HIGH DIMENSIONAL COMPUTATIONAL MODELS FOR BIOMEDICAL IMAGING DATA ANALYSIS

By

Liangliang Zhang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics - Doctor of Philosophy

2017

ABSTRACT

HIGH DIMENSIONAL COMPUTATIONAL MODELS FOR BIOMEDICAL IMAGING DATA ANALYSIS

By

Liangliang Zhang

The importance of big data does not revolve around how much data you have, but what you do with it. You can take data from any source and analyze it to find answers that enable 1) cost reductions, 2) time reductions, 3) new product development and optimized offerings, and 4) smart decision making. My thesis is mainly focused on high-dimensional image data analysis and computations. The image data I worked on are CT images of abdominal aortic aneurysm (AAA) and brain images of Alzheimer's Disease. I developed Bayesian calibration method for the former and Supervised learning with Markov Chain for the latter.

Bayesian calibration has a long history within computer modeling in general. Bayesian calibration is an iterative process of updating uncertainty distributions on the calibration parameters in a way that is consistent with the observed data. Because of the advances in complex mathematical models and fast computer codes, Bayesian calibration of computer experiments are popular in the scientific research nowadays. As we know, compared to a computer model, a complex system in real life is expensive both in time and money to observe. Therefore, computer models can be a stand-alone tool or combined with (typically smaller) data from physical experiments or field observations. And Bayesian calibration is powerful in integrating all sources of uncertainties into the model definition and calibration procedure.

For the AAA data, first I modeled only one patient (patient-specific prediction discussed in Chapter 2), and then built an advanced model which can incorporate all patients (multi-

patients prediction in Chapter 3). In the process, semi-parametric functional data analysis, covariance modeling and Bayesian methods was highly practiced and used. The contributions are as follows. First, we formulate the Bayesian calibration of our AAA G&R computation model taking into account model inadequacy, prior distributions of model parameters, measurement errors, and most importantly, longitudinal CT scan images. Next, we demonstrate how to achieve the proposed aims by solving the formulated Bayesian calibration problem using a simulation study and real data analysis. In particular, we compare and discuss the performance and computation time under different sampling cases for the computation model output data and (synthesized) patient data, both of which are synthesized by the G&R computation. We apply our Bayesian calibration to the real CT data and validate our prediction, showing the usefulness of our approach to the computational science and medical communities in aiding decision making.

For the Alzheimer's Disease data, the causes are currently being researched massively, but no definitive answers exist as yet. Genetic predisposition, abnormal protein deposits in the brain and environmental factors are suspected to play a role in the development of the disease. In Chapter 4, my main goal is to model the progression of Alzheimer's Disease by applying multi-state Markov model, and to investigate the significance of known risk factors like Age, ApoE4 and some brain structural volumetric variables getting from MRI like hippocampus, and at the same time, to predict the transitions between different clinical diagnosis states based transition rates and transition probabilities. We found that the model with age is not significant (p-value is 0.1733) according to the likelihood ratio test, while ApoE4 is a significant risk factor in our Markov model. Predictions based on transition rates and transition intensities were made and validated with the accuracy as high as 0.7849.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In this thesis, the image data I worked on are CT images of abdominal aortic aneurysm (AAA) and brain images of Alzheimer's Disease. I developed Bayesian calibration method for the former and Supervised learning with Markov Chain for the latter. Chapter 1 reviews the background, research questions and development of Bayesian calibration method. Chapter 2 introduces patient-specific prediction for single subject data by using Bayesian calibration. Chapter 3 introduces patient-specific prediction for multiple subjects data by using Bayesian calibration. Chapter 4 introduces analysis on conversion of Alzheimers Disease using a multi-state Markov model. Chapter 5 introduces Conclusions, Discussion, and Suggestions for Future Research

1.1   Computer model

Complex models (also called computer models) are used to resemble real-world systems and are thus used in many different fields, such as clinical sciences, Physics, engineering and climate change. A computer model is defined by a set of equations and is implemented as computer codes. People typically use the model to predict how the real world system may behave in the future. Running a computer model at a number of different input values is known as computer experiment. Under different backgrounds the output could be either a scalar or a vector; however, we will just consider the case with univariate output here. The

computer models are deterministic, for each time they are run with the same inputs they will produce the same output. [99]

The inputs usually are used to describe some properties of the real system which need to observed. Thus the true values of these inputs may be unknown. Although when the values of the inputs can be obtained, it is likely that some form of measurement error will have occurred. Therefore, the uncertainty in the inputs needs to be quantified as it will relate to how uncertain we are that the computer model output matches reality. In other words, uncertainty analysis is aiming at quantifying the uncertainty in model outputs caused by uncertainty in inputs. Besides, another source of uncertainty is in the model itself. Even when the true values of the inputs are known, running the model at these points will not produce an output which matches exactly the observation of the real-world system.

Complex models are computationally expensive and take a long time to run. Because a simulation of a real-world system can require the high number of dimensions. Sensitivity analysis is defined as the process of evaluating how the output of a model is modified by changes in the inputs. Performing analyses such as sensitivity and uncertainty analysis can require many runs of the computer model and this quickly becomes impractical with a computationally expensive model.

1.2   Gaussian Process Emulation

In statistics, Gaussian process emulator is one name for a general type of statistical model that has been used in contexts where the problem is to make maximum use of the outputs of a complex (often non-random) computer model. The main element of the Gaussian process emulator is that it models the outputs as a Gaussian process on a space that is defined by

the computer model inputs.

The analysis of computer models and Gaussian process emulator has been studied for about 20 years. Sacks et al. (1989) [89] describe a technique for approximating an unknown deterministic computer experiment by modeling the output as a realization of a stochastic process. A Bayesian method of predicting a computer model at untested inputs is given in Currin et al. (1991) [22]. O'Hagan (2006) [77] gives an introduction to Gaussian process emulation as well as details of Bayesian methods for various analyses.

1.3   Design of Computer Experiments

In order to obtain the training data, we need to select the design. The choice of design points is named as the design of the experiment. Without intensively increasing the computational time, we aim to gain more information from the training data, so that the emulator can approximate better the computer model. Various criteria are defined under different designs. For example, under a maximum design, points are selected such that the maximum distance between any two design points in minimized.

McKay et al. (1979) [69] compare three methods for choosing designs for Monte Carlo studies–simple random sampling, stratified sampling and Latin hypercube sampling. A Latin hypercube sample ensures that the points are spread evenly over the range of each input dimension. The three methods are assessed by comparing estimators of the mean, variance and distribution function of the output. They conclude that Latin hypercube sampling is preferred. Besides, Santner et al. (2003) [91] comment on the widespread use of this method for generating designs in other kinds of computer experiment.

1.4 Bayesian Statistics

Bayesian statistics began with a posthumous publication in 1763 by Thomas Bayes, [5] a non-conformist minister from Tunbridge Wells. [51] His work was formalised as Bayes's theorem, which is a simple and uncontroversial result in probability theory. Bayesian statistics is a system for describing epistemological uncertainty using the mathematical language of probability. [98]The methods can be outlined as formal combinations through the use of Bayes's theorem of: 1. a prior distribution about the value of a quantity of interest based on evidence not derived from the study, with 2. a summary of the information (likelihood) concerning the same quantity available from the data collected in the study, to yield 3. an updated or posterior distribution of the quantity of interest. These methods address the question of how new information should update what we currently believe. They extend naturally into making predictions, synthesizing information from different sources, and designing studies. Nevertheless, Bayesian methods are a controversial topic. Because they may involve the explicit use of subjective judgment in a rigorous scientific exercise.

Compared with traditional methods, a Bayesian perspective leads to more flexible and ethical methods of analyzing clinical trials and observational studies, [56] and to more elegant ways of handling multiple substudies, for instance when simultaneously estimating the effects of a treatment on many subgroups [10]. A Bayesian approach enables one to provide suitable conclusions for making decisions for specific patients, for planning research, or for public policy [65].

1.5 Calibration

For a conventional calibration, an expert modeler repeatedly adjusts calibration parameters to minimize the difference between observed data and model outputs (Reddy et al., 2007) [82]. One main drawback of this process is that it demands intensive manual labor and inherently introduces modeler biases into the calibrated model. To improve, optimization schemes have been employed to automate calibration (New and Chandler, 2013 [74]; Raftery et al., 2011 [81]). These schemes determine the combination of model parameters by minimizing the error between data and model outputs.

Therefore, the solution of the calibration problem can be understood as identifying the global optimum of the objective function within the feasible domain of the calibration parameters. When both the objective function and the feasible domain are convex or they can be restated as convex, it is very stable for the mathematical algorithm to find the single optimum. For problems that are not convex or cannot be recast as convex, exact solutions often become computationally expensive, and only approximate solutions may be practical [9]. For more details, please refer to [25].

In most cases, calibration is the same as statistical estimation [46], [27]. Both processes are aiming at finding input values that lead to the best possible model fit. For example, if the objective function of the calibration is a likelihood function, calibration is equivalent to maximum likelihood estimation.

1.6 Bayesian calibration

Many methods of automated calibration have been developed which reduce costs, time and biases. Among them, Bayesian methods differ significantly from the others in that inputs are

assumed to be uncertain and the main goal is not to match the prediction to the measured data as closely as possible, but to reduce the uncertainty in the inputs in a manner consistent with the measured data. When the sample size of measured data are limited and the model inputs have high sensitivity and high uncertainty, Bayesian methods are particular useful to achieve calibration. For more details, please refer to Muehleisen and Bergerson 2016 [73].

Bayesian calibration is an advanced method of calibration and thus is fundamentally different than conventional calibration methods. Bayesian calibration has a long history within computer modeling in general ([30]; [80]; [58]). Bayesian calibration is an automated process of updating uncertainty distributions on the calibration parameters in a way that is consistent with the observed data. Unlike the conventional calibration that minimizes the difference between observed data and model outputs, Bayesian calibration determines the most likely uncertainties for input parameters that yield an output uncertainty in which the observed data is most likely. Bayesian calibration is an application of Bayes Theorem, which combines prior information with future information contained in the likelihood of observed outputs from the model ([6]). As commented by Kennedy and O'Hagan (2001) [58], none of the existing methods of calibration recognize fully, all sources of uncertainty.

As similar as a regular Bayesian problem, the general process of Bayesian calibration is 1) define PDFs (probability density functions) for uncertain model parameters, 2) collect simulations and observations of the real system for known input parameter values, and 3) calibrate (assumed) prior parameters based on the observed data by iteratively using Bayes Theorem until iterations converge to an acceptable level. Establishing prior PDFs for input parameters generally requires both expert opinion solicitation and literature review. For example, a uniform distribution may be assumed, when only a range of appropriate values can be determined for a parameter(Riddle and Muehleisen, 2014 [85]). A triangular distribution

may be used, when both a range of values and a most likely value can be determined.

1.7   Advantages and disadvantages

As Bayesian calibration integrates uncertainty to the model definition and calibration procedure, one of the primary advantages is that a modeler can quantify a confidence level in the calibrated model. Another advantage is that a probabilistic risk analysis can be performed and competing retrofits can be ranked according to a desired risk level, based on mean values, variances, etc. Finally, Bayesian calibration reduces the tendency to overfit the model to the observed data. Overfitting is a problem in regression analysis, multi-objective optimization, and machine learning when the fitting/optimization routine fits the model to the noise in measured data (Dietterich, 1995 [29]). Unlike the conventional calibration that minimizes the difference between observed data and model outputs, Bayesian calibration is trying to maximize the likelihood that the model output is statistically consistent with the measured data. This consistency naturally takes into account the uncertainty in the measured data (Heckerman, 1998 [48]).

There is one common disadvantages of all calibration methods, both conventional and Bayesian. It is typically very challenging, time consuming and expensive to collect observed data from the real system. Another disadvantage of Bayesian calibration requires a significant number of iterations to converge to the most likely PDFs, especially when the prior probability density functions are poorly chosen. Finnally, because of the high computational demand, in order to reduce run time, Bayesian calibration often requires a pre-step of parameter screening and selection to obtain the most significant parameters to use in the calibration process in large models [73].

Chapter 2

Patient-Specific Prediction from Single-subject Data

2.1 Introduction

(b)
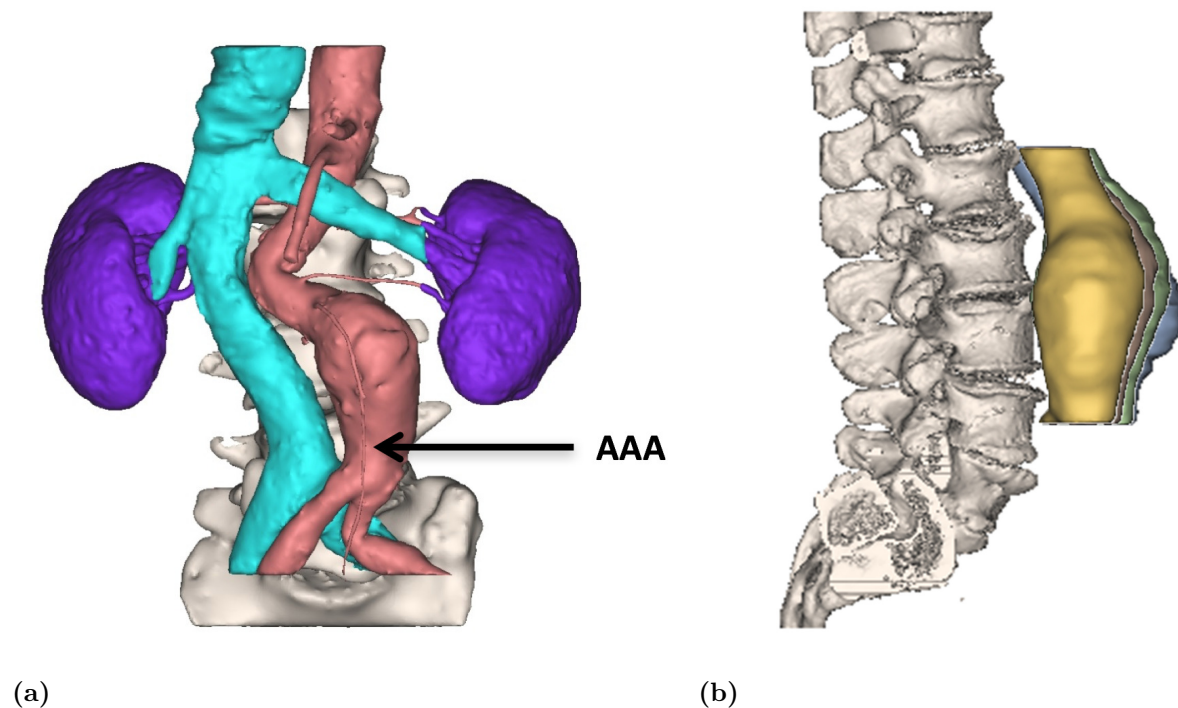
Figure 2.1. (a) The location of an abdominal aortic aneurysm colored as redis indicated by an arrow, while green region on the left indicates the inferior vena vein and (b) four longitudinal AAA CT images 1, 2, 3, and 4 which are screened at consecutive 4 times. Regarding the technique to align the CT scan images, please refer to Kwon et al. [61].

An abdominal aortic aneurysm (AAA) is an enlarged localized volume in the lower part

of the aorta, which supplies blood to a large part of the body (see Fig. 2.1). Enlargement of the aorta by more than 50% of its normal diameter is defined as an aortic aneurysm. The vast majority (over 90%) of aortic aneurysms occur in the abdominal region, specifically the infrarenal aorta [79] [114]. In general, an aorta with a diameter larger than 3 cm is considered an aneurysm.

A ruptured aneurysm can cause life threatening internal bleeding. If ruptured, patient mortality rates are greater than 80% [59, 60]. Depending on the size and rate of growth, treatment of an AAA may vary from watchful waiting to emergency surgery. Once an AAA is found, doctors will closely monitor it so that surgery can be planned if it becomes necessary. Since elective repair of AAAs can result in peri-operative deaths (4-8%) [63], performing unnecessary surgeries increases patient risk. A thorough understanding of the expansion and rupture of AAAs is thus needed in order to minimize patient risk.

While significant advances have been made in the management of AAA patients [13], this disease still carries a high mortality rate. During the last decade, bio-chemo-mechanical studies have been integrating computation modeling with increased understanding of the expansion and weakening of aneurysms. Recently, this computational platform, called a growth and remodeling (G&R) model, has been developed and incorporated with patient-specific anatomical information, which aids in treatment planning on a per-patient basis [2, 36, 108, 117]. Zeinali-Davarani *et al.* [117] developed the G&R model to take into account both elastic degeneration and stress-mediated collagen turnover during AAA development using finite element analysis (FEA). By the way there are still other models of AAAs without G&R have been shown to be potentially useful to aid in patient-specific treatment planning [31,68]. A coupled simulation of G&R with hemodynamics was conducted for studying its effects on AAA expansion [95]. Geometric, kinetic and material parameters have

been identified for individual patients using inverse optimization techniques for modeling the growth of AAAs. Furthermore, it has been shown that the same material parameters for AAA expansion can help to predict intrasac-pressure dependent vascular adaptation after endovascular repair [62].

Translating recent computational advances into a predictive tool for individualized clinical treatment, however, requires a major paradigm shift due to the incompleteness of the model, limited information, and uncertainty associated with clinical measurements with regard to each individual patient. Most importantly, the associated uncertainty in the prediction propagated from various sources needs to be correctly quantified. For example, the G&R model's internal parameters need to be carefully adjusted according to patient-specific data, e.g., longitudinal computed tomography (CT) images, in order to make better prediction and so be useful for clinical decisions.

The aims of this paper are to develop a framework that 1) first calibrates the physical AAA G&R model using patient-specific longitudinal CT scan images, 2) predicts the expansion of an AAA in future time, and 3) analyzes the associated uncertainty in the prediction.

To achieve our aims, we perform Bayesian calibration of our computational AAA G&R model [58]. In particular, Bayesian calibration will be used to incorporate the computational G&R model, patient-specific data (e.g., CT scan images), and various uncertainties as well as to compute the uncertainty level of the prediction on the AAA expansion.

As computational science advances, there is a growing interest in applying Bayesian calibration developed from the statistical community to engineering applications (see more details in [49, 50, 58, 77, 112] and references therein). To the best of our knowledge, we are the first to apply the Bayesian calibration method to the AAA G&R computation model. The successful outcomes of our aims will make the G&R computation model viable to aid

clinicians in decision making. Hawkins-Daarud A. *et al.* [47] used a Bayesian framework to address questions on validation, model selection, and uncertainty quantification for tumor growth. Biehler J. *et al.* [7] presents an uncertainty quantification framework based on multi-fidelity sampling and Bayesian formulations and analyzes the impact of the uncertainty in the input parameter on mechanical quantities typically related to abdominal aortic aneurysm rupture potential. Zhao X. and Pelegri A. [119] used a Bayesian approach to estimate a reasonable quantification of the probability distributions of soft tissue mechanical properties in the presence of measurement noise and model parameter uncertainty. While Hawkins-Daarud A. *et al.* [47] considered a set of discrete model candidates to select and validate a model, we consider a statistical model for the true physical process by introducing two Gaussian random fields for the G&R computation model and the inadequacy of the model. We adopt a Bayesian calibration technique proposed in [58] to calibrate parameters in the G&R computation model and predict the AAA expansion. The proposed statistical model and Bayesian calibration can take different sources of uncertainty; therefore, it is well suited to achieve our aims in predicting the AAA expansion process as well as in computing the propagated uncertainty given the patient-specific data.

The contributions of our paper are as follows. First, we formulate the Bayesian calibration of our AAA G&R computation model taking into account model inadequacy, prior distributions of model parameters (Seyedsalehi S and Zhang L, et al. [94]), measurement errors, and most importantly, patient-specific longitudinal CT scan images. Next, we demonstrate how to achieve the proposed aims by solving the formulated Bayesian calibration problem using a simulation study and real data analysis. In particular, we compare and discuss the performance and computation time under different sampling cases for the computation model output data and (synthesized) patient data, both of which are synthesized by the

G&R computation. We apply our Bayesian calibration to the real patient-specific CT data and validate our prediction, showing the usefulness of our approach to the computational science and medical communities in aiding decision making.

The organization of the paper is as follows. Section 2.2 introduces the AAA G&R model, the quantity of interest in making prediction, and the full statistical model with hyperparameters for Bayesian calibration. In Section 2.3, we discuss assumptions for priors and then provide the posterior and predictive distributions under the Bayesian framework. Section 2.4 describes the design of the simulation study with synthetic observation data and a case study with real patient data. In Section 2.5, we present the results of the Bayesian calibration for both simulation and real data cases. Finally, we provide some discussions and concluding remarks in Section 2.6.

2.2 Models
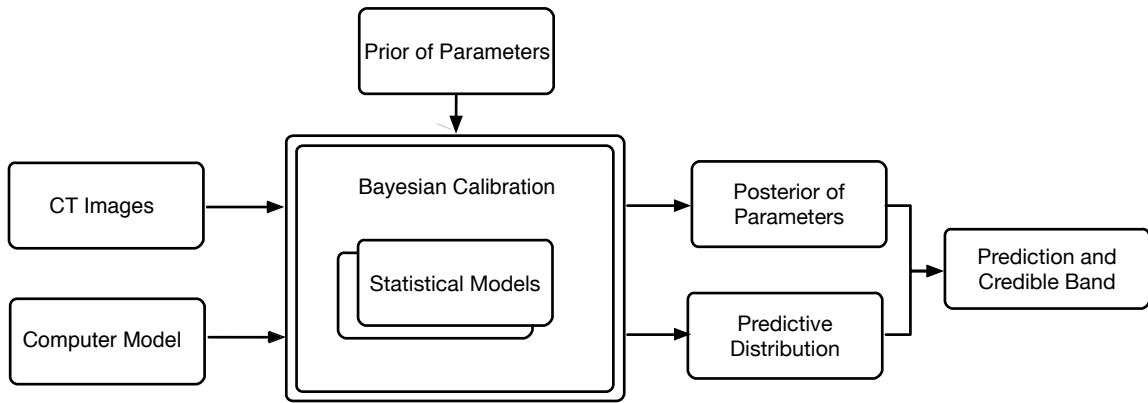
Figure 2.2. The flow-chart of Bayesian calibration.

Bayesian calibration can be defined as a Bayesian approach to calibrate investigated parameters in a theoretical model for a real complex system (here stands for the G&R computer model for AAA) that enables the incorporation of uncertainty regarding parameters,

real observations and possible simulations from the theoretical model. Fig. 2.2 shows the general framework of Bayesian calibration, which also structured all the contents needed to be discussed in the paper. The left two rectangles and one on the top list the inputs we should prepare for a Bayesian calibration, while the right three rectangles list the outputs we can get from a Bayesian calibration. The rectangle with double edges in the middle depicts that Bayesian calibration is actually a specially designed complex system capsulating a bunch of statistical models. As you can see, Bayesian calibration is including a bunch of models and integrating different sources of data together to generate products, such as calibrated parameters (posterior distribution), prediction (predictive distribution) and statistical inference (credible bands). In this section, we introduce the data and all the basic models needed in our Bayesian calibration (basically things in left two rectangles and the rectangle with double edges in the middle). The following parts in this section are organized as follows. 2.2.1 introduces a computational model of AAA growth. 2.2.2 introduces how the real data from CT images are prepared for Bayesian calibration. 2.2.3 introduces a calibration model to combine the computational model with real data. 2.2.4 introduces statistical specifications for implementing Bayesian calibration.

2.2.1 AAA G&R computation model

The Bayesian calibration framework includes a computational G&R model of AAA as a data input, where the detailed computational G&R model was described in Zeinali-Davarani et al. [117]. Briefly, the computational G&R model has three parts: constitutive relations of intrinsic material behavior, a stress-mediated production function, and a damage function. To describe material behavior, we assume that the aorta is comprised of three stress-bearing constituents, viz. elastin, collagen fiber families, and vasoactive smooth muscle cells. Each

13

constituent, in addition to its contribution to the construction and strength of the artery's wall, has its own individual properties, where the population-based material parameter distributions of abdominal aortas was given by Seyedsalehi et al. [94]. Also, the stress-mediated production functions connect the stress-state of the artery to changes of the mass rates with a stress-mediated feedback approach. Moreover, in the G&R model, the AAA is initialized by imposing damage function to the elastin of normal aorta. The initialization can be justified by the previous study [87], which shows that one of the main features of AAA is that elastin is decreased and the study [55, 115] that elaborate that the degradation of elastin can directly form patient-specific shapes of aneurysms. However, the other factors, such as alteration of intrinsic material parameters [83], disturbed collagen production [23, 24] and hemodynamics [3], are taken as the minor reason of initialization of AAA, and be considered as statistical error later.

In this paper, we aim to calibrate the computational model with the expansion rate and evolution of aneurysm shapes along the longitudinal scan images. Our focus of interest, called quantity of interest (QoI) (described in section 2.2 for the Bayesian calibration), is used to calibrate the aneurysm growth with medical data. However, those parameters related to intrinsic material function and stress-mediated production function cannot directly be used to form the patient-specific geometry, but have a strong relation with patient's health, age and other patient behavior, while with respect to [115], the degradation of elastin play a leading role in forming a patient-specific geometry, which is to say different patient-specific geometry can be achieved given different damage parameters in G&R simulation. Consequently, we prescribe the mean values of the material parameters [94] and stress-mediated parameters [116] and calibrate the damage parameters statistically with the evolution of AAA geometry. Also, in the statistical calibration model introduced later, random variations of prescribed

14

parameters will be treated as errors.

In our model, damage function works on elastin and vasoactive smooth muscle, therein elastin plays an important role in the mechanical behavior of aorta. Elastin contributes resilience and elasticity to the aortic tissue; but when the person's age is advanced, elastin cannot be replaced. For an AAA, the localized dilation of the aorta is initiated by the degradation of the elastin. This degradation (or damage) will reduce the amount of elastin, leading to the weakening of the wall. The damage will result in the increase of the diameter and wall stress of the aneurysm. The increase of the stress in constituents results in an increase in the accumulation of collagen and smooth muscles as a way to compensate for the elastin's loss and decrease stress in the wall. All the relations and details of the model have been previously reported by Zeinali-Davarani *et al.* [116] and Kwon *et al.* [62].

Although the AAA's shape has a wide variability and its expansion associates with the complexity of the G&R process, in the current study, we decided to use the 2D axisymmetric G&R model and first study the effectiveness of the Bayesian calibration. As discussed, the elastin damage in the aortic wall initiates the growth of the aneurysm. Here we define the initial elastin's damage function as

$$d(s) = \theta_1 \, \exp\left(-\left[\frac{|s - \alpha_1|^{\alpha_2}}{2\,\theta_2^2}\right]\right) \tag{2.1}$$

where s is the coordinate defined on the centerline. $g(s) = 1 - d(s)$ is the ratio of remaining elastin to the initial amount at s.

The damage functions in (2.1) contains two parameters of interest $\{\theta_1, \theta_2\}$ to be calibrated from the real data and two other quantities $\{\alpha_1, \alpha_2\}$ that can be easily identified via CT images without Bayesian calibration. They have their own specific effects on the shape of

the damage function and thus on the stress-stretch and geometrical state of the AAA at a given time. In particular, θ_1 is a scaling factor with $\theta_1 \in [0, 1)$. An increase in θ_1 toward 1 will increase the degradation of elastin and thus increase the dilatation of the artery. $\theta_1 = 0$ means no degradation; hence, the artery will retain its initial state. α_1 corresponds to the location on the centerline at which the maximum damage occurs. α_1 and α_2 control the areal distribution of damage on the sides of the peak location.

Note that α_1 and α_2 are relatively easy to be estimated from the CT images compared to θ_1 and θ_2. In particular, the maximum diameter occurs at or very close to the location of the maximum damage; hence, we estimated α_1 by finding the location of the maximum diameter using the CT images. α_2 is chosen by minimizing the error between the model simulation and CT images. Thus, we fix α_1 and α_2 with appropriate values obtained from CT images directly *a-priori* and focus on calibration of θ_1 and θ_2 in a Bayesian way.

To see our simulation results in a symmetrical and straightforward configuration, we have chosen $\alpha_1 = 7.5$ (i.e., averaged value) and $\alpha_2 = 2$ (i.e., the most simple shape). On the other hand, we used $\alpha_1 = 12.1$ and $\alpha_2 = 5$, which are estimated directly from the patient CT images for real data analysis in Section 2.5.2. Additionally, the choice of $\alpha_2 = 2$ for the simulation study covers the other end of the possible values as compared to the real observation case with $\alpha_2 = 5$.

2.2.2 Quantity of interest (QoI) for AAA G&R

The quantity of interest (QoI) of AAA G&R is what we want to predict in AAA growth in future time. The selection of the QoI will let us subsequently determine the statistical models and investigate the associated uncertainties. In our study, we select the radius of the inscribed sphere with respect to a centerline as a QoI, where a centerline is computed

by inscribed spheres in the AAA 3D images [40]. Figs. 2.3a and 2.3b show inscribed spheres and the resulting centerline, respectively, for a given 3D point cloud sampled from CT scan image data from a patient. This QoI selection is consistent with medical practice in which the diameter of the AAA is used as an important *decision variable* [59, 60].

To have a brief idea of the relationship among G&R model, 2D axisymmetric and QoI, we plot a diagram showing how the 3D images are transformed into 2D data and how the computer model simulates it. When we look at Fig. 2.4, there are two lines of processes. First, I will describe how the data is transformed, as shown in Fig. 2.4a, Fig. 2.4b and Fig. 2.4c. Fig. 2.4a shows the same process of transforming 3D images into 2D data as Fig. 2.3. Thereafter, taking the length of the centerline as the coordinate, we obtain the radius of the inscribed sphere r versus the height coordinate of the centerline s in Fig. 2.4b, whose details are explained in [40]. Then we truncate the curve in Fig. 2.4b, and get the QoI in Fig. 2.4c. In short, the 3D images are transformed into 2D data and then cut off into QoI. Second, I will describe how G&R model works to simulate the QoI. Fig. 2.4d shows the computational model of healthy aorta, then the healthy aorta evolutes into AAA following mechanisms introduced by computational G&R model, which is actually a 2D axisymetric geometric shape (Fig. 2.4e). Then we obtain the radius versus the height coordinate of the centerline in Fig. 2.4f which can be called as simulation. And finally, Fig. 2.4g shows the simulation result by minimizing the difference between the QoI and simulation.

Fig. 2.5 shows the real data needed in the Bayesian calibration procedure, which is the QoI's from scan images of patient K taken at 4 time points. To get each curve, we can repeat the same process from (a) to (c) in Fig. 2.4 on each scan image. Bayesian calibration combines the QoI's and the simulation data to calibrate the selected parameters in the computer model. And the radius in the figure are actually the target quantities we want

17

to predict finally. Therefore, we are saying that we would like to make prediction of the QoI with associated uncertainty in future times. In this paper we consider the following questions. Can we predict the radius vs. height profiles for CT scan images 3 and 4 in future time given CT scan images 1 and 2 in Fig. 2.5? What is the uncertainty associated with such prediction? We answer these questions by applying the proposed Bayesian calibration method. The answers to the questions for this particular patient-specific data set shown in Fig. 2.5 are given in Section 2.5.2 as a real data study case.

(a)   (b)
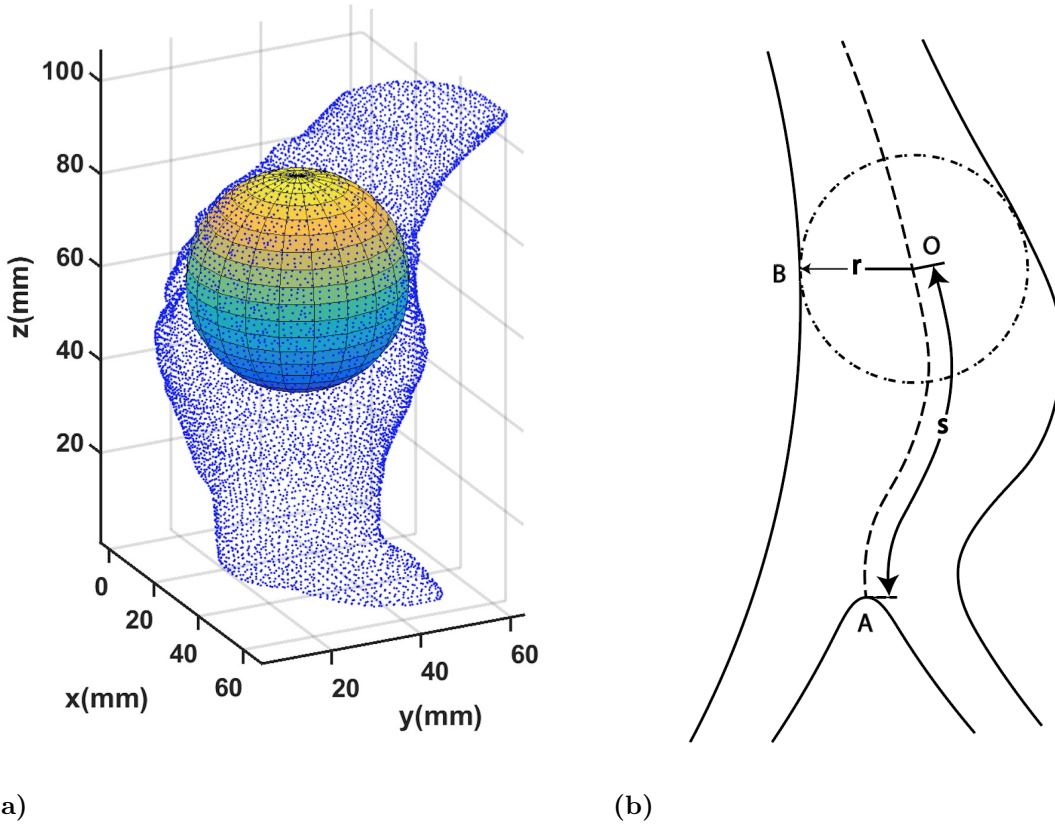
Figure 2.3. 3D point clouds sampled from a CT scan image. (a) An inscribed sphere with point clouds from the CT scan images and (b) a centerline computed from point clouds.
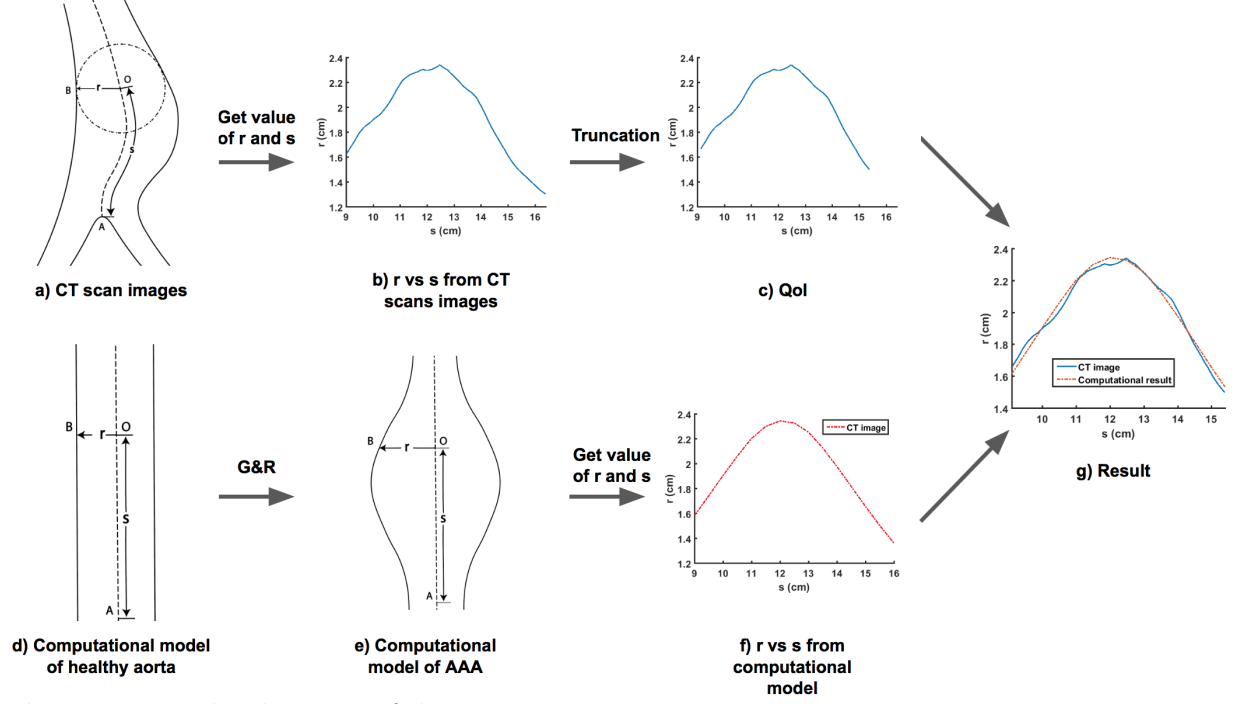
Figure 2.4. The diagram of data preparation.

2.2.3 Calibration model

Let $\zeta(\boldsymbol{x})$ be the QoI of the true AAA G&R process, the input variable of which is denoted by $\boldsymbol{x}$ and defined as $\boldsymbol{x} = [t\ s]$, where t is the time and s is the height on the centerline as illustrated in Fig. 2.3b. Suppose that we have n observations from CT scan images of a given patient, then "image n" is the "n-th" image of one patient. To model possible observation error, e.g., resolution and segmentation errors in CT scan images, we consider the noisy observations as follows.

$$z_i = \zeta(\boldsymbol{x}_i) + \varepsilon_i, \quad \forall i \in \{1, \cdots, n\}, \tag{2.2}$$

where the difference (or error) between the observation and the true process is denoted by ε_i. We further assume that each ε_i is independently distributed as $\mathcal{N}(0, \lambda)$, which represents a normal (Gaussian) random variable with mean 0 and variance λ.

19
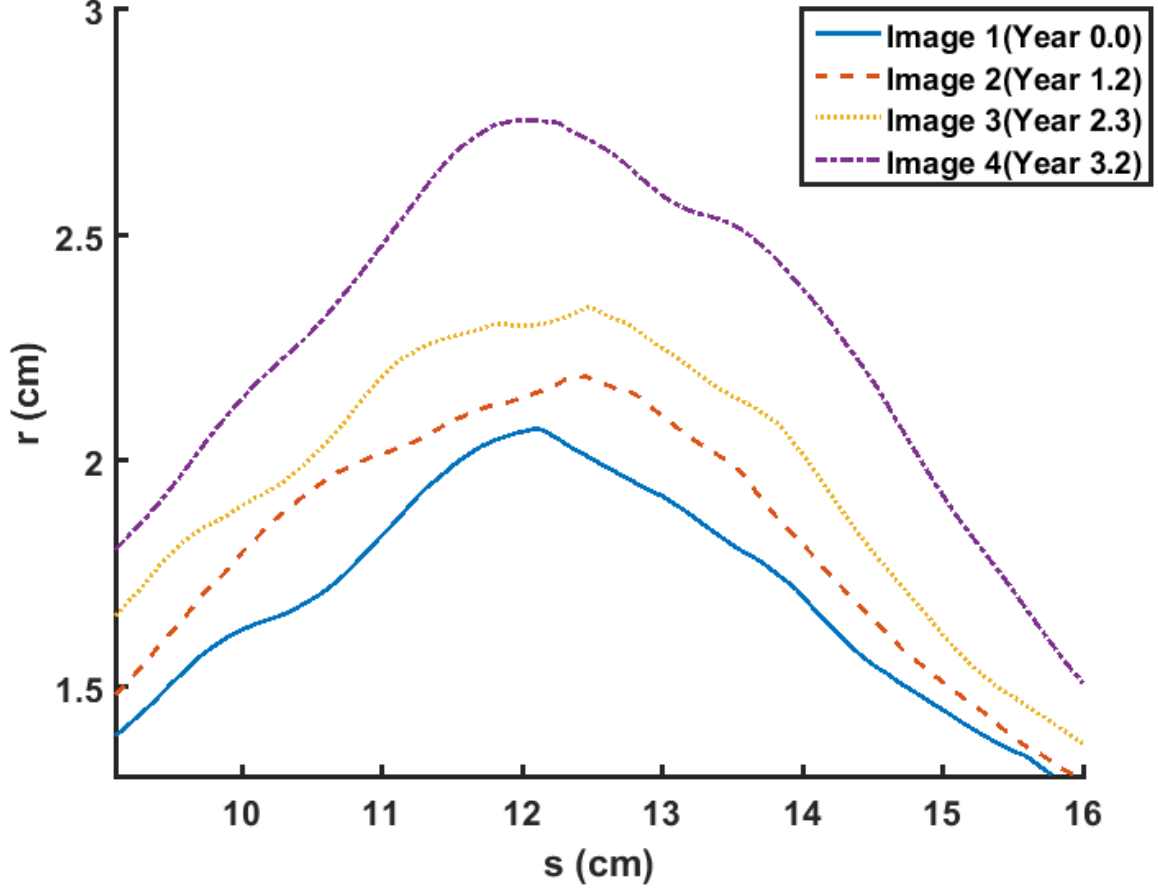
Figure 2.5. Radius vs height coordinate on the centerline measured from CT scans of a patient over the surveillant time. The lines from bottom to top represent images 1, 2, 3, and 4, respectively.

We denote the QoI of the G&R computation model output at $\boldsymbol{x}$ as $r(\boldsymbol{x}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is called a set of calibration parameters, or calibration inputs. In the G&R computation model for the AAA expansion, damage parameters serve as calibration parameters that are patient specific for the AAA growth, i.e., the calibration parameters are $\boldsymbol{\theta} = [\theta_1 \ \theta_2]$ defined in Section 2.2.1.

Given the available QoI of the G&R computation model, we model the true process as

$$\zeta(\boldsymbol{x}_i) = r(\boldsymbol{x}_i, \boldsymbol{\theta}) + \delta(\boldsymbol{x}_i), \tag{2.3}$$

where $\delta(\cdot)$ is a model inadequacy function, i.e., model error, which is independent of the computation outputs. It is natural to assume the true AAA expansion process $\zeta(\boldsymbol{x})$ cannot be fully described by the computation model, therefore we introduce $\delta(\boldsymbol{x})$ to represent the discrepancy in (2.3).

By combining (2.2) and (2.3), we have

$$z_i = r(\boldsymbol{x}_i, \boldsymbol{\theta}) + \delta(\boldsymbol{x}_i) + \varepsilon_i, \tag{2.4}$$

which gives *a calibration model* that relates G&R computation outputs with the true process and the observations. The Bayesian calibration method [58] we adopt in this paper introduces Gaussian process priors for the computational model and the model error in order to calibrate $\boldsymbol{\theta}$ and predict the QoI. Fig. 2.2 shows the flow-chart for Bayesian calibration. The statistical models in Fig. 2.2 build on two Gaussian process priors for the G&R computation model and model error, which will be discussed in the next section.

2.2.4 Statistical models

Let $\mathcal{GP}(\boldsymbol{m}(\cdot), \boldsymbol{k}(\cdot, \cdot))$ be the Gaussian process with the mean function $\boldsymbol{m}(\cdot)$ and the covariance function $\boldsymbol{k}(\cdot, \cdot)$. $\mathcal{GP}$ is flexible and popularly used as a prior model for functions [110, 111]. We introduce the following Gaussian processes as prior beliefs for the G&R computation model and the model error:

$$r(\boldsymbol{x}, \boldsymbol{\theta}) \sim \mathcal{GP}(\boldsymbol{m}_1(\boldsymbol{x}, \boldsymbol{\theta}), \boldsymbol{k}_1(\boldsymbol{x}, \boldsymbol{\theta}; \boldsymbol{x}', \boldsymbol{\theta}')),$$
$$\delta(\boldsymbol{x}) \sim \mathcal{GP}(\boldsymbol{m}_2(\boldsymbol{x}), \boldsymbol{k}_2(\boldsymbol{x}, \boldsymbol{x}')). \tag{2.5}$$

To specify Gaussian process priors in (2.5) further, we introduce mean and covariance

structures for both processes.

For a mean structure, one can consider a linear combination of basis functions to approximate the general mean structure. The linear function is always the first thing to do, although it is not perfect, since the results of linear function will give us insights for other nonlinear study. Let $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta})$ be a vector of basis functions and $\boldsymbol{\beta}$ be a vector of corresponding coefficients then we can represent $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta})\boldsymbol{\beta}^T$, where $(\cdot)^T$ is the transpose of a matrix or a vector. Since we need two mean functions, $\boldsymbol{m}_1(\boldsymbol{x}, \boldsymbol{\theta})$ and $\boldsymbol{m}_2(\boldsymbol{x})$, we introduce two sets of $\boldsymbol{h}$ and $\boldsymbol{\beta}$ such that $\boldsymbol{m}_1(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{h}_1(\boldsymbol{x}, \boldsymbol{\theta})\boldsymbol{\beta}_1^T$ and $\boldsymbol{m}_2(\boldsymbol{x}) = \boldsymbol{h}_2(\boldsymbol{x})\boldsymbol{\beta}_2^T$.

In this paper, we consider the following linear mean structures for computational efficiency:

$$\boldsymbol{m}_1(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{h}_1(\boldsymbol{x}, \boldsymbol{\theta})\boldsymbol{\beta}_1^T = \beta_{10} + \beta_{11}t + \beta_{12}\theta_1 + \beta_{13}\theta_2, \tag{2.6}$$

where $\boldsymbol{h}_1(\boldsymbol{x}, \boldsymbol{\theta}) = [1 \ t \ \theta_1 \ \theta_2]$ and $\boldsymbol{\beta}_1 = [\beta_{10} \ \beta_{11} \ \beta_{12} \ \beta_{13}]$, and

$$\boldsymbol{m}_2(\boldsymbol{x}) = \boldsymbol{h}_2(\boldsymbol{x})\boldsymbol{\beta}_2^T = \beta_{21}t + \beta_{22}s, \tag{2.7}$$

where $\boldsymbol{h}_2(\boldsymbol{x}) = [t \ s]$ and $\boldsymbol{\beta}_2 = [\beta_{21} \ \beta_{22}]$. These mean structures imply that the mean function of the G&R computation model is linear in time t and calibration parameters, $\{\theta_1, \theta_2\}$. The mean function of the model error is linear in time t and location s. $\boldsymbol{\beta} = [\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2]$ are hyperparameters for the mean functions in a Bayesian context.

For a covariance structure, we use the following exponential functions as follows.

$$\boldsymbol{k}_1(\boldsymbol{x}, \boldsymbol{\theta}; \boldsymbol{x}', \boldsymbol{\theta}') = \sigma_1^2 \exp\{-(\boldsymbol{x} - \boldsymbol{x}')\boldsymbol{\Omega}_x(\boldsymbol{x} - \boldsymbol{x}')^T\} \exp\{-(\boldsymbol{\theta} - \boldsymbol{\theta}')\boldsymbol{\Omega}_\theta(\boldsymbol{\theta} - \boldsymbol{\theta}')^T\},$$

$$\boldsymbol{k}_2(\boldsymbol{x}, \boldsymbol{x}') = \sigma_2^2 \exp\{-(\boldsymbol{x} - \boldsymbol{x}')\boldsymbol{\Omega}_x^\star(\boldsymbol{x} - \boldsymbol{x}')^T\},$$

where $\boldsymbol{\Omega}_x$, $\boldsymbol{\Omega}_\theta$, $\boldsymbol{\Omega}_x^\star$ are diagonal matrices such that

$$\boldsymbol{\Omega}_x = \begin{pmatrix} \omega_{x1} & 0 \\ 0 & \omega_{x2} \end{pmatrix}, \boldsymbol{\Omega}_\theta = \begin{pmatrix} \omega_{\theta 1} & 0 \\ 0 & \omega_{\theta 2} \end{pmatrix}, \boldsymbol{\Omega}_x^\star = \begin{pmatrix} \omega_{x1}^\star & 0 \\ 0 & \omega_{x2}^\star \end{pmatrix}$$

so that the hyperparameters for the covariance functions are σ_1^2, σ_2^2 and w_{xj}, $w_{\theta j}$, $w_{xj}^\star$'s. Note that the resulting covariance functions are not isotropic since they allow different scaling for each coordinate. Additionally, they can be regarded as separable covariance functions since the dependence due to one coordinate is independent of the dependence due to another coordinate. One can consider a more generalized covariance structure at the cost of higher computational complexity.

For the parameters controlling the covariances, we introduce $\boldsymbol{\psi} = [\boldsymbol{\psi}_1 \ \boldsymbol{\psi}_2]$, with

$$\begin{aligned} \boldsymbol{\psi}_1 &= [\omega_{x1} \ \omega_{x2} \ \omega_{\theta 1} \ \omega_{\theta 2} \ \sigma_1^2], \\ \boldsymbol{\psi}_2 &= [\omega_{x1}^\star \ \omega_{x2}^\star \ \sigma_2^2]. \end{aligned} \tag{2.8}$$

$\boldsymbol{\psi}_1$ is the set of hyperparameters related to the G&R computation model. $\boldsymbol{\psi}_2$ is the set of hyperparameters related to model error. For the remainder of the paper, the AAA G&R computation model is referred to as the computation model for notational simplicity.

2.3 Bayesian analysis for calibration

Bayesian calibration considers computation model outputs as the data in addition to the observations and combines these two to calibrate $\boldsymbol{\theta}$ under the Bayesian framework. Calibration helps improve the prediction of the QoI compared to the prediction using only the computation model outputs or using only observations. The diagram of the Bayesian cali-

bration process Fig. 2.6 gives us an overview of the whole process of Bayesian analysis which combines the information of priors, computer model, statistical model and observations to calibrate the parameter $\boldsymbol{\theta}$ to make predictions. The top right part shows that computer model data covers the whole time range of real observations and predictions. And the computer model data helps to estimate the covariance hyperparamter $\boldsymbol{\psi}_1$, which is defined as State 1. The left middle part shows the real observations helps to estimate the covariance hyperparameter $\boldsymbol{\psi}_2$, which is defined as Stage 2. The right bottom part shows the calibration and prediction which is called Stage 3. As we can see calibration parameters only exist in computer models, and we use real observations to calibrate the parameters in the computer model so that the prediction can be improved.

2.3.1 Likelihood of computer model

We need computation model outputs generated at various sets of $\boldsymbol{\theta}$ and $\boldsymbol{x}$ as the data for calibration. To avoid confusion, we use $\boldsymbol{\theta}^* = [\theta_1^* \ \theta_2^*]$ and $\boldsymbol{x}^*$ instead of $\boldsymbol{\theta}$ and $\boldsymbol{x}$, respectively, when they were used to generate computation model outputs. For example, at $(\boldsymbol{x}^*, \boldsymbol{\theta}^*)$, we obtain one computation model output $r(\boldsymbol{x}^*, \boldsymbol{\theta}^*)$. Computation model outputs which correspond to a set of various $(\boldsymbol{x}^*, \boldsymbol{\theta}^*)$ will be used in Bayesian calibration as a part of the data set. We call $\boldsymbol{\theta}^*$ calibration inputs and $\boldsymbol{x}^*$ variable inputs.

To illustrate the Bayesian calibration method [58, 77], we introduce necessary notations. Regarding the computation model outputs, let N be the total number of pairs of variable inputs and calibration inputs. Then, for the ith set of inputs, $(\boldsymbol{x}_i^*, \boldsymbol{\theta}_i^*)$, we let y_i be the corresponding computation model output, that is, $y_i = r(\boldsymbol{x}_i^*, \boldsymbol{\theta}_i^*)$. We further define $\boldsymbol{y} = [y_1 \ \cdots \ y_N]^T \in \mathbb{R}^{N \times 1}$ as the computation model output vector and $X_c = [(\boldsymbol{x}_1^*, \boldsymbol{\theta}_1^*)^T, \cdots, (\boldsymbol{x}_N^*, \boldsymbol{\theta}_N^*)^T]^T \in \mathbb{R}^{N \times 4}$ as the input matrix that corresponds to $\boldsymbol{y}$.
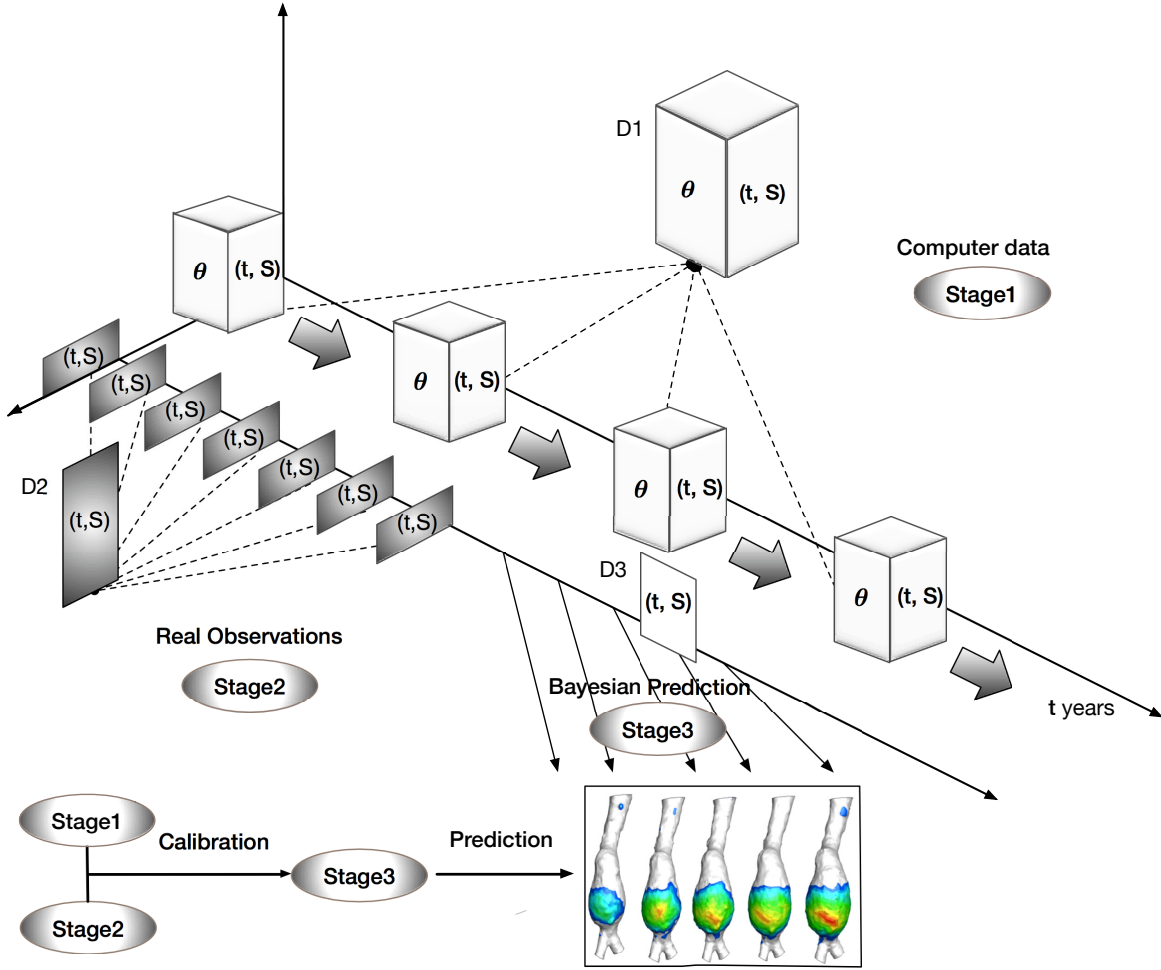
Figure 2.6. Diagram of the Bayesian calibration process.

Then, the assumption of the Gaussian process prior on the computation model gives

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{H}_1(X_c)\boldsymbol{\beta}_1^T, \boldsymbol{V}_1(X_c)),$$

where $\boldsymbol{H}_1(X_c) = [\boldsymbol{h}_1(\boldsymbol{x}_1^*, \boldsymbol{\theta}_1^*)^T, \cdots, \boldsymbol{h}_1(\boldsymbol{x}_N^*, \boldsymbol{\theta}_N^*)^T]^T$ and the (i, j) entry of $\boldsymbol{V}_1(X_c)$ is $\boldsymbol{k}_1((\boldsymbol{x}_i^*, \boldsymbol{\theta}_i^*), (\boldsymbol{x}_j^*, \boldsymbol{\theta}_j^*))$.

2.3.2 Likelihood of real observations

Let n be the number of observations and $\boldsymbol{z} = [z_1 \ \cdots \ z_n]^T \in \mathbb{R}^{n \times 1}$ be the set of observations corresponding to the variable input matrix $X_o = [\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_n^T]^T \in \mathbb{R}^{n \times 2}$. Note that X_o is not

necessarily the same as the set of variable inputs for the computation model outputs. In general, the number of observations is smaller than that of the computation model outputs since we can control the amount of the computation model outputs. To calibrate $\boldsymbol{\theta}$ from the observations, we augment variable inputs X_o with $\boldsymbol{\theta}$ such that $X_o(\boldsymbol{\theta}) = [(\boldsymbol{x}_1, \boldsymbol{\theta})^T, \cdots, (\boldsymbol{x}_n, \boldsymbol{\theta})^T]^T$. Then, from the calibration model, we have

$$\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{H}_1(X_o(\boldsymbol{\theta}))\boldsymbol{\beta}_1^T + \boldsymbol{H}_2(X_o)\boldsymbol{\beta}_2^T, \lambda \boldsymbol{I}_n + \boldsymbol{V}_1(X_o(\boldsymbol{\theta})) + \boldsymbol{V}_2(X_o)),$$

where

$$\boldsymbol{H}_1(X_o(\boldsymbol{\theta})) = [\boldsymbol{h}_1(x_1, \boldsymbol{\theta})^T, \cdots, \boldsymbol{h}_1(\boldsymbol{x}_n, \boldsymbol{\theta})^T]^T$$

and

$$\boldsymbol{H}_2(X_o) = [\boldsymbol{h}_2(x_1)^T, \cdots, \boldsymbol{h}_2(\boldsymbol{x}_n)^T]^T.$$

$\boldsymbol{V}_1(X_o(\boldsymbol{\theta}))$ is defined in a similar way to define $\boldsymbol{V}_1(X_c)$. The (i, j) entry of $\boldsymbol{V}_2(X_o)$ is $\boldsymbol{k}_2(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

2.3.3 Joint likelihood

We then combine the computation model outputs and observations, $\boldsymbol{d} = [\boldsymbol{y}^T \ \boldsymbol{z}^T]^T \in \mathbb{R}^{(N+n) \times 1}$, which we call a data vector.

$$\boldsymbol{d} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{m}_d(\boldsymbol{\theta}), \boldsymbol{V}_d(\boldsymbol{\theta})), \tag{2.9}$$

where

$$\boldsymbol{m}_d(\boldsymbol{\theta}) := \mathbb{E}(\boldsymbol{d}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\psi}) = \boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{\beta}^T,$$

with

$$\boldsymbol{H}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{H}_1(X_c) & \boldsymbol{0} \\ \boldsymbol{H}_1(X_o(\boldsymbol{\theta})) & \boldsymbol{H}_2(X_o) \end{pmatrix},$$

and

$$\boldsymbol{V}_d(\boldsymbol{\theta}) := \mathrm{var}(\boldsymbol{d}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\psi}) = \begin{pmatrix} \boldsymbol{V}_1(X_c) & \boldsymbol{C}_1(X_c, X_o(\boldsymbol{\theta}))^T \\ \boldsymbol{C}_1(X_c, X_o(\boldsymbol{\theta})) & \lambda \boldsymbol{I}_n + \boldsymbol{V}_1(X_o(\boldsymbol{\theta})) + \boldsymbol{V}_2(X_o) \end{pmatrix},$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix and $\boldsymbol{C}_1(X_c, X_o(\boldsymbol{\theta}))$ is a cross-covariance matrix whose (i, j) entry is $\boldsymbol{k}_1((\boldsymbol{x}_i^*, \boldsymbol{\theta}_i^*), (\boldsymbol{x}_j, \boldsymbol{\theta}_j))$. We note that $\boldsymbol{m}_d(\boldsymbol{\theta})$ and $\boldsymbol{V}_d(\boldsymbol{\theta})$ also depend on $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$. We drop them to reduce notational complexity.

Remark 2.3.1 *Recall that the calibration model assumes the true process as the sum of the computation model process and the model error process by assuming they are Gaussian processes as shown in (2.4) and (2.5). If we have observations of the true process only, we cannot identify these two Gaussian processes. However, the Bayesian calibration method makes use of both computation model outputs and observations as data. Computation model outputs contribute to the first Gaussian process only while observations contributes to both Gaussian processes as well as observation errors. Since the computation model outputs and observations contribute to two Gaussian processes differently, there is no identifiability issue.*

2.3.4 Calibration

To estimate $(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\psi})$ under the Bayesian framework, we consider the following prior distributions and assumptions.

A.1 $\boldsymbol{\beta}_1$ in (3.3) and $\boldsymbol{\beta}_2$ in (3.4) have non-informative priors, i.e., $p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \propto 1$.

A.2 $\boldsymbol{\theta}$ is independent of the other parameters.

A.3 $\boldsymbol{\theta}$ follows a normal distribution.

A.4 $\boldsymbol{\psi}_1$ and **$\boldsymbol{\psi}_2$** in (2.8) follow lognormal distributions.

A.5 $\log(\lambda)$ has a non-informative prior, i.e., $p(\log(\lambda)) \propto 1$.

Remark 2.3.2 *Note that based on **A.1** and **A.2**, we can get the joint prior distribution in the following form $p(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\psi}) \propto p(\boldsymbol{\theta})p(\boldsymbol{\psi})$. For **A.3**, we use the sample mean and variance of the calibration parameter inputs that were used to generate computation model outputs as mean and variance for normal prior density. **A.4** guarantees that $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ in (2.8) are positive values in the calculation since they are all hyperparameters in covariance functions.*

Together with the prior specification given above, we have the following full joint posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\psi})$ given $\boldsymbol{d}$:

$$
\begin{aligned}
p(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\psi} \mid \boldsymbol{d}) &\propto p(\boldsymbol{\theta})p(\boldsymbol{\psi})f(\boldsymbol{d}; \boldsymbol{m}_d(\boldsymbol{\theta}), \boldsymbol{V}_d(\boldsymbol{\theta})) \\
&\propto p(\boldsymbol{\theta})p(\boldsymbol{\psi})|\boldsymbol{V}_d(\boldsymbol{\theta})|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}\{(\boldsymbol{d} - \boldsymbol{m}_d(\boldsymbol{\theta}))^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1}(\boldsymbol{d} - \boldsymbol{m}_d(\boldsymbol{\theta}))\} \right],
\end{aligned}
\tag{2.10}
$$

where $f(\boldsymbol{d}; \boldsymbol{m}, V)$ is the multivariate Gaussian density function with mean m and variance V. Note that the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\psi})$ given $\boldsymbol{d}$ is proper even if we assume non-informative priors for $\boldsymbol{\beta}$.

To calibrate (estimate) $\boldsymbol{\theta}$ from the posterior distribution of $\boldsymbol{\theta} \mid \boldsymbol{d}$, we need to integrate out $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$. As shown explicitly in (2.9), $\boldsymbol{m}_d(\boldsymbol{\theta})$ depends on $\boldsymbol{\beta}$, while $\boldsymbol{V}_d(\boldsymbol{\theta})$ depends on $\boldsymbol{\psi}$. Clearly, $\boldsymbol{m}_d(\boldsymbol{\theta})$ is a linear function of $\boldsymbol{\beta}$, so we can find

$$
\boldsymbol{\beta} \mid \boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{d} \sim \mathcal{N}(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{W}(\boldsymbol{\theta})),
\tag{2.11}
$$

where

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{W}(\boldsymbol{\theta})\boldsymbol{H}(\boldsymbol{\theta})^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}\boldsymbol{d},$$

$$\boldsymbol{W}(\boldsymbol{\theta}) = (\boldsymbol{H}(\boldsymbol{\theta})^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}\boldsymbol{H}(\boldsymbol{\theta}))^{-1}.$$

Thus, we can integrate out $\boldsymbol{\beta}$ in (2.10) and get

$$
\begin{aligned}
p(\boldsymbol{\theta}, \boldsymbol{\psi} \mid \boldsymbol{d}) \propto{}& p(\boldsymbol{\theta})p(\boldsymbol{\psi})|\boldsymbol{V}_d(\boldsymbol{\theta})|^{-\frac{1}{2}}|\boldsymbol{W}(\boldsymbol{\theta})|^{\frac{1}{2}} \\
&\times \exp\left[-\frac{1}{2}\{(\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}(\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))\}\right].
\end{aligned}
\tag{2.12}
$$

While $\boldsymbol{\beta}$ is integrated out easily, $\boldsymbol{\psi}$ is not. Numerical integration can be considered but may not give an accurate result since it is the multi-dimensional integration. Thus, we use the conditional posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{d}$ and a plausible estimate of $\boldsymbol{\psi}$ plugged in, which is proposed by Kennedy and O'Hagan (2001) [58]. The plausible estimate of $\boldsymbol{\psi}$ is obtained in a two-step approach. First, we estimate $\boldsymbol{\psi}_1$ by maximizing $p(\boldsymbol{\psi}_1|\boldsymbol{y})$ (Stage 1). Then, we estimate $\boldsymbol{\psi}_2$ by maximizing $p(\boldsymbol{\psi}_2|\boldsymbol{d}, \boldsymbol{\psi}_1)$ after plugging in the estimated $\boldsymbol{\psi}_1$ (Stage 2). This process is shown in Fig. 2.7. In this study, numerical maximization is used with 500 iterations for Stage 1 and 100 iterations for Stage 2.
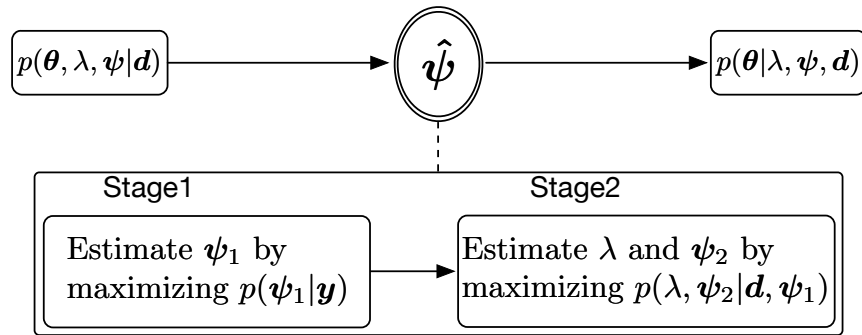
Figure 2.7. Two stages to estimate $\boldsymbol{\psi}$.

Having estimated $\boldsymbol{\psi}$ and plugging it into the distribution of $\boldsymbol{\theta} \mid \boldsymbol{d}, \boldsymbol{\psi}$, we obtain the

posterior distribution of the calibration parameter $\boldsymbol{\theta}$ to be

$$p(\boldsymbol{\theta} \mid \hat{\boldsymbol{\psi}}, \boldsymbol{d}) \propto p(\boldsymbol{\theta})|\boldsymbol{V}_d(\boldsymbol{\theta})|^{-\frac{1}{2}}|\boldsymbol{W}(\boldsymbol{\theta})|^{\frac{1}{2}}$$
$$\times \exp\left[-\frac{1}{2}\{(\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1}(\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))\}\right]$$

(2.13)

and use (2.13) to calibrate and make an inference of $\boldsymbol{\theta}$.

2.3.5 Prediction

We introduce another set of variable inputs $X_p = [\boldsymbol{x}_{n+1}^T, \cdots, \boldsymbol{x}_{n+m}^T]^T \in \mathbb{R}^{m \times 2}$ to be used for

prediction. The corresponding prediction of the QoI at given variable inputs X_p is denoted

as $\boldsymbol{P} = [P_1 \ \cdots \ P_m]^T \in \mathbb{R}^{m \times 1}$. For quality of the prediction, we can consider variable inputs

in X_c that cover X_p.

Under the Bayesian framework, a prediction can be made using the predictive distribu-

tion of the (unobserved) true process $\zeta(\boldsymbol{x})$ given full data $\boldsymbol{d}$. More specifically, we consider

the predictive distribution of $\zeta(\boldsymbol{x})$ given $\boldsymbol{d}$ and $\hat{\boldsymbol{\psi}}$ since we fix $\boldsymbol{\psi}$ using the estimate $\hat{\boldsymbol{\psi}}$ in

our approach. We first analytically obtain the distribution of $\zeta(\boldsymbol{x}) \mid \boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{d}, \boldsymbol{\beta}$, which is

conditionally normal since $\zeta(\boldsymbol{x})$ and $\boldsymbol{d}$ are normally distributed. Using the laws of total ex-

pectation and total variance to integrate out $\boldsymbol{\beta}$, we obtain the distribution of $\zeta(\cdot)$ conditional

on $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\psi}}$, which is also normal. Therefore, its mean function is given by

$$\mathbb{E}(\zeta(\boldsymbol{x})|\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta})^T \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1}(\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})), \qquad (2.14)$$

where

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{h}_1(\boldsymbol{x}, \boldsymbol{\theta}) \\ \boldsymbol{h}_2(\boldsymbol{x}) \end{pmatrix},$$

and

$$v(x, \theta) = \begin{pmatrix} V_1((x, \theta), X_o) \\ V_1\{(x, \theta), X_o(\theta)\} + V_2(x, X_o) \end{pmatrix}.$$

Additionally, its covariance function is given by

$$\begin{aligned}
\text{cov}(\zeta(x), \zeta(x') \mid \theta, \hat{\psi}, d) = {} & k_1((x, \theta), (x', \theta)) + k_2(x, x') \\
& - v(x, \theta)^T V_d(\theta)^{-1} v(x', \theta) + (h(x, \theta) \\
& - H(\theta)^T V_d(\theta)^{-1} v(x, \theta))^T W(\theta)(h(x', \theta) \\
& - H(\theta)^T V_d(\theta)^{-1} v(x', \theta)).
\end{aligned} \tag{2.15}$$

We then can obtain the predictive distribution of $\zeta(x)$ given d and $\hat{\psi}$ by integrating θ out from the distribution of $\zeta(x) \mid \theta, \hat{\psi}, d$ with respect to the posterior distribution of θ given in (2.13). From (2.14) and (2.15), we obtain the predictive expectation and variance of $\zeta(\cdot)$ evaluated at inputs X_p as follows:

$$\mathbb{E}[\zeta(X_p) \mid \hat{\psi}, d] = \mathbb{E}_\theta\{\mathbb{E}[\zeta(X_p) \mid \theta, \hat{\psi}, d]\}, \tag{2.16}$$

and

$$\text{var}[\zeta(X_p) \mid \hat{\psi}, d] = \mathbb{E}_\theta\{\text{var}[\zeta(X_p) \mid \theta, \hat{\psi}, d]\} + \text{var}_\theta\{\mathbb{E}[\zeta(X_p) \mid \theta, \hat{\psi}, d]\}, \tag{2.17}$$

where $\zeta(X_p) = [\zeta(x_{n+1}) \cdots \zeta(x_{n+m})]^T$.

Based on the posterior density (2.13) of the calibration parameter θ, we can draw a bunch of Markov Chain Monte Carlo (MCMC) samples for θ and thus predictive expectation and

variance (2.16) and (2.17) can be approximated as follows.. First, we draw M samples $\{\boldsymbol{\theta}(1), \cdots, \boldsymbol{\theta}(M)\}$ from (2.13) using the Metropolis-Hastings algorithm [15] and plug these samples into (2.14) and (2.15). Next, we get the predictions $\boldsymbol{P}$, which are the MCMC estimates of (2.16) written as

$$\boldsymbol{P} := \widehat{\mathbb{E}}[\boldsymbol{\zeta}(X_p) \mid \hat{\boldsymbol{\psi}}, \boldsymbol{d}] = \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}(i), \hat{\boldsymbol{\psi}}, \boldsymbol{d}]. \tag{2.18}$$

Similarly, the outermost expectation and variance on the right hand side of (2.17) are estimated by their corresponding sample mean and sample variance with respect to $\boldsymbol{\theta}$. Then, we obtain the predictive variances $\boldsymbol{V}_p$ as

$$\boldsymbol{V}_p := \widehat{\mathrm{var}}[\boldsymbol{\zeta}(X_p) \mid \hat{\boldsymbol{\psi}}, \boldsymbol{d}] = \widehat{\mathbb{E}}_{\boldsymbol{\theta}}\{\mathrm{var}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\} + \widehat{\mathrm{var}}_{\boldsymbol{\theta}}\{\mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\}, \tag{2.19}$$

where

$$\widehat{\mathbb{E}}_{\boldsymbol{\theta}}\{\mathrm{var}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\} = \frac{1}{M} \sum_{i=1}^{M} \mathrm{var}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}(i), \hat{\boldsymbol{\psi}}, \boldsymbol{d}],$$

$$\widehat{\mathrm{var}}_{\boldsymbol{\theta}}\{\mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\} = \frac{1}{M-1} \sum_{i=1}^{M} \{\mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}(i), \hat{\boldsymbol{\psi}}, \boldsymbol{d}] - \widehat{\mathbb{E}}[\boldsymbol{\zeta}(X_p) \mid \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\}^2.$$

We will construct the 95% credible intervals based on these predictions $\boldsymbol{P}$ and their predictive variances $\boldsymbol{V}_p$.

2.4 Data Description

We consider two cases of data sets. In the first case, we use the G&R computation model of the AAA expansion to generate synthetic observations in addition to producing computation

model outputs. This is the simulated observation case that validates our approach since we know the realized true process along with the parameters that generate it for comparison. In the second case, we use real data, i.e., observations from a patient's CT scan images. The G&R computation model is used only to obtain computation model outputs in this real-world clinical application. When implementing Bayesian analysis, we standardize all the inputs x, θ and outputs y, z to stabilize computation. In particular, we use the notation $\theta = [\theta_1 \ \theta_2]$ to denote the standardized ones while we use $\theta^{org} = [\theta_1^{org} \ \theta_2^{org}]$ to denote original values for the rest of the paper.

2.4.1 Simulated Observation Cases

For the simulated observation case, we generate $r(x, \theta)$ from the G&R computation model. To validate our approach, we realize and set patient specific values on the calibration parameters (say θ_0) for the synthesized true process so that we evaluate the estimated calibration parameters. Recall that for the damage function, we need to set θ_1, θ_2, α_1, and α_2 to produce the G&R computation model outputs. For the simulated observation case, we set $\theta_1^{org} = 0.65$, $\theta_2^{org} = 6$, $\alpha_1 = 7.5$ and $\alpha_2 = 2$. Thus, the calibration input is $\theta_0^{org} = [0.65 \ 6]$ for the true process. For the variable inputs $x = [t \ s]$, we consider 8 equally spaced s from 0 to 15 cm and a time step of 5 days, which are fixed throughout the study. The time duration is 7 years, but several different choices of sampling times are considered. Note that when we implement Bayesian calibration, the values of the QoI (the radius with respect to the centerline) are standardized as well so that its mean is 0 and its variance is 1.

Once we obtain a realization of $r(x, \theta_0)$, we add model and observation errors to get the final observations. The model error, $\delta(x)$, is realized from the model assumption (2.5) with $\beta_2 = [0.001 \ 0.001]$, $\omega_{21} = 1$, $\omega_{22} = 1$, and $\sigma_2^2 \in \{0.005, 0.001\}$. Then the standard deviation

of model error $\delta(\boldsymbol{x})$ on each height at each time is about $\{0.071, 0.032\}$. The observation error $\boldsymbol{\varepsilon}$ is generated from $\mathcal{N}(0, \lambda)$ with $\lambda = 0.001$, i.e., $\epsilon \sim \mathcal{N}(0, 0.032^2)$. The standard deviation of standardized computation outputs r is 1. To generate simulated observations, we add a model error process of 7.1% or 3.2% (with respect to r) and an observation error process of 3.2%. We chose these values for the model and observation errors in order to produce a data set that best illustrates the effects of different noise levels and sampling schemes.

We also need computation model outputs at various combinations of variable inputs and calibration inputs. For calibration inputs, we consider $\boldsymbol{\theta}^{org} = [\theta_1^{org} \ \theta_2^{org}] \in \{0.5, 0.65, 0.8\} \times \{2, 6, 10\}$ so that there are total 9 combinations of θ_1 and θ_2. For variable inputs, s will be considered at 8 equally spaced values from 0 to 15 cm. For time t, we consider three different scenarios to see the effects of different time resolutions. With two choices of σ_2^2, there are a total of 6 different scenarios. Table 2.1 gives the label of each scenario. Further details of each scenario are given in Table 2.2. With three cases of sampling time grids and the two levels of σ_2^2, we want to investigate the behavior of interaction between the computation model and Bayesian calibration. In Case 1, computation model inputs have a sparse time grid while those in Case 3 have a denser one. In Case 2, the computation model inputs do not cover the full time range we want to predict while those in Case 1 and Case 3 do.

We note that variable inputs and calibration inputs for the proposed Bayesian calibration are standardized in the actual implementation so that we can assume the realized calibration inputs are centered at zero. Then, we use 0.3 as prior means of θ_1 and θ_2 (away from the zero) to investigate the robustness of prior distributions. In addition, we consider 0.1 for prior variances of θ_1 and θ_2. The same setting for the prior of θ_1 and θ_2 is used later in the real observation case. Note that the normal prior can be used for θ_1 that is constrained to

the interval $[0, 1]$. As θ_1 was standardized (mean is 0, standard deviation is 1), the range of θ_1 becomes $[-1.3, 1.3]$ based on the parameter values we use. We set the prior of mean at 0.3, and the variance of prior at 0.1 (standard deviation is 0.32). According to the empirical rule, 99.7% of the samples from the prior will lie in the interval $[-0.66, 1.26]$ which contains in the interval $[-1.3, 1.3]$. The sensitivity of prior choice on the posterior estimates of θ_1 and θ_2 is investigated and presented in Supplemental Fig. A.3 (in Supplementary Section Appendix 2), where we can see that the posterior distributions are robust against the subjective choice of prior distributions.

Table 2.1. Differences in X_c and σ_2^2 for simulation cases

sampling time in X_c	$\sigma_2^2 = 0.005$	$\sigma_2^2 = 0.001$
every 2 years	Case 1a	Case 1b
every 1.5 years	Case 2a	Case 2b/2c
every 1 year	Case 3a	Case 3b

2.4.2   Real Observation Case

For the real observation case, we consider 4 longitudinal CT scan images of a patient as shown in Fig. 2.1. This particular patient's CT scan examination spans a period of 3 years. For each image, we measured the corresponding radius (QoI) at each height on the centerline as illustrated in Fig. 2.5 using the maximally-inscribed sphere method developed in [40]. From the preliminary study of the damage shape with respect to the CT scan images, we selected α_1 and α_2 such that $\alpha_1 = 12.6$ and $\alpha_2 = 5$ for (2.1) *a-priori*. In addition, we selected the reasonable range of the calibration parameter $\boldsymbol{\theta}$ *a-priori* to produce comparable computation model outputs for the particular observations (Fig. 2.5). More specifically, as prior information to determine the range of the calibration parameter $\boldsymbol{\theta}^{org}$, we fit the diameter of the computation model outputs to the diameter calculated from CT scan images using `fminsearch` in MAT-

LAB (Mathwork, Natick, USA) and found the optimum value to be $[\theta_1^{org}\theta_2^{org}] = [0.2\ 0.9]$.

From this information, we consider $\boldsymbol{\theta}^{org} = [\theta_1^{org}\ \theta_2^{org}] \in \{0.05, 0.2, 0.35\} \times \{0.7, 0.90, 0.99\}$.

For these 9 combinations of $\boldsymbol{\theta}^{org}$, we run the G&R model to produce the computation model

outputs. More details for this real observation case are given in Table 2.2.

Table 2.2. Case details

	Case 1a	Case 1b	Case 2a	Case 2b/2c	Case 3a	Case 3b	Real Data
X_c times	[0 2 4 6]		[0 1.5 3 4.5 6]		[0 1 2 3 4 5 6 7]		[1 2 3 4]
X_c dimension	288×4		360×4		576×4		288×4
X_o times	[0 0.5 1 1.5 2 2.5 3]		[0 0.5 1 1.5 2 2.5 3 3.5 4]				[1 2]
X_o dimension	56×2		72×2				16×2
X_p times	[3.5 4 4.5 5 5.5 6]		[4.5 5 5.5 6 6.5 7]				[3 4]
X_p dimension	48×2		48×2				16×2
$\boldsymbol{y}$ dimension	288×1		360×1		576×1		288×1
$\boldsymbol{z}$ dimension	56×1		72×1				16×1
σ_2^2	0.005	0.001	0.005	0.001	0.005	0.001	0.01
$\boldsymbol{\delta}$ % of $\boldsymbol{r}$	7.2%	3.2%	7.2%	3.2%	7.2%	3.2%	10%

2.5   Results

In this section, we present results of the Bayesian calibration method on simulated observa-

tions and real observations as described in Section 2.4. For detailed implementation steps,

readers are referred to Section Appendix 1 in the supplementary document.

For simulated observations, we show three plots, prediction results with true processes

(Fig. 2.8), 95% credible bands with true processes (Fig. 2.9) and relative errors for

prediction (Fig. A.1). Relative errors $\boldsymbol{R} = [R_1 \cdots R_m]^T$ over the height of the AAA are

calculated as

$$R_i = \frac{\mid P_i - T_i \mid}{T_i}, \quad \forall i \in \{1, \cdots, m\}, \tag{2.20}$$

where $T = [T_1 \cdots T_m]^T$ is the realized true process for the simulated observation case that

is described in Section 2.4.1.

For the real observations, the prediction and credible band graphs are given together in

Fig. 2.11. Since only observations are available for this case, relative errors for prediction are calculated by replacing $\{T_i\}$ in (2.20) with the observations (i.e., noisy true values) as shown in Fig. A.2b.

2.5.1   Results from Simulated Observation Cases

From Figs. 2.8, 2.9, and A.1 of the simulation study, we observe the following findings. Prediction quality improves when more computation model outputs are used for calibration. When we consider different sampling time resolutions, prediction quality monotonically improves from Case 1 (coarse resolution) to Case 3 (fine resolution). Smaller model errors (Cases 1b, 2b, and 3b) provide better prediction results. Compared to Cases 1a, 2a, and 3a (left column of each plot), prediction quality from Cases 1b, 2b, and 3b (right column of each plot) is better. As one can expect, prediction quality decreases at the year for which the computation model outputs or observations are not available for calibration. For example, at years 4.5, 5, and 5.5, the errors between the true processes and predictions are more pronounced in Case 1a since both computation model outputs and observations are not available at those years. In particular, predictions at years 6.5 and 7 for Case 2 are the worst in terms of relative errors. Such results are reflected in wide credible bands as well. A possible reason is that calibration for Case 2 did not have information at years 6.5 or later to predict at years 6.5 and 7. Recall that computation model outputs are up to year 6 and observations are up to year 4. On the other hand, calibration for Case 1 has computation model outputs at year 6 to predict at years 4.5, 5 and 5.5. This result suggests that we can use computation model outputs at future times during calibration for better prediction. In this regard, results from Case 3 are better than Cases 1 and 2. However, more computation model outputs in the calibration process implies higher computational burden as shown in

(a) Case 1a

(b) Case 1b

(c) Case 2a

(d) Case 2b

(e) Case 3a

(f) Case 3b
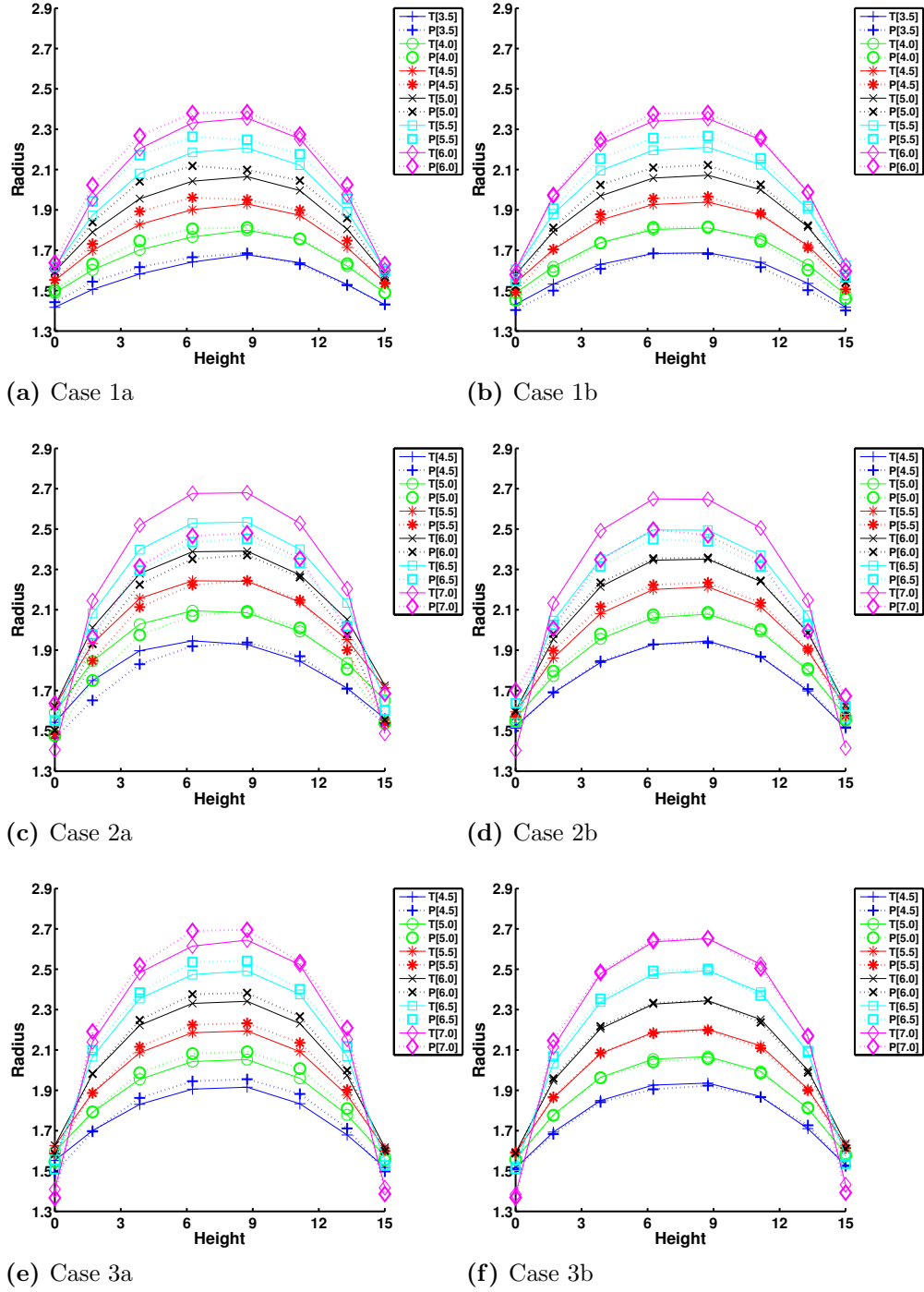
Figure 2.8. Predictions of the true processes. $T[t]$ denotes the true values at year t, while $P[t]$ denotes the predictions at year t. The solid line denotes the true values, while the dashed line denotes the predictions.

Table 2.4.

From the Bayesian calibration, we have posterior samples of calibration parameters as

(a) Case 1a

(b) Case 1b

(c) Case 2a

(d) Case 2b

(e) Case 3a

(f) Case 3b

Figure 2.9. 3D 95% credible band of predictions. The blue stars denote the true values lying inside the credible band. The red marks denote the true values lying outside the credible band.

well as hyperparameters. For the simulated observation case, true values of parameters (values that were used to generated the data) are known so that we can compare the per-

formance of Bayesian calibration for various simulation scenarios. The calibrated estimates of standardized θ_1 and θ_2 for each case are provided in Table 2.3. Posterior densities of standardized θ_1 and θ_2 are shown in Supplementary Fig. A.3. The prior distributions and estimates of hyperparameters are shown in Supplementary Tables A.1–A.8.

Comparing the posteriors of θ_1 and θ_2 in all cases, we clearly notice that the posteriors of Case 3 have the sharpest peaks around the true values, which can be utilized as point estimators. As shown in Table 2.3, the posterior estimates of θ_1 and θ_2 in Case 3b are the closest to their true values 0 and 0 among all cases. This also illustrates that more information from computation model outputs can give more accurate estimates of the calibrated parameters. Consequently, better estimation of calibrated parameters is likely to give us better quality of predictions.

(a) Predictions of the true processes. $T[t]$ denotes the true values at year t, while $P[t]$ denotes the predictions at year t. The solid line denotes the true values, while the dashed line denotes the predictions.

(b) 3D 95% credible band of predictions. The blue stars denote the true values lying inside the credible band. The red marks denote the true values lying outside the credible band.

Figure 2.10. Predictions and credible bands for Case 2c.

2.5.1.1 Robustness of prior selection

To investigate the impact of different priors on Bayesian calibration, we compare the results of Case 2b and Case 2c, where Case 2c is the same as Case 2b except the prior variances of calibration parameters are 10 times larger.

The posteriors of θ_1 and θ_2 for Case 2c (Fig. A.4a, in the supplementary document) cover the prior mean values 0 and 0, but they have larger posterior variances than those from Case 2b due to the larger prior variances. We see that the predictions (Figs. 2.8d and 2.10a) and the 95% credible bands (Figs. 2.9d and 2.10b) are similar even though their prior variances are quite different. This can be seen also in relative errors (Figs. A.1d and A.2a). This implies that our Bayesian calibration method is robust to such changes in priors. Therefore, we may use a diffuse prior that can cover all the possible values of parameters $\boldsymbol{\theta}$ by learning the statistical information from the background and the computation model. One can choose the priors for the real observation case in a similar way.

2.5.1.2 Computation time

In this study, computation model outputs for calibration were generated separately. Calibration and prediction were implemented using the computational model outputs together with observations. Our method requires computation model outputs for each set of calibration parameters. Thus, we provide computation time to generate computation model outputs per each set of calibration parameters. The time to generate computation outputs by solving the G&R computation model is 10 minutes for one combination of parameters (θ_1 and θ_2), producing outputs in 8 grid points of heights (total height is 15 cm) and every 5 days out of total 2750 days. The G&R computation has been coded and implemented using MATLAB R2012a on an Intel Core i7 3770 3.4 GHz Processor with 12 GB of RAM.

If we generate a large number of computation model outputs for densely sampled calibration parameter values, the accuracy of calibration and prediction will improve at the expense of the computational cost. Note that this computation is naturally *parallel* since we run the computation model on a given parameter vector, the procedure of which is repeated to cover a defined range of parameter combinations. As the computation model becomes increasingly complicated, this time will also increase. The total computation will still remain parallel.

In Table 2.4, we show the computation time for Bayesian calibration and prediction. Case 1, Case 2, and Case 3 took around 6.5, 12, and 30 hours, respectively. As expected, Case 3 takes the longest time yielding the best prediction results. To provide more insight, we divide the total computational time into two parts in Table 2.4: one spent in optimization to find ψ_1 and ψ_2, the other one spent in calculation of estimation, prediction, and uncertainty via sampling. As can be seen in Table 2.4, the latter part takes more time due to the sequential computation of Bayesian analysis steps. All computations for Bayesian analysis were implemented on a twelve-core CPU 2.90GHz running Windows Server 2008 using R software (R version 3.1.1 Copyright (C) 2014 The R Foundation for Statistical Computing). We used the package "calibrator" [45] in R with appropriate modifications in order to adopt our application and also to reduce computation time significantly (e.g., 100 times faster). The results from Cases 1-3 illustrate that if affordable, the best results are achieved when X_c has a dense time grid covering the whole targeted prediction range of interest.

2.5.2   Results from Real Observation Case

In this section, we discuss the calibration results from real observations using CT images 1-4 as shown in Fig. 2.5, i.e., (CT images taken at 0, 1.2, 2.3, and 3.2 years, respectively). To generate computation model outputs to be used for calibration, we consider the inputs X_c

42

Table 2.3. Posterior estimates of θ_1 and θ_2

Case	$\theta_1^{org}(mean, std^1)$	$\theta_2^{org}(mean, std)$	$\theta_1(mean, std)$	$\theta_2(mean, std)$
1a	$(0.678, 0.035)$	$(6.05, 0.105)$	$(0.225, 0.288)$	$(0.015, 0.032)$
1b	$(0.668, 0.023)$	$(5.912, 0.147)$	$(0.149, 0.281)$	$(-0.027, 0.045)$
2a	$(0.681, 0.035)$	$(6.072, 0.105)$	$(0.249, 0.079)$	$(0.022, 0.032)$
2b	$(0.672, 0.024)$	$(6.013, 0.080)$	$(0.177, 0.195)$	$(0.004, 0.025)$
3a	$(0.669, 0.016)$	$(5.997, 0.056)$	$(0.155, 0.127)$	$(-0.001, 0.017)$
3b	$(0.660, 0.010)$	$(6.003, 0.040)$	$(0.083, 0.084)$	$(0.001, 0.012)$
2c	$(0.665, 0.032)$	$(5.992, 0.115)$	$(0.122, 0.264)$	$(-0.003, 0.035)$
Real	$(0.250, 0.096)$	$(0.858, 0.062)$	$(0.406, 0.784)$	$(-0.041, 0.515)$

[1] std denotes the standard deviation.

Table 2.4. Computation times for Bayesian calibration and prediction. Estimation (time) describes the time spent in maximizing to estimate ψ_1 and ψ_2, while sampling (time) describes the time spent in calculating predictions and credible intervals via sampling from posteriors.

Case	Estimation	Sampling	Case	Estimation	Sampling
1a	1.177 hours	5.305 hours	1b	1.195 hours	5.160 hours
2a	2.828 hours	8.785 hours	2b	2.796 hours	8.941 hours
3a	7.314 hours	24.377 hours	3b	7.221 hours	24.240 hours
2c	2.875 hours	9.044 hours	Real	8.237 mins	2.998 hours

for a computation model as follows. Eight heights uniformly ranging from 9.3 to 17.5 are chosen. For calibration inputs, θ_1^{org} is chosen from $\{0.05, 0.2, 0.35\}$ and θ_2^{org} is chosen from $\{0.7, 0.90, 0.99\}$. 500 iterations are used for Stage 1 and 100 iterations are used for Stage 2. The computation time is 30 minutes as shown in Table 2.4.

Table 2.3 shows the estimates of all calibration parameters. The prior means of θ_1^{org} and θ_2^{org} are 0.2 and 0.863, and their prior standard deviations are 0.150 and 0.148, respectively. The resulting posterior means of θ_1^{org} and θ_2^{org} from Bayesian analysis are 0.250, and 0.858, and their standard deviations are 0.096 and 0.062, respectively.

We withhold observations for the last two years during the calibration procedure in order to validate our approach by comparing predictions on the last two years (CT images 3 and 4, taken at 2.3 and 3.3 years, respectively). In contrast to the simulation study, we compare the results with observations (noisy true values) since the true process is not available in

the real observations. The difference between the true process and the observation (the measurement error) is estimated to be $\mathcal{N}(0, 0.01)$ from the real observation data used for Bayesian calibration as shown in Table 2.2. There are some larger discrepancies in the lower part of predictions at 3.2 years, which can be shown by the corresponding larger relative errors in Fig. A.2b. Most of the relative errors are less than 5%. Besides, Figure 2.11a shows how close the Bayesian calibration model predicts the observation (i.e., noisy true) values. From Fig. 2.11b, we also find that only one observation value lies outside of the credible bands. These results support that our approach has a capability to predict AAA expansion using real data.
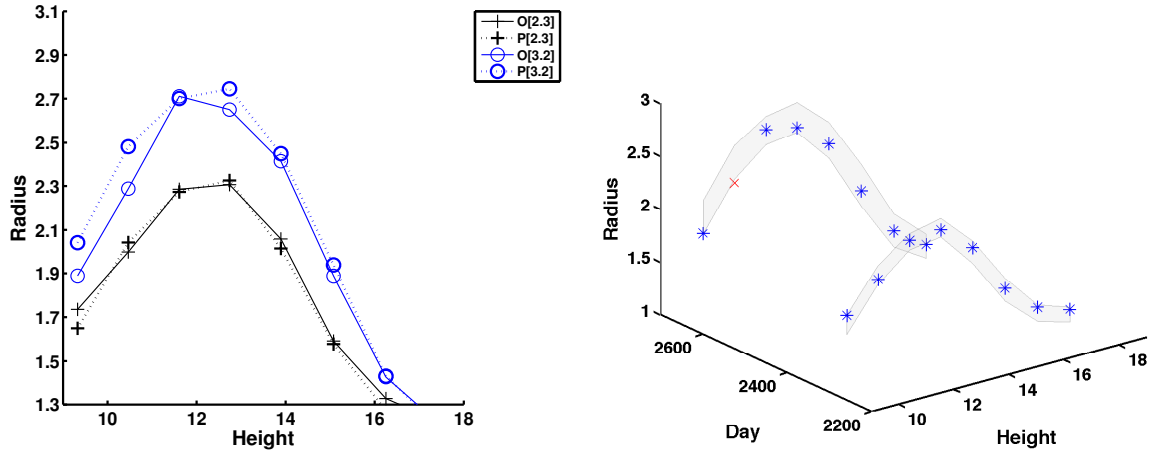
(a) Predictions of the observation processes. $O[t]$ denotes the observation values at year t, while $P[t]$ denotes the predictions at year t. The solid line denotes the observations, while the dashed line denotes the predictions.
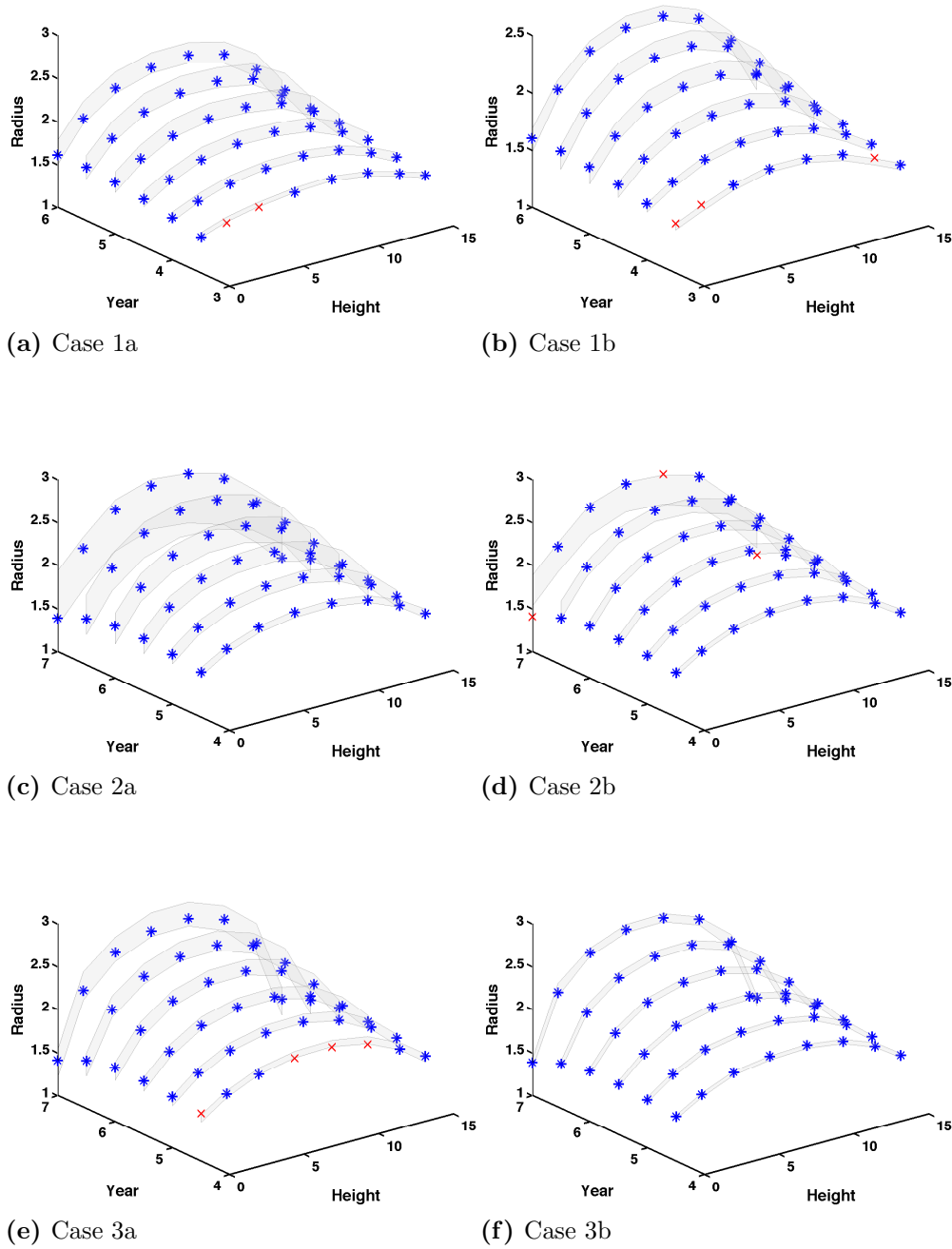
(b) 3D 95% credible band of predictions. The blue stars denote the observations lying inside the credible band. The red marks denote an observation value lying outside the credible band.

Figure 2.11. Predictions and credible bands for Real observation.

2.6 Discussion and Conclusion

From prediction graphs in Figs. 2.8e and 2.8f of Case 3 (simulation study), we notice that the diameters of proximal and distal ends of the true line at year 7 tend to be smaller than the

previous ones. When the AAA is gradually expanding, the end of both regions can be axially compressed, which means the radii of the two ends are often contracted. The computation model successfully captures this phenomenon. Currently we are studying longitudinal images of abdominal aortic aneurysms, registered with the vertebral column. We speculate that the renal vein and artery, superior mesenteric artery, and iliac bifurcation can serve as an anchor (both longitudinal and circumferential directions at the superior and inferior boundaries) to the infrarenal AAA during expansion. Hence, it may be possible that the physical constraints of the tethering of those vessels provide a strong confinement, or an anchor at the region of the aorta. During the AAA expansion, the volume of the AAA's sac will gradually increase while stretching mostly in the circumferential direction and slightly in the axial direction simultaneously. Hence, because of the AAA expansion and the confinement in axial direction, the neck and distal part of aorta (from renal branches to the AAA's sac) will be compressed in the axial direction. Using the 3D growth and remodeling simulation, Zeinali et al. (2012) [115] show the local change in the stress distribution, in which the stress of the sac is increased but the neck's stress is decreased.

Recall that we estimated α_1 and α_2 *a-priori* before Bayesian calibration since we assume that the peak location of the future AAA and its overall geometry do not change significantly from the previous scans. As illustrated in the real observation case, the AAA peak location and the aneurysm shape did not significantly change during the follow-up images, which support our approach.

At present, implementing a full Bayesian analysis especially in a large data case is not practical since Stage 2 (the process of estimating ψ_2) in Fig. 2.7 is time consuming even for one iteration. If we adopt a full Bayesian method, e.g., Gibbs sampling, we need to update ψ_2 in every iteration. In this case, the computational time will quickly accumulate to a

non-feasible numerical solution. This is the main reason to use point estimators ψ_1 and ψ_2 instead of using a full Bayesian method. The case studies presented in the paper using our approach with point estimators show this method's feasibility and good performance.

The estimates of hyperparameters in all cases are shown in Supplementary Tables A.1– A.8. Estimates of β_{11} (the coefficient for time t) are always positive, which means the radii of AAAs increase in time due to AAA expansion. Estimates of β_{12} (the coefficient for θ_1) are also positive in all cases, which means the QoI, i.e., the radii of AAA increases as θ_1 increases. This can be explained by the damage function (2.1). The QoI increases while the elastin contents decreases as the amplitude of the damage function (e.g., θ_1) increases. Estimates of β_{13} (the coefficient for θ_2) are negative in the real observation case while all β_{13}'s are positive in the simulation cases. This could be due to the fact that θ_2 values are largely different between real and simulation observations. In particular, θ_2 samples values from $\{0.7, 0.9, 0.99\}$ in the real observation case while θ_2 samples values from $\{2, 6, 10\}$ in the simulation cases (with 6 being the true value). The largely different parameters values could give different local sensitivities of the QoI with respect to parameters when the influence of the parameters is rather complex. On the other hand, the estimates of $\boldsymbol{\beta}_{21}$, $\boldsymbol{\beta}_{22}$ and σ_2^2 in all cases are small. This implies that the computation model could explain most of the linear and covariance structures of the true process.

The results from the simulation case study suggest that Bayesian calibration may be used to combine the computation model, prior and uncertainty models, and real observations to predict the QoI at future times or at unobserved locations. Computation model output data provide a deterministic trend of the expansion process, which is modeled by a Gaussian process. Additionally, computation model discrepancies obtained from real observations are modeled by another Gaussian process. When we have more information from the computa-

tion model (in the form of finer grids for inputs X_c), we achieve lower prediction errors, and the posteriors of parameters θ_1 and θ_2 are more likely to concentrate on their true values as illustrated in our simulation results. We also find that posteriors and predictions from our approach are robust to the selection of priors for θ_1 and θ_2. In the real observation case of one patient, the results of Bayesian calibration indicated that the predictions are reasonably good when compared with the unused last two observations. Most of the unused observations match reasonably well with predictions and lie inside the 95% credible bands. The model and the observation errors collectively capture the structure of the true process in a consistent manner from a Bayesian perspective.

Considering these results, we conclude that the Bayesian calibration process is capable of predicting complex G&R AAA processes. We also provide evidence that this holds true for a set of real longitudinal observations, where Bayesian calibration produces good predictions. However, there is a need for validating our approach with a large number of real observation cases to evaluate its performance and efficacy in a clinical sense. We believe that for given past CT images of a patient, Bayesian calibration can help to guide the scheduling of future CT scans according to predictions with the credible bands.

We also used the 2D axisymmetric G&R model ignoring the time evolution of the centerline in order to simplify the formulation of the Bayesian calibration process and use only two parameters to study the effectiveness of Bayesian calibration. As future work, we plan to consider complicated damage functional shapes with more parameters as well as asymmetric AAA expansion in a 3D space. Additionally, the linearity assumption on the mean functions for Gaussian processes in (2.5) will be relaxed to be more flexible for better prediction performance. We also plan to investigate how to deal with high-dimensional calibration parameters when flexible and complex functions are used for damage and the rate of mass

production. Finally, we will investigate how to incorporate other available patient-specific data for Bayesian calibration of a particular patient case.

2.7 Data Accessibility

More information about the results is summarized in the supplementary document in the forms of tables and figures. The real observation data set used for the real observation case in this paper (Fig. 2.5) is available in the websites of the corresponding authors (http: //www.egr.msu.edu/~jchoi/files/papers/index.html and http://www.egr.msu.edu/~sbaek/ Publications.html).

APPENDICES

Appendix 1   Programming steps for simulated observations

Figure 2.6 shows us the whole process of Bayesian analysis which combines the information of priors, computer model, statistical model and observations to calibrate the parameter $\boldsymbol{\theta}$ to make predictions. To be more specific, the programming steps for the simulation study are given as follows. The sections, tables, equations referred here are from the main paper.

Step 0: Specify the statistical models for $\boldsymbol{m}_1(\boldsymbol{x}, \boldsymbol{\theta})$ and $\boldsymbol{m}_2(\boldsymbol{x})$ (See Sections 2.(d)).

Step 1: Set the initials including true values of $\boldsymbol{\psi}_2$ and $\boldsymbol{\beta}_2$, normal priors of $\boldsymbol{\theta}$, log-normal priors of $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$. See prior assumptions in Section 3.

Step 2: Determine inputs X_c, X_o and X_p as described in Table 2.

Step 3: Simulate the full data $\boldsymbol{d} = [\boldsymbol{y}^T \ \boldsymbol{z}^T]^T$ by using X_c, X_o and true values of $\boldsymbol{\psi}_2$ and $\boldsymbol{\beta_2}$ as described in Section 4.

Step 4: Run Stage 1 with 500 iterations to obtain the estimate $\hat{\boldsymbol{\psi}}_\mathbf{1}$. See Section 3.

Step 5: Run Stage 2 with 100 iterations to obtain the estimate $\hat{\boldsymbol{\psi}}_\mathbf{2}$. See Section 3.

Step 6: Draw samples of $\boldsymbol{\theta}$ from the posterior distribution $p(\boldsymbol{\theta} \mid \hat{\boldsymbol{\psi}}, \boldsymbol{d})$ with estimated $\hat{\boldsymbol{\psi}} = [\hat{\boldsymbol{\psi}}_\mathbf{1}^T, \hat{\boldsymbol{\psi}}_\mathbf{2}^T]^T$ plugged in. See Section 3.

Step 7: Plug samples of $\boldsymbol{\theta}$, estimated $\boldsymbol{\psi}$, full data $\boldsymbol{d}$ and X_p into (3.10) and (3.11) to get predictions and their corresponding variances. See Section 3.

Appendix 2   Other Results

Table A.1. Priors and Estimates of Hyperparameters for Case 1a

ψ_1		ω_{x1}	ω_{x2}	$\omega_{\theta1}$	$\omega_{\theta2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.641	6.552	0.044	0.725	0.355
ψ_2			λ	$\omega_{x1}^\star$	$\omega_{x2}^\star$	σ_2^2
True values			0.001	1	1	0.005
Lognormal Prior	mean		0.01	1	1	0.005
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.0007	0.270	0.185	0.022
$\boldsymbol{\beta}$	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.008	0.626	0.096	0.290	0.016	-0.048

Table A.2. Priors and Estimates of Hyperparameters for Case 1b

ψ_1		ω_{x1}	ω_{x2}	$\omega_{\theta1}$	$\omega_{\theta2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.703	6.406	0.046	0.702	0.290
ψ_2			λ	$\omega_{x1}^\star$	$\omega_{x2}^\star$	σ_2^2
True values			0.001	1	1	0.001
Lognormal Prior	mean		0.01	1	1	0.001
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.002	0.442	0.791	0.0007
$\boldsymbol{\beta}$	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.0008	0.628	0.096	0.291	0.024	0.017

Table A.3. Priors and Estimates of Hyperparameters for Case 2a

ψ_1		ω_{x1}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.803	6.815	0.059	0.675	0.261

ψ_2			λ	$\omega_{x1}^\star$	$\omega_{x2}^\star$	σ_2^2
True values			0.001	1	1	0.005
Lognormal Prior	mean		0.01	1	1	0.005
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.001	0.384	0.272	0.122

β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.009	0.637	0.101	0.315	-0.086	-0.016

Table A.4. Priors and Estimates of Hyperparameters for Case 2b

ψ_1		ω_{x1}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.700	6.836	0.054	0.990	0.304

ψ_2			λ	$\omega_{x1}^\star$	$\omega_{x2}^\star$	σ_2^2
True values			0.001	1	1	0.001
Lognormal Prior	mean		0.01	1	1	0.001
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.004	0.331	1.103	0.003

β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.030	0.638	0.102	0.313	-0.014	-0.009

Table A.5. Priors and Estimates of Hyperparameters for Case 2c

ψ_1		ω_{x1}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.794	6.467	0.054	0.818	0.283

ψ_2			λ	$\omega_{x1}^\star$	$\omega_{x2}^\star$	σ_2^2
True values			0.001	1	1	0.001
Lognormal Prior	mean		0.01	1	1	0.001
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.001	0.353	2.925	0.008

β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.009	0.641	0.101	0.312	0.006	-0.003

Table A.6. Priors and Estimates of Hyperparameters for Case 3a

ψ_1		ω_{x1}	ω_{x2}	$\omega_{\theta2}$	$\omega_{\theta2}$	σ^2_1
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.898	15.920	0.250	1.640	0.533
ψ_2			λ	$\omega^\star_{x1}$	$\omega^\star_{x2}$	σ^2_2
True values			0.001	1	1	0.005
Lognormal Prior	mean		0.01	1	1	0.005
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.003	0.210	0.054	0.032
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.070	0.621	0.094	0.341	-0.030	0.004

Table A.7. Priors and Estimates of Hyperparameters for Case 3b

ψ_1		ω_{x1}	$\omega_{\theta1}$	$\omega_{\theta1}$	$\omega_{\theta2}$	σ^2_1
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.951	15.786	0.217	1.971	0.526
ψ_2			λ	$\omega^\star_{x1}$	$\omega^\star_{x2}$	σ^2_2
True values			0.001	1	1	0.001
Lognormal Prior	mean		0.01	1	1	0.001
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.001	0.637	1.705	0.0006
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.066	0.616	0.094	0.345	-0.020	0.015

Table A.8. Priors and Estimates of Hyperparameters for real observations

ψ_1		ω_{x2}	ω_{x2}	$\omega_{\theta1}$	$\omega_{\theta2}$	σ^2_1
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.242	1.702	0.082	0.008	0.729
ψ_2			λ	$\omega^\star_{x1}$	$\omega^\star_{x2}$	σ^2_2
Lognormal Prior	mean		0.01	1	1	0.1
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.01	1	1	0.001
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.182	0.537	0.163	-0.189	0.021	-0.058
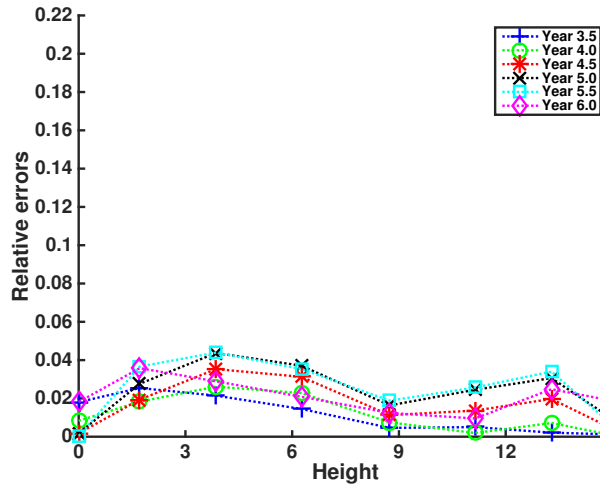
(a) Case 1a
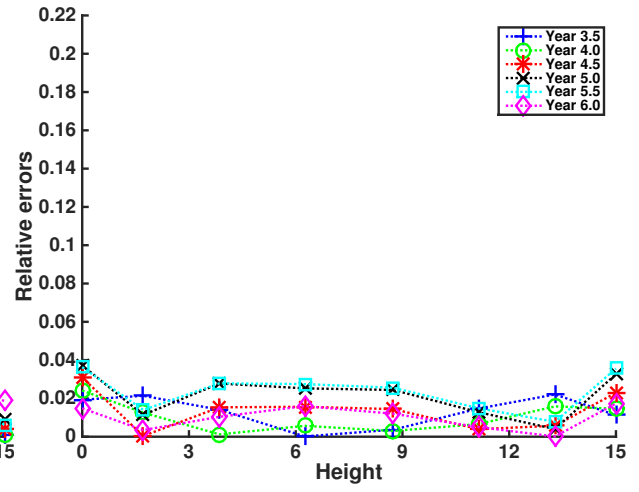
(b) Case 1b
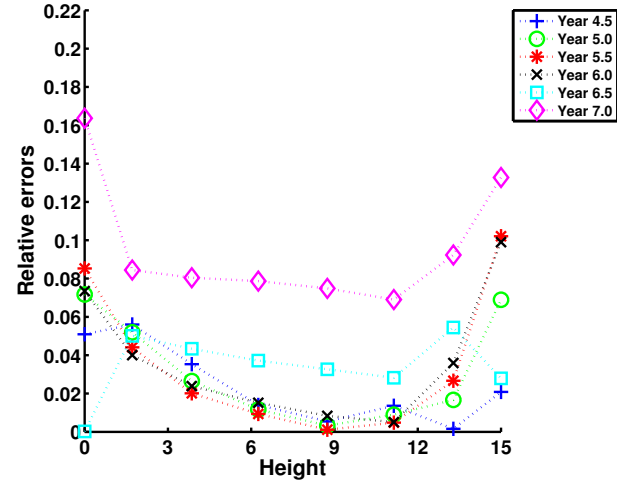
(c) Case 2a
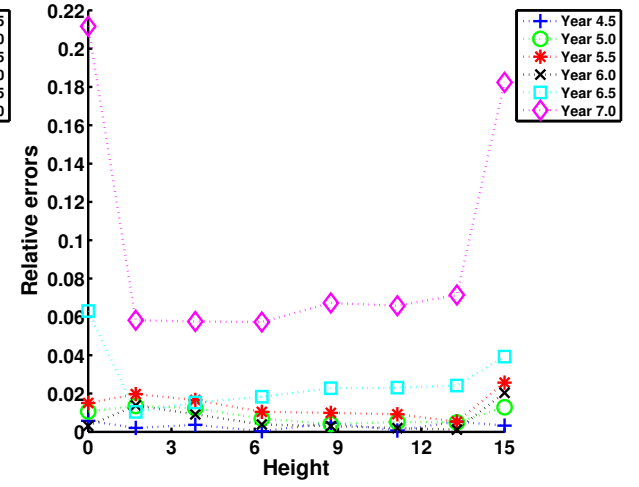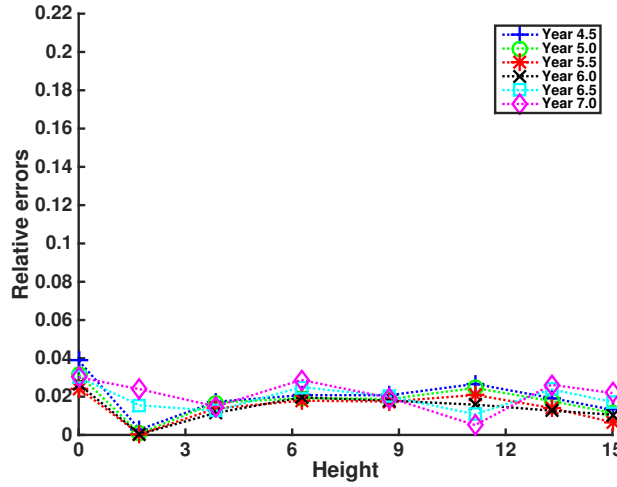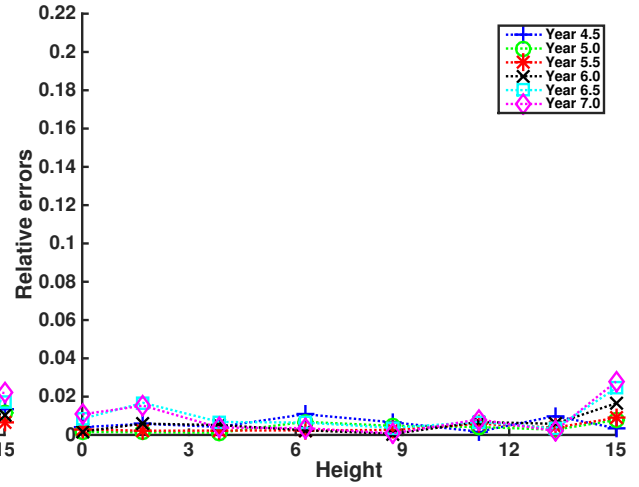
(d) Case 2b

(e) Case 3a

(f) Case 3b

Figure A.1. Relative errors between predictions and tmrue values.
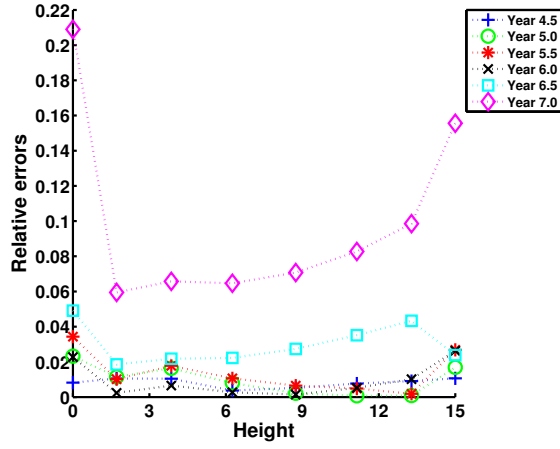
(a) Case 2c
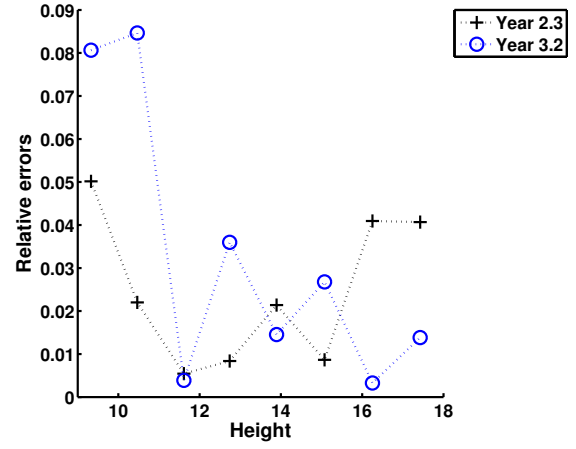
(b) Real observation

Figure A.2. Relative errors for Case 2c (a) and for real observation case (b).

55
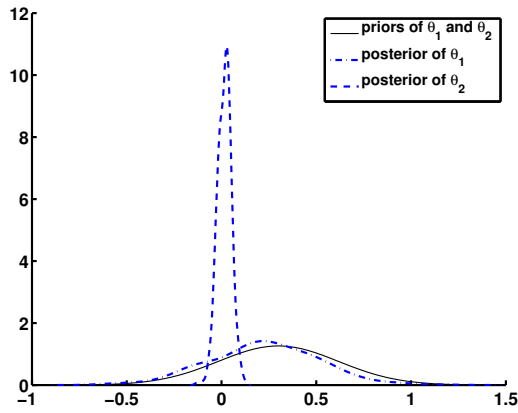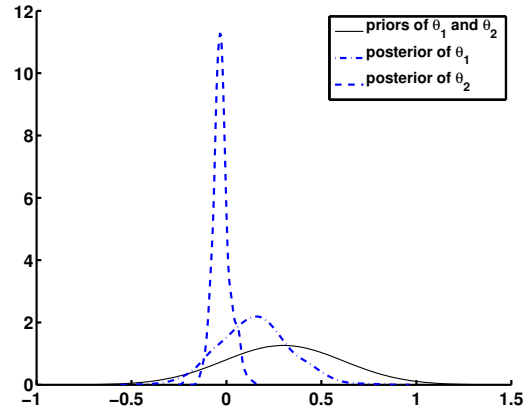
(a) Case 1a

(b) Case 1b
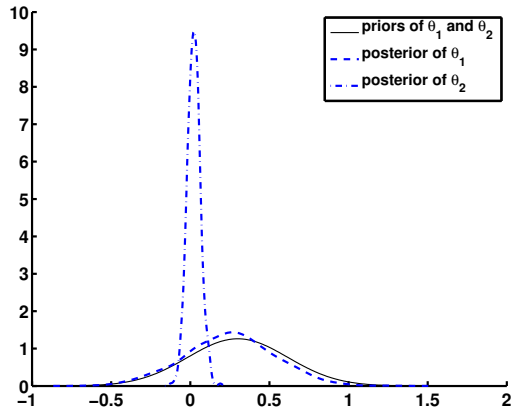
(c) Case 2a
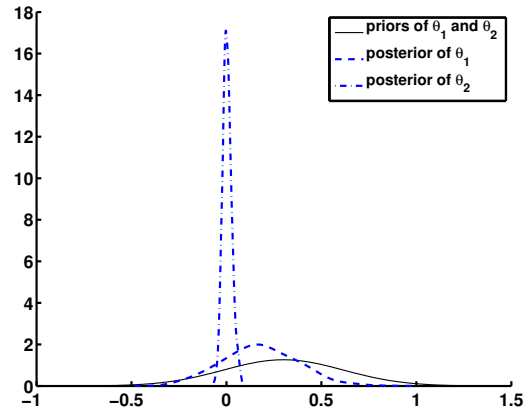
(d) Case 2b
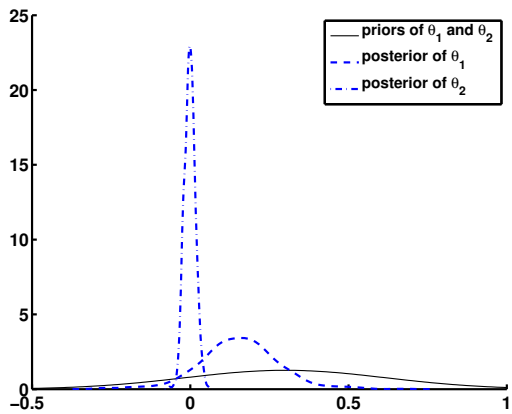
(e) Case 3a
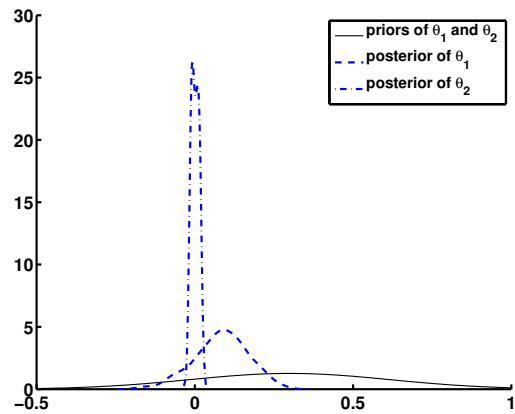
(f) Case 3b

Figure A.3. Prior and posterior densities of θ_1 and θ_2
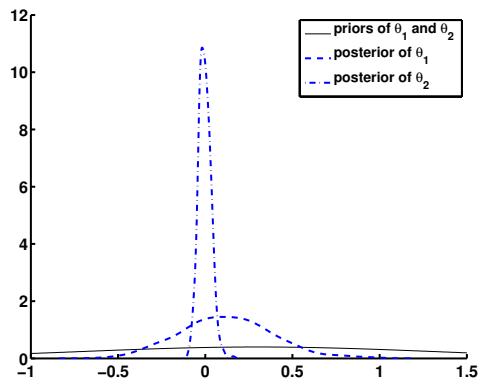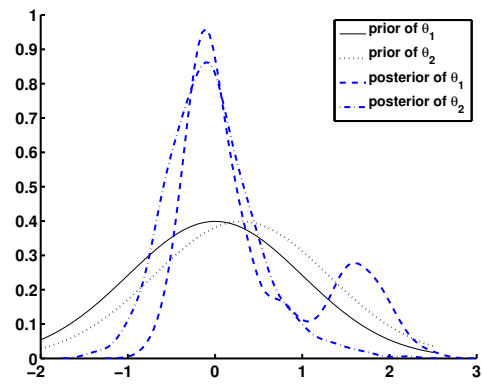
(a) Case 2c (b) Real observation

Figure A.4. Prior and posterior densities of θ_1 and θ_2.

Chapter 3

Patient-Specific Prediction from Multi-subject Data

3.1 Introduction

The computer model is built based on all the 26 subjects which determine the biomedical system of AAA enlargement. Therefore, the resembling mechanism of computer model considers both the within-subject and between-subject effects. But the Patient-specific Bayesian calibration calibrates the unobserved parameters in the computer model by consider real observations from only one subject, which might actually cause the calibration results to distort itself more on the mechanism determined by the specific patient. Therefore, it has practical meaning to build the Bayesian calibration model for multi-subject data so that the calibrated parameter could consider uncertainties from real observations across all the subjects.

Bayesian calibration analysis of longitudinal imaging data it self is very challenging and interesting. On the one hand, the dimension of imaging data for a single patient at only one time point is very high, not to say multi-subject data collected at multiple time points. Thus, whatever models built for the data requires a heavy computational burden. On the other hand, variability of real observations happens both within-subject and between-subject across time. And as we know, calibration parameters are unknown inputs in the computer

model proposed by the biomedical engineers. Thus, how uncertainties in real observations across all the subjects affect uncertainties of calibration parameters is an important problem in Bayesian analysis, since no others have done it yet. But in Bayesian analysis, mixing and separating so many uncertainties with different properties is quite a challenging and worthwhile work to do.

Given the above reasons, we model the Bayesian calibration for multi-subject data as follows. Mostly, we keep the same notation and symbols as Single-subject modeling case in Chapter 2.

3.2 Calibration model

Suppose the total number of subjects is I, then for the i-th subject, the real observation z_{ij} can be modeled as

$$\boldsymbol{z}_{ij} = r(\boldsymbol{x}_{ij}, \boldsymbol{\theta}_i) + \delta(\boldsymbol{x}_{ij}) + \boldsymbol{\varepsilon}_{ij}, \quad \forall i \in \{1, \cdots, I\}, \tag{3.1}$$

where $j = 1, \cdots, n_i$, which means we have n_i observations from CT scan images of a given subject i. $r(\cdot, \cdot)$ is the computer model. $\delta(\cdot)$ is the model error. $\boldsymbol{\epsilon}_{ij}$ is the observation error. We assume that each ε_{ij} is independently distributed as $\mathcal{N}(0, \lambda)$, which represents a normal (Gaussian) random variable with mean 0 and variance λ. We further assume that observational errors are independent among and within subjects.

3.3 Statistical models

Let $\mathcal{GP}(\boldsymbol{m}(\cdot), \boldsymbol{k}(\cdot, \cdot))$ be the Gaussian process with the mean function $m(\cdot)$ and the covariance function $\boldsymbol{k}(\cdot, \cdot)$. We introduce the following Gaussian processes as prior beliefs for the G&R computation model and the model error:

$$r(\boldsymbol{x}_i, \boldsymbol{\theta}_i) \sim \mathcal{GP}(\boldsymbol{m}_1(\boldsymbol{x}_i, \boldsymbol{\theta}_i), \boldsymbol{k}_1(\boldsymbol{x}_i, \boldsymbol{x}_i')),$$
$$\delta(\boldsymbol{x}_i) \sim \mathcal{GP}(\boldsymbol{m}_2(\boldsymbol{x}_i), \boldsymbol{k}_2(\boldsymbol{x}_i, \boldsymbol{x}_i')). \tag{3.2}$$

To specify Gaussian process priors in (3.2) further, we introduce mean and covariance structures for both processes.

For a mean structure, one can consider a linear combination of basis functions to approximate the general mean structure.

$$\boldsymbol{m}_1(\boldsymbol{x}_i, \boldsymbol{\theta}_i) = \boldsymbol{h}_1(\boldsymbol{x}_i, \boldsymbol{\theta}_i)^T \boldsymbol{\beta}_1$$
$$= \beta_{10} + \beta_{11} t_i + \beta_{12} \theta_{i1} + \beta_{13} \theta_{i2}, \tag{3.3}$$

where $\boldsymbol{h}_1(\boldsymbol{x}_i, \boldsymbol{\theta}_i) = [1 \ t_i \ \theta_{i1} \ \theta_{i2}]^T$ and $\boldsymbol{\beta}_1 = [\beta_{10} \ \beta_{11} \ \beta_{12} \ \beta_{13}]^T$.

$$\boldsymbol{m}_2(\boldsymbol{x}_i) = \boldsymbol{h}_2(\boldsymbol{x}_i)^T \boldsymbol{\beta}_2 = \beta_{21} t_i + \beta_{22} s_i, \tag{3.4}$$

where $\boldsymbol{h}_2(\boldsymbol{x}_i) = [t_i \ s_i]^T$ and $\boldsymbol{\beta}_2 = [\beta_{21} \ \beta_{22}]^T$.

These mean structures imply that the mean function of the G&R computation model is linear in time t_i and calibration parameters, $\{\theta_{i1}, \theta_{i2}\}$. The mean function of the model error is linear in time t_i and location s_i. $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T \ \boldsymbol{\beta}_2^T]^T$ are hyperparameters for the mean functions in Bayesian context.

For a covariance structure, we use the following exponential functions as follows.

$$\boldsymbol{k}_1(\boldsymbol{x}_i, \boldsymbol{x}_i') = \sigma_1^2 \exp\{-(\boldsymbol{x}_i - \boldsymbol{x}_i')\boldsymbol{\Omega}_x(\boldsymbol{x}_i - \boldsymbol{x}_i')^T\} \exp\{-(\boldsymbol{\theta}_i - \boldsymbol{\theta}_i')\boldsymbol{\Omega}_\theta(\boldsymbol{\theta}_i - \boldsymbol{\theta}_i')^T\},$$

$$\boldsymbol{k}_2(\boldsymbol{x}_i, \boldsymbol{x}_i') = \sigma_2^2 \exp\{-(\boldsymbol{x}_i - \boldsymbol{x}_i')\boldsymbol{\Omega}_x^\star(\boldsymbol{x}_i - \boldsymbol{x}_i')^T\},$$

where $\boldsymbol{\Omega}_x$, $\boldsymbol{\Omega}_x^\star$ are diagonal matrices such that

$$\boldsymbol{\Omega}_x = \begin{pmatrix} \omega_{x1} & 0 \\ 0 & \omega_{x2} \end{pmatrix}, \boldsymbol{\Omega}_\theta = \begin{pmatrix} \omega_{\theta1} & 0 \\ 0 & \omega_{\theta2} \end{pmatrix}, \boldsymbol{\Omega}_x^\star = \begin{pmatrix} \omega_{x1}^\star & 0 \\ 0 & \omega_{x2}^\star \end{pmatrix}$$

so that the hyperparameters for the covariance functions are σ_1^2, σ_2^2 and w_{xj}, $w_{\theta j}$, $w_{xj}^\star$'s.

For the parameters controlling the covariances, we introduce $\boldsymbol{\psi} = [\boldsymbol{\psi}_1 \ \boldsymbol{\psi}_2]$, with

$$\boldsymbol{\psi}_1 = [\omega_{x1} \ \omega_{x2} \ \omega_{\theta1} \ \omega_{\theta2} \ \sigma_1^2],$$

$$\boldsymbol{\psi}_2 = [\omega_{x1}^\star \ \omega_{x2}^\star \ \sigma_2^2].$$

(3.5)

$\boldsymbol{\psi}_1$ is the set of hyperparameters related to the G&R computation model. $\boldsymbol{\psi}_2$ is the set of hyperparameters related to model error.

3.4 Computer outputs

Let N_i be the number of computer outputs for each subject i and $\boldsymbol{y}_i = [y_{i1} \cdots z_{iN_i}]^T \in \mathbb{R}^{N_i \times 1}$. Then given the N_i-th inputs for subject i, $(\boldsymbol{x}_{iN_i}^*, \boldsymbol{a}_{iN_i})$, we have the corresponding computer output $y_{iN_i} = r(\boldsymbol{x}_{iN_i}^*, \boldsymbol{a}_{iN_i})$. The corresponding variable input matrix is $X_{i,c} = [(\boldsymbol{x}_{i1}^*)^T, \cdots, (\boldsymbol{x}_{iN_i}^*)^T]^T \in \mathbb{R}^{N_i \times 2}$. We further define $\boldsymbol{y} = [\boldsymbol{y}_1^T \ \cdots \ \boldsymbol{y}_I^T]^T \in \mathbb{R}^{N \times 1}$ as the computer model output vector. The final variable input matrix is $X_c = [(X_{1,c})^T, \cdots, (X_{I,c})^T]^T \in$

$\mathbb{R}^{N \times 2}$. We augment variable inputs X_c with parameter inputs $\boldsymbol{a}_i$ such that $X_{i,c}(\boldsymbol{a}_i) = [(\boldsymbol{x}_{i1}^*, \boldsymbol{a}_i)^T, \cdots, (\boldsymbol{x}_{iN_i}^*, \boldsymbol{a}_i)^T]^T \in \mathbb{R}^{N_i \times 2}$. The final input matrix for computer outputs is $X_c(\boldsymbol{a}) = [X_{1,c}(\boldsymbol{a}_1)^T, \cdots, X_{I,c}(\boldsymbol{a}_I)^T]^T \in \mathbb{R}^{N \times 4}$.

Then, the $\mathcal{GP}$ assumption for the computer model gives

$$\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}_i))\boldsymbol{\beta}_1, \boldsymbol{V}_1(X_{i,c})),$$

where $\boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}_i))$ denotes the matrix with rows $\boldsymbol{h}_1(\boldsymbol{x}_{i1}^*, \boldsymbol{a}_i)^T, \cdots, \boldsymbol{h}_1(\boldsymbol{x}_{iN_i}^*, \boldsymbol{a}_i)^T$ and the (j, k) entry of $\boldsymbol{V}_1(X_{i,c})$ for the i-th subject is $\boldsymbol{k}_1(\boldsymbol{x}_{ij}^*, \boldsymbol{x}_{ik}^*)$, for $j, k \in \{1, 2, \cdots, N_i\}$. We denote the mean and covariance as $\boldsymbol{\mu_{y_i}} = \boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}_i))\boldsymbol{\beta}_1$ and $\boldsymbol{V_{y_i}} = \boldsymbol{V}_1(X_{i,c})$ respectively.

Finally, the computer output is $\boldsymbol{y} = [\boldsymbol{y}_1^T \ \cdots \ \boldsymbol{y}_I^T]^T \in \mathbb{R}^{N \times 1}$. The distribution of $\boldsymbol{y}$ is

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu_y}, \boldsymbol{V_y}),$$

where

$$\boldsymbol{\mu_y} = \begin{pmatrix} \boldsymbol{\mu_{y_1}} \\ \boldsymbol{\mu_{y_2}} \\ \vdots \\ \boldsymbol{\mu_{y_I}} \end{pmatrix},$$

and

$$\boldsymbol{V_y} = \begin{pmatrix} \boldsymbol{V_{y_1}} & 0 & \cdots & 0 \\ 0 & \boldsymbol{V_{y_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{V_{y_I}} \end{pmatrix}.$$

3.5 Observations

Let n_i be the number of observations for each subject i. Let $\boldsymbol{z}_i = [z_{i1} \;\; \cdots \;\; z_{in_i}]^T \in \mathbb{R}^{n_i \times 1}$ be the set of observations corresponding to the variable input matrix $X_{i,o} = [\boldsymbol{x}_{i1}^T, \cdots, \boldsymbol{x}_{in_i}^T]^T \in \mathbb{R}^{n_i \times 2}$. Note that $X_o = [X_{1,o}^T, \cdots, X_{I,o}^T]^T \in \mathbb{R}^{n \times 2}$ is not necessarily the same as the set of variable inputs for the computer model outputs. In general, the number of observations is smaller than that of the computer model outputs since we can control the amount of the computer model outputs. To calibrate $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_I]$ from the observations, we augment variable inputs X_o with $\boldsymbol{\theta}_i$ such that $X_{i,o}(\boldsymbol{\theta}_i) = [(\boldsymbol{x}_{i1}, \boldsymbol{\theta}_i)^T, \cdots, (\boldsymbol{x}_{in_i}, \boldsymbol{\theta}_i)^T]^T$.

Then, from the calibration model, we have

$$\boldsymbol{z}_i \sim \mathcal{N}(\boldsymbol{H}_1(X_{i,o}(\boldsymbol{\theta}_i))\boldsymbol{\beta}_1 + \boldsymbol{H}_2(X_{i,o})\boldsymbol{\beta}_2, \lambda_i \boldsymbol{I}_{n_i} + \boldsymbol{V}_1(X_{i,o}) + \boldsymbol{V}_2(X_{i,o})),$$

where $\boldsymbol{H}_1(X_{i,o}(\boldsymbol{\theta}_i))$ denotes the matrix with rows $\boldsymbol{h}_1(\boldsymbol{x}_{i1}, \boldsymbol{\theta}_i)^T, \cdots, \boldsymbol{h}_1(\boldsymbol{x}_{in_i}, \boldsymbol{\theta}_i)^T$, $\boldsymbol{H}_2(X_{i,o})$ denotes the matrix with rows $\boldsymbol{h}_2(x_{i1})^T, \cdots, \boldsymbol{h}_2(\boldsymbol{x}_{in_i})^T$. $\boldsymbol{V}_1(X_{i,o})$ is defined in a similar way to define $\boldsymbol{V}_1(X_{i,c})$. The (j, k) entry of $\boldsymbol{V}_2(X_{i,o})$ is $\boldsymbol{k}_2(\boldsymbol{x}_{ij}, \boldsymbol{x}_{ik})$, for $j, k \in \{1, 2, \cdots, n_i\}$. Denote $\boldsymbol{\mu}_{\boldsymbol{z}_i} = \boldsymbol{H}_1(X_{i,o}(\boldsymbol{\theta}_i))\boldsymbol{\beta}_1 + \boldsymbol{H}_2(X_{i,o})\boldsymbol{\beta}_2$ and $\boldsymbol{V}_{\boldsymbol{z}_i} = \lambda_i \boldsymbol{I}_{n_i} + \boldsymbol{V}_1(X_{i,o}) + \boldsymbol{V}_2(X_{i,o}))$

Then the observation is $\boldsymbol{z} = [\boldsymbol{z}_1^T \cdots \boldsymbol{z}_I^T]^T \in \mathbb{R}^{n \times 1}$. The distribution of $\boldsymbol{z}$ is

$$\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{z}}, \boldsymbol{V}_{\boldsymbol{z}}),$$

where

$$\boldsymbol{\mu_z} = \begin{pmatrix} \boldsymbol{\mu_{z_1}} \\ \boldsymbol{\mu_{z_2}} \\ \vdots \\ \boldsymbol{\mu_{z_I}} \end{pmatrix},$$

and

$$\boldsymbol{V_z} = \begin{pmatrix} \boldsymbol{V_{z_1}} & 0 & \cdots & 0 \\ 0 & \boldsymbol{V_{z_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{V_{z_I}} \end{pmatrix}.$$

3.6   Full data

We then combine the computer model outputs and observations, $\boldsymbol{d} = [\boldsymbol{d}_1^T \cdots \boldsymbol{d}_I^T]^T \in \mathbb{R}^{(N+n)\times 1}$ which we call a data vector.

$$\boldsymbol{d}_i = \begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{z}_i \end{pmatrix} \sim \mathcal{N}(\boldsymbol{m}_{d_i}(\boldsymbol{\theta}_i), \boldsymbol{V}_{d_i}), \tag{3.6}$$

where

$$\boldsymbol{m}_{d_i}(\boldsymbol{\theta}_i) := \mathbb{E}(\boldsymbol{d}_i | \boldsymbol{\theta}_i, \boldsymbol{\beta}, \boldsymbol{\psi}) = \boldsymbol{H}(\boldsymbol{\theta}_i)\boldsymbol{\beta},$$

with

$$\boldsymbol{H}(\boldsymbol{\theta}_i) = \begin{pmatrix} \boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}_i)) & \boldsymbol{0} \\ \boldsymbol{H}_1(X_{i,o}(\boldsymbol{\theta}_i)) & \boldsymbol{H}_2(X_{i,o}) \end{pmatrix},$$

and

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T.$$

The full mean is

$$\boldsymbol{m}_d(\boldsymbol{\theta}) = [\boldsymbol{m}_{d_1}(\boldsymbol{\theta}_1)^T, \cdots, \boldsymbol{m}_{d_I}(\boldsymbol{\theta}_I)^T]^T.$$

Denote $\boldsymbol{H}(\boldsymbol{\theta}) = [\boldsymbol{H}(\boldsymbol{\theta}_1)^T, \cdots, \boldsymbol{H}(\boldsymbol{\theta}_I)^T]^T$.

We assume different subjects are independent, then the full covariance is

$$\boldsymbol{V}_d(\boldsymbol{\psi}) = \begin{pmatrix} \boldsymbol{V}_{d_1}(\boldsymbol{\psi}) & 0 & \cdots & 0 \\ 0 & \boldsymbol{V}_{d_2}(\boldsymbol{\psi}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{V}_{d_I}(\boldsymbol{\psi}) \end{pmatrix},$$

where

$$\boldsymbol{V}_{d_i}(\boldsymbol{\psi}) := \mathrm{var}(\boldsymbol{d}_i|\boldsymbol{\psi}) = \begin{pmatrix} \boldsymbol{V}_{\boldsymbol{y}_i} & \boldsymbol{C}_{\boldsymbol{y}_i\boldsymbol{z}_i}^T \\ \boldsymbol{C}_{\boldsymbol{y}_i\boldsymbol{z}_i} & \boldsymbol{V}_{\boldsymbol{z}_i} \end{pmatrix},$$

where $\boldsymbol{C}_{\boldsymbol{y}_i\boldsymbol{z}_i} \in \mathbb{R}^{n_i \times N_i}$ is a cross covariance matrix whose (j, k) entry is $\boldsymbol{k}_1(\boldsymbol{x}_{ij}^*, \boldsymbol{x}_{ik})$, for $j \in \{1, 2, \cdots, N_i\}$ and $k \in \{1, 2, \cdots, n_i\}$. Note that $\boldsymbol{C}_{\boldsymbol{y}_i\boldsymbol{z}_i}$ is not necessary to be a square matrix, since usually $N_i \neq n_i$.

3.7 Bayesian analysis

3.7.1 Joint posterior distribution

We note that $\boldsymbol{m}_d(\boldsymbol{\theta})$ also depend on $\boldsymbol{\beta}$. We drop it to reduce notational complexity. We have the following full joint posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\psi})$ given $\boldsymbol{d}$,

$$
\begin{aligned}
p(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\psi} \mid \boldsymbol{d}) \propto & p(\boldsymbol{\theta})p(\boldsymbol{\psi})f(\boldsymbol{d}; \boldsymbol{m}_d(\boldsymbol{\theta}), \boldsymbol{V}_d(\boldsymbol{\psi})) \\
\propto & \prod_{i=1}^{I} p(\boldsymbol{\theta}_i)p(\boldsymbol{\psi})|\boldsymbol{V}_d(\boldsymbol{\psi})|^{-\frac{1}{2}} \\
& \times \exp\left[-\frac{1}{2}\{(\boldsymbol{d} - \boldsymbol{m}_d(\boldsymbol{\theta}))^T \boldsymbol{V}_d(\boldsymbol{\psi})^{-1}(\boldsymbol{d} - \boldsymbol{m}_d(\boldsymbol{\theta}))\}\right],
\end{aligned}
\tag{3.7}
$$

where $f(d; m, V)$ is the multivariate Gaussian density function with mean $\boldsymbol{m}$ and variance $\boldsymbol{V}$.

Clearly, $\boldsymbol{m}_d(\boldsymbol{\theta})$ is a linear function of $\boldsymbol{\beta}$, so we can find

$$
\boldsymbol{\beta} \mid \boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{d} \sim \mathcal{N}(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}, \boldsymbol{\psi}), \boldsymbol{W}(\boldsymbol{\theta}, \boldsymbol{\psi})),
\tag{3.8}
$$

where

$$
\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}, \boldsymbol{\psi}) = \boldsymbol{W}(\boldsymbol{\theta})\boldsymbol{H}(\boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\psi})^{-1}\boldsymbol{d},
$$

$$
\boldsymbol{W}(\boldsymbol{\theta}, \boldsymbol{\psi}) = (\boldsymbol{H}(\boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\psi})^{-1}\boldsymbol{H}(\boldsymbol{\theta}))^{-1},
$$

We can then integrate out $\boldsymbol{\beta}$ in (3.7) and get

$$
\begin{aligned}
p(\boldsymbol{\theta}, \boldsymbol{\psi} \mid \boldsymbol{d}) \propto & \, p(\boldsymbol{\psi})|\boldsymbol{W}(\boldsymbol{\theta}, \boldsymbol{\psi})|^{\frac{1}{2}} \prod_{i=1}^{I} p(\boldsymbol{\theta}_i)|\boldsymbol{V}_{d_i}(\boldsymbol{\psi})|^{-\frac{1}{2}} \\
& \times \exp\left[-\frac{1}{2}\{(\boldsymbol{d}_i - \boldsymbol{H}(\boldsymbol{\theta}_i)\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}, \boldsymbol{\psi}))^T \boldsymbol{V}_{d_i}(\boldsymbol{\psi})^{-1}(\boldsymbol{d}_i - \boldsymbol{H}(\boldsymbol{\theta}_i)\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}, \boldsymbol{\psi}))\}\right]
\end{aligned}
\tag{3.9}
$$

Because of independence, $\hat{\beta}$ and $\boldsymbol{W}$ in (3.9) can be simplified as

$$\boldsymbol{W}(\boldsymbol{\theta}, \boldsymbol{\psi}) = \left(\sum_{i=1}^{I} \boldsymbol{H}(\boldsymbol{\theta}_i)^T \boldsymbol{V}_{d_i}(\boldsymbol{\psi})^{-1} \boldsymbol{H}(\boldsymbol{\theta}_i) \right)^{-1}$$

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}, \boldsymbol{\psi}) = \boldsymbol{W}(\boldsymbol{\theta}, \boldsymbol{\psi}) \left(\sum_{i=1}^{I} \boldsymbol{H}(\boldsymbol{\theta}_i)^T \boldsymbol{V}_{d_i}(\boldsymbol{\psi})^{-1} \boldsymbol{d}_i \right)$$

3.7.2   Full conditional of ψ_1

We can estimate $\boldsymbol{\psi}_1$ by maximizing

$$p(\boldsymbol{\psi}_1|\boldsymbol{y}) \propto |\boldsymbol{V}_1(X_c)|^{-1/2} p(\boldsymbol{\psi}_1) |\boldsymbol{W}_1(X_c(\boldsymbol{a}))|^{1/2}$$
$$\times \exp\left\{ -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{H}_1(X_c(\boldsymbol{a}))\hat{\boldsymbol{\beta}}_1)^T \boldsymbol{V}_1(X_c)^{-1}(\boldsymbol{y} - \boldsymbol{H}_1(X_c(\boldsymbol{a}))\hat{\boldsymbol{\beta}}_1) \right\},$$

(3.10)

where

$$\hat{\boldsymbol{\beta}}_1 = \boldsymbol{W}_1(X_c(\boldsymbol{a})) \boldsymbol{H}_1(X_c(\boldsymbol{a}))^T \boldsymbol{V}_1(X_c) \boldsymbol{y},$$

$$\boldsymbol{W}_1(X_c(\boldsymbol{a})) = (\boldsymbol{H}_1(X_c(\boldsymbol{a}))^T \boldsymbol{V}_1(X_c)^{-1} \boldsymbol{H}_1(X_c(\boldsymbol{a})))^{-1}.$$

3.7.3   Full conditional of ψ_2

We can estimate $\boldsymbol{\psi}_2$ by maximizing $p(\boldsymbol{\psi}_2|\boldsymbol{d}, \boldsymbol{\psi}_1)$. To obtain this density, we can first write

$$p(\boldsymbol{\beta}_2, \boldsymbol{\psi}_2|\boldsymbol{d}, \boldsymbol{\psi}_1) \propto p(\boldsymbol{\beta}_2, \boldsymbol{\psi}_2) p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\psi}),$$

since $\boldsymbol{y}$ is independent of the second stage hyper-parameters.

We can't obtain the distribution of $[\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\psi}]$ analytically, but starting with $[\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\psi}, \boldsymbol{\theta}]$,

which is normally distributed.

$$z_{ij} \mid \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\psi}, \boldsymbol{\theta}_i \sim \mathcal{N}(\boldsymbol{h}_1(\boldsymbol{x}_{ij}, \boldsymbol{\theta}_i)^T \boldsymbol{\beta}_1 + \boldsymbol{h}_2(x_{ij})^T \boldsymbol{\beta}_2, \lambda + \boldsymbol{V}_1(\boldsymbol{x}_{ij}) + \boldsymbol{V}_2(\boldsymbol{x}_{ij}))$$

67

$$E(\boldsymbol{z}_{ij}|\boldsymbol{y}_i,\boldsymbol{\beta}_2,\boldsymbol{\psi},\boldsymbol{\theta}_i,\boldsymbol{\beta}_1) = \boldsymbol{h}_2(x_{ij})^T\boldsymbol{\beta}_2 + \boldsymbol{h}_1(\boldsymbol{x}_{ij},\boldsymbol{\theta}_i)^T\boldsymbol{\beta}_1$$
$$+\boldsymbol{t}(\boldsymbol{x}_{ij})^T\boldsymbol{V}_1(X_{i,c})^{-1}(\boldsymbol{y}_i - \boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}))\boldsymbol{\beta}_1),$$

where the k-th element of $\boldsymbol{t}(\boldsymbol{x}_{ij})$ is $k_1(\boldsymbol{x}_{ij},\boldsymbol{x}_{ik}^*)$, for $k = 1,\ldots,N_i$.

We can obtain expressions for its mean vector.

$$E(\boldsymbol{z}_{ij}|\boldsymbol{y}_i,\boldsymbol{\beta}_2,\boldsymbol{\psi})$$
$$= \int E(\boldsymbol{z}_{ij}|\boldsymbol{y}_i,\boldsymbol{\beta}_2,\boldsymbol{\psi},\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i$$
$$= \boldsymbol{h}_2(x_{ij})^T\boldsymbol{\beta}_2 + \int \boldsymbol{h}_1(\boldsymbol{x}_{ij},\boldsymbol{\theta}_i)^Tp(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i\hat{\boldsymbol{\beta}}_1$$
$$+t(\boldsymbol{x}_{ij})^T\boldsymbol{V}_1(X_{i,c})^{-1}(\boldsymbol{y}_i - \boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}_i))\hat{\boldsymbol{\beta}}_1)$$

where the k-th element of $\boldsymbol{t}(\boldsymbol{x}_{ij},\boldsymbol{\theta}_i)$ is $k_1(\boldsymbol{x}_{ij},\boldsymbol{x}_{ik}^*)$, for $k = 1,\ldots,N_i$.

We can obtain expressions for its variance matrix. The covariance matrix is $\boldsymbol{V}_{i,\psi_2} = \lambda\boldsymbol{I}_i + \boldsymbol{V}_2(X_{i,o}) + \boldsymbol{C}_i$, where the (j,k) element of $\boldsymbol{C}_i$ is

$$\int \mathrm{cov}(r(\boldsymbol{x}_{ij},\boldsymbol{\theta}_i),r(\boldsymbol{x}_{ik},\boldsymbol{\theta}_i)|\boldsymbol{y}_i,\boldsymbol{\psi}_1)p(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i$$
$$= k_1(\boldsymbol{x}_{ij},\boldsymbol{x}_{ik})$$
$$-\mathrm{tr}\{\boldsymbol{V}_1(X_{i,c})^{-1}\boldsymbol{t}(\boldsymbol{x}_{ij})\boldsymbol{t}(\boldsymbol{x}_{ik})^T\}$$
$$+\mathrm{tr}\{\boldsymbol{W}_1(X_{i,c}(\boldsymbol{a}_i))\int \boldsymbol{h}_1(\boldsymbol{x}_{ij},\boldsymbol{\theta}_i)\boldsymbol{h}_1(\boldsymbol{x}_{ik},\boldsymbol{\theta}_i)^Tp(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i\}$$
$$-\mathrm{tr}\{\boldsymbol{W}_1(X_{i,c}(\boldsymbol{a}_i))\boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}_i))^T\boldsymbol{V}_1(X_{i,c})^{-1}\boldsymbol{t}(\boldsymbol{x}_{ij})\int \boldsymbol{h}_1(\boldsymbol{x}_{ik},\boldsymbol{\theta}_i)^Tp(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i\}$$
$$-\mathrm{tr}\{\boldsymbol{V}_1(X_{i,c})^{-1}\boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}_i))\boldsymbol{W}_1(X_{i,c}(\boldsymbol{a}_i))\int \boldsymbol{h}_1(\boldsymbol{x}_{ij},\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i\boldsymbol{t}(\boldsymbol{x}_{ik})^T\}$$
$$+\mathrm{tr}\{\boldsymbol{V}_1(X_{i,c})^{-1}\boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}_i))\boldsymbol{W}_1(X_{i,c}(\boldsymbol{a}_i))\boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}_i))^T\boldsymbol{V}_1(X_{i,c})^{-1}$$
$$\times\boldsymbol{t}(\boldsymbol{x}_{ij})\boldsymbol{t}(\boldsymbol{x}_{ik})^T\}$$

$$\text{var}\left(r(\boldsymbol{x}_{ij}, \boldsymbol{\theta}_i), r(\boldsymbol{x}_{ik}, \boldsymbol{\theta}_i), \boldsymbol{y}_i | \boldsymbol{\psi}_1, \boldsymbol{\theta}_i\right)$$

$$= \left[\begin{array}{cc|c} \text{var}(r(\boldsymbol{x}_{ij}, \boldsymbol{\theta}_i)) & \text{cov}\left(r(\boldsymbol{x}_{ij}, \boldsymbol{\theta}_i), r(\boldsymbol{x}_{ik}, \boldsymbol{\theta}_i)\right) & \boldsymbol{t}(\boldsymbol{x}_{ij})^T \\ \text{cov}\left(r(\boldsymbol{x}_{ij}, \boldsymbol{\theta}_i), r(\boldsymbol{x}_{ik}, \boldsymbol{\theta}_i)\right) & \text{var}(r(\boldsymbol{x}_{ik}, \boldsymbol{\theta}_i)) & \boldsymbol{t}(\boldsymbol{x}_{ik})^T \\ \hline \boldsymbol{t}(\boldsymbol{x}_{ij}) & \boldsymbol{t}(\boldsymbol{x}_{ik}) & \boldsymbol{V}_1(X_{i,c}) \end{array}\right]$$

$$= \left[\begin{array}{cc|c} \boldsymbol{k}_1(\boldsymbol{x}_{ij}, \boldsymbol{x}_{ij}) & \boldsymbol{k}_1(\boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}) & \boldsymbol{t}(\boldsymbol{x}_{ij})^T \\ \boldsymbol{k}_1(\boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}) & \boldsymbol{k}_1(\boldsymbol{x}_{ik}, \boldsymbol{x}_{ik}) & \boldsymbol{t}(\boldsymbol{x}_{ik})^T \\ \hline \boldsymbol{t}(\boldsymbol{x}_{ij}) & \boldsymbol{t}(\boldsymbol{x}_{ik}) & \boldsymbol{V}_1(X_{i,c}) \end{array}\right]$$

Mean

$$\text{E}\left(r(\boldsymbol{x}_{ij}, \boldsymbol{\theta}_i) | \boldsymbol{y}_i, \boldsymbol{\psi}_1, \boldsymbol{\theta}_i\right)$$

$$= \boldsymbol{h}_1(\boldsymbol{x}_{ij}, \theta_i)^T \boldsymbol{\beta}_1 + \boldsymbol{t}(\boldsymbol{x}_{ij})^T \boldsymbol{V}_1(X_{i,c})^{-1}(\boldsymbol{y}_i - \boldsymbol{H}_1(X_{i,c}(\boldsymbol{a}_i)\boldsymbol{\beta}_1)$$

Covariance

$$\text{cov}\left(r(\boldsymbol{x}_{ij}, \boldsymbol{\theta}_i), r(\boldsymbol{x}_{ik}, \boldsymbol{\theta}_i) | \boldsymbol{y}_i, \boldsymbol{\psi}_1, \boldsymbol{\theta}_i\right)$$

$$= \boldsymbol{k}_1(\boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}) - \boldsymbol{t}(\boldsymbol{x}_{ij})^T \boldsymbol{V}(X_{i,c})^{-1} \boldsymbol{t}(\boldsymbol{x}_{ik})$$

$$= \boldsymbol{k}_1(\boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}) - \text{tr}\left\{\boldsymbol{V}(X_{i,c})^{-1}\boldsymbol{t}(\boldsymbol{x}_{ij})\boldsymbol{t}(\boldsymbol{x}_{ik})^T\right\}$$

Plug the above two equations into the law of total covariance, we can get $\boldsymbol{C}_i$.

We then approximate $[\boldsymbol{z}_i | \boldsymbol{y}_i, \boldsymbol{\beta}_2, \boldsymbol{\psi}]$ by a normal distribution with the above moments for the purpose of estimating ψ_2. To simplify the notation we write the mean vector as $\boldsymbol{\mu}_{i,\psi_2} = \boldsymbol{H}_2(X_{i,o})\boldsymbol{\beta}_2 + \hat{\boldsymbol{\eta}}(X_{i,o})$, where $\hat{\boldsymbol{\eta}}(X_{i,o})$ is the vector with elements $E(r(\boldsymbol{x}_{ij}, \boldsymbol{\theta}_i) | \boldsymbol{y}_i)$, $i = 1, \ldots, n$. we can now integrate out $\boldsymbol{\beta}_2$ to obtain the approximation.

$$\begin{aligned} p(\boldsymbol{\psi}_2 | \boldsymbol{d}_i, \boldsymbol{\psi}_1) \propto{}& p(\boldsymbol{\psi}_2) |\boldsymbol{V}_{i,\psi_2}|^{1/2} |\boldsymbol{W}_{i,2}|^{1/2} \\ & \times \exp\left\{-\tfrac{1}{2}(z_i - \boldsymbol{H}_2(X_{i,o})\hat{\boldsymbol{\beta}}_{i,2} - \hat{\boldsymbol{\eta}}(X_{i,o}))^T \boldsymbol{V}_{i,\psi_2}^{-1} \right. \\ & \left. (z_i - \boldsymbol{H}_2(X_{i,o})\hat{\boldsymbol{\beta}}_{i,2} - \hat{\boldsymbol{\eta}}(X_{i,o}))\right\}, \end{aligned} \qquad (3.11)$$

where

$$\hat{\boldsymbol{\beta}}_{i,2} = \boldsymbol{W}_{i,2}\boldsymbol{H}_2(X_{i,o})\boldsymbol{V}_{i,\psi_2}^{-1}(z_i - \hat{\boldsymbol{\eta}}(X_{i,o}))$$

$$\boldsymbol{W}_{i,2} = (\boldsymbol{H}_2(X_{i,o})^T\boldsymbol{V}_{i,\psi_2}^{-1}\boldsymbol{H}_2(X_{i,o}))^{-1}.$$

Full mean

$$\boldsymbol{\mu}_{\psi_2} = \boldsymbol{H}_2(X_o)\boldsymbol{\beta}_2 + \hat{\boldsymbol{\eta}}(X_o)$$

where $\hat{\boldsymbol{\eta}}(X_o) = [\hat{\boldsymbol{\eta}}(X_{1,o})^T, \cdots, \hat{\boldsymbol{\eta}}(X_{I,o})]$

Full covariance

$$\boldsymbol{V}_{\psi_2} = \begin{pmatrix} \boldsymbol{V}_{1,\psi_2} & 0 & \cdots & 0 \\ 0 & \boldsymbol{V}_{2,\psi_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{V}_{I,\psi_2} \end{pmatrix},$$

$$p(\boldsymbol{\psi}_2|\boldsymbol{d}, \boldsymbol{\psi}_1) \propto p(\boldsymbol{\psi}_2)|\boldsymbol{V}_{\psi_2}|^{1/2}|\boldsymbol{W}_2|^{1/2}$$
$$\times \exp\left\{-\tfrac{1}{2}(\boldsymbol{z} - \boldsymbol{H}_2(X_o)\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\eta}}(X_o))^T\boldsymbol{V}_{\psi_2}^{-1}(\boldsymbol{z} - \boldsymbol{H}_2(X_o)\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\eta}}(X_o))\right\}, \tag{3.12}$$

where

$$\hat{\boldsymbol{\beta}}_2 = \boldsymbol{W}_2\boldsymbol{H}_2(X_o)^T\boldsymbol{V}_{\psi_2}^{-1}(\boldsymbol{z} - \hat{\boldsymbol{\eta}}(X_o))$$

$$\boldsymbol{W}_2 = (\boldsymbol{H}_2(X_o)^T\boldsymbol{V}_{\psi_2}^{-1}\boldsymbol{H}_2(X_o))^{-1}.$$

Notice the following expression

$$(\boldsymbol{z} - \boldsymbol{H}_2(X_o)\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\eta}}(X_o))^T\boldsymbol{V}_{\psi_2}^{-1}(\boldsymbol{z} - \boldsymbol{H}_2(X_o)\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\eta}}(X_o)) \tag{3.13}$$

Denote $\boldsymbol{u} = \boldsymbol{z} - \hat{\boldsymbol{\eta}}(X_o)$.

Then expression 3.13 can be simplified as

$$[(\boldsymbol{I} - \boldsymbol{H}_2(\boldsymbol{H}_2^T\boldsymbol{V}^{-1}\boldsymbol{H}_2)^{-1}\boldsymbol{H}_2^T\boldsymbol{V}^{-1})\boldsymbol{u}]^T\boldsymbol{V}^{-1}[(\boldsymbol{I} - \boldsymbol{H}_2(\boldsymbol{H}_2^T\boldsymbol{V}^{-1}\boldsymbol{H}_2)^{-1}\boldsymbol{H}_2^T\boldsymbol{V}^{-1})\boldsymbol{u}]$$

$$=\boldsymbol{u}^T[(\boldsymbol{I} - \boldsymbol{V}^{-1}\boldsymbol{H}_2(\boldsymbol{H}_2\boldsymbol{V}^{-1}\boldsymbol{H}_2)^{-1}\boldsymbol{H}_2^T)\boldsymbol{V}^{-1}(\boldsymbol{I} - \boldsymbol{H}_2(\boldsymbol{H}_2^T\boldsymbol{V}^{-1}\boldsymbol{H}_2)^{-1}\boldsymbol{H}_2^T\boldsymbol{V}^{-1})]\boldsymbol{u}$$

$$=\boldsymbol{u}^T[\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{H}_2(\boldsymbol{H}_2^T\boldsymbol{V}^{-1}\boldsymbol{H}_2)^{-1}\boldsymbol{H}_2^T\boldsymbol{V}^{-1}]\boldsymbol{u}$$

$$=\boldsymbol{u}^T\boldsymbol{R}\boldsymbol{u}$$

where $\boldsymbol{R} = [\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{H}_2(\boldsymbol{H}_2^T\boldsymbol{V}^{-1}\boldsymbol{H}_2)^{-1}\boldsymbol{H}_2^T\boldsymbol{V}^{-1}]$

Since $\boldsymbol{V}_{\psi_2}$, $\boldsymbol{W}_2$ and $\boldsymbol{R}$ are symmetric, therefore we can decompose equation 3.12 into the products of patient-wise likelihoods.

3.7.4   Posterior distribution of calibration parameter θ

$$p(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{d}) \propto \prod_{i=1}^{I} p(\boldsymbol{\theta}_i)|\boldsymbol{W}(\boldsymbol{\theta},\boldsymbol{\psi})|^{\frac{1}{2}}$$
$$\times \exp\left[-\frac{1}{2}\{(\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta},\boldsymbol{\psi}))^T\boldsymbol{V}_d(\boldsymbol{\psi})^{-1}(\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta},\boldsymbol{\psi}))\}\right] \tag{3.14}$$

3.7.5   Predictive distribution of true process $\zeta(\cdot)$

Its mean function is given by

$$\mathbb{E}(\zeta(\boldsymbol{x})|\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{d}) =\boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta})^T\hat{\boldsymbol{\beta}}(\boldsymbol{\theta},\boldsymbol{\psi})$$
$$+ \boldsymbol{t}(\boldsymbol{x},\boldsymbol{\theta})^T\boldsymbol{V}_d(\boldsymbol{\psi})^{-1}(\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta},\boldsymbol{\psi})), \tag{3.15}$$

where

$$\boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{h}_1(\boldsymbol{x},\boldsymbol{\theta}) \\ \boldsymbol{h}_2(\boldsymbol{x}) \end{pmatrix},$$

and

$$t(x) = \begin{pmatrix} V_1(x, X_c) \\ V_1\{x, X_o\} + V_2(x, X_o) \end{pmatrix}.$$

Additionally, its covariance function is given by

$$\begin{aligned}
\mathrm{cov}(\zeta(\boldsymbol{x}_i), \zeta(\boldsymbol{x}'_i) \mid \boldsymbol{\theta}_i, \boldsymbol{\psi}, \boldsymbol{d}_i) = \ & k_1(\boldsymbol{x}_i, \boldsymbol{x}'_i) + k_2(\boldsymbol{x}, \boldsymbol{x}') \\
& - \boldsymbol{t}(\boldsymbol{x})^T \boldsymbol{V}_{d_i}(\boldsymbol{\psi})^{-1} \boldsymbol{t}(\boldsymbol{x}', \boldsymbol{\theta}) + (\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta}) \\
& - \boldsymbol{H}(\boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\psi})^{-1} t(\boldsymbol{x}, \boldsymbol{\theta}))^T \boldsymbol{W}(\boldsymbol{\theta})(\boldsymbol{h}(\boldsymbol{x}', \boldsymbol{\theta}) \\
& - \boldsymbol{H}(\boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\psi})^{-1} t(\boldsymbol{x}', \boldsymbol{\theta})).
\end{aligned} \tag{3.16}$$

3.8 Data Analysis

As people are quite familiar with the classical method regression, people are tending to compare statistical methods with regression. In terms of modeling, calibration is a reverse process to regression, where instead of a future dependent variable being predicted from known explanatory variables, a known observation of the dependent variables is used to predict a corresponding explanatory variable. In our application problem, the observation of dependent variables is the Bio-geometrical shape of AAA enlargement, while the corresponding explanatory variables the calibration parameters. In terms of applications, Bayesian calibration is actually resembling the mechanisms of a physical system (here is a biomedical system) by calibrating the internal parameters (unobserved) in a way that the model outputs can enhance the similarity of the model outputs and real observations. But obviously, for example, in medical science, a regression model is sometimes used to identify the significant covariate or treatments. Bayesian calibration integrates the computer model

as an independent source of useful information and expertise besides the real observations, while a regular regression usually just has one data source which is just real observations.

Bayesian calibration combines two sources of data–Computer data and real observations. The latter is usually more expensive to collect than the former one, as a result, computer data has much higher dimension than the real observations. Thus, the patient specific prediction modeling is usually more computationally intensive than the regular regression using one sources of data. And we can imagine, the computation of Bayesian calibration model for multi-subject data can be much more demanding. Therefore, we make two parts of the data analysis. One part is the analysis result for small sample (only two patients). Another part is the analysis result for big sample (which includes 6 patient.)

3.8.1 Small sample

First, I run data analysis on the real observations from two subjects. For each subject, the computer data covers the time span of real data and future inputs. We use the real data to calibrate the model at one time point, while we use future inputs to make predictions at a second time point. Besides, we choose equally spaced 8 points along the center line to describe the AAA shape. The prediction results are shown as follows.

Priors and estimates of hyperparameters are shown in Table 3.1. As we can see, the estimates of ψ_2 are close to the prior means which are actually equal to true values. As shown in Fig. 3.1, all the real observations are captured by the credible intervals and the maximum relative error is 2.5%. The results shown in Fig. 3.2 is not as good as the results in Fig. 3.1. But it is expected as discussed before, there is an acceleration property of AAA expansion which sometimes can not be completely captured by the computer model. As the values of calibration parameters from the two subjects are close, we use the same prior for

both subject 1 and subject 2 (Figures 3.1d and 3.2d).

(a) Predictions of the observation processes. The solid line denotes the observations, while the dashed line denotes the predictions.

(b) 95% credible band of predictions. The blue stars denote the observations lying inside the credible band. The red marks denote an observation value lying outside the credible band.

(c) Relative errors between predictions and true values.

(d) Prior and posterior densities of θ_1 and θ_2.

Figure 3.1. Prediction results for subject 1.

3.8.2 Big sample

The design of Bayesian calibration is given as follows. For each subject, the computer data covers the time span of real data and future inputs. The real data are used to calibrate the model at the first two time points, while future inputs are used to make predictions at
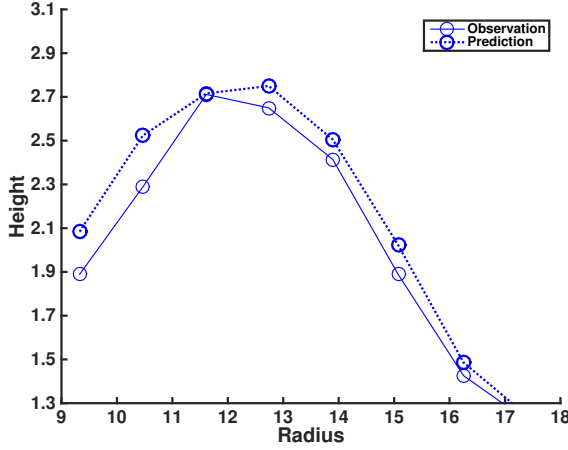
(a) Predictions of the observation processes. The solid line denotes the observations, while the dashed line denotes the predictions.
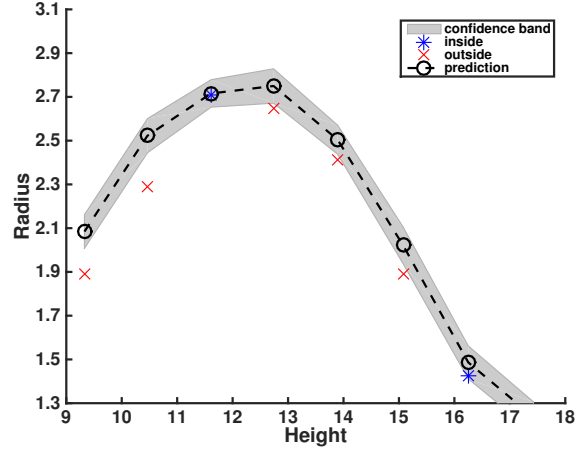
(b) 95% credible band of predictions. The blue stars denote the observations lying inside the credible band. The red marks denote an observation value lying outside the credible band.
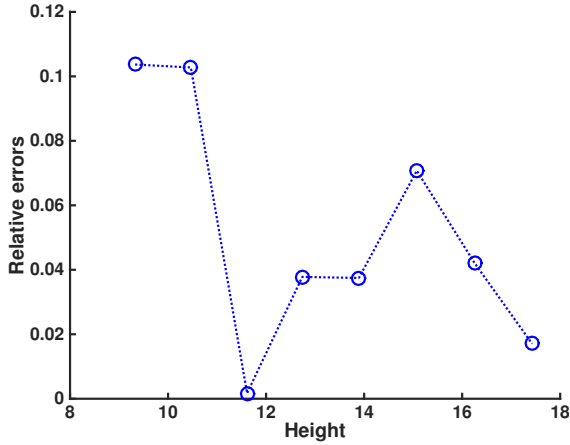
(c) Relative errors between predictions and true values.
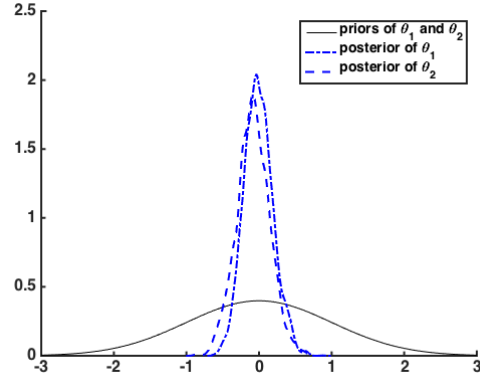
(d) Prior and posterior densities of θ_1 and θ_2.

Figure 3.2. Prediction results for subject 2.

last two time points. Therefore, the computer data totally have 4 time points. Besides, we choose equally spaced 8 points along the center line to describe the AAA shape. The results are shown as follows.

When we have 6 patients, there are totally 12 calibration parameters to be inferred by Bayesian analysis since each patient has two calibration parameters. And we can imaging, drawing samples from an unknown density with 12 parameters is not an easy task to

Table 3.1. Priors and Estimates of Hyperparameters

ψ_1		ω_{x2}	ω_{x2}	$\omega_{\theta1}$	$\omega_{\theta2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.186	1.738	0.0832	0.007	0.640
ψ_2			λ	$\omega_{x1}^{\star}$	$\omega_{x2}^{\star}$	σ_2^2
Lognormal Prior	mean		0.01	1	1	0.1
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.01	1	1	0.001
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.006	0.418	0.146	-0.176	0.053	-0.016

accomplish.

As discussed above, Bayesian calibration is an iterative process of updating uncertainty distributions on the calibration parameters in a way that is consistent with the observed data. Therefore, it is more important for the model to update uncertainty distributions on the calibration parameters by considering all the parameter information of all the subjects. Therefore, we assume all the subjects have the same calibration parameter values with the same distributions. From the computer simulations, we chose values of θ_1 ranging from 0.5 to 0.8 and values of θ_2 ranging from 1 to 9 as the calibration inputs. Then I used normal priors which can cover the range of calibration parameters (See Figure 3.3). Finally the posterior will determine a smaller range of values as the update of the values obtained from computer simulations.

3.9 Conclusion

Based on the results, we can see individualized version of Bayesian calibration is powerful in integrating individual information (uncertainty) of calibration parameters to update the range of them. So these new ranges and distributions of calibration parameters will improve
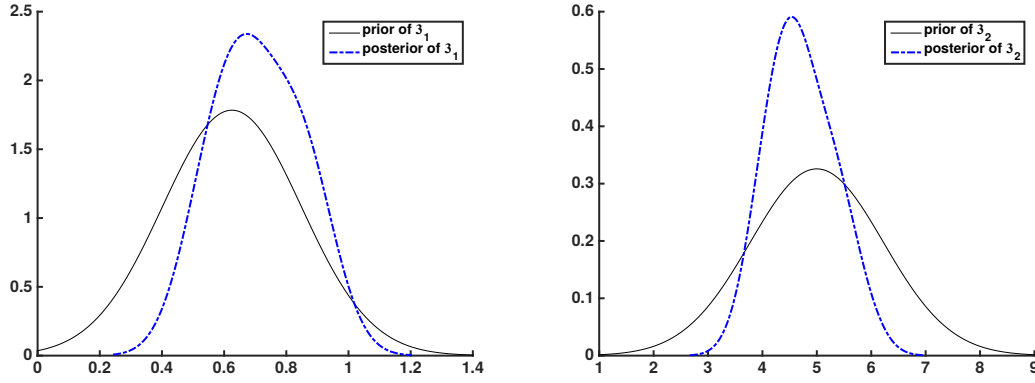
(a) Prior and posterior densities of θ_1 **(b)** Prior and posterior densities of θ_2.

Figure 3.3. Prior and posterior results for the data consisting of 6 patients.

in some way the consistency between estimated and observed, whenever researchers are generating simulations from computer code or making predictions of future observations.

Computational difficulties mainly come from drawing samples of calibration parameters from posteriors by using Metropolis Hasting algorithms. As we can see from equation 3.14, posterior density is not tractable of calibration parameters analytically. Therefore, once we incorporate large number of subjects into the model which means the number of calibration parameters is high, then the computation of high dimensional Metropolis Hasting Algorithms is very hard to converge and control. Therefore, we have to think of better ways to tackle this problems.

Additionally, as introduced at the beginning of this dissertation, we should think of using Latin hypercube sampling to improve the design of Computer Experiments, especially in the Bayesian calibration model for multiple subjects. Because the computational time is increasing significantly when we include more subjects into the model and make predicitons.

Chapter 4

Alzheimer's Disease Progression Analysis Using Multi-state Markov Model

4.1   Introduction

With the rapid aging of the world population, approximately 10 million people worldwide are affected by Alzheimer's Disease. It is a leading cause of death behind cardiovascular disease and cancer. Nearly 10% of people 65 years of age and older are affected by Alzheimer's Disease. While the disease is most common in the elderly, it has been diagnosed in patients in their 40s and 50s.

Alzheimer's Disease (AD) is a progressive, irreversible, degenerative brain disorder, causing impaired memory, thinking and behavior [76]. These impairments are related to the death of brain cells and the breakdown of the connections between them. From the onset of the symptoms, the disease usually progresses over a period of eight years. But the disease can last for up to 20 years. Three stages are defined in the advance of AD, from early signs, mild cognitive impairment to severe dementia. Generally speaking, the early signs of AD appear after the age of 60, which often include loss of recent memory, faulty judgment,

and changes in personality. Individuals in the mild stage may even forget how to do simple routine tasks such as dressing or bathing. In the final stages of the disease patients are bedridden and frequently develop coexisting illnesses. Most commonly, individuals with AD die from pneumonia. Although the risk of developing AD increases with age, AD is quite different from the normal aging process, because AD is a disease, while the normal aging process is not. Without this disease, the human brain often can operate to the age of 100 and beyond.

The research of Alzheimer's Disease has been divided into three broad categories: causes and risk factors, diagnosis, and treatment. Although researchers cannot provide definitive answers to the causes of Alzheimer's Disease until now, they've made progress to approach the truth in many ways. For example, genetic predisposition, abnormal protein deposits in the brain and environmental factors are suspected to play a role in the development of the disease. In order to understand the pathology of the AD brain, scientists need to understand the structure and function of the aging nervous system. Particularly to find the causes and risk factors in AD, they should understand the loss of communication between neurons and the death of neurons in the AD brain. The research on diagnosis research is focused on finding ways to identify markers (indicators) of dementias, develop and improve ways to test patients, determine causes and assess risk factors, and improve methods for finding cases and sampling populations. When it comes to treatment, researchers are working to slow AD's progression, delay its onset, or eventually, prevent it altogether. They are also developing ways to improve a patient's ability to function, and to support caregivers of people with AD. For full descriptions, please refer to the Laboratory of Neuro Imaging website [76].

The data we are using in this paper to understand the structure of the brain system is magnetic resonance imaging (MRI). MRI is currently the imaging modality of choice

for assessing neurologic disorders, because of the exclusive capacity of providing excellent anatomical detail in a safe setting. MRI is an imaging technique based on the principles of nuclear magnetic resonance that detects proton signals from water molecules and that allows us to produce high quality images of the internal structure of the living brain. To distinguish between gray matter, white matter, and cerebrospinal fluid (CSF), certain protocols are designed to create anatomical images of the brain. As a result, MRI is able to differentiate between tissue types. By providing the contrast needed to distinguish these, MRI allows researchers to measure the sizes as well as various other properties of different parts of the brain [71]. Among all the measurement, quantitative assessment of brain volumes, obtained through volumetric MRI, has been increasingly applied in recent years to a wide range of neurologic conditions due to advances in computational technology. Typical current anatomical images of the brain captured by MRI have a spatial resolution of approximately 1 cubic millimeter (mm^3). The volume of an adult human brain is, on average, between 1,131,0001,273,000 mm^3 , with substantial variation between individuals [1]. Although the spatial resolution of modern MRI protocols is very high, $1mm^3$ of cortical gray matter can contain between 10,000 and 60,000 neurons, up to four times as many glial cells per neuron [84], as well as neuronal processes, blood vessels, intracortical myelin and dendritic spines. Measures of brain volumes have been shown to be valid biomarkers of the clinical state and progression by offering high reliability and robust inferences on the underlying disease-related mechanisms [42]. In general, atrophy rates in different brain structures of AD patients are related to deterioration of cognitive performance [34].

During all these years the clinical progression of AD has been studied for subjects in different stages. Subjects with the traditional amnestic MCI (mild cognitive impairment) have higher annual rate of conversion to AD than NC (normal controls) [78]. In particular,

MCI subjects with abnormalities in both CSF and medial temporal lobe has a very high risk of progression to AD [8], where the hippocampal atrophy rate is in the order of 3.5% per year compared to about 1% per year of MCI subjects not converting to AD [28]. However, in addition to structures of the medial temporal lobe, volumes of cortical regions in parietal and lateral temporal lobes are also used to predict the likelihood of progression to AD in MCI subjects [64]. Moreover, an MRI score reflecting the degree of AD-like brain atrophy features showed slightly higher predictive value for future cognitive impairment than CSF measures [105].

In longitudinal medical studies, patients are observed over time and covariate information is collected at several occasions. The analysis in such studies where individuals may experience several events is often performed using multi-state models. A multi-state model (MSM) is a model for a continuous time stochastic process allowing individuals to move among a finite number of states [70]. For example, the survival analysis is actually a two-state model, while the continuous time Markov model is one special kind of multi-state model. Particularly, in our study, each subject is scanned at each visit and thus MRI data are collected longitudinally. The changing volumes in each brain region plotted by these MRI scans can be considered as covariate information to predict how AD individuals progress from early signs, mild cognitive impairment to severe dementia. Commenges, D. [17, 18] have given reviews on the analysis of interval-censored observations from multi-state models. Sukkar, Rafid, et al. [100] used Hidden Markov Model (HMM) to help improve the assessment of Alzheimer's Disease progression, while Wang Ying, et al [107] proposed a spatio-temporal methodology based on Hidden Markov Models (HMM), and apply it on four-dimensional structural brain magnetic resonance imaging series of older individuals. Yu, Hong-mei, et al. [113] applied Multi-state Markov model to explore known risk factors' effects on each of

the possible transitions in the progression of Alzheimer's Disease. Besides, a bunch of other references [14, 35, 43, 93, 101, 118] focus on various research interests by using different prediction techniques. Nevertheless, before doing any modeling of AD progression, two things should be emphasized and understood carefully: measurement of progression and limitations of prediction. Gelb, Douglas J. [37], from a clinician's perspective, reviewed questions that can be answered by the natural course of AD, and specifically, information regarding measures of functional impairment and how they change over time, whereas Barnes, Deborah E. [4] pointed out the limitations of most AD risk prediction strategies and suggested that future techniques should simultaneously model the risk of mortality as well as the risk of AD over the full preclinical spectrum and to consider the potential harm as well as the benefit of identifying and treating high-risk older patients.

In this paper, we applied the multi-state Markov model to resemble the progression of Alzheimer's Disease. We treated the clinical diagnosis as the response variable (state space), which consists of 3 levels (states)– NC (Normal control), MCI (Mild cognitive impairment) and AD (Alzheimer's Disease). Meanwhile, we treated the screening time (the time when MRI scans are collected)) as the timeline along which the Markov chain undergoes transitions from one state to another. Besides, we assumed that transition rates are assumed to be independent of time (i.e., Markov model is homogeneous), but they depend on patient related variables, such as MRI regional volumetric information, cognitive assessment information and demographical information. More specifically, we incorporate them into the Markov model by considering the logarithm of an transition intensity as a linear function of relevant predictors. Then we performed likelihood ratio tests (LRT) to compare which model fits the data better. Finally, we made predictions based on transition rates and transition probabilities by using a supervised learning method. And leave-one-out cross validation results are provided. As

a whole, our contributions are listed as follows. First, for this data set, age is recognized as not a statistically significant risk factor in our Markov model. Second, different sources of information are integrated together to describe the overall trend of the progression of Alzheimer's Disease. Third, transition rates and transition probabilities from the Markov model provide us a potential direction for predicting the course of Alzheimer's Disease.

The remainder of this paper is structured as follows: Section 4.2 introduces how we got and processed the data. Section 4.3 explains in details the multi-state Markov model used for prediction, the Likelihood ratio test (LRT) for comparison and the survival analysis for interpretation. Section 2.5 presents with 4 models, the model without any covariates, the model with a single covariate Age, the model with a single covariate ApoE4 and the model for predictions with all the known risk factors in the data. Section 2.6 summarizes the implications of our findings and provide future perspectives.

4.2 Data

ADNI is an ongoing, longitudinal, multi-center study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's Disease (AD). The ADNI study began in 2004 and included 400 subjects diagnosed with mild cognitive impairment (MCI), 200 subjects with early AD and 200 elderly control subjects. This initial phase of the study is known as ADNI1. In 2009, ADNI1 was extended with ADNI GO which assessed the existing ADNI1 cohort and added 200 participants identified as having early mild cognitive impairment (EMCI). The objective of this phase was to examine biomarkers in an earlier stage of disease. In 2011, as ADNI GO was ending, ADNI2 began. ADNI2 assesses participants from the ADNI1/ADNI GO cohort in addition to the

following new participants–150 elderly controls, 100 EMCI participants, 150 LMCI (late "mild cognitive impairment") participants and 150 mild AD patients.

We downloaded the "adnimerge" data from ADNI in 12/8/2014. The "adnimerge" data merges together several of the key variables from various case report forms and biomarker lab summaries across the ADNI protocols (ADNI1 , ADNIGO, and ADNI2). It is updated daily directly from the Alzheimer's Disease Cooperative Study (ADCS; http://adcs.org/) data system and posted to LONI. We separate out all the 570 subjects in ADNI1 as our target data set. To reduce the impact of unbalanced longitudinal design and characterize the way the response changes in time, we only keep the 456 subjects who have more than 4 observations and remove all the others. Among them, 139 subjects, 105 subjects and 66 subjects stay in NC (Normal control), MCI (Mild cognitive impairment) and AD (Alzheimer's Disease) all the time respectively, while 24 subjects, 8 subjects, 105 subjects and 2 subjects convert from NC to MCI, from MCI to NC, from MCI to AD and from AD to MCI respectively. Besides, 3 subjects convert from NC to MCI to AD. 2 subjects oscillate between NC and MCI, while another 2 subjects oscillate between MCI and AD.

This cleaned dataset has 272 male subjects and 184 female subjects. Each observation on every subject is comprised of 3 sources of information – longitudinal MRI regional volumetric information about each subject including "Ventricles", "Hippocampus", "WholeBrain", "Entorhinal", "Fusiform", "MidTemp", "ICV" (intracranial volume), cognitive assessment information like "MMSE" (Mini Mental State Exam [21]) and "CDR-SB" (Clinical Dementia Rating-Sum of Boxes [12]), demographical information like "Age", "Gender", "Race", "Education", "Marital Status" and one pivotal genetic characteristic "ApoE4". Up to 9 structural MRI scans are available for each subject. The average MRI data collecting time among all the 456 subjects throughout the ADNI1 study is about 3.5 years, while the min-

Table 4.1. Distribution of age in each diagnosis group

Diagnosis	Age					MMSE				
	mean	median	sd	min	max	mean	median	sd	min	max
NC	75	75	5.09	60	90	29	29	1.15	22	30
NC-MCI	76	76	4.58	63	85	29	29	1.26	24	30
MCI-NC	71	74	7.16	62	80	29	29	1.33	24	30
MCI	75	75	7.18	58	87	28	28	2.14	18	30
MCI-AD	73	73	7.09	55	88	24	25	4.09	0	30
AD-MCI	83	83	0.11	83	83	26	26	2.10	22	29
AD	74	74	7.18	56	88	22	23	4.42	5	30

imum is about 1.5 years and maximum is about 8 years. For patients who are diagnosed as NC or MCI at baseline (first scan), MRI scans are collected every half year within the beginning 2 years and every year after 2 years. For the patients who are diagnosed as AD at baseline, MRI scans are collected every half year within the first year and once at the second year, therefore we get 4 time points for subjects staying in AD all the time.

Based on the age of each patient at baseline, the 456 subjects can be divided into 4 age groups – 13 subjects in 50-60, 74 subjects in 60-70, 275 subjects in 70-80 and 94 in 80-90. In other words, almost 60.31% of the subjects are between 70 and 80. This percentage is consistent with the natural history timeline that most Alzheimer's Disease patients were diagnosed between the ages of 70 and 79 [54]. Additionally, distributions of age in each diagnosis group are given in Table 4.1. But we haven't observed any significant differences of age distribution between different diagnosis groups. As a matter of fact, the relationship between brain aging and Alzheimer's Disease (AD) is contentious. One view holds AD results when brain aging surpasses a threshold. The other view postulates AD is not a consequence of brain aging. For more details, please refer to Swerdlow, H [102].

To describe and visualize the all the 456 subjects, I made the longitudinal trace plot (Figure 4.1) and longitudinal interaction plot (Figure 4.2). As shown in Figure 4.1, the percentage of hippocampus volume to ICV is decreasing with age which is known to all.
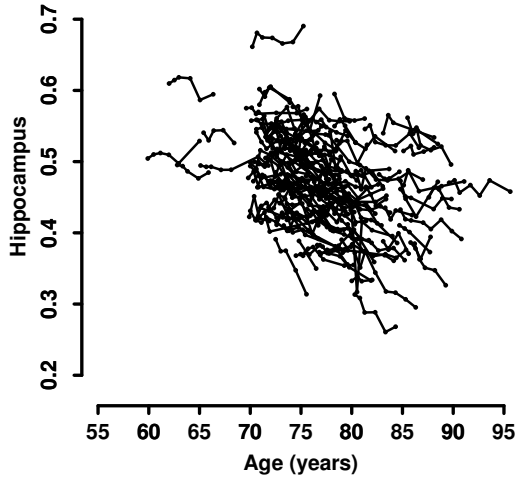
(a) Subjects within NC
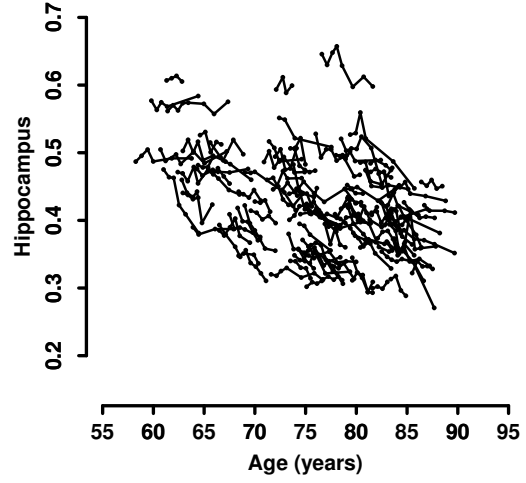
(b) Subjects within MCI
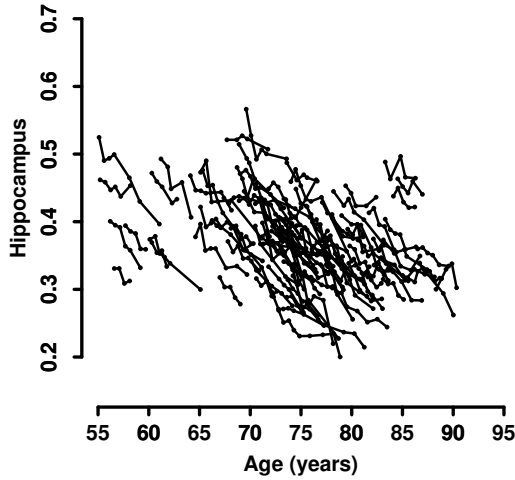
(c) Subjects ever transit from MCI to AD
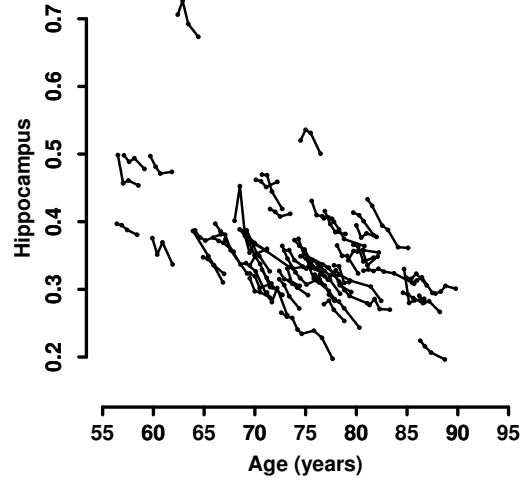
(d) Subjects within AD

Figure 4.1. Individual plot of percentage of hippocampus volume to ICV vs. age. In each graph, each black line segment with solid points in it denotes how the hippocampus of each patient changes with age. The solid point denotes the age when the patient takes screening. The x-axis is the real age of patients. The y-axis is the percentage of hippocampus volume to ICV (intracranial volume).

Besides, the slope of the average decreasing trend of AD subjects is higher than those of

NC subjects and MCI subjects. As shown by Figure 4.2a and Figure 4.2b, the paths of

Hippocampus and MMSE aggregated by age at baseline in different diagnosis groups can

not be clearly separated, while the paths aggregated by screening time (months) can be separated very well in Figure 4.2c and Figure 4.2d. It implies that natural aging process interact with the progression process of Alzheimer's Disease and thus it is better to position the observations on the time axis with screening time since we aim to detect how Alzheimer's Disease affect the transition between different states of severity.

4.3   Methods

A multi-state Markov model is used to model the progression of diseases (see, e.g. Jackson *et al.*, 2003 [53]) which can provide information of disease progression by transition probabilities between different states and mean sojourn time at one state. Specifically, we used a homogeneous model by assuming a transition rate (also called transition intensity), the rate of a transition probability, is independent of time. In other words, this model allows a transition probability that changes over time but the rate remains constant. Generally speaking, the transition rate here can be interpreted as the overall changing speed (or risk) of the progression of Alzheimer's Disease. Commenges, Daniel, et al. [19] proved that for large sample size, the incidence (the number of new cases) in time period Δt approximately equals to the transition rates times Δt. If Δt is the time unit, for instance one year, then they are the same. Therefore, the transition rate is not just a concept in the statistical model, but it does have a direct interpretation in clinical science.

Although transition rates are assumed to be independent of time, actually as we know, changing speed or risk should depend on patient related variables, such as MRI regional volumetric information, cognitive assessment information and demographical information. So we incorporate them into the model by considering the logarithm of an transition intensity

(a) Fraction of hippocampus volum to ICV aggregated by age at baseline

(b) MMSE aggregated by age at baseline

(c) Fraction of hippocampus volume to ICV aggregated by screening months

(d) MMSE aggregated by screening months
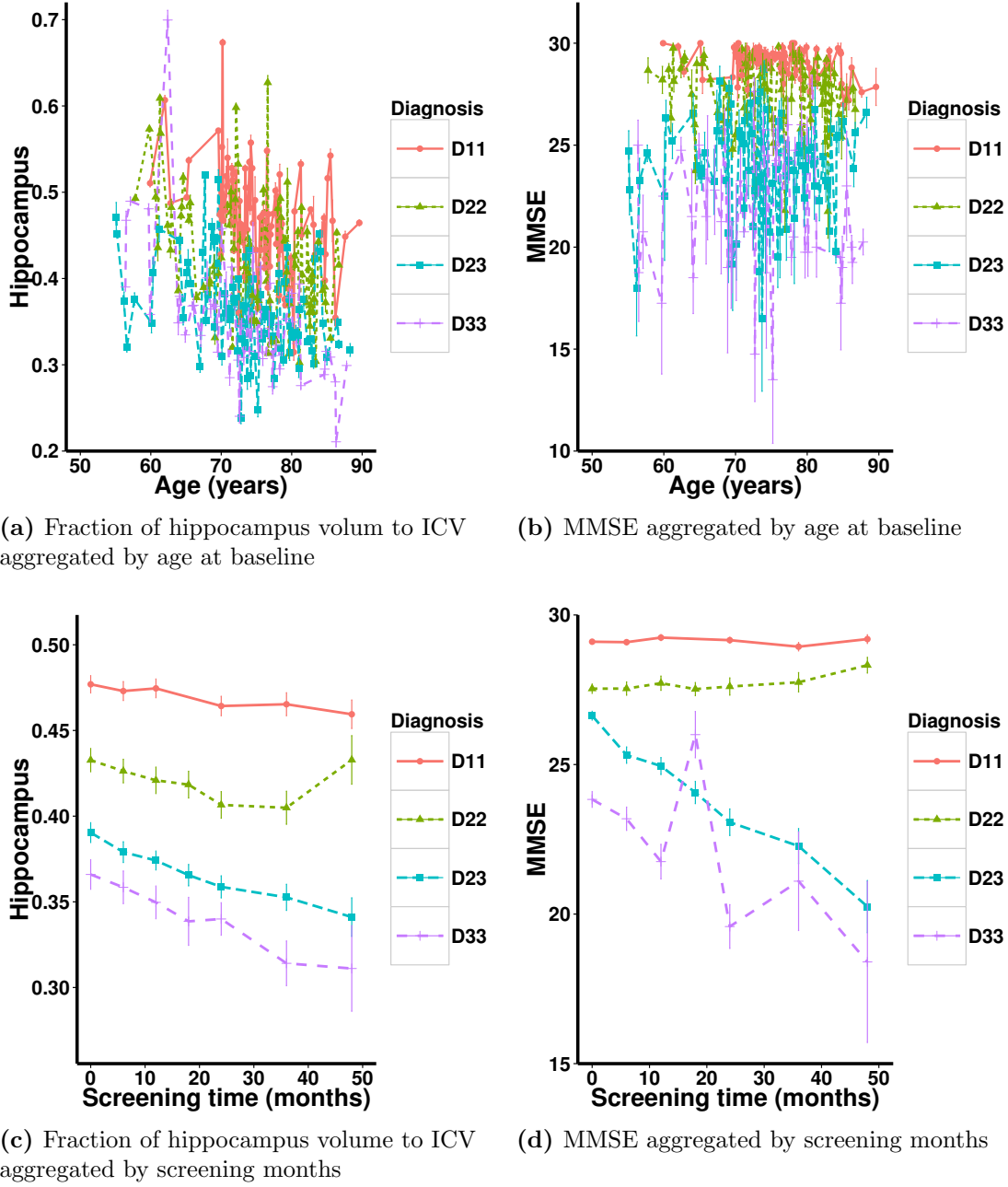
Figure 4.2. Longitudinal plots of fraction of hippocampus volume to ICV and MMSE in 4 different diagnosis groups. 'D11' denotes the group of patients who stay within NC all the time. 'D22' denotes the group of patients who stay within MCI all the time. 'D23' denotes the group of patients who ever transit from MCI to AD. 'D33' denotes the group of patients who stay within AD all the time.

as a linear function of relevant predictors (known as Cox proportional hazard model [67]).

For detailed explanation of a multi-state Markov model, please see Appendices Appendix 1

and Appendix 2.

When comparing the longitudinal observations from different clinical groups, at different aging stages, it is crucial to correctly position the observations on the time axis. This is not straightforward since the disease appears at different ages and chronologically older brains may have greater structural integrity than younger ones affected by the pathology ([66]). And obviously, in our longitudinal study, there are two timelines: screening time and age. Therefore, the way of accommodating the two timelines into our Markov model should be clearly described. Actually, we treat the screening time as the timeline along which the Markov chain undergoes transitions from one state to another, as observations in the original data are aligned by screening time. Meanwhile, we are thinking about including age as a covariate into the Markov model so that the continuous effect of age on the progression of Alzheimer's Disease will be fully identified.

To investigate which model fits the data better, we perform likelihood ratio tests (LRT). It helps to determine whether known risk factors increased the risk of progression from one state to another. The LRT is a statistical hypothesis test in which test statistics is the ratio of two likelihoods (or -2 log likelihood ratio), where the likelihood in the numerator is from the null model and the likelihood in the denominator is from the alternative model. The larger value of the likelihood ratio (or smaller value of -2 log likelihood ratio) indicates the data support the null model.

We can also describe results with a survival probability curve. Survival probability in the multi-state Markov model implies the chance that a patient is not entering the final state (i.e. survived when the final state is 'death'). Note that the final state in this study is not 'death' but instead the clinical diagnosis "Alzheimer's Disease". Compared with Kaplan-Meier (KM) curve [44] which generally models a process with two states and one possible

transition from an alive state to a dead state, a fitted non-AD probability curve from the Markov model is more powerful and flexible in modeling longitudinal data with multiple states and any possible transitions among them.

4.4   Results

In this section, we explore several homogeneous continuous-time Markov models to fit the growth curve of the severity of Dementia (clinical diagnosis) by including different sets of covariates into transition rates . We first consider the model without any covariates, then the model with relevant covariates in it. In particular, we incorporate the real age, ApoE4, MMSE, CDR-SB, and some regional volumetric variables, such as "Ventricles", "Hippocampus", "WholeBrain", "Entorhinal", "Fusiform", "MidTemp", "ICV". The data we used here includes all the subjects whose diagnosis is NC or MCI at baseline, and then become MCI or AD in the end. And we treat AD state as the absorbing state, which means once a subject enters AD state, s/he will never transit to other states.

4.4.1   Model without covariates

A homogeneous Markov model assumes that transition rate from state r to state r', $q_{rr'}$, is independent of time (Appendix Appendix 1). The model without any covariates serves as a baseline model which barely describes how the growth curve of the transition probability changes from previous state to the next state between two screening times. In other words, the transition rates (changing speed of the severity progression) do not depend on the other conditions but remain constant within each pair of transitions across time of observation. The estimate of the baseline intensity $(q_{0,rr'})$ and its 95% confidence interval (CI) are given

in Table 4.1.

Transition rate is the changing speed of transition probabilities with respect to time (equation (A.2) in Appendix Appendix 1). From the Table 4.1, transition from 'NC' to 'MCI' is about 2% (0.0441-0.0269) higher in rate compared to transition from 'MCI' to 'NC'. Moving from 'MCI' state to 'AD' state (0.1968) is about 15% higher in rate compared to moving from 'NC' state to 'MCI' state (0.0441). More clearly, fitted survival probability plots in Figure 4.1 shows that subjects in state NC has the higher fitted survival probabilities (red solid line) than empirical survival probabilities (blue dashed line) while subjects in state MCI has the lower survival probabilities. The empirical survival curve describes the mean trend of Alzheimer's Disease deterioration along with time. In fact, more than half of the subjects never have a chance to convert into MCI or AD, so the lowest value that empirical survival probability curve can reach is around 0.6 rather than 0 in common sense. Note that an empirical survival probability curve is calculated directly from the data without assuming any model, while the homogeneous continuous-time Markov model gives a smooth fitted survival probability curve which is different for the patient in different state as shown in Figure 4.1.

An estimated transition probability describes a chance of a subject moving from one state to another state for the duration of given time. We provide estimated transition probabilities for 2 years and 4 years as shown in Tables 4.1. Overall speaking, compared to year 2, the subjects in year 4 tend to have higher transition probabilities from NC to MCI, from NC to AD, from MCI to AD. In other words, with the time being, naturally there will be more and more subjects converting into MCI and AD. Specially, we will observe more and more complete progression traces that convert from NC to MCI then to AD, although there are only 3 of them right now. More precisely, the result indicates that a chance of staying in

Table 4.1. Table of transition rates and probabilities for the model without covariates.

Transition	Rates(95% Confidence Intervals)	Probabilities(95% Confidence Intervals)	
		Year 2	Year 4
NC-NC	-	0.9176 (0.8836,0.9417)	0.8447 (0.7794,0.8908)
NC-MCI	0.0441 (0.0306, 0.0635)	0.0679 (0.0478,0.0970)	0.1058 (0.0739,0.1477)
NC-AD	0	0.0146 (0.0100,0.0221)	0.0495 (0.0337,0.0719)
MCI-NC	0.0269 (0.0159, 0.0454)	0.0414 (0.0250,0.0716)	0.0645 (0.0383,0.1071)
MCI-MCI	-	0.6410 (0.5858,0.6886)	0.4136 (0.3494,0.4745)
MCI-AD	0.1968 (0.1629, 0.2379)	0.3176 (0.2720,0.3670)	0.5218 (0.4546,0.5856)

the same state is higher than moving to a following progression state within 2 years when a patient is in either 'NC' or 'MCI' (higher remaining in the same state values 0.9176 and 0.6410 than the other values). Once the subject is in MCI' state, there is a 32% chance that s/he will move to the 'AD' state within 2 years. However, chance of moving from MCI to AD (52%) is getting higher than staying in the same state MCI within 4 years, which is reasonable to expect.

The estimated mean sojourn time in Table 4.5 shows that a subject stays in 'NC' state about 23 years while a subject stays in 'MCI' state for about 4.5 years. This implies the progression of Alzheimer's Disease is a slowly changing process especially for the subjects in NC. This is also confirmed by the estimated transition rates in Table 4.1, where transition rates at early states are just less than 5%. But once a subject becomes MCI, the converting speed to AD is much faster (0.2611) and the mean sojourn time staying within MCI is much shorter (4.5 years).

4.4.2   Model with a single covariate

In this section, we will investigate the effects of two main risk factors: age and apolipoprotein E e4 (simplified as ApoE4) by building each of them as a single covariate into two separate models. Details are introduced as follows.
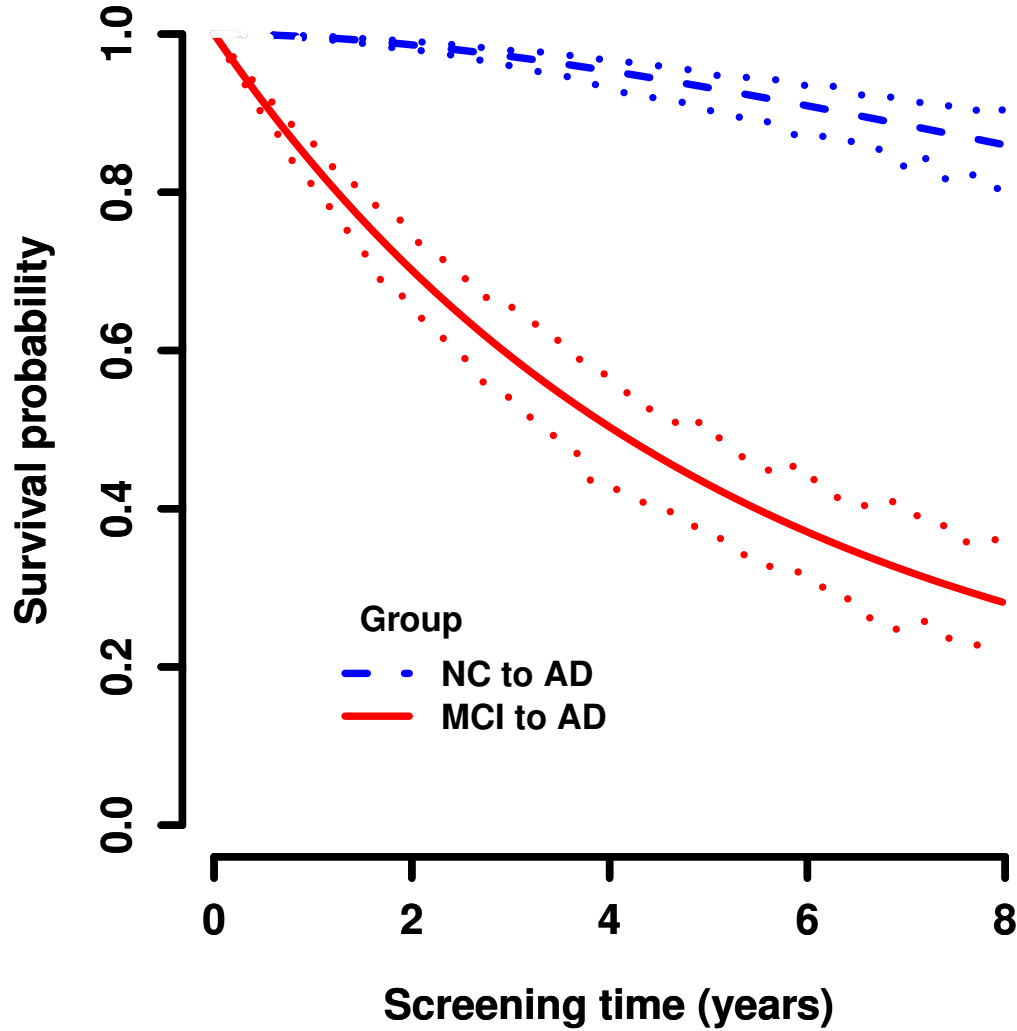
Figure 4.1. Plot of survival curves (not entering to the final state AD). The dashed blue line denotes fitted survival curve from NC to AD. The solid red line denotes fitted survival curve from MCI to AD. The dotted lines denote the 95% confidence band of corresponding fitted survival curve.

4.4.2.1 Age

The major risk factor for Alzheimer's Disease (AD) is age, with a sharp increase in incidence after 60 years (Kawas et al., 2000 [57]). Thus, the inherent link between aging and AD

should be understood thoroughly and clearly. Just as we know, the morphology observed in the brains of patients affected by Alzheimer's Disease (AD) is a combination of different biological processes, such as normal aging and the pathological matter loss specific to AD (Lorenzi, Marco, et al. [66]). Therefore, aging might be a contributing effect that accelerate the progression of Alzheimer's Disease as well as the transitions from mild status (MCI) to severe status (AD). All these considerations motivate us to include age (real age when MRI scans are collected) as a continuous covariate into our Markov model.

To have a brief view of the contributing effect of age, we plot changing curves of fitted transition rates by age for 3 different transition groups. As shown in Figure 4.2, in the average sense, transition rates from NC to MCI are increasing with age which is consistent with the common sense that Brain function deteriorates more likely for older people. But meanwhile, transition rates from MCI to AD are decreasing with age, which actually contradict with the fact. The reason might be due to limitations of cross-sectional and longitudinal study designs [32]. Cross-sectional studies may suffer from cohort effects, and, potentially more seriously, different recruitment bias across age-groups, while for longitudinal data the selective attrition and test-retest effects that may be larger than the change across time points (Salthouse, 2012 [90]). And thus it is not surprising to find that compared to model without any covariates, the model with age is not significant (p-value is 0.1733) according to the likelihood ratio test Table 4.6.

4.4.2.2 ApoE4

The apolipoprotein E (APOE) e4 allele is the most prevalent genetic risk factor for Alzheimer's Disease (AD) (Corder et al., 1993 [20]; Saunders et al., 1993 [92]) and is present in roughly 2025% of North Americans and Europeans (Gerdes et al., 1992 [39]). The presence of an

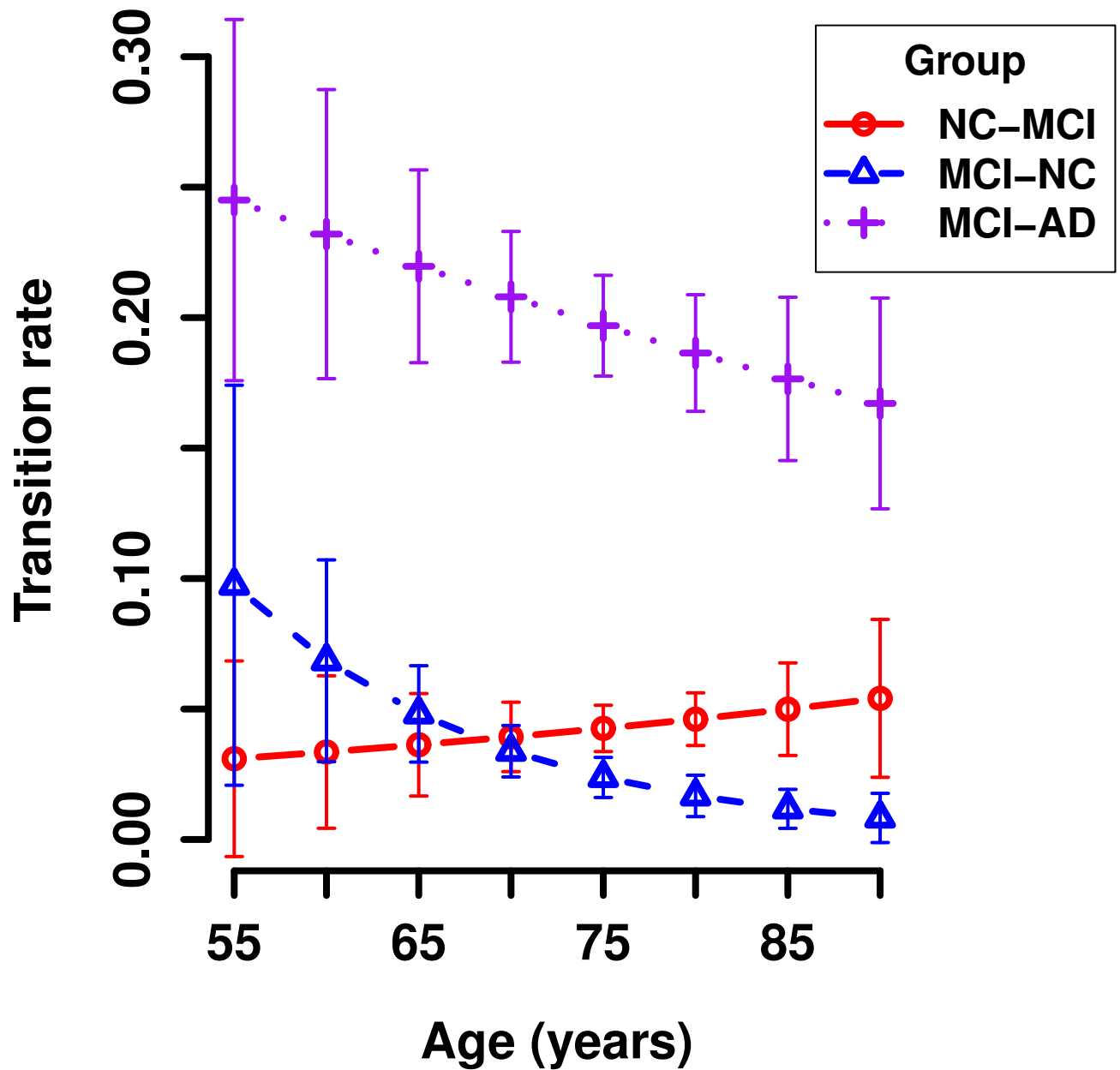**Figure 4.2.** Fitted transition rates changing with age in different groups. The red solid line with circle marks denotes the transition rates from NC state to MCI state. The blue dashed line with triangular marks denotes the transition rates from MCI state to NC state. The purple dotted line with plus marks denotes the transition rates from MCI state to AD state. x-axis denotes the age, while y-axis denotes the transition rates.

e4 allele confers a significantly higher likelihood of developing AD [86]. Shi, Jie, et al. [96] found significant hippocampal morphological deformation in APOE e4 carriers relative to non-carriers in the entire cohort as well as in the non-demented (pooled MCI and control) subjects. Spampinato, Maria Vittoria, et al. [97] found that, in a longitudinal magnetic resonance imaging (MRI) study using voxel-based morphometry (VBM), carriers of the APOE e4 allele with MCI developed atrophy in hippocampus, insula, temporal, and parietal cortex before converting to AD, while structural changes underlying the conversion to dementia in non-carriers did not become apparent.

In our study, ApoE4 is a categorical variable, taking values at 0, 1, 2. We group the ApoE4 into two levels – ApoE4 negative (ApoE4 taking value at 0) and ApoE4 positive (ApoE4 taking value at 1 or 2). Because there are only 7.51% out of all the subjects whose ApoE4 takes value at 2 and we won't get any powerful conclusions if we compare it with other levels (APoE 4 is 0 or 1). Then we include ApoE4 as a two level categorical covariate into our Markov model. As expected from the model fitting, carriers of the ApoE4 (ApoE4 is positive) are more likely to transit from NC to MCI and from MCI to AD. Evidences are shown in Table 4.2, Table 4.3, Table 4.4 and Figure 4.3. In the transition rates table (Table 4.2), ApoE4 carriers (0.0906) have 3 times higher transition rates from NC to MCI than ApoE4 non-carriers (0.0299), while ApoE4 carriers (0.2611) have 2 times higher transition rates from MCI to AD than ApoE4 non-carriers (0.1338). In the table of transition probabilities at year 2 (Table 4.3), ApoE4 carriers (0.1247) have more than 2 times higher transition probabilities from NC to MCI than ApoE4 non-carriers (0.0495), while ApoE4 carriers (0.3973) have almost 2 times higher transition rates from MCI to AD than ApoE4 non-carriers (0.2291). When comparing transition probabilities tables at different years Table 4.4 (year 4) and Table 4.3 (year 2), the same pattern maintains between

Table 4.2. Table of transition rates (95% confidence intervals) for the model with covariate ApoE4.

Transition	ApoE4 is positive	ApoE4 is negative
NC-MCI	0.0906 (0.0535, 0.1527)	0.0299 (0.0180, 0.0495)
NC-AD	0	0
MCI-NC	0.0277 (0.0132, 0.0582)	0.0265 (0.0126, 0.0556)
MCI-AD	0.2611 (0.2067, 0.3297)	0.1338 (0.0969, 0.1848)

Table 4.3. Table of transition probabilities (95% confidence intervals) at year 2 for the model with covariate ApoE4.

Transition	ApoE4 is positive	ApoE4 is negative
NC-NC	0.8383 (0.7422,0.9016)	0.9434 (0.9070,0.9653)
NC-MCI	0.1247 (0.0760,0.1993)	0.0495 (0.0307,0.0825)
NC-AD	0.0369 (0.0220,0.0615)	0.0071 (0.0039,0.0125)
MCI-NC	0.0382 (0.0182,0.0747)	0.0439 (0.0214,0.0900)
MCI-MCI	0.5645 (0.4920,0.6264)	0.7270 (0.6445,0.7889)
MCI-AD	0.3973 (0.3297,0.4699)	0.2291 (0.1684,0.3022)

ApoE4 carriers and non-carriers. And in both tables, transition probabilities from MCI to AD are higher than those from NC to MCI. In mean sojourn time table (Table 4.5), ApoE4 carriers have much shorter mean sojourn time than ApoE4 non-carriers which means that ApoE4 has the effect of accelerating the progression of Alzheimer's Disease. Compared to model without any covariates, the model with APOE4 is significant (p-value is 0.0002) according to the likelihood ratio test Table4.6.

4.4.3   Model with multiple covariates

In this section, we made predictions of possible transitions based on transition probabilities and transition rates by including known risk factors as covariates into the model. The covariates we incorporate into the model are the real age, ApoE4, MMSE, CDR-SB, and some regional volumetric variables, such as "Ventricles", "Hippocampus", "WholeBrain", "Entorhinal", "Fusiform", "MidTemp", "ICV". Note that we believe that the real age contributes in prediction and thus should be included in the model, although it is not recognized

Table 4.4. Table of transition probabilities (95% confidence intervals) at year 4 for the model with covariate ApoE4.

Transition	ApoE is positive	ApoE is negative
NC-NC	0.7076 (0.5558,0.8161)	0.8922 (0.8312,0.9331)
NC-MCI	0.1750 (0.1101,0.2610)	0.0827 (0.0505,0.1278)
NC-AD	0.1175 (0.0732,0.1885)	0.0251 (0.0143,0.0423)
MCI-NC	0.0536 (0.0240,0.1111)	0.0734 (0.0366,0.1495)
MCI-MCI	0.3235 (0.2411,0.3969)	0.5307 (0.4331,0.6196)
MCI-AD	0.6229 (0.5423,0.7087)	0.3959 (0.3069,0.4989)

Table 4.5. Table of mean sojourn time (95% confidence intervals) for the model with covariate ApoE4.

Transition	Model without covariates	Model with covariate ApoE4	
		ApoE is positive	ApoE is negative
NC	22.6864 (15.7601, 32.6567)	11.0653 (6.5486,18.6972)	33.5035 (20.1918,55.5913)
MCI	4.4695 (3.7402, 5.3411)	3.4632 (0.3939, 2.7712)	6.2389 (4.6385, 8.3914)

(a) ApoE4 is positive   **(b)** ApoE4 is negative
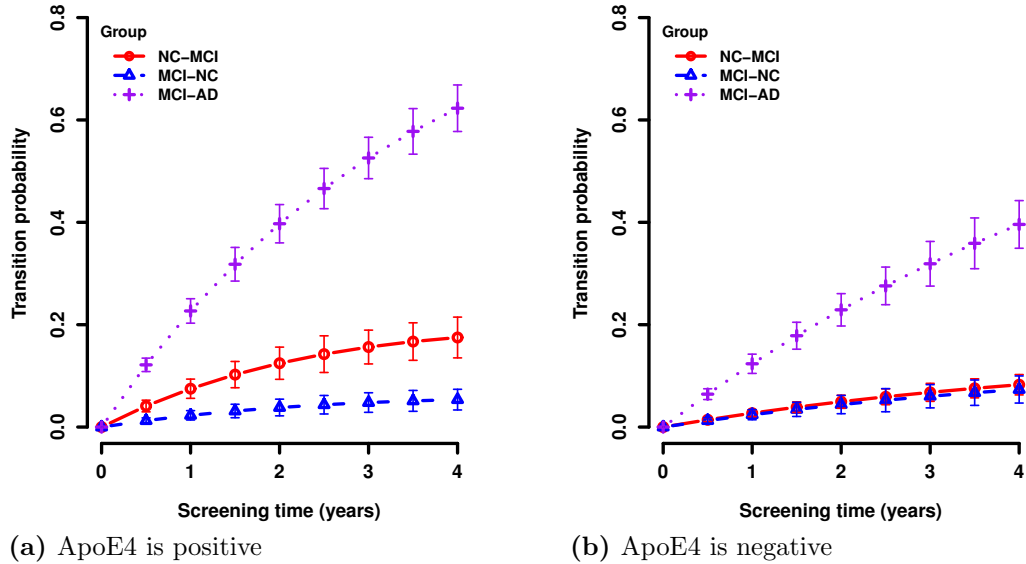
Figure 4.3. Transition probabilities changing with time for the model with covariate ApoE4. The red solid line with circle marks describes transition probabilities from NC state to MCI state. The blue dashed line with triangle marks describes transition probabilities from MCI state to NC state. The purple dotted line with plus marks describes transition probabilities from MCI state to AD state. x-axis denotes the scanning time, while y-axis denotes the transition probabilities. The left panel describes the ApoE4 carriers (positive), while the right panel describes the ApoE4 non-carriers (negative).

as a significant risk factor in the previous section (p-value is 0.1733 in the Likelihood Ratio Test).

Table 4.6. Likelihood ratio test between different models.

Models			Age	ApoE4	Volumes	Scores	Full
without covariates	-2 log LR		4.9796	19.7598	69.10305	129.8185	166.6650
	df		3	3	18	6	30
	p-value		0.1733	0.0002	0	0	0
Volumes	-2 log LR		-64.1235	-49.3432	-	60.7155	97.5619
	df		15	15	-	12	12
	p-value		1	1	-	0	0

4.4.3.1 Prediction and validation

As described in the previous sections, MCI subjects have much higher risk of developing into AD. Therefore, we are more interested in making predictions of transitions from MCI to AD (only two- state model). The data we used here includes only the subjects whose diagnosis is MCI at baseline, and then become AD in the end, which means we are not going to predict the transition from NC to MCI. And we still treat AD state as the absorbing state.

Cross validation is performed by processing the testing set with the trained Markov model. More specifically, a general two-step training and testing procedure is described as follows. All the transition rates and probabilities here stand for the transition from MCI to AD, and the logistic regression is just used as a tool for discriminating observations with lower or higher transition rates and transition probabilities into MCI state or AD state. First, our Markov model is trained by using the training set. Then we collect the outputs including the estimated transition rates and transition probabilities. Afterwards, a logistic regression model of diagnosis states (MCI or AD) was trained on the estimated transition rates and transition probabilities. Second, by plugging the testing set into the trained Markov model, we get the new estimated transition rates and transition probabilities. Then, we plug them into the trained logistic regression model, and get the predicted diagnosis state. Note that predictions are performed within each subjects, and it is one time step prediction, in

other words, each subject's information in the previous scanning time is used to predict the diagnosis state in the next scanning time.

As known to all, correct registration and alignment of subjects on the timeline is crucial to longitudinal medical studies. Therefore, to investigate how registration and alignment will affect prediction of Alzheimer's Disease progression, we prepared two data sets for cross validation, Data1 and Data2. Data1 keeps all the observations before AD state for each subject so that all the subjects are aligned to the baseline of scanning time and the numbers of replications (observations) of different subjects are unbalanced. Data2 only keeps two observations before AD state for each subject so that all the subjects are aligned to the last scanning time point (AD state) and the numbers of replications (observations) of different subjects are balanced (totally 3 observations). And we believe that, in Data2, subjects are sharing more similarities in structural volumetric changing patterns during the two time points before becoming AD state. But things are usually not as perfect as expected, because the two time points before AD state can not be fully aligned for all the subjects. The reason is that, as prespecified by ADNI, for subjects who are diagnosed as NC or MCI at baseline (first scan), MRI scans are collected every half year within the beginning 2 years and every year after 2 years. Thus it is highly likely that, within two years before AD, one subject is observed every half year while the other is observed every year. Furthermore, based on the analysis from previous section, we know that subjects with higher transition rates and transition probabilities tend to develop higher risk of accelerating the deterioration of the Alzheimer's Disease. This phenomenon motivates us to divide the subjects in each data sets (Data1 and Data2) into groups according to the scale of transition rates and transition probabilities. And Group1 (Table 4.7) stands for the higher risk group, while Group2 (Table 4.7) stands for the lower risk group. The performance of prediction between the two groups should be

100

different, just as they should be treated in different ways in clinical treatment.

Table 4.7 presents us with all the leave-one-out cross validation results for different groups in different data sets. Note that the true positives here stand for the true AD is predicted as AD, while the true negatives here stand for the true MCI is predicted as MCI. Comparing the Overall model based on Data1 with the Overall model based on Data2, the latter has a bigger sensitivity (0.5158 vs 0.4000), a smaller specificity (0.6947 vs 0.8930) and a smaller accuracy (0.6053 vs 0.7313). It implies that there are more cases that true MCI or AD are predicted as AD (for Data2, both true positives and false positives are increasing). In other words, the trained model based on Data2 is more likely to transit to AD than trained model on Data1. We consider two reasons. One reason is quite clear. On average, there are more MCI state observations in Data1 than in Data2, thus the observations of transitions from MCI to AD is relatively diluted in Data1. Hence, it turns out that the trained Markov model based on Data1 reduces the rates and probability of transitions. Another reason is that, for Data2, there are only 3 observations for each subject and also observations for MCI state are not perfectly aligned on the timeline as explained in the last paragraph. These in some sense will amplify the effect of disalignment on increasing the number of false positives and false negatives. When comparing Group1 with Group2 in either data set, Group1 has much higher sensitivity, specificity, precision and accuracy. It implies that transition rates and transition probability not only provides us integrated information to do prediction, but also provides a possible way to cluster individuals into different groups sharing similarities in progression of brain functional impairment.

Table 4.7. Prediction of two data sets with different alignments in time. Data1 includes all the observations before AD for each subject, while Data 2 includes only 2 observations before Ad for each subject. Data1 is divided into Group1 and Group2. Group1 in Data1 includes all the patients whose transition probability is ever more than 0.45 or whose transition rate is more than 1.2. Also, Data2 is divided into Group1 and Group2. Group1 in Data2 include all the patients whose transition probability is ever more than 0.6 or whose transition rate is more than 1.2.

Data1	Subjects	Sensitivity	Specificity	Precision	Accuracy	AUC
Group1	51	0.5098	0.9008	0.6842	0.7849	0.8583
Group2	54	0.4074	0.8404	0.5946	0.6824	0.6843
Overall	105	0.4000	0.8930	0.6461	0.7313	0.7404
Data2	Subjects	Sensitivity	Specificity	Precision	Accuracy	AUC
Group1	59	0.7119	0.6610	0.6774	0.6864	0.7672
Group2	36	0.5556	0.6389	0.6061	0.5972	0.5829
Overall	95	0.5158	0.6947	0.6282	0.6053	0.6504

4.4.3.2 Impact of transition rates and transition probabilities

To have a clearer picture of how both transition rates and transition probabilities work as integrated predictors helping with predictions in progression of Alzheimer's Disease, we are trying to make prediction on the original data set. The training set we used here includes all the subjects whose diagnosis is NC or MCI at baseline, and then become MCI or AD in the end. And we treat AD state as the absorbing state. The prediction set we used here is the most original one which allows transitions from AD to AD. After training our Markov model by using the training set, we can get one time step predictions of transition rates and transition probabilities (from NC to MCI and from MCI to AD) based on the testing set.Then we aggregate them by screening time (months) and real diagnosis state and get the longitudinal plots of predicted transition rates and probabilities in different diagnosis groups (Figure 4.4). Note that we took logarithm of transition rates, since they are in a wide range. When comparing plots of predicted transition probabilities from NC to MCI (Figure 4.4b) with plots of predicted transition probabilities from MCI to AD (Figure 4.4d), we observe that neither the diagnosis group 'D23' nor 'D33' separated very well from all the

other diagnosis groups in Figure 4.4b, which tells us that when doing prediction of transition from MCI to AD, we'd better use the information of transition intensities from MCI to AD instead of from NC to MCI. The reason of the difference is obvious that transition probabilities from NC to MCI mostly reflects the overall risk in the progression of disease severity from NC state to MCI state. In the diagnosis group 'D23' and 'D33', with the time being, the transition probabilities from NC to MCI will be decreasing finally since most of subjects in 'D23' and 'D33' have been getting involved in converting from MCI or AD to AD state instead of converting from NC to MCI. To the opposite, transition probabilities within four groups in Figure 4.4d have a continuously increasing trend, since all the diagnosis states will convert to AD state (with the order from NC to MCI to AD) in the end. When comparing plots of Predicted transition rates from MCI to AD (Figure 4.4c) with plots of predicted transition probabilities from MCI to AD (Figure 4.4d), we find that in both cases 'D23' and 'D33' are not separated very well with each other, which is consistent with the fact that the boundary in the assessment of diagnosis between transition from MCI to AD and always staying in AD are not very clear. However, all the other diagnosis group 'NC', 'MCI', 'AD' are separated very well, which indicates that transition rates and transition probabilities could help to discriminate transition states and thus predict progression of brain functional impairment. To put it differently, our Markov model is capable of integrating different sources of information into two simple indices to resemble and predict the mechanism of progression of Alzheimer's Disease. One index is the transition rate, which reflects the overall speed of progression. Another is the transition probability, which tells us the possibility of a certain transition at a given time.
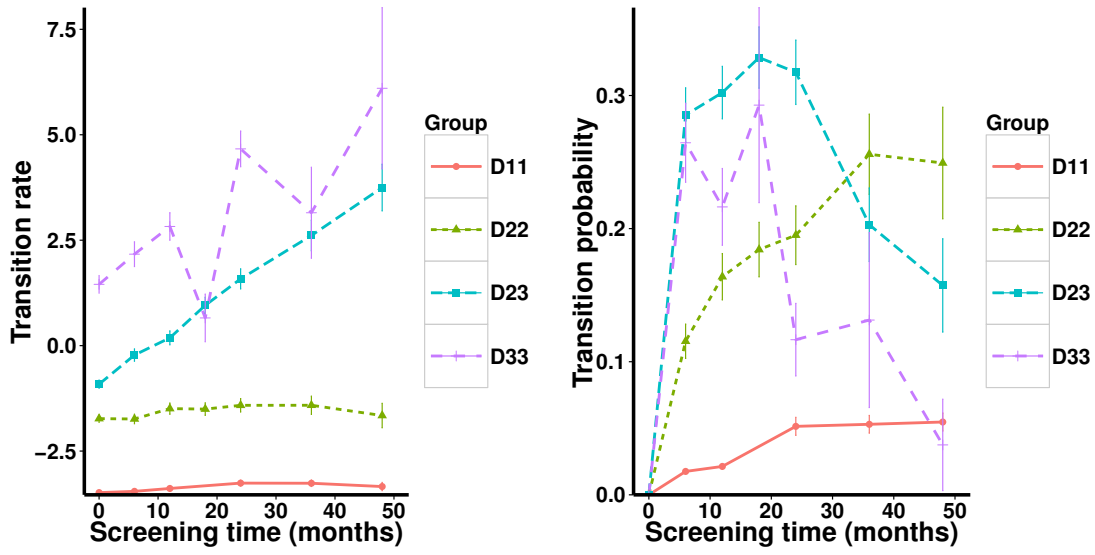
(a) Predicted transition rates from NC to MCI by screening months **(b)** Predicted transition probabilities from NC to MCI by screening months
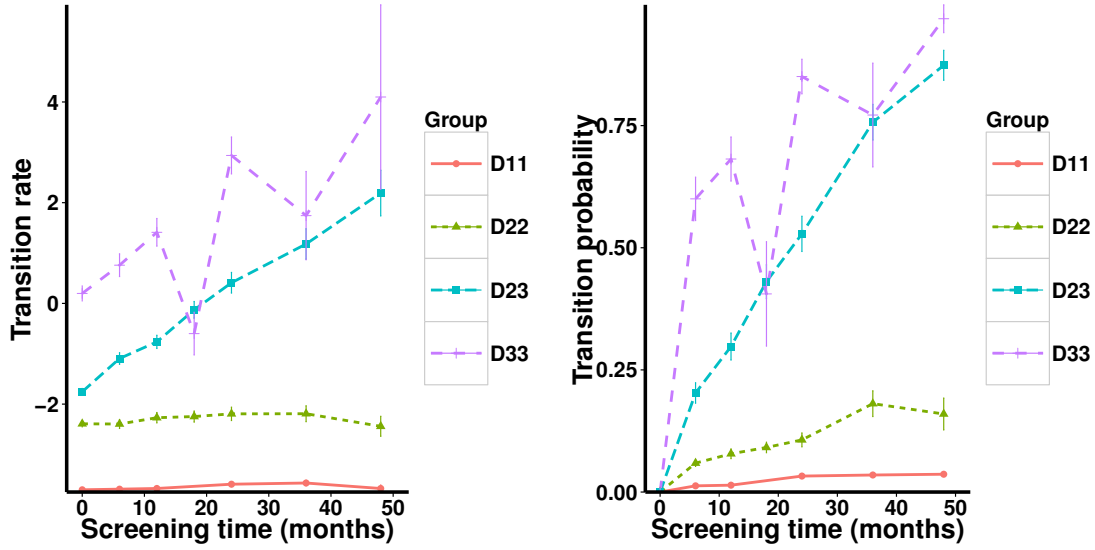
(c) Predicted transition rates from MCI to AD by screening months **(d)** Predicted transition probabilities from MCI to AD by screening months

Figure 4.4. Longitudinal plots of predicted transition rates and probabilities in different diagnosis groups. 'D11' denotes the group of patients who stay within NC all the time. 'D22' denotes the group of patients who stay within MCI all the time. 'D23' denotes the group of patients who ever transit from MCI to AD. 'D33' denotes the group of patients who stay within AD all the time.

4.5 Conclusion and discussion

When positioning the observations on the time axis, we found that aligning the observations by screening time instead of age provides us more distinct pathways to differentiate the transition pattern between different transition groups (diagnosis). Technically, we treated the screening time as the timeline along which the Markov chain undergoes transitions from one state to another. Meanwhile, we included age as a covariate into the Markov model to make prediction so that the continuous effect of age on the progression of Alzheimer's Disease will be fully identified.

Results from model without covariates imply the progression of Alzheimer's Disease is a slowly changing process especially for the subjects in NC. The transition speed from MCI to AD is much faster than the transition speed from NC to MCI or from NC to AD, which means MCI subjects take higher risk to develop into Alzheimer's Disease. Actually, the average sojourn time staying MCI is 4.47 year. In other words, after 4.47 years on average, MCI subjects will transit to other states, most likely to AD. Compared with Kaplan-Meier (KM) curve which generally models a process with two states and one possible transition from an alive state to a dead state, a fitted non-AD probability curve from the Markov model is more powerful and flexible in modeling longitudinal data with multiple states and any possible transitions among them.

Compared to model without any covariates, the model with age is not significant (p-value is 0.1733), while the model with ApoE4 (p-value is 0.0002), the model with Volumes (p-value is 0), the model with scores (p-value is 0) and the full model (p-value is 0) are all significant. Compared to model with volumes, the model with age (p-value is 1) and the model with ApoE4 (p-value is 1) are not significant, while the model with cognitive testing

105

scores (p-value is 0) and the full model (p-value is 0) are significant.

Besides survival probability and mean sojourn time, Markov Chain model gives us two main products can be used to do predictions: transition rates and transition probabilities. First, we fitted the Markov model and got the corresponding transition rates and transition probabilities and then use some machine learning methods to built the connection between the two main products and the transition states. From the leave-one-out cross validation results, the highest AUC (area under curve) we can get is 0.8583.

The key property of a continuous-time Markov chain model is that the state of the next step only depends on the state information on the previous step. Because of this property, Markov chain is more powerful in making one step prediction instead of multiple steps prediction. What is more, the longitudinal data we used here is obviously interval censored. We didn't design specially a scheme to model the interval censoring problem in this paper. There are some other limitations. On the one hand, the results and findings are based only on this particular data set. On the other hand, this is the largest available longitudinal study on ADNI1, but we didn't consider the data from ADNIGo or ADNI2.

APPENDICES

Appendix 1 Multi-state Markov model

Let $\{S(t), t \geq 0\}$ be a multi-state Markov process with n different states. Let $\mathcal{S}$ be the set of states. For simplicity, assume that $\mathcal{S} = \{1, \cdots, n\}$. The process is Markovian if for all $t, t' > 0$ and $r, r' \in \mathcal{S}$,

$$P(S(t+t') = r' \mid S(t) = r, \mathcal{F}_u, u < t) = P(S(t+t') = r' \mid S(t) = r), \qquad \text{(A.1)}$$

where $\mathcal{F}_u$ is the observation history of the Markov process up to time u. That is, the state of the process at a future time $t + t'$ given the previous history of the process up to the present time t depends only through the state of the present time t. Given the initial distribution of the process at time 0, a multi-state Markov process can be characterized by the matrix of transition probabilities, $\boldsymbol{P}(t, t+t') = ((p_{rr'}(t, t+t')))$, where $p_{rr'}(t, t+t') = P(S(t+t') = r' \mid S(t) = r)$. Note that the sum of each row of $\boldsymbol{P}(t, t+t')$ is always 1. Alternatively, we can introduce the matrix of transition intensities, $\boldsymbol{Q}(t) = ((q_{rr'}(t)))$,

$$q_{rr'}(t) = \lim_{t' \to 0} [P(S(t+t') = r' \mid S(t) = r)/t']. \qquad \text{(A.2)}$$

$q_{rr'}$ can be interpreted as the instantaneous hazard of progressio n to stage r' from stage r (see, e.g. Jackson *et al.*, 2003 [53]). If the process $S(t)$ is assumed to be homogeneous in time, $q_{rr'}(t)$ is independent of t.

Equation (A.2) implies that the intensity matrix $\boldsymbol{Q}$ behaves like the derivatives of transition probability matrix $\boldsymbol{P}$. Thus, the sum of each row of $\boldsymbol{Q}$ is 0 so that $q_{rr}(t) = -\sum_{r' \neq r} q_{rr'}(t)$ for each $r \in \mathcal{S}$. From the Chapman-Kolmogorov equations, we can show that $\boldsymbol{P}(t) = \exp(t\boldsymbol{Q})$ for a homogeneous continuous-time Markov process (see, e.g. Whitt, 2012, [109]).

It is of interest to incorporate additional information of each subject in the model for AAA progression. This can be done by introducing a regression component into the intensity matrix $\boldsymbol{Q}$. Let $X(t)$ be the set of explanatory variables available at time t. Then, we model the intensity matrix $\boldsymbol{Q}$ by

$$q_{rr'}(t, X(t)) = q_{0,rr'} \exp\left(\boldsymbol{\beta}_{rr'}^T X(t)\right), \tag{A.3}$$

where $\boldsymbol{\beta}_{rr'}$ is the set of regression parameters corresponding to the explanatory variable $X(t)$ for $r \neq r'$. Note that we consider positive $q_{rr'}$ for $r \neq r'$ while $X(t)$ may be not. Thus, an exponential function is natural to link these two quantities. q_{rr} will be specified using $q_{rr}(t) = -\sum_{r' \neq r} q_{rr'}(t)$. $q_{0,rr'}$ can be interpreted as a baseline intensity when $X(t)$ does not affect the process. Note that the term of 'homogeneous' in the homogeneous Markov chains describes that the transition intensities q_{rs} are independent of time t. Also in the homogeneous Markov model, the sojourn time in each state r is exponentially distributed with mean $-1/q_{rr}$ and the probability that an individual in state r moves to state s is $-q_{rs}/q_{rr}$. For more details, please refer to [75].

Appendix 2 Specification for AAA progression

For AAA stages, we consider the following intensity matrix,

$$\boldsymbol{Q} = \begin{pmatrix} -q_{12} & q_{12} & 0 & 0 \\ 0 & -q_{23} & q_{23} & 0 \\ 0 & q_{32} & -(q_{32}+q_{34}) & q_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{A.4}$$

The structure in (A.4) represents a multi-state model. Also, the model assumes that the last stage is an absorbing stage, i.e. one can not go back to the other stage once it enters that absorbing stage. Note that once an AAA has been treated with a surgery or been ruptured, state 4 (fatal) can not go back to the previous state. Thus, the absorbing stage assumption is reasonable for AAA data.

We estimate q_{12}, q_{23}, q_{32} and q_{34} using the maximum likelihood method via EM algorithm (Jackson *et al.*, 2003 [53]). For the EM algorithm, we need to set the initial values. Mean sojourn time, $-1/q_{rr}$, can be approximated by

$$\frac{\text{total years spent in state } r}{\text{total transitions from state } r},$$

so that we can obtain initial values for $q_{rr'}$ from $q_{rr} = -\sum_{r' \neq r} q_{rr'}$. For AAA data, the mean period in state 1 before moving to state 2 is about 2 year. Thus, we approximate $q_{11} = -2$ and $q_{12} = 0.5$ for their initial values.

Jackson (2011) [16] developed the **R** package, **msm** (available at `www.r-project.org`) for a multi-state hidden Markov model. We use this package for out analysis.

Chapter 5

Conclusions, Discussion, and Directions for Future Research

5.1 Conclusions and discussion

Data analytics is playing a more and more crucial role in this information explosion era. Tons of thousands of events, situations, systems and phenomenon are waiting for statisticians to explore. All kinds of efforts and intelligence have been put into exploring the intractable sources of uncertainties, the complex structures hidden behind observations, the unknown interactions and dependencies across space and time, the ultra-sparse signals immersed in noises, the evolutionary path of a high dimensional network and so much other information.

In this dissertation, Bayesian calibration is used to explore the intractable sources of uncertainties and the complex structures hidden in the system of AAA enlargement. Markov model is used to explore the unknown interactions and dependencies across space and time in the Brain images of patients having Alzheimer's Disease. But still many research questions and aspects need to be taken care of in the future. Although Bayesian calibration has been widely applied in analyzing big and complex data from Engineering, clinical sciences, climate change and Physics, etc., theoretical results of convergence and consistency has not been fully developed yet. For Brain imaging analysis, although we used Markov chain to analyze the

regional volumes getting from MRI of patients having Alzheimer's Disease, further effort should be dedicated to other types of data such as genetic data, PET scan, DTI, fMRI, etc. and related research questions.

Another important concept that has not been discussed is dimension reduction. To tackle the high dimensionality, variability and complexity of big data, I mainly seek two directions of statistical research to approach innovative theories and methodologies that imply huge impacts in data science–hierarchical modeling and dimension reduction (penalization).

Therefore, the future research is mainly focused on three aspects: consistency theory of Bayesian calibration, Bayesian variable selection, dynamic network and sparsity regulation.

5.2   Directions for future research

5.2.1   Bayesian consistency

In most cases, the true value of the calibration parameters cannot be measured or observed physically. Therefore, it is necessary to define the true parameter value and then provide a theoretical study showing that the estimate of the calibration parameter converges to the true value in probability. Rui Tuo and C. F. Jeff Wu 2016 [103] has provided a theoretical framework for calibration in computer models by considering simplifications like dropping the prior. So first I would like to introduce Rui Tuo and C. F. Jeff Wu's work as follows, and then propose the future directions of research.

Real observation and computer model are related through

$$z^p(x) = r^s(x, \theta_0) + \delta(x) + e, \tag{5.1}$$

where θ_0 is the true value of the calibration parameter and $\delta(\cdot)$ is an unknown discrepancy function. Suppose the physical system is deterministic, then the observational error is neglected and we get

$$z^p(x) = r^s(x, \theta_0) + \delta(x), \tag{5.2}$$

where θ_0 is the true value of the calibration parameter and δ is the discrepancy between z^p and $r^s(\cdot, \theta_0)$. Usually δ is a nonzero function because the computer code is built based on certain assumptions or simplifications which do not match the reality exactly. Define

$$\varepsilon(x, \theta) := z^p(x) - r^s(x, \theta), \tag{5.3}$$

The discrepancy function δ is a realization of a Gaussian process with mean 0 and the covariance function $\sigma^2 \Phi$, where σ^2 is an unknown parameter and the function Φ is known.

We adopt the L_2 norm in defining the distance. The L_2 distance projection of θ is given by

$$\theta^\star = \text{argmin}_{\theta \in \Theta} \|\varepsilon(\cdot, \theta)\|_{L_2(\Omega)}, \tag{5.4}$$

The L_2 norm comes from the common use of the quadratic loss criterion. The expected quadratic loss between the prediction and the computer model given θ is

$$\int_\Omega (z^p(x) - r^s(x, \theta_0))^2 dx = \|\varepsilon(\cdot, \theta)\|_{L_2(\Omega)}$$

Because we don't know the true value of θ, the authors define the true calibration parameter as the value $\theta^\star$ minimizes the average predictive error **??**.

Maximum likelihood estimation (MLE) is used to estimate (θ, σ^2) instead of Bayesian

analysis. Under these assumptions, the functions $\varepsilon(x_i, \cdot)$ is known for $i = 1, \ldots, n$. Then it can be shown that the log-likelihood function for (θ, σ^2) is given by

$$l(\theta, \sigma^2; Y) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |\Phi| - \frac{1}{\sigma^2} \varepsilon(\boldsymbol{x}, \theta)^T \Phi^{-1} \varepsilon(\boldsymbol{x}, \theta), \tag{5.5}$$

where $\varepsilon(\boldsymbol{x}, \theta) = (\varepsilon(x_1, \theta), \ldots, \varepsilon(x_n, \theta))^T$ and $\Phi = (\Phi(x_i, x_j))_{ij}$. And the authors discuss the consistency of the maximum likelihood estimate to the defined true parameter.

I would like to generalize the above theory from the frequency setup into Bayesian setup. Following the same logic, first I need to define what is the true value. Define true process as

$$\zeta(\cdot) = r^s(\cdot, \theta) + \delta(x),$$

The conditional posterior predictive density of true process is $f(\zeta(x)|\theta, \hat{\psi}, d)$, while the posterior distribution of parameter θ is $f(\theta|\hat{\psi}, d)$. I would like to show that when $f(\zeta(x)|\theta, \hat{\psi}, d)$ is consistent to $f(\zeta(x)|\theta_0, \hat{\psi}, d)$, then $f(\theta|\hat{\psi}, d)$ is consistent to $f(\theta_0|\hat{\psi}, d)$.

Two popular methods to prove Bayesian consistency are Kullback-Leibler divergence and Hellinger consistency. Stephen Walker, 2004 [106] derives sufficient conditions for both Hellinger and Kullback-Leibler consistency and proves equivalence between them.

Hellinger distance is defined as

$$d_H(f, f_0) = \left\{ \int (\sqrt{f}) - \sqrt{f_0} \right\}^{1/2} = \left\{ 2(1 - \int (\sqrt{f f_0})) \right\}^{1/2}$$

For Hellinger consistency, the required result we are aiming for is

$$\Pi^n(\{f : d_H(f, f_0) > \epsilon\}) \to 0 \ a.s. \ [F_0^\infty]$$

114

Kullback-Leibler divergence between f and f_0 is $d_K(f, f_0) = \int f_0 \log(f_0/f)$. The Kullback-Leibler property is given by $\Pi(\{f : d_K(f, f_0) < \epsilon\}) > 0$.

For more details about the consistency and convergence of posterior distribution, please refer to Subhashis Ghosal, 1997 [41].

5.2.2 Variable selection in Bayesian statistics

Penalization methods in high dimensional problems, being the flavor of the last decade, have enjoyed much attention resulting in a rich theoretical literature establishing their optimality properties. There are also fast algorithms and compelling applied results underlining the success story of these methods. In the regression setup, it is common to view the negative log prior density as a penalty function. As a result, a Gaussian prior is equivalent to a L_2 penalty regularization, while a Laplacian prior is equivalent to a L_1 penalty regularization. Besides, the idea of a weighted mixture prior is similar to that of an elastic net regularization. Every time, model improvement happens when prior put more probability around zero, and heavier tails of the non-zero regions. Because variable selection techniques always aim to balance the efficiency of penalizing true zero coefficients to be exact zero and pulling true nonzero coefficients away from zero. In terms of these consideration, a horseshoe prior (Carvalho, 2010 [11]) comes to the scene, since it has infinite mass at zero and Cauchy-like tails.

Although exact zeros can be achieved in spike and slab (George,and McCulloch, 1997 [38]), the computational burden is high since sampling from the posterior of binary variables is not an easy task especially in the high dimension case. There are several optimality results available for the horseshoe and horseshoe+ priors (Carvalho et al., 2010 [11]); (Datta et al., 2013 [26]), (van der Pas et al., 2014 [104])). However, to the best of my knowledge, all of

these results are concerned with the estimation of the multivariate normal mean. A satisfactory theory for uncertainty quantification for penalized estimators is not yet in place. I hope to be able to choose a default shrinkage prior that leads to similar optimality properties to those shown for L_1 penalization and other approaches, but with the added advantage of the entire posterior distribution concentrating at the optimal rate, instead of just focusing on the point estimate.

In the following, I am going to introduce the Bayesian variable selection using Inverted Beta distribution (Special case is Horse shoe prior). For the regression situation involving the observation of a dependent variable Y and a set of potential predictors $X_1, \ldots, X_p$, we consider the canonical regression setup

$$Y|X, \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 I), \tag{5.6}$$

where Y is $n \times 1$, $X = [X_1, \ldots, X_p]$ is $n \times p$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, and σ^2 is a scalar. Both $\boldsymbol{\beta}$ and σ^2 are considered unknown.

We describe the prior of $\boldsymbol{\beta}$ as a multivariate normal

$$\pi(\boldsymbol{\beta}|\sigma^2, \lambda) = \mathcal{N}_p(0, \Upsilon_{(\sigma, \lambda)}). \tag{5.7}$$

The residual variance σ^2 is conveniently modeled as a realization from an inverse gamma prior.

$$\pi(\sigma^2) = \mathcal{IG}(\nu/2, \nu\omega/2), \tag{5.8}$$

which is equivalent to $\nu\omega/\sigma^2 \sim \chi^2_\nu$. It might be desirable to have ω decrease with p.

We consider the special case of $\Upsilon_{(\sigma, \lambda)}$ in (5.7) is of the form

$$\pi(\boldsymbol{\beta}|\sigma, \lambda) = \mathcal{N}_p(0, \sigma^2 D(\lambda) R D(\lambda)), \tag{5.9}$$

where $D(\lambda)$ is diagonal and R is a correlation matrix. We denote the i-th diagonal element of $D^2(\lambda)$ by

$$(D^2(\lambda))_{ii} = \lambda^2.$$

Each component of $\boldsymbol{\beta}$ is modeled as a normal distribution

$$\pi(\beta_i|\sigma^2, \lambda^2) = \mathcal{N}(0, \lambda^2 \sigma^2). \tag{5.10}$$

The scale parameter λ^2 follows an inverted-beta distribution :

$$p(\lambda^2|a, b) = \frac{(\lambda^2)^{b-1}(1 + \lambda^2)^{-(a+b)}}{\text{Beta}(a, b)}, \tag{5.11}$$

where $\text{Beta}(a, b)$ denotes the beta function, and where a and b are positive reals.

The Inverted Beta distribution is plotted when parameters a and b take different (Figure 5.1). When $b \leq 1$, the density is infinite around origin. When $b > 1$, the density is 0 at origin. As you might notice, the behaviors (convergence rates) at the tail and around origin are different with the change of parameters a and b. It gives us ideas that Inverted Beta distribution can serve as a flexible and versatile prior in Bayesian variable selection and it suffices to model different level sparsity if parameters a and b are carefully chosen. As we know, the horseshoe prior is just a special case when $a = 0.5$ and $b = 0.5$.
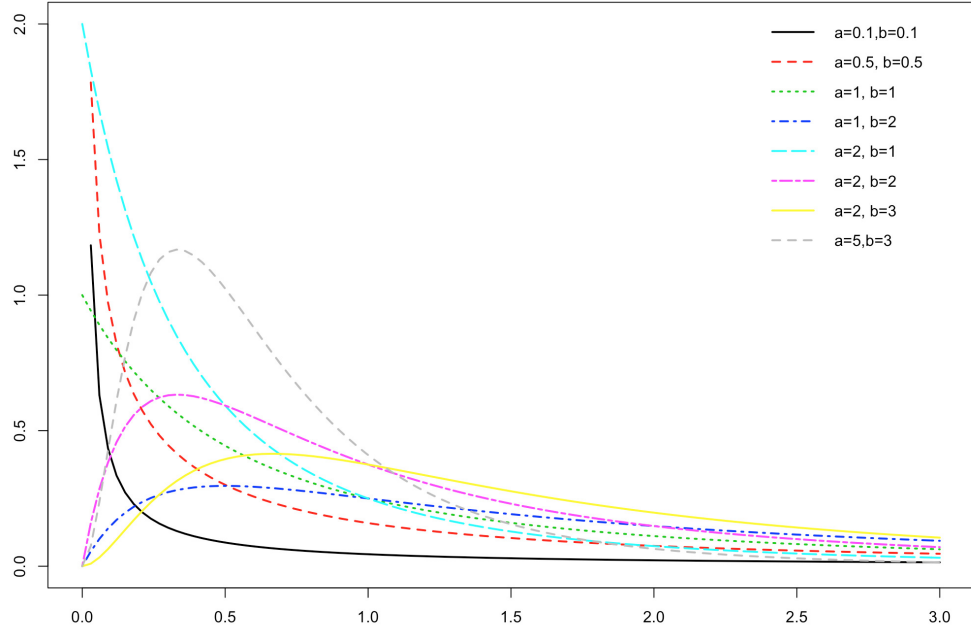
Figure 5.1. Inverted Beta distribution when parameters a and b take different values.

The generalized beta prime distribution has the probability density function as

$$f(x; \alpha, \beta, p, q) = \frac{p(\frac{x}{q})^{\alpha p - 1}(1 + (\frac{x}{q})^p)^{-(\alpha + \beta)}}{q\text{Beta}(\alpha, \beta)},$$

where $p > 0$ controls the shape and $q > 0$ controls the scale.

Shown below are the two modeling options I am aiming to do, if we would like to use the global-local prior. The first one is putting the global parameter τ as an upper hierarchy of the local parameter λ_i. Each regression coefficient β_i is controlled by local parameter λ_i,

while each λ_i is dominated by the global parameter τ.

$$Y|\boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 I),$$

$$\boldsymbol{\beta}|\sigma^2, \lambda_i \sim \mathcal{N}_p(0, \sigma^2 D(\lambda_i) R D(\lambda_i)),$$

$$\beta_i|\sigma^2, \lambda_i^2 \sim \mathcal{N}(0, \lambda_i^2 \sigma^2),$$

$$\pi(\lambda_i|\tau, a_1, b_1) = \frac{(\frac{\lambda_i}{\tau})^{a_1-1}(1+(\frac{\lambda_i}{\tau}))^{-(a_1+b_1)}}{\tau \mathrm{Beta}(a_1, b_1)},$$

$$\pi(\tau|a_0, b_0) = \frac{(\tau)^{a_0-1}(1+\tau)^{-(a_0+b_0)}}{\mathrm{Beta}(a_0, b_0)}.$$

(5.12)

The second one is putting the global parameter τ and the local parameter λ_i at the same hierachy. But we use the product of them to control the magnitude each regression coefficient β_i.

$$Y|\boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 I),$$

$$\boldsymbol{\beta}|\sigma^2, \lambda_i, \tau \sim \mathcal{N}_p(0, \sigma^2 D(\lambda_i, \tau) R D(\lambda_i, \tau)),$$

$$\beta_i|\sigma^2, \lambda_i^2, \tau^2 \sim \mathcal{N}(0, \lambda_i^2 \tau^2 \sigma^2),$$

$$\pi(\lambda_i|a_1, b_1) = \frac{(\lambda_i)^{a_1-1}(1+\lambda_i)^{-(a_1+b_1)}}{\mathrm{Beta}(a_1, b_1)},$$

$$\pi(\tau|a_0, b_0) = \frac{(\tau)^{a_0-1}(1+\tau)^{-(a_0+b_0)}}{\mathrm{Beta}(a_0, b_0)}.$$

(5.13)

5.2.3 Bayesian Dynamic network

When applied to the brain, the term connectivity refers to several different and interrelated aspects of brain organization (Horwitz, 2003 [52]). A fundamental distinction is that between structural connectivity, functional connectivity and effective connectivity. Functional connectivity is the connectivity between brain regions that share functional properties. More specifically, it can be defined as the temporal correlation between spatially remote neurophysiological events, expressed as deviation from statistical independence across these events

in distributed neuronal groups and areas. This applies to both resting state and task-state studies.

Let $G = (V, E)$ be a graph with the vertex (node) set V, and the edge set E. A node $u \in V$ represents an entity (intensity of a voxel) and an edge $(u, v) \in E$ represents a connectivity. The precision matrix (inverse covariance matrix) is a popular measure used to describe an undirected graph, since the zero entry is used to model the conditional independence between the corresponding two nodes. Friedman, 2008 [33] developed a simple algorithm to solve the estimation of sparse undirected graphical models through the use of L_1 (lasso) regularization.

For each patient, the data $Y = (y_{t,v})$ is a matrix, with each row t ($t \in \{1, \ldots, T\}$) denotes the time, and each column v ($v \in \{1, \ldots, V\}$) denotes a voxel. And the values of each entry $y_{t,v}$ denotes the intensity. Each column is time-ordered. Then the precision matrix estimated from data Y must be a function of time as the data is spatial and temporal. From the perspective of application, we actually get a time varying undirected graph which is called dynamic network. As shown in Figure [?], the network topology at time $t - 1$ is different from that at time t. The topology includes several aspects–organization of voxels and nodes, number of nodes, connectivities between nodes. When considering the changes of network between two time points, we might consider changes in the intensity of nodes, changes in the number of nodes and changes between network connectivities. We call them all together as topology transitions. For more details, please refer to Zhou, et al. 2010 [120] and Monti, et al. 2014 [72].

Because the precision matrix is positive definite, modeling and penalizing the time-varying precision matrix directly is difficult. We have to decompose the precision matrix into parts that do not have positive definite requirements. Suppose each precision matrix
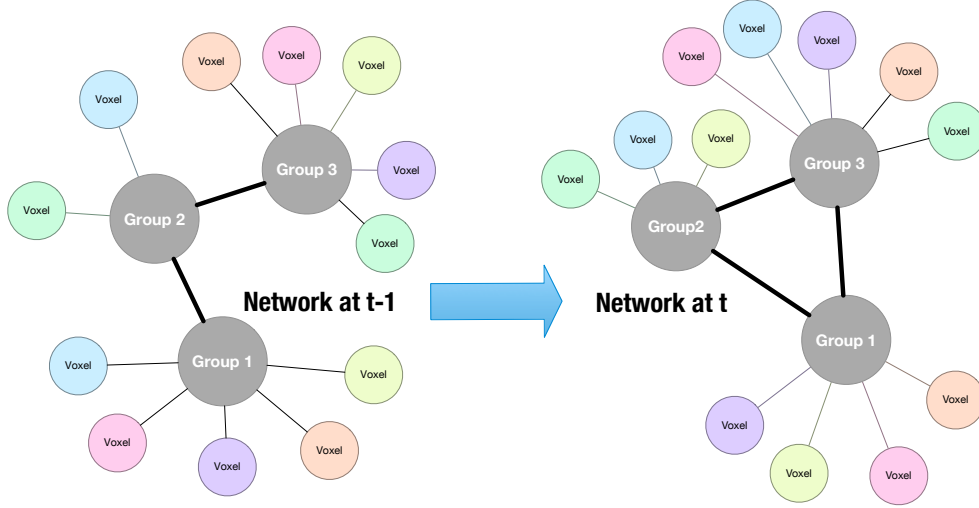
Figure 5.2. Group 1, Group 2 and Group 3 (big circles) denote the nodes of the brain network. The thick solid line denotes the connectivity between nodes. As we can see the connectivity changes with time. Voxels (small circles) denote the real observations.

$\mathbf{\Omega}^t$ has a "LDL" type of matrix decomposition

$$\mathbf{\Omega}^t = \boldsymbol{L}\boldsymbol{D}^t\boldsymbol{L}',$$

where $\boldsymbol{D}^t$ is a diagonal matrix and $\boldsymbol{L}$ is a upper triangular matrix, which is an unique form given by Cholesky Decomposition.

And we assume that the diagonal matrix $\boldsymbol{D}^t$ evolutes with time. We model the sparsity of the dynamic network by penalizing the Cholesky factor. But the sparse Cholesky factor does not guarantee to construct a sparse symmetric positive-definite matrix. Rothman, 2010 [88] discovered a subtle relationship between the sparsity of Cholesky factor and the sparsity of the covariance matrix. When connectivity representative L changes with time, then it is more likely that every Ω^t can be decomposed exactly into $L^t D^t D^t L^t$. Then the graphical

model likelihood with penalty and other hierarchies becomes

$$
Q(\boldsymbol{L}, \boldsymbol{D}, \beta, \sigma)
$$
$$
= \sum_{t=1}^{T} \left(\text{tr}(S^t L^t D^t D^t (L^t)') - \log \det L^t D^t D^t (L^t)' - \lambda_1 \|L^t\|_1 \right)
$$
$$
+ \sum_{t=2}^{T} \left(\frac{1}{2(\sigma^t)^2} \sum_{i=1}^{p} \left(\log(d_i^t) - \beta_i^t \log(d_i^{t-1}) \right)^2 + p \log(\sigma^t) \right),
$$

where d_i^t is the $i - th$ diagonal entry of $\boldsymbol{D}^t$.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Allen, J. S., Damasio, H., and Grabowski, T. J. Normal neuroanatomical variation in the human brain: An mri-volumetric study. *American journal of physical anthropology 118*, 4 (2002), 341–358.

[2] Aparicio, P., Mandaltsi, A., Boamah, J., Chen, H., Selimovic, A., Bratby, M., Uberoi, R., Ventikos, Y., and Watton, P. Modelling the influence of endothelial heterogeneity on the progression of arterial disease: application to abdominal aortic aneurysm evolution. *International Journal for Numerical Methods in Biomedical Engineering 30*, 5 (2014), 563–586.

[3] Arzani, A., Les, A., Dalman, R., and Shadden, S. Effect of exercise on patient specific abdominal aortic aneurysm flow topology and mixing. *International Journal for Numerical Methods in Biomedical Engineering 30*, 2 (2014), 280295.

[4] Barnes, D. E., and Lee, S. J. Predicting alzheimer's risk: why and how? *Alzheimer's Research & Therapy 3*, 6 (2011), 1–3.

[5] Berger, J. Statistical decision theory and bayesian inference. *Springer0Verlag (New York)* (1985).

[6] Bergerson, J., and Muehleisen, R. Bayesian large model calibration using simulation and measured data for improved predictions. *SAE International Journal of Passenger Cars-Mechanical Systems 8*, 2015-01-0481 (2015), 415–420.

[7] Biehler, J., Gee, M. W., and Wall, W. A. Towards efficient uncertainty quantification in complex and large-scale biomechanical problems based on a bayesian multifidelity scheme. *Biomechanics and modeling in mechanobiology 14*, 3 (2015), 489–513.

[8] Bouwman, F., Schoonenboom, S., van Der Flier, W., Van Elk, E., Kok, A., Barkhof, F., Blankenstein, M., and Scheltens, P. Csf biomarkers and medial temporal lobe atrophy predict dementia in mild cognitive impairment. *Neurobiology of aging 28*, 7 (2007), 1070–1074.

[9] Boyd, S., and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004.

[10] Breslow, N. Biostatistics and bayes. *Statistical Science* (1990), 269–284.

[11] Carvalho, C. M., Polson, N. G., and Scott, J. G. The horseshoe estimator for sparse signals. *Biometrika 97*, 2 (2010), 465–480.

[12] Cedarbaum, J. M., Jaros, M., Hernandez, C., Coley, N., Andrieu, S., Grundman, M., Vellas, B., Initiative, A. D. N., et al. Rationale for use of the clinical dementia rating sum of boxes as a primary outcome measure for alzheimers disease clinical trials. *Alzheimer's & Dementia 9*, 1 (2013), S45–S55.

[13] Chaikof, E. L., Brewster, D. C., Dalman, R. L., Makaroun, M. S., Illig, K. A., Sicard, G. A., Timaran, C. H., Upchurch, G. R., and Veith, F. J. The care of patients with an abdominal aortic aneurysm: the society for vascular surgery practice guidelines. *Journal of Vascular Surgery 50*, 4 (2009), S2–S49.

[14] Chen, Y., and Pham, T. D. Development of a brain mri-based hidden markov model for dementia recognition. *Biomedical engineering online 12*, Suppl 1 (2013), S2.

[15] Chib, S., and Greenberg, E. Understanding the Metropolis-Hastings Algorithm. *American Statistician 49*, 4 (1995), 327–335.

[16] Christopher, J. Multi-state models for panel data: The msm package for r. *Journal of Statistical Software 38*, 8 (1 2011), 1–28.

[17] Commenges, D. Multi-state models in epidemiology. *Lifetime data analysis 5*, 4 (1999), 315–327.

[18] Commenges, D. Inference for multi-state models from interval-censored data. *Statistical methods in medical research 11*, 2 (2002), 167–182.

[19] Commenges, D., Joly, P., Letenneur, L., and Dartigues, J.-F. Incidence and mortality of alzheimer's disease or dementia using an illness-death model. *Statistics in medicine 23*, 2 (2004), 199–210.

[20] Corder, E., Saunders, A., Strittmatter, W., Schmechel, D., Gaskell, P., Small, G., Roses, A. D., Haines, J., and Pericak-Vance, M. A. Gene dose of apolipoprotein e type 4 allele and the risk of alzheimer's disease in late onset families. *Science 261*, 5123 (1993), 921–923.

[21] Crane, P. K., Carle, A., Gibbons, L. E., Insel, P., Mackin, R. S., Gross, A., Jones, R. N., Mukherjee, S., Curtis, S. M., Harvey, D., et al. Development and assessment of a composite score for memory in the alzheimers disease neuroimaging initiative (adni). *Brain imaging and behavior 6*, 4 (2012), 502–516.

[22] Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association 86*, 416 (1991), 953–963.

[23] Cyron, C., and Humphrey, J. Vascular homeostasis and the concept of mechanobiological stability. *International journal of engineering science 85* (2014), 203–223.

[24] CYRON, C., WILSON, J., AND HUMPHREY, J. Mechanobiological stability: a new paradigm to understand the enlargement of aneurysms? *Journal of The Royal Society Interface 11*, 100 (2014), 20140680.

[25] DAHABREH, I., CHAN, J., EARLEY, A., MOORTHY, D., AVENDANO, E., TRIKALINOS, T., BALK, E., AND WONG, J. Modeling and simulation in the context of health technology assessment: Review of existing guidance, future research needs, and validity assessment.

[26] DATTA, J., GHOSH, J. K., ET AL. Asymptotic properties of bayes risk for the horseshoe prior. *Bayesian Analysis 8*, 1 (2013), 111–132.

[27] DAWKINS, C., SRINIVASAN, T. N., AND WHALLEY, J. Calibration. *Handbook of econometrics 5* (2001), 3653–3703.

[28] DESIKAN, R., FISCHL, B., CABRAL, H., KEMPER, T., GUTTMANN, C., BLACKER, D., HYMAN, B., ALBERT, M., AND KILLIANY, R. Mri measures of temporoparietal regions show differential rates of atrophy during prodromal ad. *Neurology 71*, 11 (2008), 819–825.

[29] DIETTERICH, T. Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR) 27*, 3 (1995), 326–327.

[30] DUNSMORE, I. A bayesian approach to calibration. *Journal of the Royal Statistical Society. Series B (Methodological) 30*, 2 (1968), 396–405.

[31] ERHART, P., HYHLIK-DÜRR, A., GEISBÜSCH, P., KOTELIS, D., MÜLLER-ESCHNER, M., GASSER, T. C., VON TENGG-KOBLIGK, H., AND BÖCKLER, D. Finite element analysis in asymptomatic, symptomatic, and ruptured abdominal aortic aneurysms: in search of new rupture risk predictors. *European Journal of Vascular and Endovascular Surgery 49*, 3 (2015), 239–245.

[32] FJELL, A. M., MCEVOY, L., HOLLAND, D., DALE, A. M., WALHOVD, K. B., INITIATIVE, A. D. N., ET AL. What is normal in normal aging? effects of aging, amyloid and alzheimer's disease on the cerebral cortex and the hippocampus. *Progress in neurobiology 117* (2014), 20–40.

[33] FRIEDMAN, J., HASTIE, T., AND TIBSHIRANI, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics 9*, 3 (2008), 432–441.

[34] FRISONI, G. B., FOX, N. C., JACK, C. R., SCHELTENS, P., AND THOMPSON, P. M. The clinical use of structural mri in alzheimer disease. *Nature Reviews Neurology 6*, 2 (2010), 67–77.

[35] GASER, C., FRANKE, K., KLÖPPEL, S., KOUTSOULERIS, N., SAUER, H., INITIA-TIVE, A. D. N., ET AL. Brainage in mild cognitive impaired patients: predicting the conversion to alzheimers disease. *PloS one 8*, 6 (2013), e67346.

[36] GASSER, T. C. An irreversible constitutive model for fibrous soft biological tissue: a 3-d microfiber approach with demonstrative application to abdominal aortic aneurysms. *Acta biomaterialia 7*, 6 (2011), 2457–2466.

[37] GELB, D. J. Measurement of progression in alzheimer's disease: a clinician's perspective. *Statistics in Medicine 19*, 11-12 (2000), 1393–1400.

[38] GEORGE, E. I., AND MCCULLOCH, R. E. Approaches for bayesian variable selection. *Statistica sinica* (1997), 339–373.

[39] GERDES, L. U., KLAUSEN, I. C., SIHM, I., FAERGEMAN, O., AND VOGLER, G. Apolipoprotein e polymorphism in a danish population compared to findings in 45 other study populations around the world. *Genetic epidemiology 9*, 3 (1992), 155–167.

[40] GHARAHI, H., ZAMBRANO, B., LIM, C., CHOI, J., LEE, W., AND BAEK, S. On growth measurements of abdominal aortic aneurysms using maximally inscribed sphere, 2015.

[41] GHOSAL, S. A review of consistency and convergence of posterior distribution. In *Varanashi Symposium in Bayesian Inference, Banaras Hindu University* (1997).

[42] GIORGIO, A., AND DE STEFANO, N. Clinical use of brain volumetry. *Journal of Magnetic Resonance Imaging 37*, 1 (2013), 1–14.

[43] GODBEY, M. W. The use of variable-bagging and the cross-validation selector in the prediction of alzheimers using the adni database.

[44] GOEL, M., KHANNA, P., AND KISHORE, J. Understanding survival analysis: Kaplan-meier estimate. *International journal of Ayurveda research 1*, 4 (2010), 274.

[45] HANKIN, R. K. Introducing BACCO, an R bundle for Bayesian Analysis of Computer Code Output. *Journal of Statistical Software 14*, 16 (2005), 1–21.

[46] HANSEN, L. P., AND HECKMAN, J. J. The empirical foundations of calibration. *The Journal of Economic Perspectives 10*, 1 (1996), 87–104.

[47] HAWKINS-DAARUD, A., PRUDHOMME, S., VAN DER ZEE, K. G., AND ODEN, J. T. Bayesian calibration, validation, and uncertainty quantification of diffuse interface models of tumor growth. *Journal of mathematical biology 67*, 6-7 (2013), 1457–1485.

[48] HECKERMAN, D., ET AL. A tutorial on learning with bayesian networks. *Nato Asi Series D Behavioural And Social Sciences 89* (1998), 301–354.

[49] HIGDON, D., GATTIKER, J., WILLIAMS, B., AND RIGHTLEY, M. Computer model calibration using high-dimensional output. *Journal of the American Statistical Association 103*, 482 (2008).

[50] HIGDON, D., KENNEDY, M., CAVENDISH, J. C., CAFEO, J. A., AND RYNE, R. D. Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing 26*, 2 (2004), 448–466.

[51] HOLLAND, J. The reverend thomas bayes, frs (1702-61). *Journal of the Royal Statistical Society. Series A (General) 125*, 3 (1962), 451–461.

[52] HORWITZ, B. The elusive concept of brain connectivity. *Neuroimage 19*, 2 (2003), 466–470.

[53] JACKSON, C. H., SHARPLES, L. D., THOMPSON, S. G., DUFFY, S. W., AND COUTO, E. Multistate markov models for disease progression with classification error. *Journal of the Royal Statistical Society: Series D (The Statistician) 52*, 2 (2003), 193–209.

[54] JOST, B. C., AND GROSSBERG, G. T. The natural history of alzheimer's disease: A brain bank study. *Journal of the American Geriatrics Society 43*, 11 (1995), 1248–1255.

[55] JS, W., BAEK, S., AND HUMPHREY, J. Importance of initial aortic properties on the evolving regional anisotropy, stiffness and wall thickness of human abdominal aortic aneurysms. *Journal of the Royal Society Interface 9*, 74 (2012), 2047–2058.

[56] KADANE, J. B. Prime time for bayes. *Controlled Clinical Trials 16*, 5 (1995), 313–318.

[57] KAWAS, C., GRAY, S., BROOKMEYER, R., FOZARD, J., AND ZONDERMAN, A. Age-specific incidence rates of alzheimers disease the baltimore longitudinal study of aging. *Neurology 54*, 11 (2000), 2072–2077.

[58] KENNEDY, M. C., AND O'HAGAN, A. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*, 3 (2001), 425–464.

[59] KLINK, A., HYAFIL, F., RUDD, J., FARIES, P., FUSTER, V., MALLAT, Z., MEILHAC, O., MULDER, W. J., MICHEL, J.-B., RAMIREZ, F., ET AL. Diagnostic and therapeutic strategies for small abdominal aortic aneurysms. *Nature Reviews Cardiology 8*, 6 (2011), 338–347.

[60] KNIEMEYER, H., KESSLER, T., REBER, P., RIS, H., HAKKI, H., AND WIDMER, M. Treatment of ruptured abdominal aortic aneurysm, a permanent challenge or a waste of resources? prediction of outcome using a multi-organ-dysfunction score. *European Journal of Vascular and Endovascular Surgery 19*, 2 (2000), 190–196.

[61] KWON, S., BUREK, W., DUPAY, A., FARSAD, M., BAEK, S., PARK, E., AND LEE, W. Interaction of expanding abdominal aortic aneurysm with surrounding tissue: Retrospective ct image studies. *Journal of nature and science 1*, 8 (2015), e150.

[62] KWON, S., RECTENWALD, J., AND BAEK, S. Intrasac pressure changes and vascular remodeling after endovascular repair of abdominal aortic aneurysms: review and biomechanical model simulation. *Journal of biomechanical engineering 133*, 1 (2011), 011011.

[63] LEDERLE, F. A., WILSON, S. E., JOHNSON, G. R., REINKE, D. B., LITTOOY, F. N., ACHER, C. W., BALLARD, D. J., MESSINA, L. M., GORDON, I. L., CHUTE, E. P., ET AL. Immediate repair compared with surveillance of small abdominal aortic aneurysms. *New England Journal of Medicine 346*, 19 (2002), 1437–1444.

[64] LIKEMAN, M., ANDERSON, V. M., STEVENS, J. M., WALDMAN, A. D., GODBOLT, A. K., FROST, C., ROSSOR, M. N., AND FOX, N. C. Visual assessment of atrophy on magnetic resonance imaging in the diagnosis of pathologically confirmed young-onset dementias. *Archives of neurology 62*, 9 (2005), 1410–1415.

[65] LILFORD, R., AND BRAUNHOLTZ, D. For debate: The statistical basis of public policy: a paradigm shift is overdue. *Bmj 313*, 7057 (1996), 603–607.

[66] LORENZI, M., PENNEC, X., FRISONI, G. B., AYACHE, N., INITIATIVE, A. D. N., ET AL. Disentangling normal aging from alzheimer's disease in structural magnetic resonance images. *Neurobiology of aging 36* (2015), S42–S52.

[67] MACHIN, D., CHEUNG, Y. B., AND PARMAR, M. K. Cox's proportional hazards model. *Survival Analysis: A Practical Approach, Second Edition* (2006), 121–153.

[68] MAIER, A., GEE, M., REEPS, C., PONGRATZ, J., ECKSTEIN, H.-H., AND WALL, W. A comparison of diameter, wall stress, and rupture potential index for abdominal aortic aneurysm rupture risk prediction. *Annals of biomedical engineering 38*, 10 (2010), 3124–3134.

[69] MCKAY, M. D., BECKMAN, R. J., AND CONOVER, W. J. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics 21*, 2 (1979), 239–245.

[70] MEIRA-MACHADO, L. F., DE UÑA-ÁLVAREZ, J., CADARSO-SUÁREZ, C., AND ANDERSEN, P. Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research* (2008).

[71] MILLS, K. L., AND TAMNES, C. K. Methods and considerations for longitudinal structural brain imaging analysis across development. *Developmental cognitive neuroscience 9* (2014), 172–190.

[72] Monti, R. P., Hellyer, P., Sharp, D., Leech, R., Anagnostopoulos, C., and Montana, G. Estimating time-varying brain connectivity networks from functional mri time series. *NeuroImage 103* (2014), 427–443.

[73] Muehleisen, R. T., and Bergerson, J. Bayesian calibration-what, why and how.

[74] New, J., and Chandler, T. Evolutionary tuning of building models to monthly electrical consumption. *ASHRAE Transactions 119* (2013), 89.

[75] Norris, J. R. *Markov Chains.* No. 7. Cambridge University Press, 1998.

[76] of Neuro Imaging, L. Alzheimer's disease.

[77] O'Hagan, A. Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering & System Safety 91*, 10 (2006), 1290–1300.

[78] Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., and Kokmen, E. Mild cognitive impairment: clinical characterization and outcome. *Archives of neurology 56*, 3 (1999), 303–308.

[79] Porth, C. *Essentials of pathophysiology: Concepts of altered health states.* Lippincott Williams & Wilkins, 2010. Edition 3.

[80] Racine-Poon, A. A bayesian approach to nonlinear calibration problems. *Journal of the American Statistical Association 83*, 403 (1988), 650–656.

[81] Raftery, P., Keane, M., and ODonnell, J. Calibrating whole building energy models: An evidence-based methodology. *Energy and Buildings 43*, 9 (2011), 2356–2364.

[82] Reddy, T. A., Maor, I., and Panjapornpon, C. Calibrating detailed building energy simulation programs with measured datapart i: General methodology (rp-1051). *Hvac&R Research 13*, 2 (2007), 221–241.

[83] Reeps, C., Maier, A., Pelisek, J., Härtl, F., Grabher-Meier, V., Wall, W. A., Essler, M., Eckstein, H.-H., and Gee, M. W. Measuring and modeling patient-specific distributions of material properties in abdominal aortic aneurysm wall. *Biomechanics and Modeling in Mechanobiology 12*, 4 (2013), 717–733.

[84] Ribeiro, P. F., Ventura-Antunes, L., Gabi, M., Mota, B., Grinberg, L. T., Farfel, J. M., Ferretti-Rebustini, R. E., Leite, R. E., Herculano-Houzel, S., et al. The human cerebral cortex is neither one nor many: neuronal distribution reveals two quantitatively different zones in the gray matter, three in the white matter, and explains local variations in cortical folding. *Front. Neuroanat 7*, 28 (2013), 10–3389.

[85] RIDDLE, M., AND MUEHLEISEN, R. T. A guide to bayesian calibration of building energy models. In *Building Simulation Conference* (2014).

[86] RISACHER, S. L., KIM, S., SHEN, L., NHO, K., FOROUD, T., GREEN, R. C., PETERSEN, R. C., JACK JR, C. R., AISEN, P. S., KOEPPE, R. A., ET AL. The role of apolipoprotein e (apoe) genotype in early mild cognitive impairment (e-mci).

[87] RIZZO, R. J., MCCARTHY, W. J., DIXIT, S. N., LILLY, M. P., SHIVELY, V. P., FLINN, W. R., AND YAO, J. S. Collagen types and matrix protein content in human abdominal aortic aneurysms. *Journal of Vascular Surgery 10*, 4 (1989), 365 – 373.

[88] ROTHMAN, A. J., LEVINA, E., AND ZHU, J. A new approach to cholesky-based covariance regularization in high dimensions. *Biometrika 97*, 3 (2010), 539–550.

[89] SACKS, J., WELCH, W. J., MITCHELL, T. J., AND WYNN, H. P. Design and analysis of computer experiments. *Statistical science* (1989), 409–423.

[90] SALTHOUSE, T. A. Robust cognitive change. *Journal of the International Neuropsychological Society 18*, 04 (2012), 749–756.

[91] SANTNER, T. J., WILLIAMS, B. J., AND NOTZ, W. I. *The design and analysis of computer experiments.* Springer Science & Business Media, 2013.

[92] SAUNDERS, A., STRITTMATTER, W., SCHMECHEL, D., GEORGE-HYSLOP, P. S., PERICAK-VANCE, M., JOO, S., ROSI, B., GUSELLA, J., CRAPPER-MACLACHLAN, D., ALBERTS, M., ET AL. Association of apolipoprotein e allele $\epsilon4$ with late-onset familial and sporadic alzheimer's disease. *Neurology 43*, 8 (1993), 1467–1467.

[93] SCHMITTER, D., ROCHE, A., MARÉCHAL, B., RIBES, D., ABDULKADIR, A., BACH-CUADRA, M., DADUCCI, A., GRANZIERA, C., KLÖPPEL, S., MAEDER, P., ET AL. An evaluation of volume-based morphometry for prediction of mild cognitive impairment and alzheimer's disease. *NeuroImage: Clinical 7* (2015), 7–17.

[94] SEYEDSALEHI, S., ZHANG, L., CHOI, J., AND BAEK, S. Prior distributions of material parameters for bayesian calibration of growth and remodeling computational model of abdominal aortic wall. *J Biomech Eng 137*, 10 (2015), 101001.

[95] SHEIDAEI, A., HUNLEY, S., ZEINALI-DAVARANI, S., RAGUIN, L., AND BAEK, S. Simulation of abdominal aortic aneurysm growth with updating hemodynamic loads using a realistic geometry. *Medical engineering & physics 33*, 1 (2011), 80–88.

[96] SHI, J., LEPORE, N., GUTMAN, B. A., THOMPSON, P. M., BAXTER, L. C., CASELLI, R. J., AND WANG, Y. Genetic influence of apolipoprotein e4 genotype on hippocampal morphometry: An n= 725 surface-based alzheimer's disease neuroimaging initiative study. *Human brain mapping 35*, 8 (2014), 3903–3918.

[97] Spampinato, M. V., Rumboldt, Z., Hosker, R. J., and Mintzer, J. E. Apolipoprotein e and gray matter volume loss in patients with mild cognitive impairment and alzheimer disease. *Radiology 258*, 3 (2011), 843–852.

[98] Spiegelhalter, D. J., Myles, J. P., Jones, D. R., and Abrams, K. R. Bayesian methods in health technology assessment: a review.

[99] Stephenson, G. Statistical analysis of computer models and the use of derivative information.

[100] Sukkar, R., Katz, E., Zhang, Y., Raunig, D., and Wyman, B. T. Disease progression modeling using hidden markov models. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* (2012), IEEE, pp. 2845–2848.

[101] Sweeting, M., Farewell, V., and De Angelis, D. Multi-state markov models for disease progression in the presence of informative examination times: An application to hepatitis c. *Statistics in medicine 29*, 11 (2010), 1161–1174.

[102] Swerdlow, R. H. Brain aging, alzheimer's disease, and mitochondria. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1812*, 12 (2011), 1630–1639.

[103] Tuo, R., and Jeff Wu, C. A theoretical framework for calibration in computer models: parametrization, estimation and convergence properties. *SIAM/ASA Journal on Uncertainty Quantification 4*, 1 (2016), 767–795.

[104] Van Der Pas, S., Kleijn, B., Van Der Vaart, A., et al. The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics 8*, 2 (2014), 2585–2618.

[105] Vemuri, P., Wiste, H., Weigand, S., Shaw, L., Trojanowski, J., Weiner, M., Knopman, D., Petersen, R., Jack, C., Initiative, A. D. N., et al. Mri and csf biomarkers in normal, mci, and ad subjects predicting future clinical change. *Neurology 73*, 4 (2009), 294–301.

[106] Walker, S. New approaches to bayesian consistency. *Annals of Statistics* (2004), 2028–2043.

[107] Wang, Y., Resnick, S. M., and Davatzikos, C. Analysis of spatio-temporal brain imaging patterns by hidden markov models and serial mri images. *Human brain mapping 35*, 9 (2014), 4777–4794.

[108] Watton, P. N., Huang, H., and Ventikos, Y. Multi-scale modelling of vascular disease: Abdominal aortic aneurysm evolution. *Computational Modeling in Tissue Engineering 7* (2013), 309–339.

[109] WHITT, W. Continuous-time markov chains. *INTERNET: www. columbia. edu/ ww2040* (2006).

[110] WILLIAMS, C. K., AND RASMUSSEN, C. E. Gaussian processes for machine learning. *MIT Press* (2006).

[111] XU, Y., CHOI, J., DASS, S., AND MAITI, T. Sequential Bayesian prediction and adaptive sampling algorithms for mobile sensor networks. *IEEE Transactions on Automatic Control 57*, 8 (2012), 2078–2084.

[112] YAMAMOTO, S., MIYAZAKI, M., IWANO, T., AND KITAZAWA, S. Bayesian calibration of simultaneity in audiovisual temporal order judgments. *PLoS ONE 7* (2012), 1–9.

[113] YU, H.-M., YANG, S.-S., GAO, J.-W., ZHOU, L.-Y., LIANG, R.-F., AND QU, C.-Y. Multi-state markov model in outcome of mild cognitive impairments among community elderly residents in mainland china. *International Psychogeriatrics 25*, 05 (2013), 797–804.

[114] ZANKL, A., SCHUMACHER, H., KRUMSDORF, U., KATUS, H., JAHN, L., ET AL. Pathology, natural history and treatment of abdominal aortic aneurysms. *Clinical Research in Cardiology 96*, 3 (2007), 140–151.

[115] ZEINALI-DAVARANI, S., AND BAEK, S. Medical image-based simulation of abdominal aortic aneurysm growth. *Mechanics Research Communications 42* (2012), 107–117.

[116] ZEINALI-DAVARANI, S., RAGUIN, L. G., VORP, D. A., AND BAEK, S. Identification of in vivo material and geometric parameters of a human aorta: toward patient-specific modeling of abdominal aortic aneurysm. *Biomechanics and modeling in mechanobiology 10*, 5 (2011), 689–699.

[117] ZEINALI-DAVARANI, S., SHEIDAEI, A., AND BAEK, S. A finite element model of stress-mediated vascular adaptation: application to abdominal aortic aneurysms. *Computer methods in biomechanics and biomedical engineering 14*, 9 (2011), 803–817.

[118] ZHANG, D., SHEN, D., INITIATIVE, A. D. N., ET AL. Predicting future clinical changes of mci patients using longitudinal and multimodal biomarkers. *PloS one 7*, 3 (2012), e33182.

[119] ZHAO, X., AND PELEGRI, A. A. A bayesian approach for characterization of soft tissue viscoelasticity in acoustic radiation force imaging. *Int. J. Numer. Meth. Biomed. Engng.* (2015).

[120] ZHOU, S., LAFFERTY, J., AND WASSERMAN, L. Time varying undirected graphs. *Machine Learning 80*, 2 (2010), 295–319.