption with order statistics (Link, 1988), or assuming constant model parameters (Foschi & Yao, 1986). Under the Bayesian framework, the parameter identifiability problem can usually be solved by introducing informative priors, good initials and boundary constraints. However, informative priors require more information on parameters. Since we do not have these information, the usual mathematical techniques will not work in our study.

In order to solve the identifiability problem, we proposed a new experimental design for load duration tests, where each pair of parameters  $A_i$  and  $B_i$  can have two breakage time observations. We proposed that the wood specimens for load-duration tests shall be prepared in pairs and each pair should be either side-matched or end-matched. With this preparation, it is reasonable to assume that the  $i^{th}$  pair of matched test specimens share the same parameters  $A_i$  and  $B_i$ . Since, in practice, both ramp load and constant load tests are needed to estimate duration of load factors, their observations are available. In result, a simultaneous parameter estimation using both the short-term and long-term tests results can be conducted.

### 3.3.2 Hierarchical Bayesian Approach

It is sensible to believe that, instead of assuming constant, the breakage time observations of wood specimens shall have some underlying distributions. Further, from previous discussion, we should consider the random effects for model parameters  $A_i$  and  $B_i$ . Therefore, the underlying distributions of both short-term and long-term failure times can be reconstructed with parameter estimates. Detail modeling setup and methodology implementation are provided as the followings.

**Set up Bayesian models for parameter estimation**

Firstly, let  $y_{ri}$  and  $y_{ci}$  denote the observations for the  $i^{th}$  pair of wood specimens obtained from ramp-load and constant-load tests respectively. Moreover, let  $\boldsymbol{\theta}_i = (A_i, B_i)^T$  be the vector of model parameters, and let  $\epsilon_{ri}$  &  $\epsilon_{ci}$  denote the measurement errors from normal distributions with mean zero in ramp-load and constant-load test respectively. Finally, let  $\mu_A$  and  $\mu_B$  be the overall mean values of model parameters A and B respectively, and let  $\eta_{A_i}$  and  $\eta_{B_i}$  be their corresponding mixed effects from normal distributions with mean zero. Thus, the hierarchical Bayesian nonlinear mixed effects model can be set up in the following three stages.

**Stage 1:** Observed level:

$$\log(y_{ri}) = \log(T_{ri} | \boldsymbol{\theta}_i) + \epsilon_{ri}, \quad \epsilon_{ri} \sim \mathbf{Normal}(0, \sigma_r^2)$$

$$\log(y_{ci}) = \log(T_{ci} | \boldsymbol{\theta}_i) + \epsilon_{ci}, \quad \epsilon_{ci} \sim \mathbf{Normal}(0, \sigma_c^2)$$

**Stage 2:** Unobserved level:

$$\log(A_i) = \mu_A + \eta_{A_i}, \quad \eta_{A_i} \sim \mathbf{Normal}(0, \sigma_A^2)$$

$$\log(B_i) = \mu_B + \eta_{B_i}, \quad \eta_{B_i} \sim \mathbf{Normal}(0, \sigma_B^2)$$

Logarithm-transformation technique has been employed in Stage 1 to make sure time observations are positive. Further, from the dynamic load-duration model (3.3), it is easy to see that parameter  $A_i$  should be positive because the time-dependent damage variable  $\alpha_i(t)$  is an increasing function over time. Moreover, equation (3.4) implies that parameter  $B_i$  should also be positive as breakage time must be non-negative. Thus, logarithm-transformation is also used in Stage 2 for both dynamic parameters.

**Stage 3:** Hyper-prior distributions:

$$[\sigma_r^{-2}] \sim \mathbf{Gamma}(a, b)$$

$$[\sigma_c^{-2}] \sim \mathbf{Gamma}(c, d)$$

$$[\sigma_A^{-2}] \sim \mathbf{Gamma}(e, f)$$

$$[\sigma_B^{-2}] \sim \mathbf{Gamma}(g, h)$$

$$[\mu_A] \sim \mathbf{Normal}(0, j)$$

$$[\mu_B] \sim \mathbf{Normal}(0, k)$$

Here, hyper-parameters  $a, b, c, d, e, f, g, h, j$  and  $k$  are known values specified using prior information or researcher's preference. All prior distributions adopted are conjugate priors, which have the advantage of computational convenience (Gelman et al., 2003). Parameters that need to be estimated are dynamic parameters  $A_i$  and  $B_i$ , and other six distribution parameters  $\sigma_r^2, \sigma_c^2, \sigma_A^2, \sigma_B^2, \mu_A$ , and  $\mu_B$ .

### Implement Gibbs sampling for parameter estimation

In order to perform parameter estimation, we need to specify the conditional distributions, such as  $[y_{ri} | \boldsymbol{\theta}_i]$  and  $[y_{ci} | \boldsymbol{\theta}_i]$ . From our statistical models set up in Stage 1 and Stage 2, the conditional distributions for the failure times and model parameters can be designed as the followings.

$$\begin{aligned}
 [y_{ri} | \log(T_{ri} | \boldsymbol{\theta}_i), \sigma_r^2] &\sim \mathbf{LogNormal}(\log(T_{ri} | \boldsymbol{\theta}_i), \sigma_r^2) \\
 [y_{ci} | \log(T_{ci} | \boldsymbol{\theta}_i), \sigma_c^2] &\sim \mathbf{LogNormal}(\log(T_{ci} | \boldsymbol{\theta}_i), \sigma_c^2) \\
 [A_i | \mu_A, \sigma_A^2] &\sim \mathbf{LogNormal}(\mu_A, \sigma_A^2) \\
 [B_i | \mu_B, \sigma_B^2] &\sim \mathbf{LogNormal}(\mu_B, \sigma_B^2)
 \end{aligned} \tag{3.5}$$

As discussed previously, data collected from constant-load test, in practice, usually contain right censored observations. Therefore, the joint likelihood function of the breakage times from both ramp-load and constant-load tests can be written as

$$f(y_r, y_c | \boldsymbol{\psi}, \delta) = \prod_{i=1}^n f(y_{ri}, y_{ci} | \boldsymbol{\psi})^{1-\delta} S(y_{ri}, y_{ci} | \boldsymbol{\psi})^\delta$$

where,  $\delta$  is the censoring indicator that takes value of 0 or 1.  $\delta = 1$  indicates the corresponding observation is a censored time, such that  $T_{ci} > y_{ci}$ ; whereas,  $\delta = 0$  represents obtaining a breakage time  $T_{ci}$ .  $\boldsymbol{\psi} = (A, B, \sigma_r^2, \sigma_c^2, \mu_A, \sigma_A^2, \mu_B, \sigma_B^2)$  is the vector of all parameters.  $S(\cdot)$  function is called the survival function that models the probability of  $P(T_{ci} \geq y_{ci})$ .

We need to specify the hyper-prior distributions by assigning values to the hyper-parameters. It is sensible that estimation may be benefited from using relatively informative priors with small variances. However, an informative prior usually requires reliable prior information. The load-duration research studies that have been conducted are not under the Bayesian framework. Moreover, most of the studies have not considered the random effects of the dynamic parameters. Therefore, we do not have strong prior information about these parameters and have to select non-informative diffuse priors with large variance instead. Diffuse priors, however, are often chosen for parameters of interest in practice (Carlin & Louis, 1996) in order to "let the data speak for themselves" (Gelman et al., 2003).

In Bayesian statistics, MCMC method refers to a general technique to draw samples from a proposal distribution and improve them towards the targeted posterior distribution of the unknown parameter by accepting or rejecting the simulated draws. Sample of the Markov chain is drawn sequentially and is depended on the last drawn value. Thus, the targeted posterior distribution can be eventually revealed by MCMC method (Gelman et al., 2003). Among all MCMC methods, Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953) and the Gibbs sampling (Geman & Geman, 1984) are two most popular algorithms that have been adopted in numerous studies. In Gibbs sampling, samples are drawn from the full conditional posterior distributions with 100% acceptance rate. Thus, Gibbs sampling does not require specific proposal distributions (Geman & Geman, 1984), (Ntzoufras, 2009). We used Gibbs

sampling to estimate all unknown parameters because full conditional posterior distributions are available. Let us introduce a new notation  $(-m)$  here to present the all elements of the corresponding parameter vector in its posterior distribution except the  $m^{th}$  element. For example,  $\sigma_r^2 = (\sigma_{r_m}^2, \sigma_{r_{(-m)}}^2)$ , where  $\sigma_{r_m}^2$  represents the  $m^{th}$  scalar element in the vector  $\sigma_r^2$ , and  $\sigma_{r_{(-m)}}^2$  represents the rest elements of the vector  $\sigma_r^2$ . Therefore, the full conditional distributions can be listed as the followings.

$$\begin{aligned}
 [\sigma_r^{-2} | y_r, \sigma_{r_{(-m)}}^2, \sigma_c^2, A, B, \sigma_A^2, \sigma_B^2, \mu_A, \mu_B] &\sim \mathbf{Gamma}(a + \frac{n}{2}, \frac{\sum_{i=1}^n (\log(y_{ri}) - \log(T_{ri}))^2 + 2b}{2}) \\
 [\sigma_c^{-2} | y_c, y_r, \sigma_r^2, \sigma_{c_{(-m)}}^2, A, B, \sigma_A^2, \sigma_B^2, \mu_A, \mu_B] &\sim \mathbf{Gamma}(c + \frac{n}{2}, \frac{\sum_{i=1}^n (\log(y_{ci}) - \log(T_{ci}))^2 + 2d}{2}) \\
 [\sigma_A^{-2} | y_c, y_r, \sigma_r^2, \sigma_c^2, A, B, \mu_A, \sigma_{A_{(-m)}}^2, \mu_B, \sigma_B^2] &\sim \mathbf{Gamma}(e + \frac{n}{2}, \frac{\sum_{i=1}^n (\log(A_i) - \mu_A)^2 + 2f}{2}) \\
 [\sigma_B^{-2} | y_c, y_r, \sigma_r^2, \sigma_c^2, A, B, \sigma_A^2, \mu_A, \mu_B, \sigma_{B_{(-m)}}^2] &\sim \mathbf{Gamma}(g + \frac{n}{2}, \frac{\sum_{i=1}^n (\log(B_i) - \mu_B)^2 + 2h}{2}) \\
 [\mu_A | y_c, y_r, \sigma_r^2, \sigma_c^2, A, B, \mu_{A_{(-m)}}, \sigma_A^2, \sigma_B^2, \mu_B] &\sim \mathbf{Normal}(\frac{j \sum_{i=1}^n \log(A_i)}{nj + \sigma_A^2}, \frac{j \sigma_A^2}{nj + \sigma_A^2}) \\
 [\mu_B | y_c, y_r, \sigma_r^2, \sigma_c^2, A, B, \mu_A, \sigma_{A_{(-m)}}, \mu_{B_{(-m)}}, \sigma_B^2] &\sim \mathbf{Normal}(\frac{k \sum_{i=1}^n \log(B_i)}{nk + \sigma_B^2}, \frac{k \sigma_B^2}{nk + \sigma_B^2})
 \end{aligned}$$

Although the unobserved dynamic model parameters A and B are not of our interest, they still need to be estimated implicitly in order to estimate the six distribution parameters. Using the similar procedure as in developing the full conditional distributions above, at the  $m^{th}$  draw in the Markov chain, the posterior distributions for the  $i^{th}$  pair of parameters are as follows:

$$\begin{aligned}
 [A_i | y_c, y_r, \sigma_r^2, \sigma_c^2, A_{i_{(-m)}}, B_i, \mu_A, \sigma_A^2, \mu_B, \sigma_B^2] &\propto \\
 \frac{1}{A_i} \exp\left(-\frac{\sum_{i=1}^n (\sigma_c^2 \sigma_A^2 (\log(y_{ri}) - \log(T_{ri}(A_i, B_i)))^2 + \sigma_r^2 \sigma_A^2 (\log(y_{ci}) - \log(T_{ci}(A_i, B_i)))^2 + \sigma_r^2 \sigma_c^2 (\log(A_i) - \mu_A)^2)}{2 \sigma_r^2 \sigma_c^2 \sigma_A^2}\right) \\
 [B_i | y_c, y_r, \sigma_r^2, \sigma_c^2, A_i, B_{i_{(-m)}}, \mu_A, \sigma_A^2, \mu_B, \sigma_B^2] &\propto \\
 \frac{1}{B_i} \exp\left(-\frac{\sum_{i=1}^n (\sigma_c^2 \sigma_B^2 (\log(y_{ri}) - \log(T_{ri}(A_i, B_i)))^2 + \sigma_r^2 \sigma_B^2 (\log(y_{ci}) - \log(T_{ci}(A_i, B_i)))^2 + \sigma_r^2 \sigma_c^2 (\log(B_i) - \mu_B)^2)}{2 \sigma_r^2 \sigma_c^2 \sigma_B^2}\right)
 \end{aligned}$$

In order to implement the hierarchical model more efficiently, we chose the "bugs" function in the R2WinBUGS package in R. The "bugs" function will call WinBUGS program repeatedly to conduct the Gibbs sampling algorithms. WinBUGS is a black box style program for Bayesian analysis of complex statistical models using MCMC techniques (Spiegelhalter et al., 2003). In addition, in WinBUGS, instead of specifying the survivor function  $S(\cdot)$ , right censoring can be handle easily by using the command,  $I(y_{censor}, \cdot)$ , with the value of censored observation,  $y_{censor}$ .

Further, we employed two techniques introduced by Geman & Geman (1984) to improve the efficiency of the MCMC runs. Firstly, with the "burn-in" technique, we discarded a number of the initial simulated samples as they may not come from the stationary targeted distribution. Secondly, with the thin-in technique, we only kept the every  $k^{th}$  simulated samples from the Markov chain. The values of burn-ins and thinning can be decided depend on researcher's preference. Details are discussed in the simulation studies section. Mean values of the collected samples in the posterior distributions were calculated as the estimates of the unknown parameters.

### 3.3.3 Bayesian approach for constant-load breakage time prediction

Under the Bayesian framework, predictions of future observations or missing observations through the conditional predictive distributions (Ntzoufras, 2009). Let  $y^*$  and  $y$  represent predicted observation and observed data respectively. Hence, the conditional predictive distribution can be calculated as

$$\begin{aligned} f(y^*|y) &= \int f(y^*, \boldsymbol{\psi}|y) d\boldsymbol{\psi} \\ &= \int f(y^*|y, \boldsymbol{\psi}) f(\boldsymbol{\psi}|y) d\boldsymbol{\psi}, \end{aligned}$$

which is to average the likelihood of predicted data over the posterior distribution  $f(\boldsymbol{\psi}|y)$ . It is clear that given the parameters  $\boldsymbol{\psi}$ , predicted data and observed data are independent. Hence, after obtaining the estimates of our parameters of interest, prediction can be conducted through MCMC algorithms.



## 3.4 Simulation Studies

In order to demonstrate our proposed hierarchical Bayesian methodology, we conducted two simulation studies with one hundred simulation runs in each study. In the first study, breakage times were simulated from the EDRM model and simulated  $T_{c,s}$  were right censored with 10% and 20% censoring rates respectively. These percentage values, however, can be adjusted by expert opinion or real situation in practice. Parameter estimation was then conducted using our proposed approach on these censored data. In the second study, ramp test breakage times were simulated from the same EDRM model. Predictions of constant-load breakage times were performed using parameter estimates. Detail procedures including the initial values and realization of the hyper-parameters specified are explained in the following sections.

### 3.4.1 Simulation 1

In each simulation run, 200 simulated failure times, with 100 in short-term and 100 in long-term, were first simulated from the EDRM model. We right censored the long-term breakage times to 10% and 20% respectively. Moreover, since the preselected constant load  $\tau_c$  is usually set to be some percentile of the average maximum short-term strength  $\bar{\tau}_s$ , we equated  $\tau_c = 0.65\bar{\tau}_s$ . Two types of parameter estimation were thereafter conducted using our proposed hierarchical Bayesian model by considering censoring or treating censored observations as breakage time. The general simulation procedures are given as the following.

*Step 1. Generating data:* Initialize hyper-parameters and simulate 100 short-term failure times and 100 long-term breakage times respectively using the conditional distributions introduced in equation 3.5.

*Step 2. Censoring data:* Rank simulated long-term breakage times in descending order and calculate the 90<sup>th</sup> and 80<sup>th</sup> percentile values respectively. Construct the binary 10% and 20% right censoring indices by assigning value "1" to indicate those long-term simulated data which exceed the 90<sup>th</sup> and 80<sup>th</sup> percentile values respectively. Replace the readings of long-term breakage times whose censoring indices equal to "1" to the corresponding percentile values. The completed data sets with either 10% or

20% censored observations will then be used in parameter estimation.

*Step 3. Implementing MCMC algorithms:* Perform parameter estimation through Gibbs sampling in R using R2WinBUGS package. Hyper-prior distributions are listed in Stage 3. Apply this step twice on each data set by incorporating censoring observations with the censoring indication function or treating the censored observations as observed breakage times.

*Step 4.* Repeat steps 1 - 3 for 100 runs. Calculate bias, standard deviation, and root mean squared error for each parameter estimate.

The true values of the hyper-parameter we have chosen are  $(\sigma_r^2, \sigma_c^2, \mu_A, \sigma_A^2, \mu_B, \sigma_B^2) = (0.00025, 0.00025, 2.58, 0.00025, 2.45, 0.00025)$ . Figure 3.2 gives an example of one set of the simulated observations.

Diffuse prior distributions with large variance were employed by equating the hyper-parameters  $a, b, c, d, f, h$  to 0.001,  $e$  and  $g$  to 0.01, and  $j, k$  to 10000. Thus, the prior distributions for  $\sigma_r^{-2}$  and  $\sigma_c^{-2}$  become Gamma(0.001, 0.001) with mean 1 and variance 1000. Prior distributions for  $\sigma_A^{-2}$  and  $\sigma_B^{-2}$  become Gamma(0.01, 0.001) with mean 10 and variance 10000. Prior distributions for  $\mu_A$  and  $\mu_B$  are Normal(0,10000) with mean 0 and variance 10000. The values of these hyper-parameters, however, can be adjusted by expert's opinion or available prior information if necessary.

In order to cope with the right censoring, we employed the WinBUGS command  $I(y_{censor}, \quad)$ , which is an indicator function, to distinguish constant-load breakage times and censored observations. Since we use R2WinBUGS to implement MCMC algorithms, censored observations are treated as missing data to be predicted in WinBUGS. This indication function will perform as a constraint to ensure predicted values beyond the given censoring readings (Spiegelhalter et al., 2003).

We conducted 100,000 MCMC simulations in each parameter estimation. Among these MCMC samples, we discarded the first 75,000 and kept the every 5<sup>th</sup> samples to construct its posterior distribution. Posterior mean from the stored MCMC simulated samples was calculated as its estimate. Figure 3.3 illustrates one example of six parameters' posterior distributions.

Bias, standard deviations, and root mean squared errors (RMSE) were calculated for the EDRM estimates. Table 3.1 shows the parameter estimation performance using

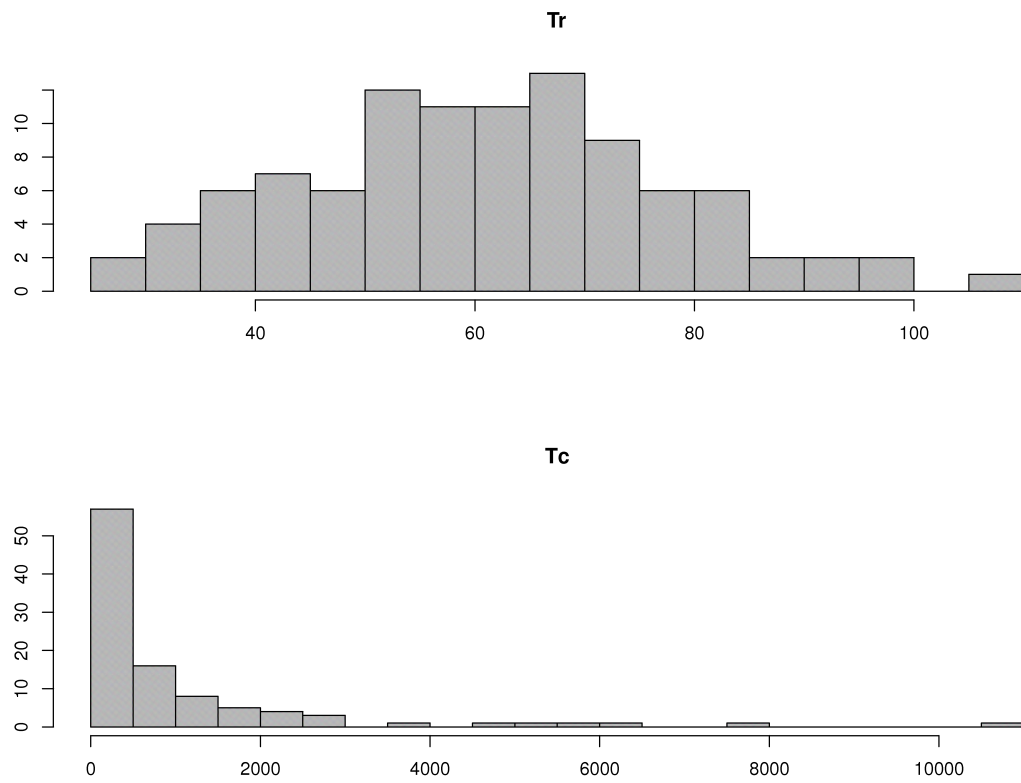


Figure 3.2: Histograms of one set of short-term ( $T_r$ ) and long-term ( $T_c$ ) breakage times simulated from the EDRM model.

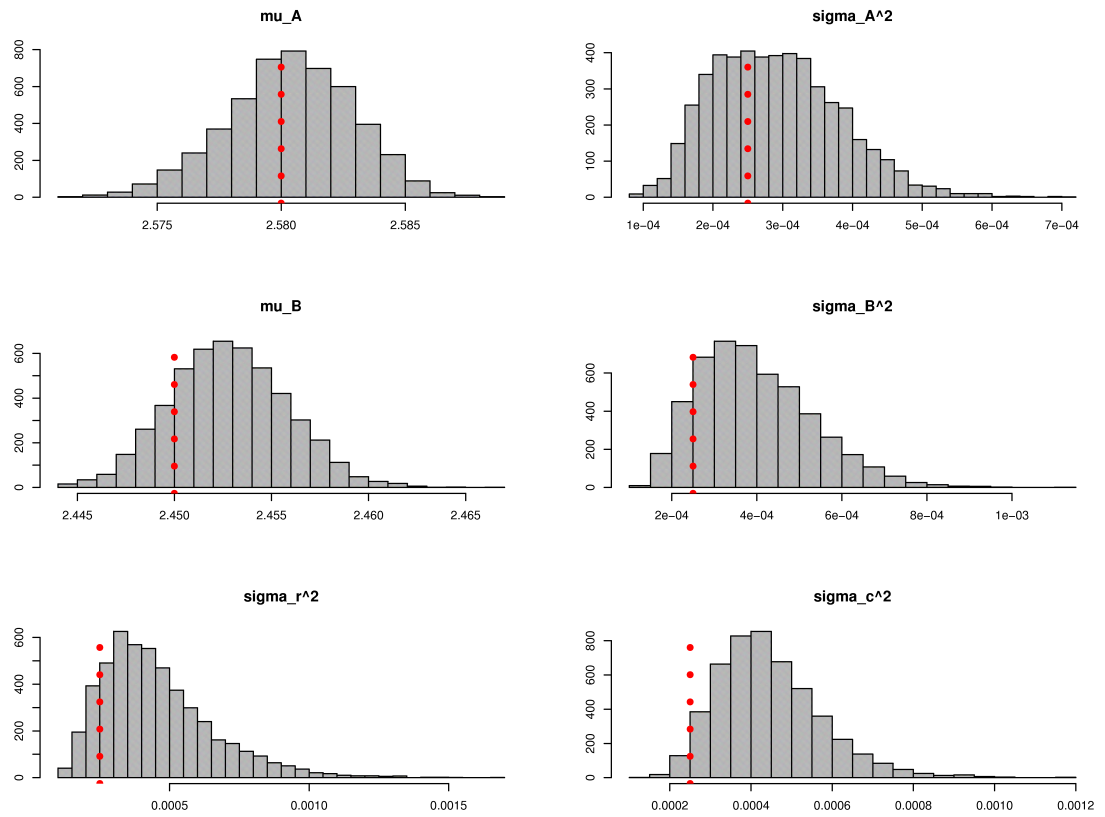


Figure 3.3: Histograms of one set of six parameters' posterior distributions. Red dotted line in each histogram indicates the location of parameter's true value we have assigned.

data set with 10% right censoring rate and Table 3.2 shows the parameter estimation performance using data set with 20% right censoring rate. It is worth notice that with data set containing 10% right censored data, RMSEs from parameter estimation having considered censoring are about 10 times to 400 times smaller than those from estimation without considering censoring. Moreover, with the data set containing 20% censored data, these differences can rise up to 1000 times more. Therefore, considering censoring is very crucial in obtaining "good" estimates in load-duration problem.

Table 3.1: The averaged bias(Bias), standard deviations (SD), and root mean squared errors (RMSE) of the estimates of six parameters using data set with 10% right censoring on the constant-load breakage times. "Cen" denotes estimation considering right censoring and "NoCen" indicates not consider censoring.

	Bias $\times 100$		SD $\times 100$		RMSE $\times 100$	
	Cen	NoCen	Cen	NoCen	Cen	NoCen
$\mu_A$	0.02	-2.36	0.18	1.41	0.18	2.75
$\mu_B$	0.01	-3.04	0.23	1.81	0.23	3.54
$\sigma_A^2$	0.0015	0.0273	0.0049	0.0401	0.0051	0.0485
$\sigma_B^2$	0.0045	0.1034	0.0058	0.1101	0.0073	0.1510
$\sigma_r^2$	0.0292	0.9914	0.0113	1.1363	0.0313	1.5080
$\sigma_c^2$	0.0072	3.140	0.0063	3.3558	0.0096	4.5959

Table 3.2: The averaged bias(Bias), standard deviations (SD), and root mean squared errors (RMSE) of the estimates of six parameters using data set with 20% right censoring on the constant-load breakage times. "Cen" denotes estimation considering right censoring and "NoCen" indicates not consider censoring.

	Bias $\times 100$		SD $\times 100$		RMSE $\times 100$	
	Cen	NoCen	Cen	NoCen	Cen	NoCen
$\mu_A$	0.02	-8.35	0.19	2.09	0.19	8.61
$\mu_B$	0.01	-10.69	0.25	2.71	0.25	11.03
$\sigma_A^2$	0.0017	-0.0040	0.0048	0.003	0.0051	0.005
$\sigma_B^2$	0.0047	0.0221	0.0052	0.0086	0.0070	0.0237
$\sigma_r^2$	0.0295	1.6648	0.0112	0.5393	0.0316	1.7500
$\sigma_c^2$	0.0076	12.36	0.0065	4.3215	0.01	13.10

### 3.4.2 Simulation 2

In the prediction study, 200 breakage times were simulated from the EDRM model using the same true values as indicated in the first study for all parameters. We then considered the 100 ramp test breakage times as known observations and treated the 100 constant-load breakage times as missing data. Predictions of these missing data were performed using the parameter estimates obtained in the first study.

In order to demonstrate the prediction performance, we plotted the predicted values against the true simulated values with reference to the 45° line. If prediction is good, the predicted value should be very closed to the true value, and its point in the graph should fall onto the reference line. By this plotting method, one can visualize the fitted performance. Further, we define the mean squared prediction error (MSPE) as the averaged value of the squared differences between predicted value and true value. Let  $T_{c_{pred}}$  be the predicted constant-load breakage time and  $T_{c_{true}}$  be the simulated true breakage time. MSPE has the following equation.

$$MSPE = \frac{\sum_{i=1}^{100} (\log(T_{c_{pred}})_i - \log(T_{c_{true}})_i)^2}{100}$$

It is sensible that smaller MSPE implies better fitting. The general prediction procedures are given as the following.

*Step 1. Generating data:* Initialize hyper-parameters and simulate 100 short-term failure times and 100 long-term breakage times respectively using the conditional distributions introduced in equation (3.5).

*Step 2. Setting up predictive distribution:* Specify the conditional distributions at observed and unobserved levels with the likelihood introduced in equation (3.5). Specify the parameters using estimates obtained in simulation 1. Hence,  
 $\boldsymbol{\psi} = (\hat{\sigma}_r^2, \hat{\sigma}_c^2, \hat{\mu}_A, \hat{\sigma}_A^2, \hat{\mu}_B, \hat{\sigma}_B^2)^T$ .

*Step 3. Predicting missing data:* Conduct predictions through implementing MCMC algorithms in R with R2WinBUGS package. Ramp-load breakage time are treated as known observations and constant-load breakage time as missing. Plot predictions against their true simulated values and calculate mean squared prediction errors (MSPE).

*Step 4.* Repeat step 3 for each set of parameter estimates obtained in study 1 and repeat steps 1 - 3 for 100 runs.

In each breakage time prediction run, 100,000 MCMC samples were drawn, where the first 75,000 were discarded and the every 5<sup>th</sup> samples were stored. Posterior mean of the stored samples was calculated as the predicted value. Figure 3.4 gives two examples of the prediction performance. In figure 3.4a, predictions were drawn with estimates from a simulated data set that had 10% right censored observations. In figure 3.4b, predictions were drawn with estimates from a simulated data set that had 20% right censored observations. Dots in both figures represent predictions using estimates that were obtained with the consideration for censoring; whereas crosses(“+”) represent predictions using estimates that were obtained without considering censoring. It is obvious that predictions using estimates considering censoring in data are much accurate than those without considering censoring.

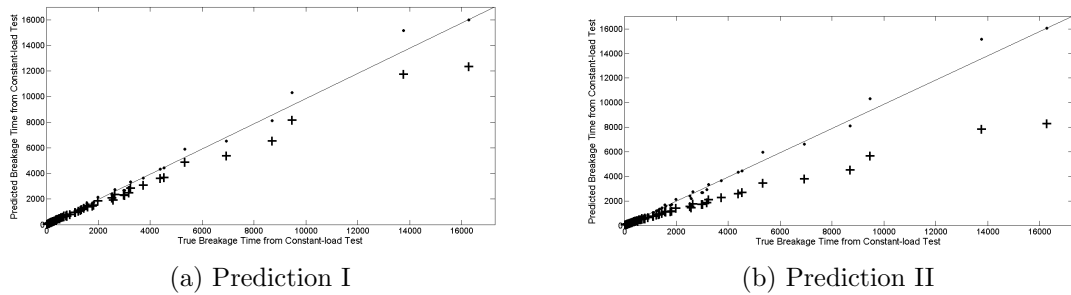


Figure 3.4: Plots of predicted values against true values. In prediction I, data used to estimate parameters had 10% right censored observations. In prediction II, data used to estimate parameters had 20% right censored observations. Dots represent predicted values that used parameter estimates obtained by considering the censoring condition. Crosses (“+”) present predicted values that used parameter estimates obtained by without considering the censoring condition

The descriptive statistics of MSPE values from 100 simulation runs are listed in Table 3.3. Again, averaged MSPE is larger if the prediction has not taken into account that censored observations in the original data set. Further, averaged MSPE is larger with higher censoring rate.



Table 3.3: Table of descriptive statistics of MSPEs, with readings of 2.5% and 97.5% percentiles, median and mean. "Cen" denotes using parameter estimates considering right censoring and "NoCen" indicates using parameter estimates not consider censoring.

	10% Censoring		20% Censoring	
	Cen	NoCen	Cen	NoCen
2.5%	0.0013	0.0028	0.0013	0.0387
97.5%	0.0030	0.0358	0.0030	0.1684
Median	0.0018	0.0145	0.0018	0.0820
Mean	0.0019	0.0149	0.0019	0.0902

### 3.5 Conclusions

In our paper, we have proposed a hierarchical Bayesian mixed-effects model to estimate dynamic parameters and their random effects. Using a Bayesian approach can relax the requirement of obtaining closed-form solution from the dynamic engineering equation. Further, this Bayesian approach is more flexible and can incorporate expert's opinion and information from previous studies into the prior distributions. Posterior distributions from the MCMC simulated samples can graphically illustrate the inferential statistics as well. Our consideration of dynamic model parameters having random effects can make the duration of load model being more closely to reality. It is clear that many properties and environmental conditions, such as types of the wood specimens, moisture contents, knots, lumber grading, and rate-of-loading, can influence the strength and breakage time (Laufenberg et al., 1999; Madsen, 1975; Rosowsky & Reinhold, 1999). Thus, random effects of the dynamic parameters should be considered to present the uniqueness of each wood specimen. We also proposed a matched pair design for short-term and long-term duration of load tests. With this experimental design, both short-term and long-term results can be employed to estimate the dynamic parameters simultaneously. Hence, the identifiability problems can be solved.

We have presented two simulation studies to demonstrate our hierarchical Bayesian approach in parameter estimation and prediction. In the parameter estimation section, it can be seen from Table 3.1 and Table 3.2 that estimates of unknown parameters were much closer to the true values if we have taken into account that the data contain right censored observations. Further, if we perform parameter estimation without considering the censoring conditions, the higher the censoring rate, the poorer estimates we will obtain. From this presentation of our study in parameter estimation, our hierarchical Bayesian approach is practical because it can handle right censored conditions. Right censoring is common in practice due to the cost and standards of the load-duration experiment. The percentage of the right censoring in our approach can be adjusted by experimenter's preference.

In the prediction section, we also illustrated predicting long-term breakage time by

constructing its underlying conditional posterior distribution,  $f(y^*|y)$ . Figure 3.4 has shown the goodness of fit of our prediction performance. It is clear that our predicted fitted values are very closed to the true values. Moreover, this prediction study can also be considered as the validation to the estimation performance. Figure 3.4 also implies that predictions taking censoring into account are more accurate than those not considering censoring in the original data set, which is consistent to the finding obtained in parameter estimation section.

There are several considerations, however, shall be suggested for future studies. First, we have considered the random effects of dynamic parameters  $A_i$  and  $B_i$  as from normal distribution. However, the symmetric property of normal distribution may be too restricted. Second, we have considered log-normal distribution for breakage times because observations should be positive. However, Weibull distribution can be considered appropriate in describing the underlying distributions too because it can handle censored data as well. Future study can be conducted to compare efficiency gain between these two models. Third, connection of the dynamic parameters  $A_i$  and  $B_i$  to real wood properties can also be considered in future studies. For example,  $A_i$  can be modeled with generalized linear relationships to moisture contents, wood specimen types, and lumber gradings, such that prediction on breakage time can be performed by using these wood characteristics. Last, employing our hierarchical Bayesian approach to real test results may be helpful to find the optimal model in different situations.

# Bibliography

- The Council of Forest Industries (retrieved on July 2011). *Forestry and BC's Economy*. The Council of Forest Industries. <http://www.forest-education.info/forest-facts/index.html>.
- Province of British Columbia (retrieved on July 2011). *A Guide to the BC Economy and Labor Market*. Province of British Columbia. [http://www.guidetobceconomy.org/major\\\_industries/forestry.htm](http://www.guidetobceconomy.org/major\_industries/forestry.htm).
- ASTM-D6815-02 (2002). Standard specification for evaluation of duration of load and creep effects of wood and wood-based products. *Annual Book of ASTM Standards, D 6815-02a*.
- Barrett, J. (1996). Duration of load-the past, present and future. *International COST 508 Wood Mechanics Conference*.
- Barrett, J. D. & Foschi, R. O. (1978). Duration of load and probability of failure in wood. *Canadian Journal of Civil Engineering*, 5(4), 505–532.
- Bates, D. M. & Watts, D. B. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.
- Bennett, J. E., Racine-Poon, A., & Wakefield, J. C. (1996). Markov chain monte carlo for nonlinear hierarchical models. In Gilks WR, S. & D.J.Spiegelhalter (Eds.), *Markov chain Monte Carlo in Practice*, (pp. 39–357). Chapman & Hal, London.
- Bock, R. D. & Thissen, D. (1980). Statistical problems of fitting individual growth curves. *Human Physical Growth and Maturation: Methodologies and Factors*, (F.E.Johnston, A.F.Roche, and C.Susanne, Eds), Plenum.
- Cao, J. & Ramsay, J. O. (2010). Linear mixed effects modeling by parameter cascading. *Journal of the American Statistical Association*, 105, 365–374.
- Carlin, B. P. & Louis, T. A. (1996). *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall, London.

- de Boor, C. (2001). *A Practical Guide to Splines*. New York: Springer.
- Dufour, D. (2002). The lumber industry: Crucial contribution to Canada's prosperity. *Research Paper, Minister of Industry, Statistics Canada, Catalogue no. 31F0027XIE No. 01*.
- Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- Eubank, R. L. (1988). *Spline smoothing and nonparametric regression*. New York and Basil: Marcel Dekker.
- Foschi, R. O. & Barrett, J. D. (1982). Load-duration effects in western hemlock lumber. *Journal of the Structural Division, American Society of Civil Engineers, 108(ST7)*, 1494–1510.
- Foschi, R. O. & Yao, Z. C. (1986). Another look at three duration of load models. In *In: Proceedings of international Council for Building Research Studies and Documentation, Meeting 19 in Florence, Italy*, volume CIB-W18.
- Friedman, J. H. (1991). Multivariate adaptive regression splines (with discussion). *The Annals of Statistics, 19*, 1–141.
- Friedman, J. H. & Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling (with discussion). *Technometrics, 31*, 3–39.
- Gasser, T., Muller, H., Kohler, W., Largo, R., Molinari, L., & Prader, A. (1985). Human height growth: Correlational and multivariate structure of velocity and acceleration. *Annals of Human Biology, 12*, 501–515.
- Gasser, T., Muller, H., Kohler, W., Molinari, L., & Prader, A. (1984). Nonparametric regression analysis of growth curves. *The Annals of Statistics, 12*, 210–229.
- Gelman, A., Carlin, J. B., Stern, H., & Rubin, D. B. (2003). *Bayesian Data Analysis, Second Edition*. Texts in Statistical Science. Chapman & Hall, New York.
- Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741.
- Gerhards, C. C. (1979). Time-related effects of loading on wood strength: A linear cumulative damage theory. *Wood Science, 11(3)*, 139–144.
- Gerhards, C. C. & Link, C. L. (1987). A cumulative damage model to predict load duration characteristics of lumber. *Wood and Fiber Science, 19(2)*, 147–164.

- Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*, 97–109.
- Jolicoeur, P., Pontier, J., & Abidi, H. (1992). Asymptotic models for the longitudinal growth of human stature. *American Journal of Human Biology*, *4*, 461–468.
- Jolicoeur, P., Pontier, J., Pernin, M., & Sempe, M. (1988). A lifetime asymptotic growth curve for human height. *Biometrics*, *44*, 995–1003.
- Kanefuji, K. & Shohoji, T. (1990). On a growth model of human height. *Growth Development and Aging*, *54*, 155–165.
- Laufenberg, T. L., L.C.Palka, & McNatt, J. (1999). Creep and creep-rupture behavior of wood-based structural panels. *Research Paper*, *FPL-RP-574*.
- Link, C. L. (1988). Statistical considerations in duration of load research. *Research Paper*, *No. R1916*.
- Madsen, B. (1975). Moisture content strength relationship for lumber subjected to bending. *Canadian Journal of Civil Engineering*, *2*, 466–473.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., & Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of chemical Physics*, *21*, 1087–1092.
- Nelson, L. F. (1985a). Wood as a cracked viscoelastic material, part i: Theory and applications. In *In: Proceedings of international workshop on duration of load in lumber and wood products*, (pp. 67–78). Forintek Canada Corporation.
- Nelson, L. F. (1985b). Wood as a cracked viscoelastic material, part ii: Sensitivity and justification of a theory. In *In: Proceedings of international workshop on duration of load in lumber and wood products*, (pp. 79–89). Forintek Canada Corporation.
- Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Wiley Series in Computational Statistics. John Wiley & Sons, Inc., New Jersey.
- Preece, M. A. & Baines, M. J. (1978). A new family of mathematical models describing the human growth curve. *Annals of Human Biology*, *5*, 1–24.
- Ramsay, J. O., Altman, N., & Bock, R. D. (1994). Variation in height acceleration in the Fels growth data. *The Canadian Journal of Statistics*, *22*, 89–102.
- Ramsay, J. O., Hooker, G., Campbell, D., & Cao, J. (2007). Parameter estimation for differential equations: a generalized smoothing approach (with discussion). *Journal of the Royal Statistical Society, Series B*, *69*, 741–796.

- Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis* (Second ed.). New York: Springer.
- Rosowsky, D. V. & Reinhold, T. A. (1999). Rate-of-load and duration-of-load effects for wood fasteners. *Journal of Structural Engineering*, *125*, 719–724.
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). Winbugs user manual, version 1.4. *Technical report*.
- Stone, C. J., Hansen, M., Kooperberg, C., & Truong, Y. K. (1997). Polynomial splines and their tensor products in extended linear modeling (with discussion). *The Annals of Statistics*, *25*, 1371–1470.
- Tuddenham, R. & Snyder, M. (1954). Physical growth of california boys and girls from birth to eighteen years. *University of California Publications in Child Development*, *1*, 183–364.
- Wahba, G. (1990). *Spline models for observational data*. Philadelphia: Society for Industrial and Applied Mathematics.
- Yao, F. Z. & Foschi, R. O. (1992). Duration of load in wood: Canadian results and implementation in reliability-based design. *Canadian Journal of Civil Engineering*, *20*, 358–365.