12-2023

Creation and Validation of a Novel Bayesian Calibration Method with Griddy Gibbs Sampling

Hannah Stewart
Clemson University, hstew213@gmail.com

Creation and Validation of a Novel Bayesian Calibration Method with Griddy Gibbs Sampling

---

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Civil Engineering

---

by
Hannah Stewart
December 2023

Accepted by:
Dr. Laura Redmond, Committee Chair
Dr. Christopher McMahan
Dr. Weichiang Pang

Abstract

Computer models of physical systems are widely used in lieu of, or in tandem with, experimental testing. It is critical to verify the accuracy of computer models through the process of calibration. Typical calibration methods are often computationally expensive and therefore cannot be performed in real time. This thesis presents a novel Bayesian calibration method using a Griddy Gibbs sampling algorithm to improve calibration speeds. This method was verified in two applications: a location-dependent dataset in the heat transfer analysis of an engine piston, and time-dependent tire forces in a drum test. The proposed method was directly compared to a traditional Bayesian calibration method in the engine piston application. It was found that the two methods were close in accuracy with large amounts of calibration data, and the Griddy Gibbs method was significantly less computationally expensive; it could calibrate in less than a minute, while the traditional method took several days.

Dedication

This thesis is dedicated to my family, my parents in particular, for their love and support throughout my education.

Acknowledgments

I would like to acknowledge my advisors, Dr. Laura Redmond and Dr. Chris McMahan, for their support, expertise, and encouragement in broadening my experience as an engineer. I'd also like to thank Stephen Wright for his contributions in generating and comparing data used in this thesis, and Clemson's VIPR-GS center for funding this research. And finally, I would like to thank my family and friends for supporting me during tough moments and celebrating the exciting ones.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

When experimental testing is difficult or expensive to perform, mathematical computer models can be used to approximate physical processes. A computer model can act as a digital twin to a physical model and update continuously according to physical data, or it can operate and update separately. These models can vary in complexity, computational time, and accuracy to the physical system, but most simply, they take in inputs and return outputs. These inputs often have physical meaning, for example describing the setup or operating conditions of the system, that cannot always be directly measured or controlled. The outputs are the modeled behavior of the physical system.

It is important to calibrate a computer model to experimental data to ensure it accurately represents the physical system. To this end, it is necessary to estimate model input parameters which would generate a given set of experimental data. With this information, one could predict how that specific system would perform in a different test scenario. Model calibration can be performed using a variety of methods, but most basically they aim to match the experimental data with simulation data via the tuning of unknown input parameters. Many existing methods for model

calibration require many simulations to be run in a linear manner until convergence to the best set of parameters is reached, precluding the ability to perform model calibration in real time.

This thesis describes a novel model calibration method which utilizes Griddy Gibbs sampling from a library of simulation data. The use of Griddy Gibbs sampling allows the calibration to run in real time, as the algorithm samples only from the existing library of simulation data, rather than having to construct and run a Gaussian process model at each iteration of the sampler. The method is validated and directly benchmarked against a traditional Bayesian calibration method in calibrating a piston thermal model with location-dependent temperature data. After, the approach is used to calibrate a tire model with time-dependent response data.

The thesis content is organized as follows: Chapter 2 reviews the existing literature related to model calibration. Chapter 3 presents the methodology for the proposed novel calibration method and an initial verification for this method. Chapter 4 presents a case study in which the proposed method is directly compared to a traditional Bayesian calibration method when applied to location-based calibration data. In chapter 5, the proposed method is applied to time-based calibration data. Chapter 6 provides conclusions and recommendations for future work. More detailed information about the codes used in calibration can be found in the appendices.

Chapter 2

Literature Review

2.1 Overview of Model Calibration Methods

There are many methods for model calibration, using Bayesian statistics, frequentist statistics, and combinations and variations thereof. Frequentist model calibration methods tend to be conceptually simple to understand and implement. Calibration is typically performed using a single metric, e.g. mean error [6], mean absolute error [3, 19], mean squared error [6, 1], maximum likelihood estimation [5, 17], method of moments [21], or root mean squared error [3, 19], to name a few. The maximum likelihood estimator (MLE) is the most common method for model calibration in the frequentist paradigm and involves locating a combination of parameters which maximize the likelihood function. As a simple example (from [16]), say we wanted to estimate the underlying probability of a coin landing on heads. We flip the coin 100 and find it lands on heads 55 times. The likelihood function, or the likelihood that the coin lands on heads for a given probability p, for this binomial example is given

by its probability density function:

$$L(p) = \binom{100}{55} p^{55}(1-p)^{55} \tag{2.1}$$

To find the maximum value of this equation, we take its derivative and set it equal to zero.

$$\frac{d}{dp}(L(p)) = \binom{100}{55} (55p^{54}(1-p)^{45} - 45p^{55}(1-p)^{44}) = 0 \tag{2.2}$$

When we solve this equation we see that our estimate for the probability of heads $\hat{p} = 0.55$.

Using a single frequentist calibration metric like the MLE makes it difficult to quantify the many sources of uncertainty in the calibration problem. While MLE does allow for the calculation of standard error, it does not include an estimate of the uncertainty of the parameter estimates or model discrepancy. Because of these limitations of frequentist methods, many advocate for the use of multiple frequentist methods in evaluating model performance [3], or frequentist methods in tandem with Bayesian methods [20]. For example, if the desired frequentist metric is sensitive to outliers, the use of an additional metric may provide a better understanding of the data.

Although most frequentist methods neglect the fundamental difference or discrepancy between a computer model and the real physical process, there is some work in this area. No computer model will be exactly equal to a physical process, and it is useful for a calibration methodology to acknowledge this fact through the addition of a discrepancy term.

In contrast, Bayesian calibration methods utilize Bayes' theorem (equation 2.3) to approximate the distribution of a parameter given some data. Bayesian meth-

ods require the user to define prior distributions on each parameter. This is useful in allowing expertise or previous work to inform the calibration, but if the priors are specified incorrectly, it can bias the posterior estimate. To avoid this, one can specify an uninformative prior for a parameter. Another possible drawback of Bayesian calibration methods is that they can be computationally expensive because they require large numbers of samples from the parameter distribution to converge. However, Bayesian calibration is widely used because it allows for robust quantification of uncertainty.

2.2 Bayesian Calibration Methods

Bayesian calibration methods utilize the relationship between the posterior ($p(\theta|y)$) and prior ($p(\theta)$) distributions as established by Bayes' Theorem:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{2.3}$$

This relationship states that the posterior probability that a model parameter θ takes on a certain value given some data y is proportional to the probability that data was generated using a certain value of that parameter, multiplied by the prior probability of the parameter taking on that value, up to a constant. This proportionality constant is given by

$$\int p(y|\theta)p(\theta)\, d\theta \tag{2.4}$$

This integral is often very difficult or impossible to evaluate directly. Instead, many Bayesian methods sample from the posterior distribution. With enough samples, the full posterior distribution can be approximated without needing to calculate the proportionality constant. Depending on the complexity of the sampling distribution,

obtaining sufficient samples can be a time intensive process.

The prior probability $p(\theta)$ is one key difference between Bayesian and other methods of model calibration: it allows the user to impart his or her prior knowledge about a parameter, its range and/or its distribution, into the calibration process.

Most often, model calibration is performed using a Bayesian methodology, particularly that developed by Kennedy and O'Hagan in 2001 [11]. Significantly, this approach accounts for the inherent discrepancy between a computer model and the physical process it aims to represent, as well as random measurement error. Kennedy and O'Hagan model the relationship between experimental observations z_i, the true physical process $\zeta(x_i)$, the computer model $\eta(x_i, \theta)$, and unknown regression parameter ρ as:

$$z_i = \zeta(x_i) + e_i = \rho\eta(x_i, \theta) + \delta(x_i) + e_i \tag{2.5}$$

Traditional Bayesian calibration methods typically employ emulators [12] as a means to bypass the generation of computer simulation data ($\eta(x_i, \theta)$) at each parameter configuration. This is useful when running these simulations are time intensive and an emulator can accurately represent the simulation, but should be avoided when an accurate emulator is difficult to obtain.

Many researchers have presented work building on the Kennedy O'Hagan framework, including efforts to calibrate model using a combination of simulation data and small amounts of experimental data [10]; when only a small number of realizations from the model are available [15]; when a traditional Gaussian discrepancy still does not accurately represent the physical process [8]

Bayesian calibration methods require a way to take large numbers of samples from the posterior distribution. This is most often done using Monte Carlo methods [13], usually Markov chain Monte Carlo (MCMC) methods. The theory behind Monte

Carlo methods is using large numbers of random simulations to approximate a desired quantity, rather than calculating it directly. A Markov chain is a series of samples in which each value depends on the previous state in the chain. Together, an MCMC sampling algorithm samples from a posterior distribution by creating Markov chains using Monte Carlo simulations [2]. These algorithms usually require large numbers of samples to converge, and they often have a "burn-in" period in which the sampler vacillates between different values until it eventually converges.

The two most common MCMC methods are the Metropolis-Hastings algorithm and Gibbs sampling. The Metropolis-Hastings algorithm uses a random walk with some acceptance-rejection criteria [4] to generate its Markov chains of samples from the posterior joint distribution, and Gibbs sampling instead takes a random walk taking samples from conditional distributions.

2.3 Gibbs Sampling

The Gibbs sampler was introduced by Geman and Geman in 1984 [7] as a novel method for sampling from a posterior distribution in a Bayesian analysis. In this method, the joint posterior distribution is approximated through successive sampling from conditional distributions, each time holding all but one parameter at its previous value, beginning at an arbitrary first value. In other words, to approximate a joint distribution $p(x_i...x_n)$, one would sample:

1. $x_1^{(i+1)}$ from $p(x_1|x_2^{(i)}, \ldots, x_n^{(i)})$
2. $x_2^{(i+1)}$ from $p(x_2|x_1^{(i+1)}, \ldots, x_n^{(i)})$

$\vdots$

3. $x_n^{(i+1)}$ from $p(x_n|x_1^{(i+1)}, \ldots, x_{n-1}^{(i+1)})$

The Gibbs sampler is a type of Markov chain Monte Carlo algorithm. Each

sample is not independent, as it depends on the previously calculated values for the parameters. It does follow the Markov property, meaning that future states rely only on the current state of the algorithm, not previous states. This fact, noted by Geman and Geman, means that the Gibbs sampler is a parallel algorithm which can be run on multiple computers simultaneously.

An update to the Gibbs sampler, the Griddy Gibbs sampler, was presented by Ritter and Tanner in 1992 [18]. They present this method for use when the conditional distributions are difficult to sample from. Instead, they suggest evaluating these conditional distributions at a discrete grid of points. Their methodology is as follows:

1. Evaluate the full conditional distribution at a discrete grid of points

2. Approximate the inverse CDF (cumulative distribution function) of the full conditional distribution (e.g. piecewise linear or piecewise constant).

3. Generate a random number between 0 and 1 and inverse transform sample from the inverse CDF

Ritter and Tanner pay special attention to the density and distribution of the sample grid, in particular advocating for a more densely packed grid of points in more highly visited areas. For simplicity, in this study we used either a uniformly spaced grid of points or a latin hypercube sampling of points, but an adaptive grid, which grows more dense in areas of repeated samples, could present a promising opportunity for future research.

Chapter 3

Griddy Gibbs Calibration Methodology and Initial Verification

3.1 Introduction

The following describes the novel Bayesian calibration method using Griddy Gibbs sampling.

3.2 Methodology

3.2.1 Data Model

The joint probability distribution of the experimental results with respect to independent variable x and parameters θ is often expressed as a function of the

deterministic computer model, discrepancy $\delta(x)$ and error term ϵ.

$$y(x,\theta) = g(x,\theta) + \delta(x) + \epsilon \tag{3.1}$$

The discrepancy term δ represents the inherent differences that would exist between experimental conditions and the data model. In the proposed model calibration method, this discrepancy term δ is not included. While it is important when training a surrogate model to account for the differences between the model and the true process, in a model calibration problem where one is only in search of a best fit, including the discrepancy in the data model is not necessary and does not improve the accuracy of calibration. The effect of random error ϵ is also taken as negligible in this formulation.

The joint distribution $y(x,\theta) = g(x,\theta)$ is difficult to sample from, and therefore it is convenient to sample from the conditional distribution. This distribution is multivariate normal with mean structure $g(\theta)$ and correlation structure $\sigma^2 I$, where I is the identity matrix.

$$y|x,\theta \sim MVN(g(x,\theta), \sigma^2 I) \tag{3.2}$$

Because it is assumed that the experimental data follows a normal distribution, the likelihood function $L(\theta)$ can be taken as the probability density function of a multivariate normal distribution:

$$L(\theta) = P(y|\theta) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} exp\left(-\frac{1}{2\sigma^2}[y(x_i,\theta_i) - g(x_i,\theta_i)]^2\right) \tag{3.3}$$

Where n denotes the total number of simulation observations. For computational efficiency, we let

$$C_\theta = (y(x,\theta) - g(x,\theta))^2 \tag{3.4}$$

and precompute the quantity for each (x, θ). This reduces computation time, as no redundant operations are performed in the MCMC draws.

Let $\tau = 1/\sigma^2$. The likelihood function becomes

$$L(\theta) = \left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} exp\left(-\frac{\tau}{2}\sum_{i=1}^{n} C_\theta\right) \tag{3.5}$$

Because n is large, the likelihood function tends towards infinity. For numerical stabilization, calculate the log likelihood:

$$ln(L(\theta)) = \frac{n}{2} ln\left(\frac{\tau}{2\pi}\right) - \frac{\tau}{2}\sum_{i=1}^{n} C_\theta \tag{3.6}$$

3.2.2 Prior Distributions

We specify a flat, uninformative prior distribution for each θ. The design of experiments is normalized in the range of each θ, giving us:

$$\pi(\theta) \sim U(0, 1) \tag{3.7}$$

The prior distribution $\pi(\tau)$ is a gamma distribution with $\alpha = 1 \times 10^{-9}$ and $\beta = 1 \times 10^{-9}$. It is designed to be a relatively weak prior distribution.

$$\pi(\tau) \sim Gamma(1 \times 10^{-9}, 1 \times 10^{-9}) \tag{3.8}$$

3.2.3 Griddy Gibbs Sampling

The full conditional distributions from which we sample are as follows:

$$p(\theta|y, \tau) \propto L(y|\theta, \tau) \times U(0, 1) \tag{3.9}$$

$$p(\tau|y,\theta) \propto L(y|\theta,\tau) \times Gamma(1\times10^{-9}, 1\times10^{-9}) \tag{3.10}$$

Inverse transform sampling is used to sample from these full conditional distributions at our discrete grid of points. As mentioned previously, to save computation time we reference previously calculated quantity $C\theta$ (equation 3.4) In this process, for each sample i of n total, we:

1. Divide the vector of full conditional distributions by their sum:

$$\frac{L(y_i|\theta_i,\tau_i)\pi(\theta_i)}{\sum_{i=1}^{n} L(y_i|\theta_i,\tau_i)\pi(\theta_i)} \tag{3.11}$$

   Since the full conditional distribution is proportional to the posterior distribution up to a constant, this normalization process removes the constant. It also converts the quantities into probabilities.

2. Find the cumulative density function (CDF) of the normalized full conditional distributions

3. Generate a random number u from a uniform distribution between 0 and 1.

For the first iteration, we assume an arbitrary first value for θ and τ.

The data $y(x,\theta)$ used in calibration can be generated either directly from a finite element model of the system, or a surrogate model trained on FE data. The latter approach is useful when the computer model is computationally expensive and an accurate surrogate model is possible and easy to generate. However, if these two conditions are not met–if the computer model runs quickly or an accurate model emulator cannot be trained–the computer model alone should be used to generate the data used in calibration.

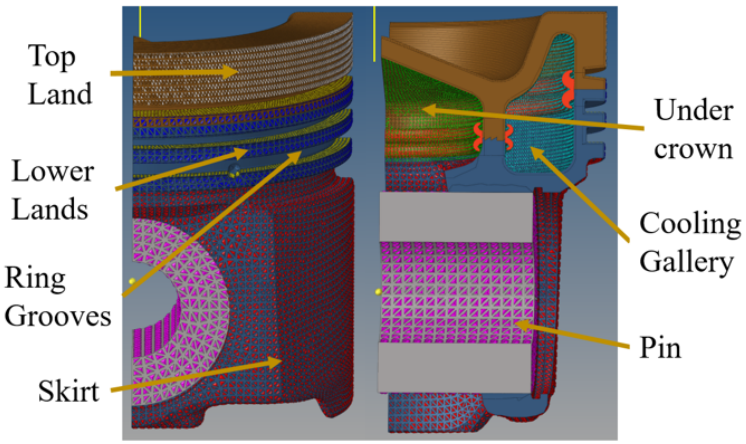

Figure 3.1: Diesel engine finite element model [22]

3.3 Initial Verification

The Griddy Gibbs calibration methodology was first used to calibrate three coefficients in a heat transfer analysis of an engine piston. These heat transfer coefficients correspond to the heat flux on the backside of the piston in three locations–the undercrown, cooling gallery 1, and cooling gallery 2–which are difficult to determine experimentally. These locations can be seen in the finite element model of the piston in figure 3.1.

These parameters were calibrated using temperature data at different locations in the piston, temporally averaged during the analysis. The locations were chosen on the surface of each boundary condition where the temperature change would be most sensitive to each parameter. They can be seen in figure 3.2. The amount of data points was varied in this study, but for each variation it was ensured that data was taken from each of the three locations of interest. For more information about this data, see [22].

The data used in each test case was generated using a deterministic finite element model of the piston. A uniform grid of test points was used to ensure that

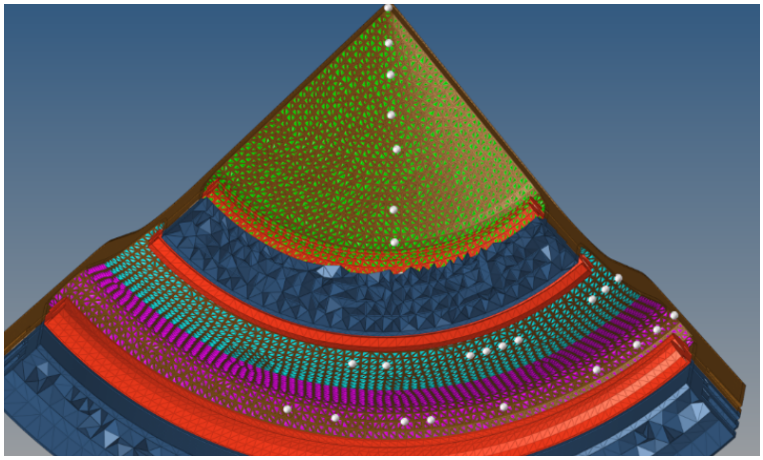

Figure 3.2: Underside of piston model, with 27 output locations shown. The undercrown is shown in green, cooling gallery 1 in magenta, and cooling gallery 2 in cyan.

the entire range of each parameter was represented, and the true value for each parameter would not exist in a sparsely populated area. This grid is represented in figure 3.3

Because running this FE model is extremely time intensive, it is infeasible to run for the hundreds of simulations necessary to calibrate the piston model using a discrete Griddy Gibbs methodology incapable of interpolation. Therefore, the finite element model was run the 343 times required to train an accurate surrogate model, and this surrogate model was used to generate the remainder of the temperature data used in calibration. The computation times reported in the results of this paper do not include the time needed to train the surrogate model; they only reflect the time required to complete the calibration.

The data for which parameters are being estimated is not true experimental temperature data, but rather FE model data. In order to mimic experimental data, error has been added to these values. This error follows the following form:

$$\epsilon \sim MVN(0, 0.4I) \tag{3.12}$$

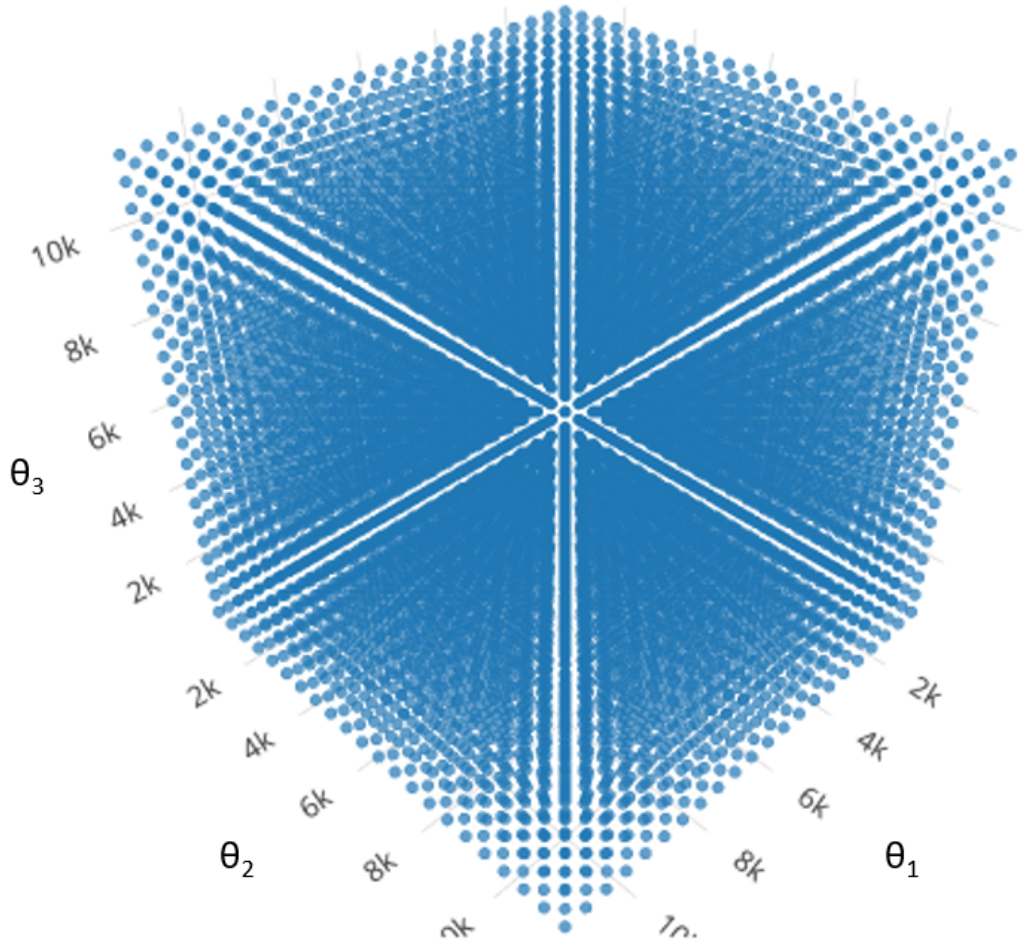

Figure 3.3: Fixed grid design of experiments

The term I indicates the identity matrix. While error exists in the experimental data, it should be noted that for simplicity the formulation of the calibration does not specifically include error in its formulation. The initial verification will determine if the calibration is reasonably accurate even without explicitly accounting for error. A similar verification will then be performed to investigate the effect of a discrepancy between the experimental data and simulation data. This discrepancy will take the form:

$$\delta \sim MVN(0, 40C) \tag{3.13}$$

where C represents the correlation structure of the discrepancy.

$$C = \sigma^2 \prod_{k=1}^{n_x} \rho_k^{4[x_k - x_{k-1}]^2} \tag{3.14}$$

Here, σ^2 is the variance and ρ is the correlation parameter. The correlation parameter used here was 0.8. The variance was set as 40, which is a reasonable amount of variance one would expect for this application. A comparison of the response data with and without error and discrepancy can be found in figure 3.4.

The Griddy Gibbs methodology was initially assessed for its response to varying two values: the number of simulations, or different values and combinations of the unknown parameters; and the amount of experimental data to which to calibrate. The goals of this verification were to determine the optimum number of simulations and experimental data points which provided enough data for the Griddy Gibbs sampler to consistently select the correct value for each parameter, while minimizing run time.

To this end, the Griddy Gibbs calibration methodology was used to calibrate piston data using 14 different values for the number of simulations: 27, 64, 125, 216,

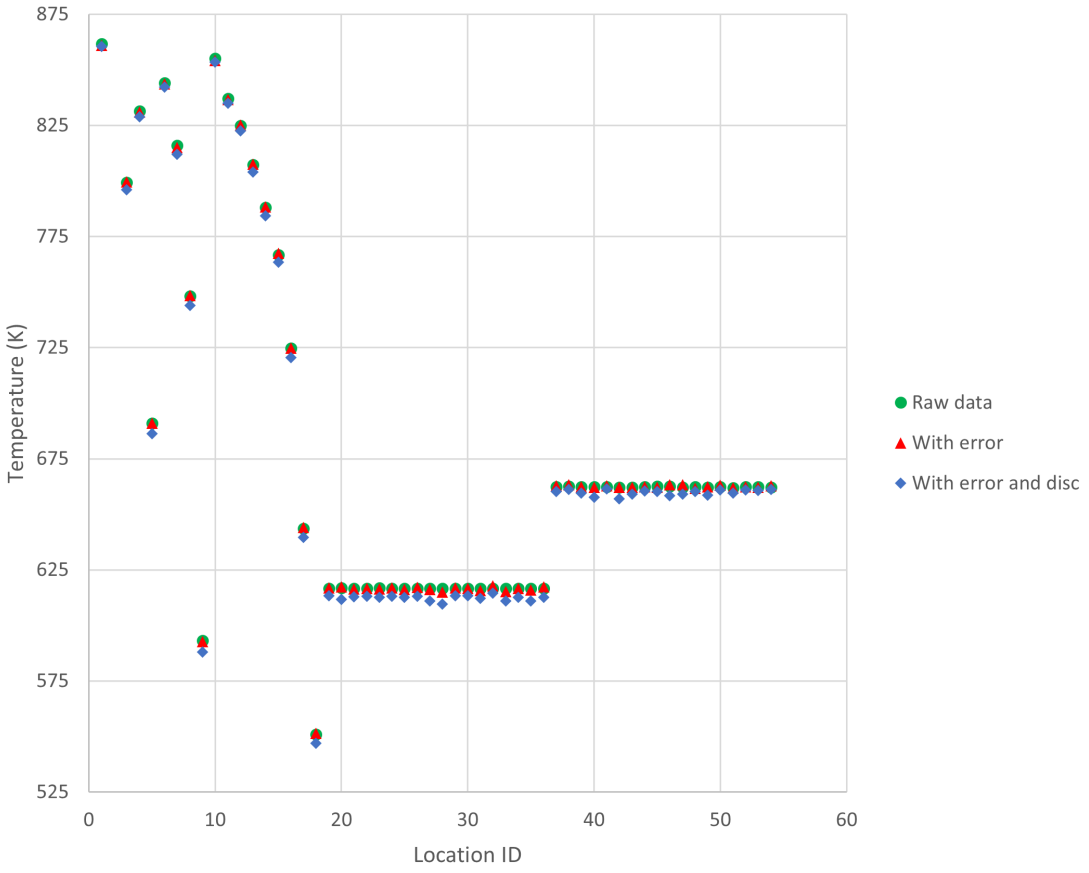

Figure 3.4: Experimental data with and without error and discrepancy

343, 512, 729, 1000, 3375, 10648, 19683, 32768, 59319, and 103823. As there were three parameters being calibrated, these values correspond to 3, 4, 5, 6, 7, 8, 9, 10, 15, 22, 27, 32, 39, and 47 samples in the range of each calibration parameter. These ranges are summarized in table 3.1.

The true value for θ_1, θ_2, and θ_3 for each run was 1000, 3000, and 2000 $\frac{W}{m^2K}$, respectively. For each number of simulations used, separate cases were run using 6, 15, 27, and 64 experimental data points. The calibration consisted of 35,000 runs of the MCMC with a burn-in of 15,000.

	θ_1	θ_2	θ_3
True value	1000	3000	2000
Minimum	13.85	8.61	35.33
Maximum	11984.70	11977.45	12002.49
Range	11970.85	11968.84	11967.16

Table 3.1: Parameter ranges and true values for piston model calibration

3.3.1 Results Without Discrepancy

The results of the initial verification are shown in figures 3.5 through 3.8 below. For each setting of experimental data, the impact of number of simulation on parameter estimate and 95% confidence interval is shown. In figure 3.9, the impact of the number of simulations on calibration runtime is shown. For this study, only random error was added to the experimental data.

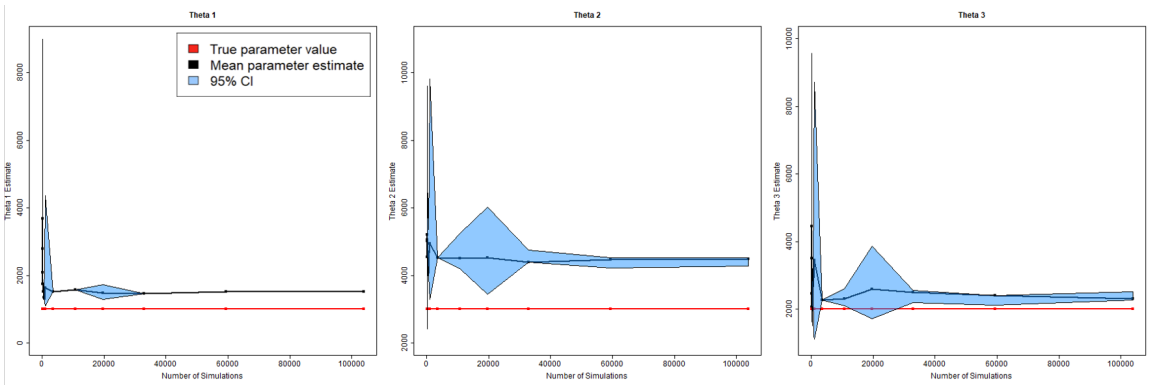

Figure 3.5: Varying number of simulations, 6 experimental data points, no discrepancy

As expected, the 95% confidence intervals tend to narrow as the amount of experimental data increases. Increasing the number of simulations also improves the accuracy of calibration. This can be seen in a comparison of the posterior predictive checks for the lowest (27) and highest (103823) number of simulations as well as the lowest (6) and highest (54) number of data points included in this study in figures 3.10 through 3.13.

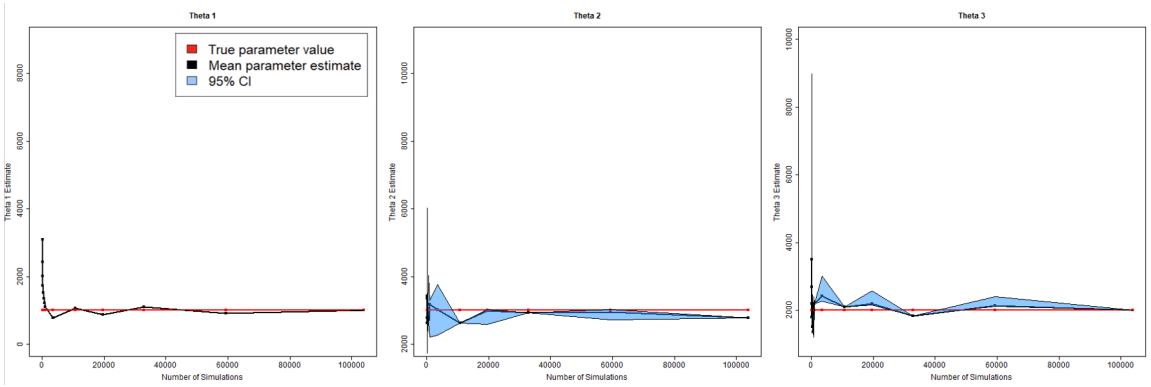

Figure 3.6: Varying number of simulations, 15 experimental data points, no discrepancy

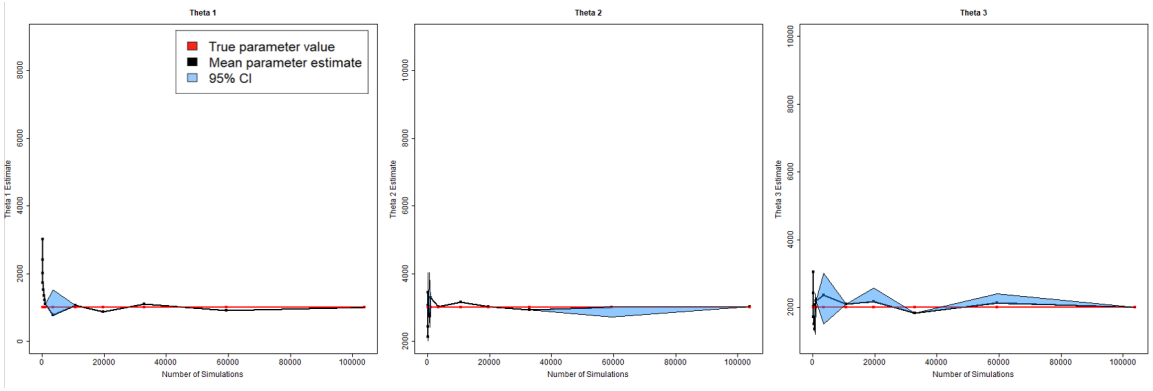

Figure 3.7: Varying number of simulations, 27 experimental data points, no discrepancy

For a calibration using only six data points, the true parameter values (shown in red) are not captured in the 95% confidence interval. As the number of experimental data points increases to 27 and 54, each parameter value becomes contained within this interval. The mean parameter estimates tend to stabilize within the confidence interval at the 59319 simulation case. Therefore, this value was selected as the optimum number of simulations within the given range to inform the Griddy Gibbs calibration. This number corresponds to 39 samples in each of the three parameter dimensions. Even at the lowest amount of data points used in this study (6), at 59319 simulations an accurate prediction of temperature data can be produced.

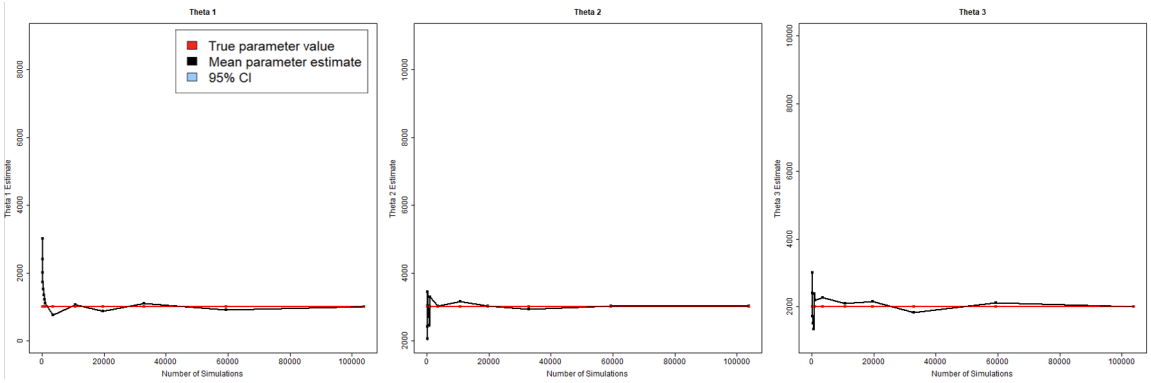

Figure 3.8: Varying number of simulations, 54 experimental data points, no discrepancy

3.3.2 Results With Discrepancy

The test cases from the previous section were then repeated, adding both error and discrepancy to the experimental data. The results from this study are produced in figures 3.14 through 3.22. The time taken for calibration is summarized in figure 3.18.

The results from this study do not differ notably from those in the previous study. It appears that the presence of systematic discrepancy does not significantly impact the quality of calibration, and neglecting the discrepancy term in the formulation of the data model used in calibration is an acceptable simplification.

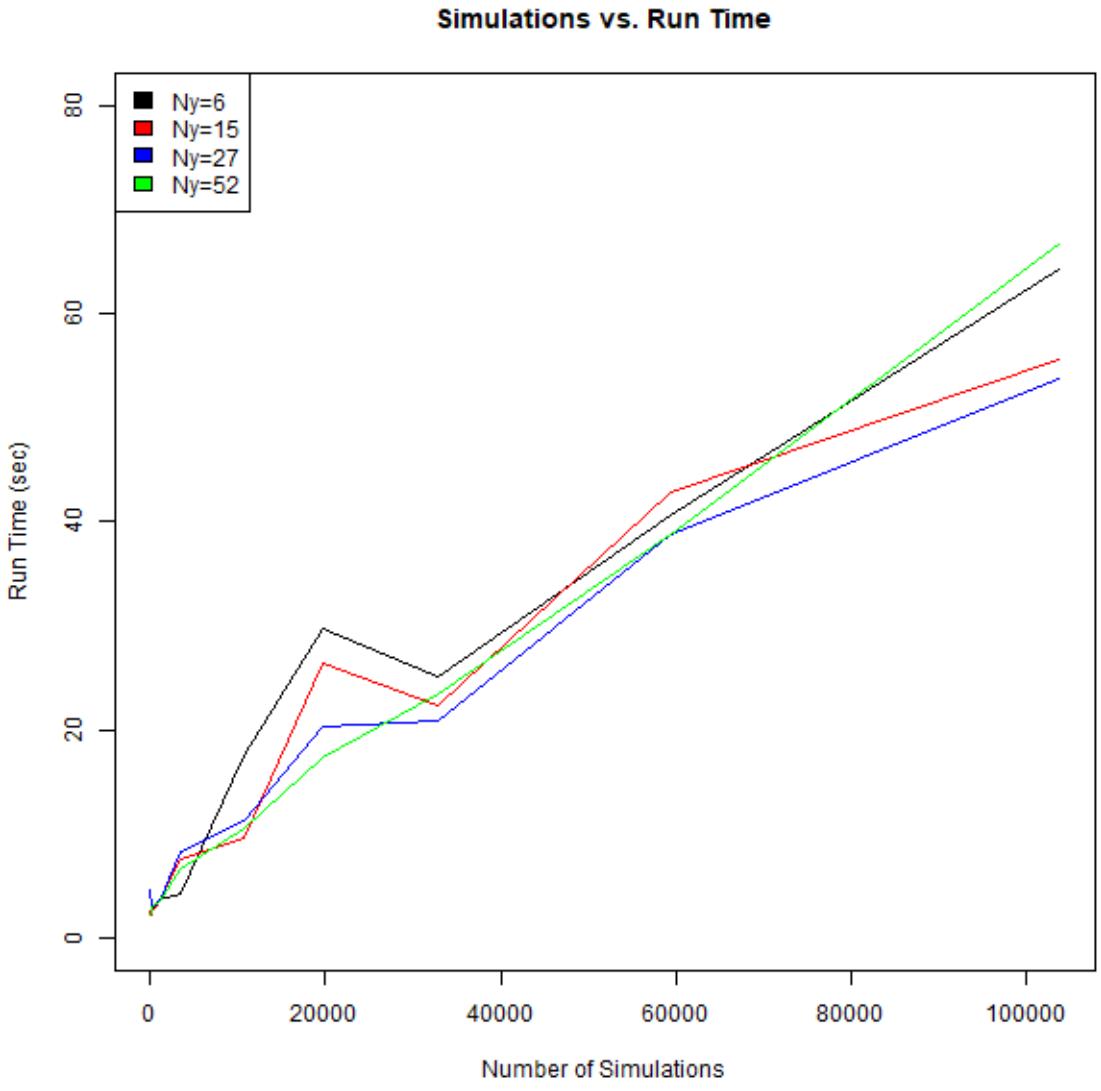

Figure 3.9: Runtime comparison, no discrepancy

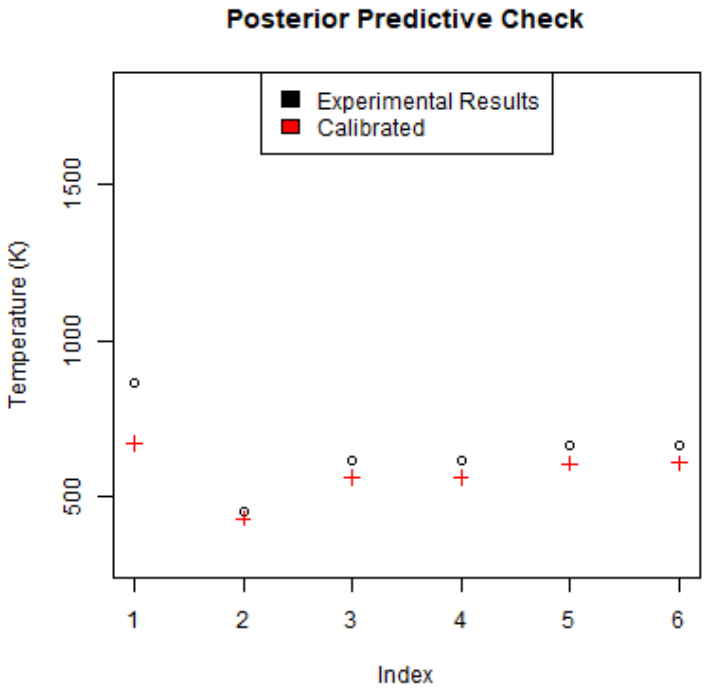

Figure 3.10: Posterior predictive check, 6 data points, 27 simulations, no discrepancy

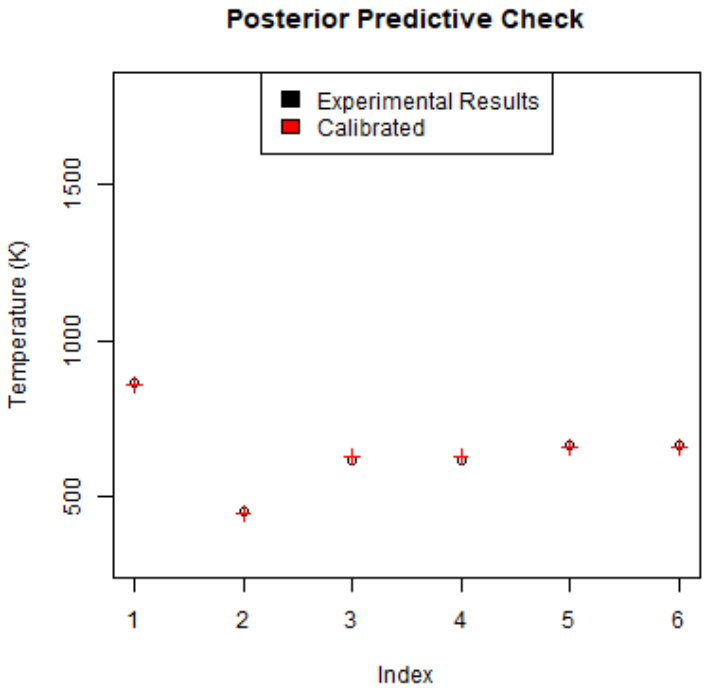

Figure 3.11: Posterior predictive check, 6 data points, 59319 simulations, no discrepancy

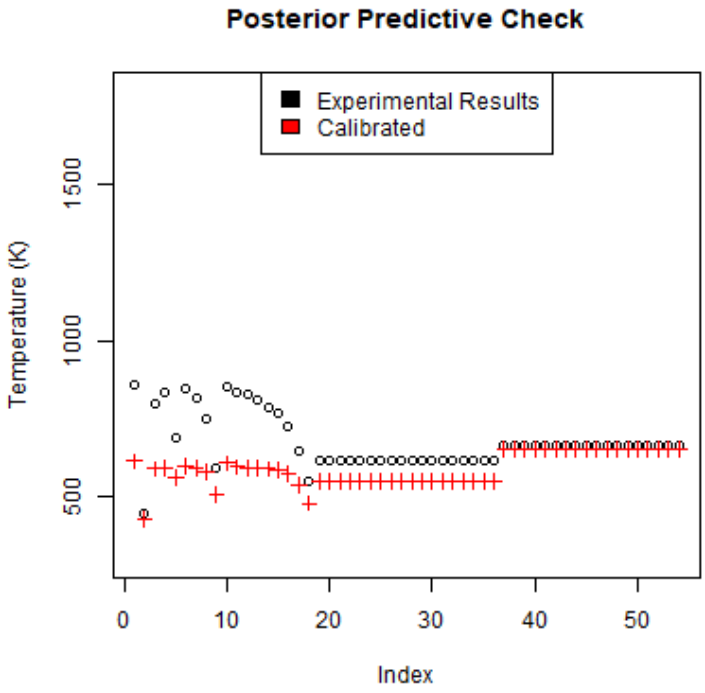

Figure 3.12: Posterior predictive check, 54 data points, 27 simulations, no discrepancy

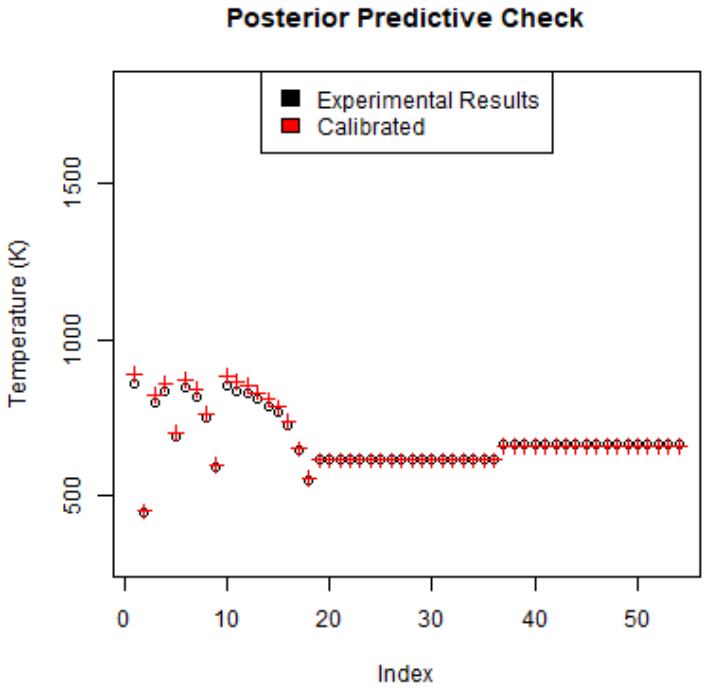

Figure 3.13: Posterior predictive check, 54 data points, 59319 simulations, no discrepancy

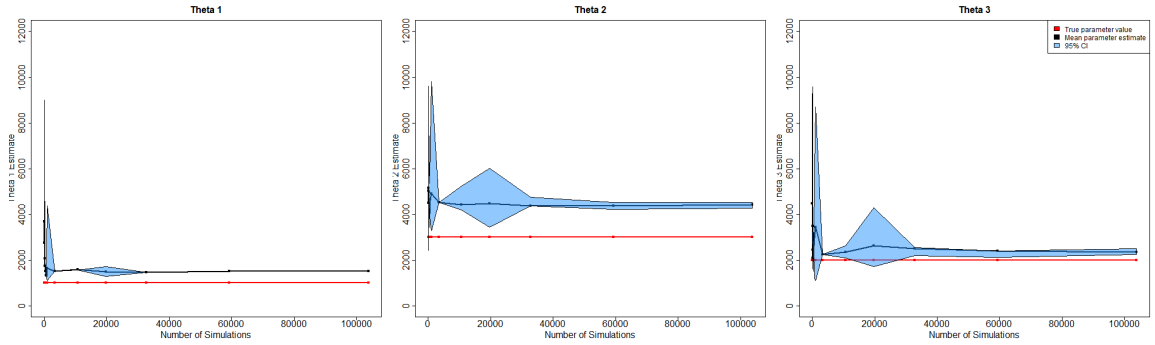

Figure 3.14: Varying number of simulations, 6 experimental data points, with discrepancy

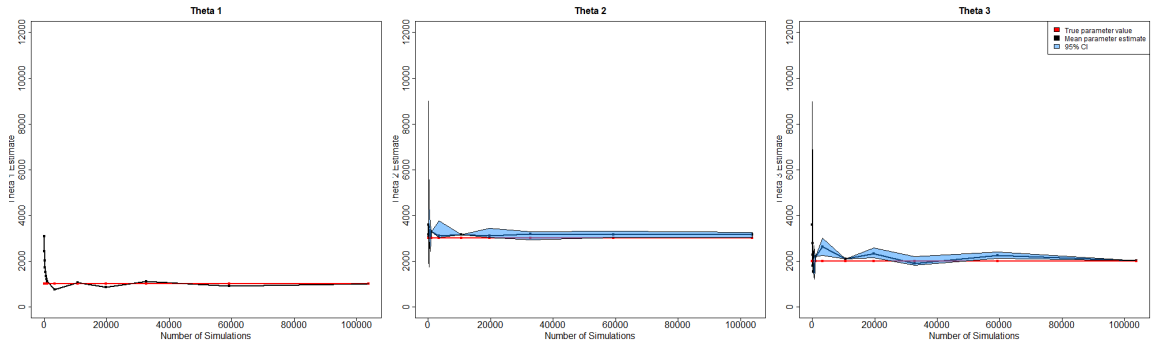

Figure 3.15: Varying number of simulations, 15 experimental data points, with discrepancy

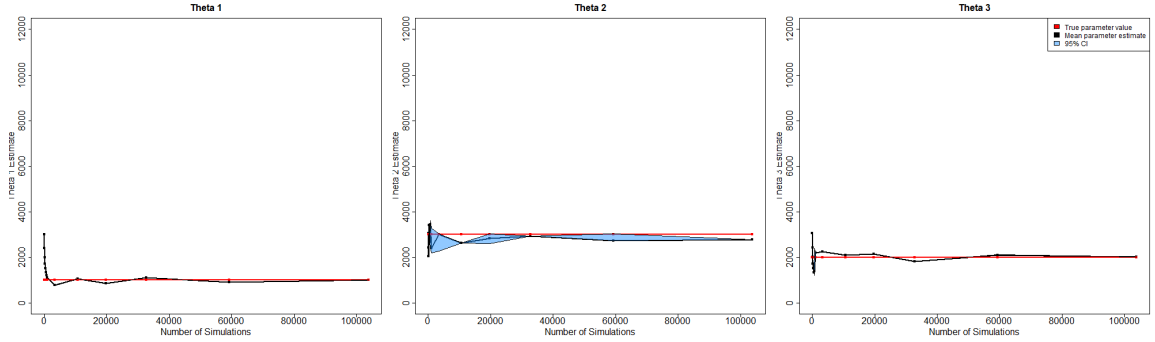

Figure 3.16: Varying number of simulations, 27 experimental data points, with discrepancy

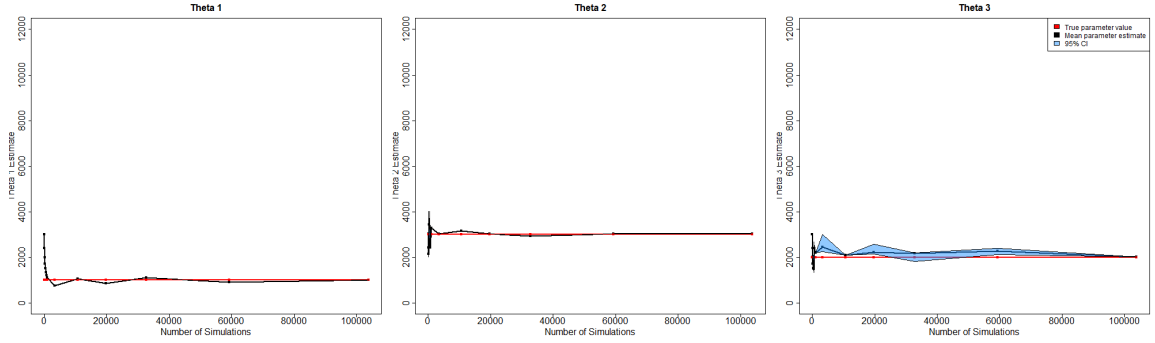

Figure 3.17: Varying number of simulations, 54 experimental data points, with discrepancy

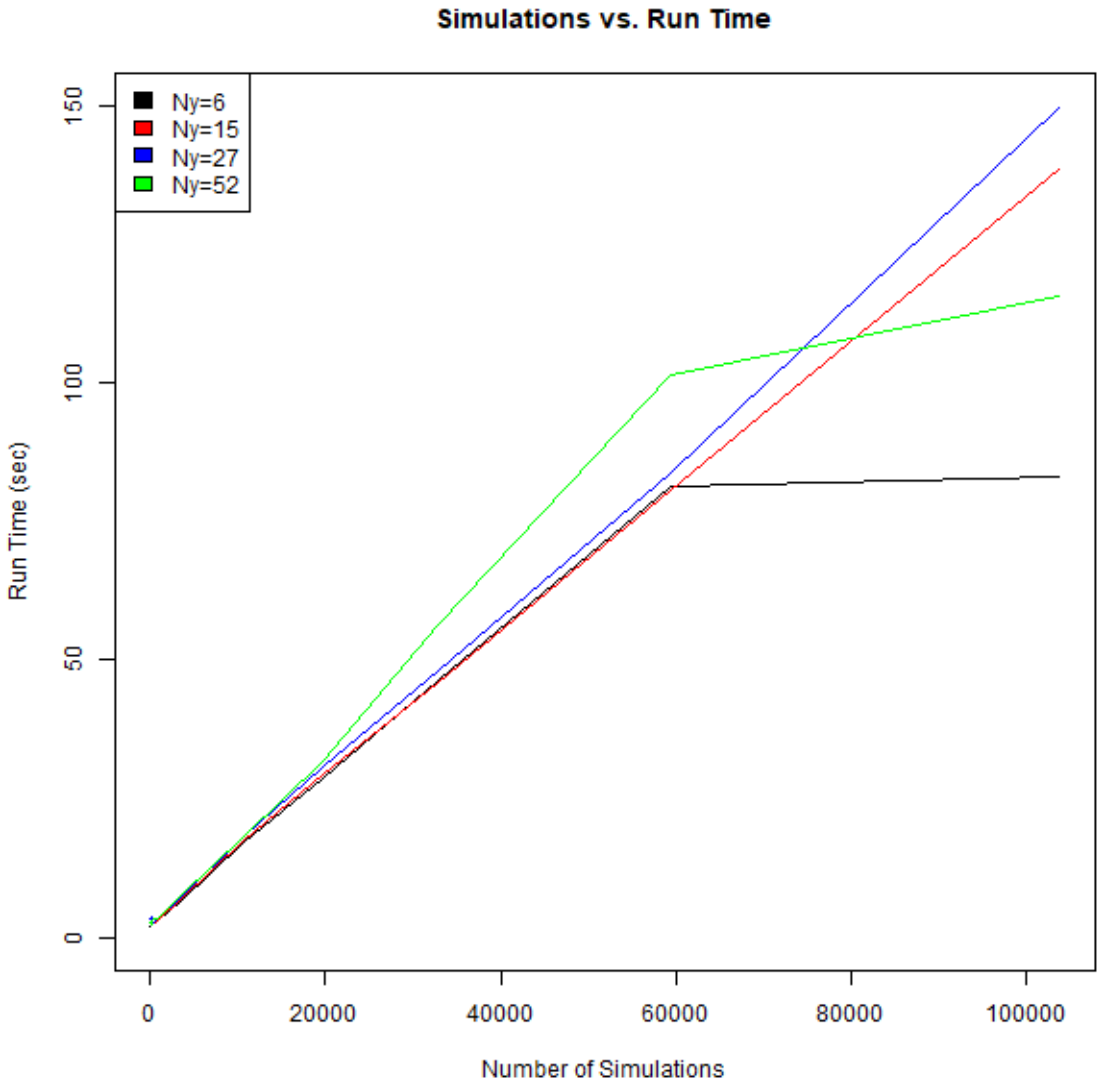

Figure 3.18: Time comparison, with discrepancy

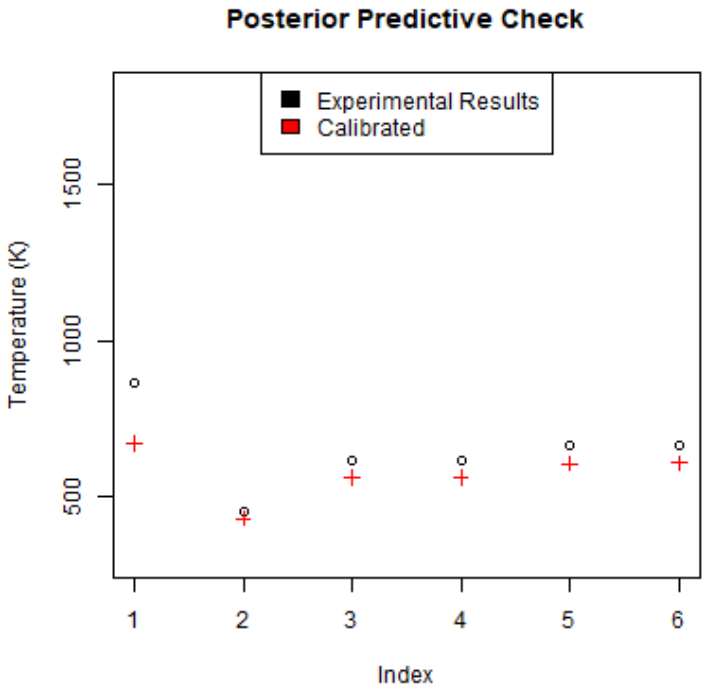

Figure 3.19: Posterior predictive check, 6 data points, 27 simulations, with discrepancy

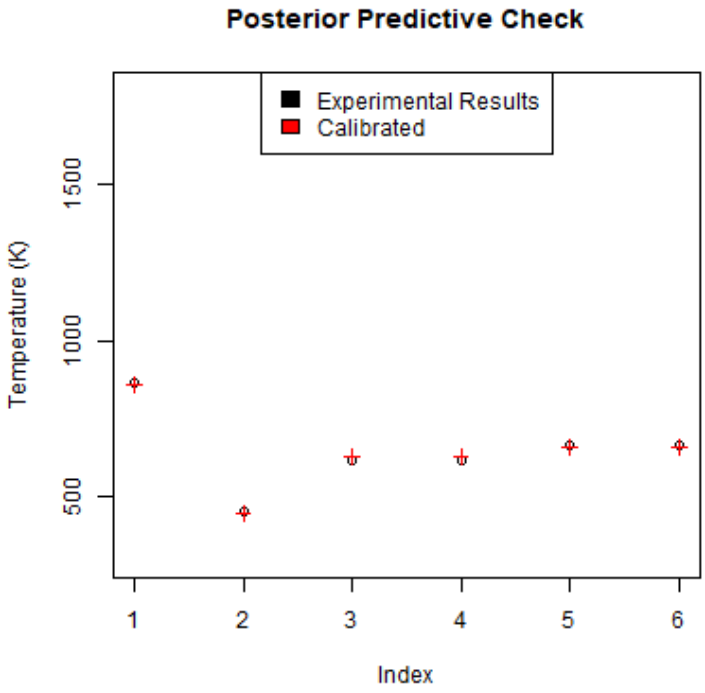

Figure 3.20: Posterior predictive check, 6 data points, 59319 simulations, with discrepancy

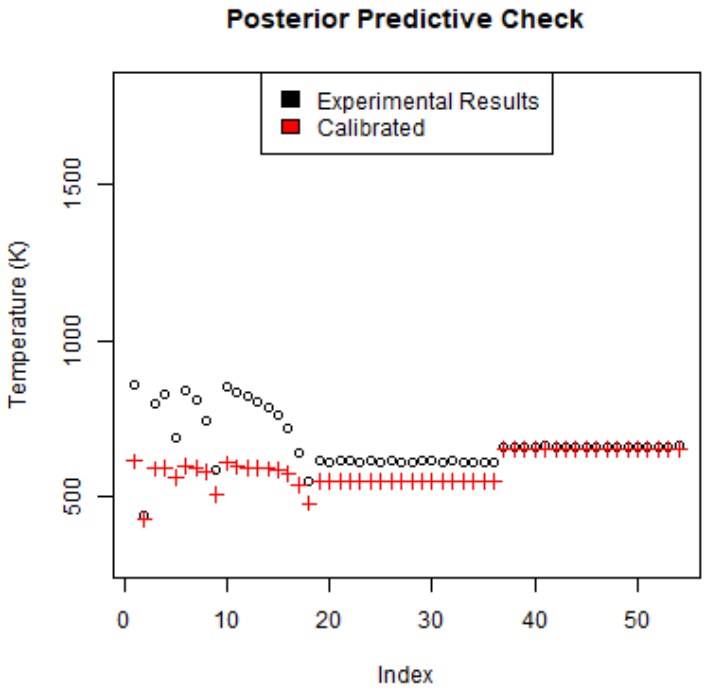

Figure 3.21: Posterior predictive check, 54 data points, 27 simulations, with discrepancy

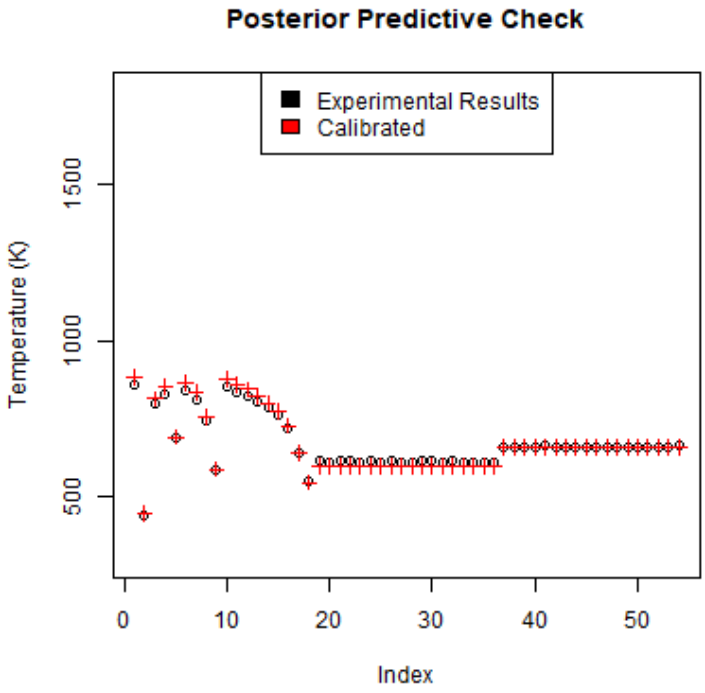

Figure 3.22: Posterior predictive check, 54 data points, 59319 simulations, with discrepancy

Chapter 4

Griddy Gibbs Compared to Traditional Bayesian Calibration

The Griddy Gibbs model calibration method was compared to a traditional Bayesian method in the same calibration of three backside boundary conditions in a heat transfer analysis of an engine piston. The relative accuracy of each method was tested in calibrating to 6, 15, 27, and finally 54 experimental data points. The Griddy Gibbs method used 59319 simulations (39 per parameter), which was the optimum value determined in its initial verification. The Griddy Gibbs results for each case were compared to a traditional Bayesian calibration methodology trained on 343 simulations (7 per calibration parameter). 343 simulations were used because additional simulations did not improve the efficacy of the traditional methodology. Discrepancy was included in the experimental data, but was not sampled or included in the data model used in calibration, as in section 3.3.2.

Posterior distributions for each parameter and posterior predictive checks are produced in figures 4.1 through 4.7 and tables 4.1 through 4.4. Computation times were compared in table 4.5.

In the posterior distributions, the blue bins represent results from the Griddy Gibbs code, and the red is the traditional Bayesian code. Areas in which they overlap are shaded purple. Each blue bin and square along the x-axis represents a different discrete available parameter setting in the design of experiments for the Griddy Gibbs calibration. As expected, as the number of simulations increase, these bins will become more narrow, and there will be more settings of each parameter represented along the x-axis.

		Traditional Bayesian Calibration	Griddy Gibbs Calibration
θ_1	95% LB	980.4	1511.2
	Estimate	1004.8	1511.3
	Truth	1000	1000
	95% UB	1027.7	1511.2
θ_2	95% LB	2868.7	4220.0
	Estimate	2959.7	4457.2
	Truth	3000	3000
	95% UB	3058.3	4519.1
θ_3	95% LB	1986.1	2113.2
	Estimate	2078.2	2401.0
	Truth	2000	2000
	95% UB	2170.9	2411.9

Table 4.1: Parameter summary for 6 data points

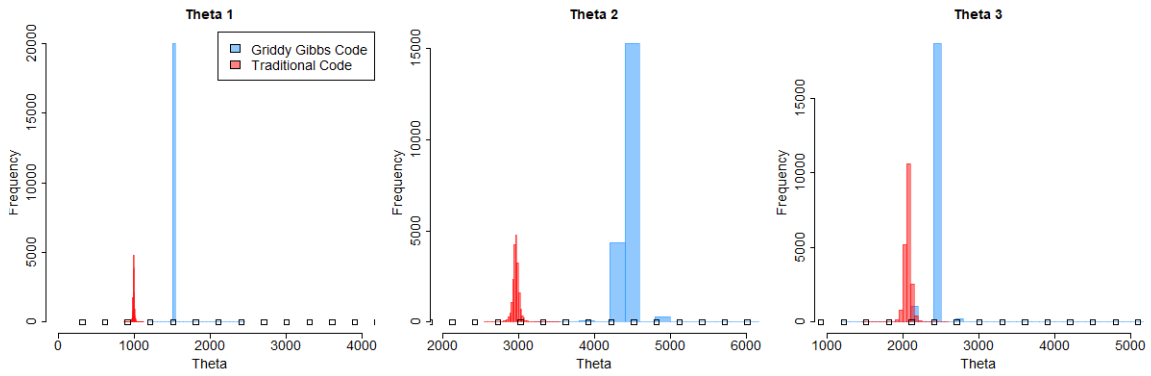

Figure 4.1: Posterior distribution, 6 data points

		Traditional Bayesian Calibration	Griddy Gibbs Calibration
	95% LB	959.2	912.5
θ_1	Estimate	971.8	912.7
	Truth	1000	1000
	95% UB	985.0	912.5
	95% LB	3129.5	2724.4
θ_2	Estimate	3188.9	2957.0
	Truth	3000	3000
	95% UB	3253.5	3023.5
	95% LB	2108.6	2113.2
θ_3	Estimate	2159.8	2144.1
	Truth	2000	2000
	95% UB	2212.2	2411.9

Table 4.2: Parameter summary for 15 data points

		Traditional Bayesian Calibration	Griddy Gibbs Calibration
	95% LB	997.8	912.5
θ_1	Estimate	1005.4	913.2
	Truth	1000	1000
	95% UB	1013.1	912.5
	95% LB	2835.2	2724.4
θ_2	Estimate	2864.5	3010.5
	Truth	3000	3000
	95% UB	2894.8	3023.5
	95% LB	1899.9	2113.2
θ_3	Estimate	1926.8	2136.7
	Truth	2000	2000
	95% UB	1855.7	2411.9

Table 4.3: Parameter summary for 27 data points

		Traditional Bayesian Calibration	Griddy Gibbs Calibration
θ_1	95% LB	1007.0	912.5
	Estimate	1010.5	912.5
	Truth	1000	1000
	95% UB	1014.1	912.5
θ_2	95% LB	3013.1	3023.5
	Estimate	3027.7	3023.1
	Truth	3000	3000
	95% UB	3041.9	3023.5
θ_3	95% LB	2092.6	2113.2
	Estimate	2106.1	2279.9
	Truth	2000	2000
	95% UB	2120.0	2411.9

Table 4.4: Parameter summary for 54 data points

Data Points	**Traditional Bayesian Calibration**	**Griddy Gibbs Calibration**
6	653.8	40.6
15	15321.5	42.8
27	79817.2	38.8
54	610182.3	38.8

Table 4.5: Comparison of runtime, in seconds

As the number of experimental data points increases, The posterior distributions are able to more closely match those generated using traditional Bayesian methods, and simulation results generated using calibrated values for each parameter (i.e. the posterior predictive check) more closely matches the data to which we were calibrating.

4.1 Summary of Results

The primary difference between the novel model calibration technique presented in this paper and traditional Bayesian model calibration is that the Griddy Gibbs technique samples only from existing simulation data. Therefore, there is no need to create and train a Gaussian process model, which allows for a dramatic reduction in computation time, particularly if the number of data points is high. Increasing experimental data points from 6 to 54 increased runtime for the traditional Bayesian calibration by over 90,000%, from 10 minutes to 169 hours, the Griddy Gibbs calibration runtime stayed effectively constant at around 40 seconds.

The Griddy Gibbs calibration methodology was less accurate than the traditional Bayesian methodology when data was scarce. Even when more data was available with which to calibrate, the traditional methods were still able to provide a slightly more accurate calibration. It is hypothesized that the traditional Bayesian calibration may better handle situations in which some or all of the parameters are confounding, that is they are not completely independent.

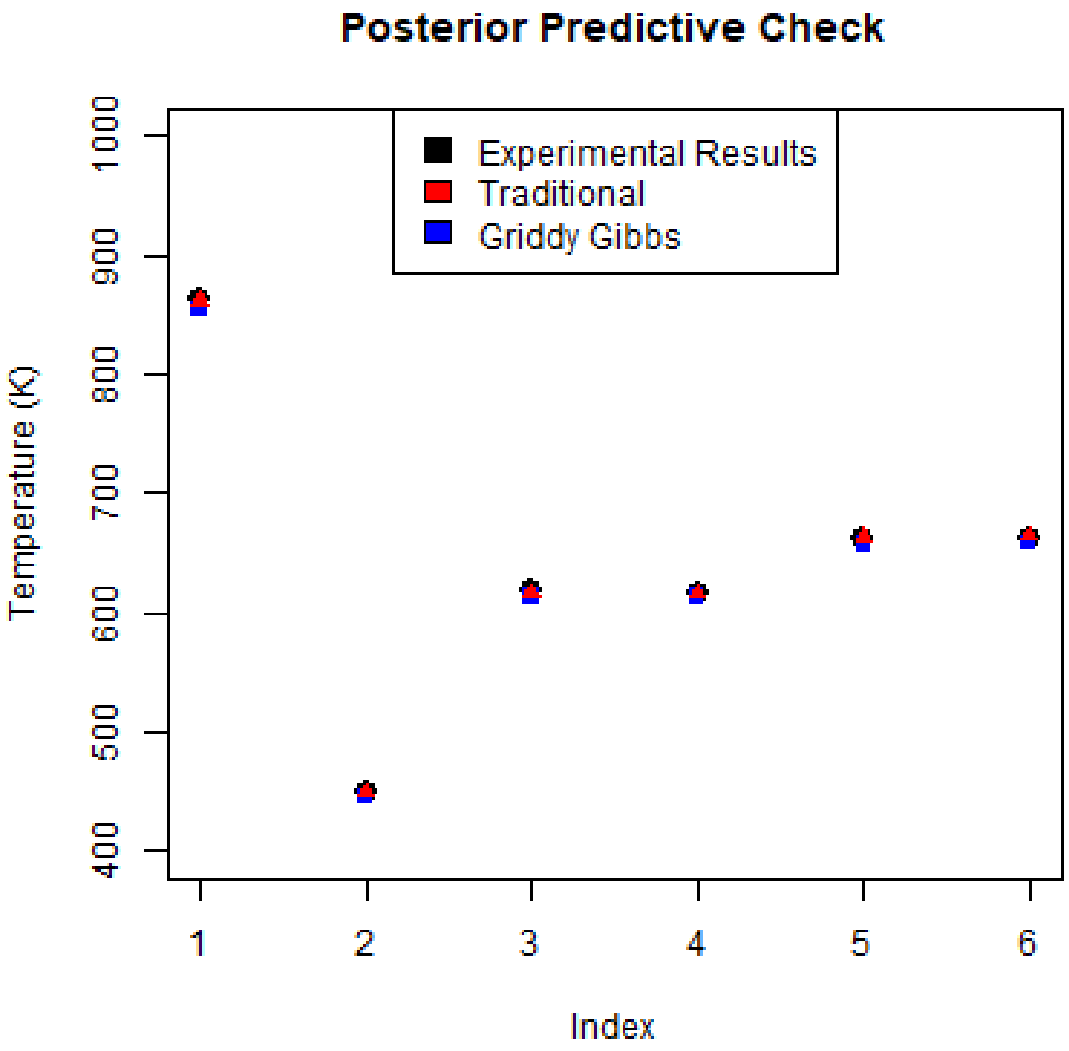

Figure 4.2: Posterior predictive check, 6 data points

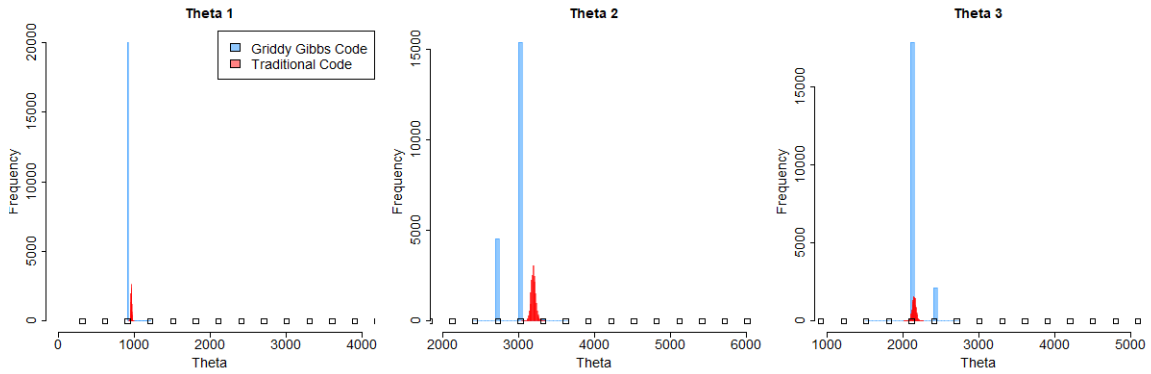

Figure 4.3: Posterior distribution, 15 data points

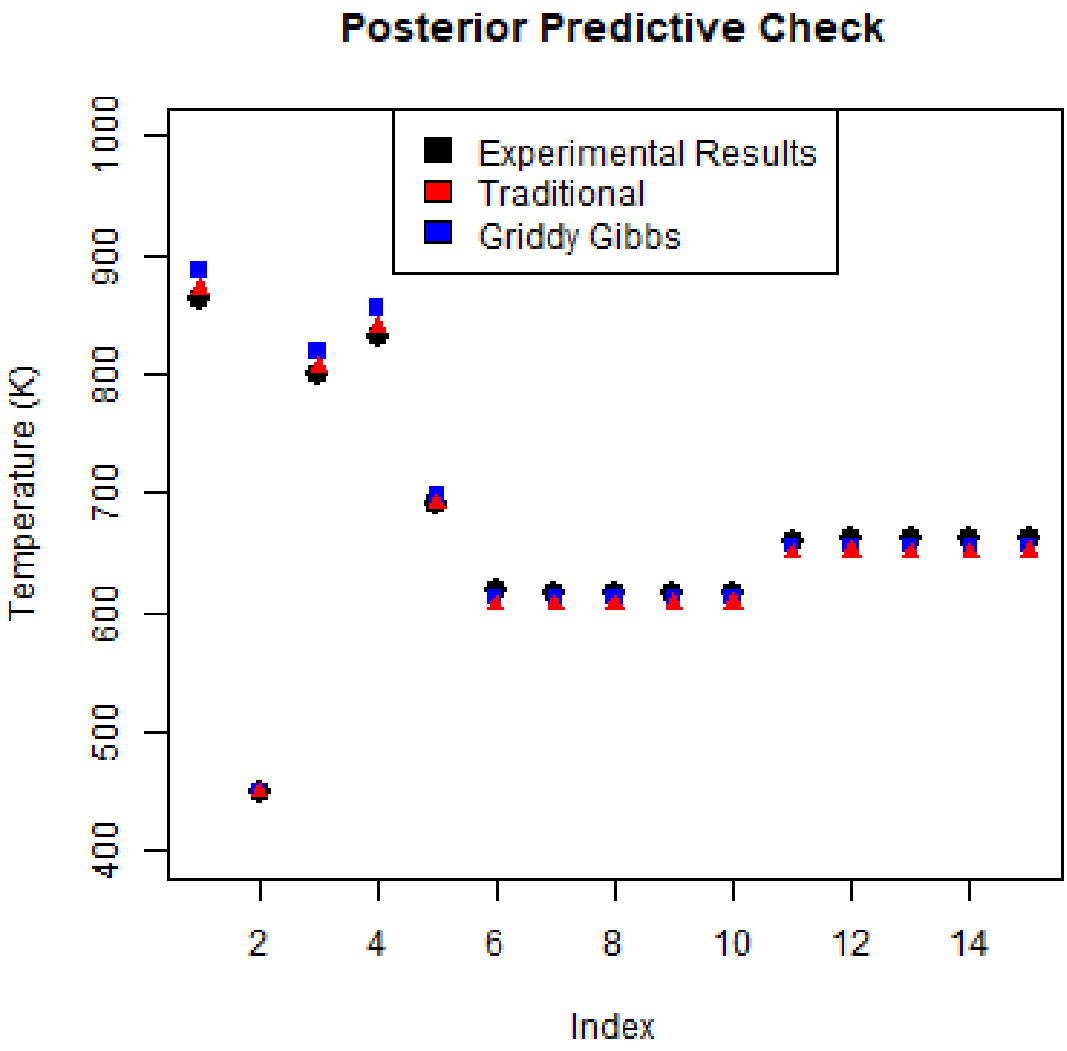

Figure 4.4: Posterior predictive check, 15 data points

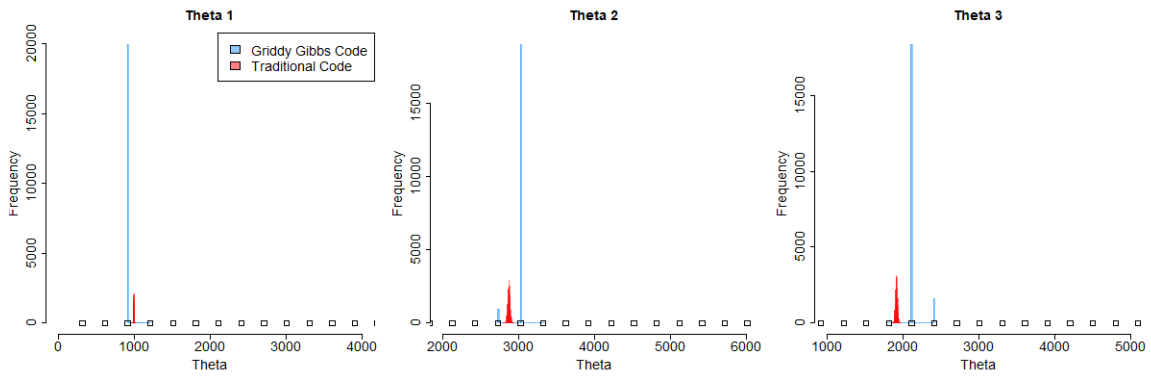

Figure 4.5: Posterior distribution, 27 data points

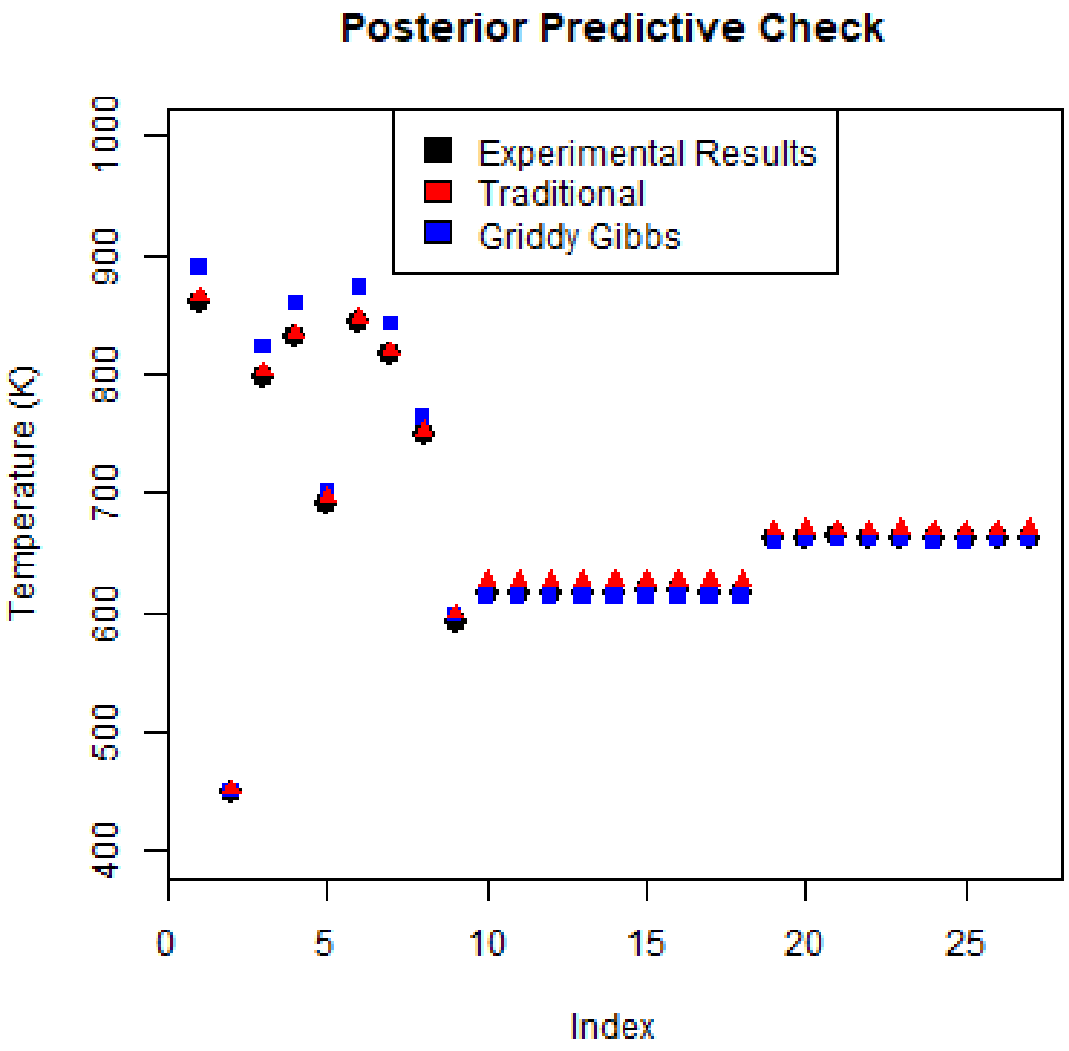

Figure 4.6: Posterior predictive check, 27 data points

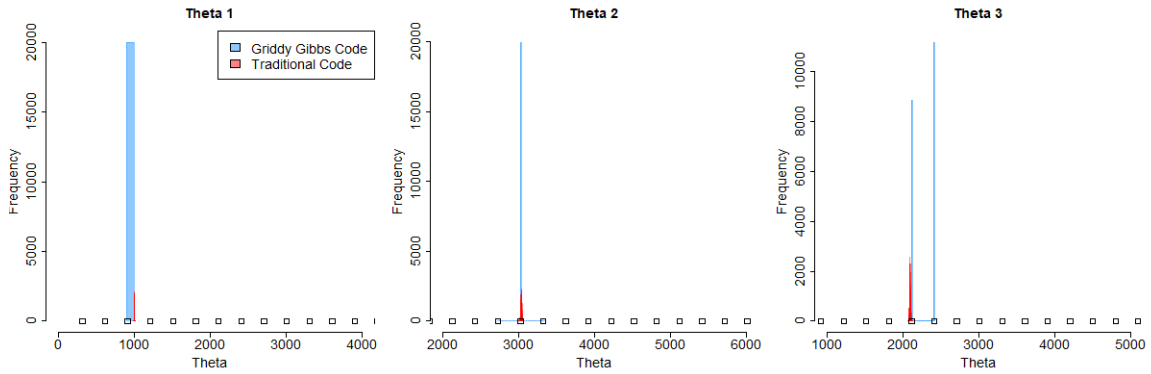

Figure 4.7: Posterior distribution, 54 data points

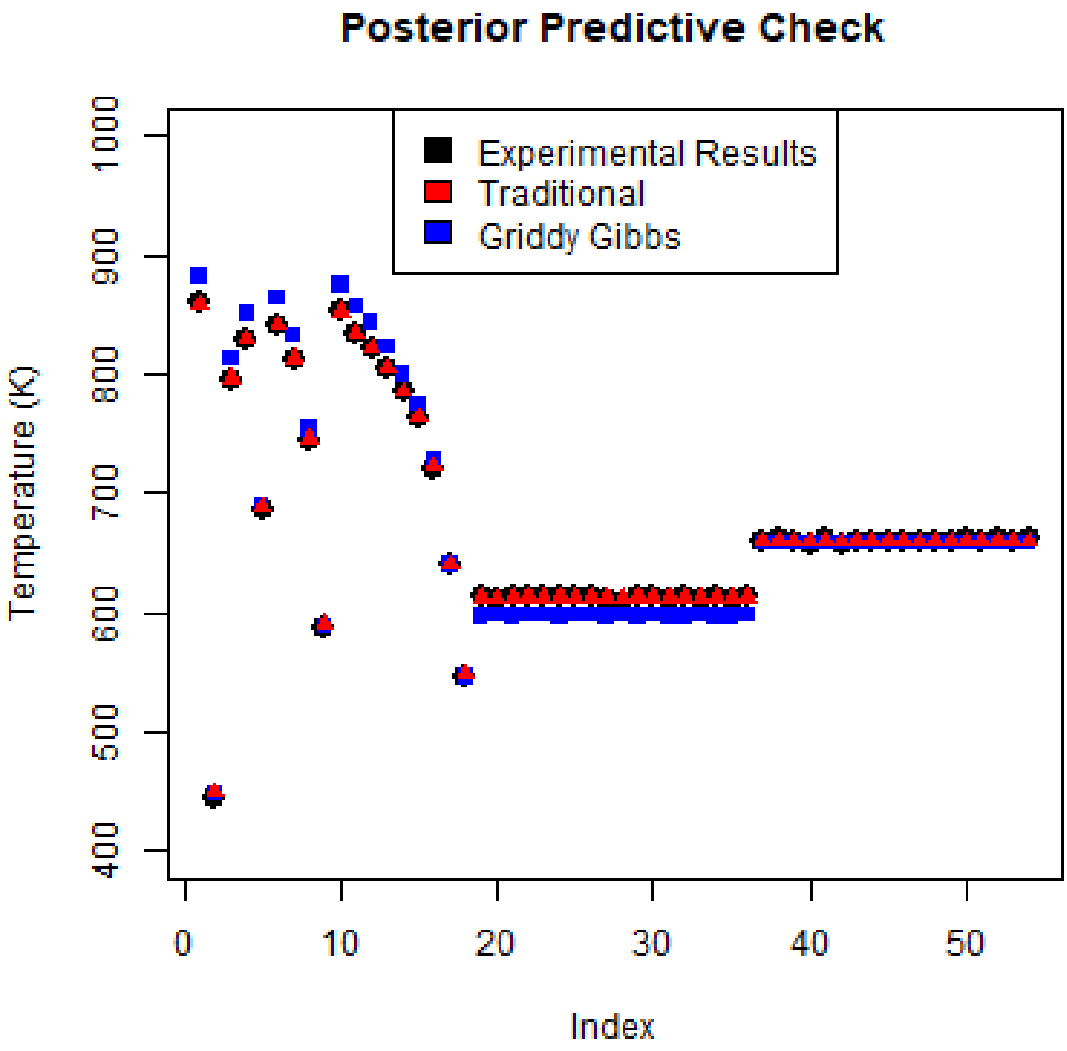

Figure 4.8: Posterior predictive check, 54 data points

Chapter 5

Tire Model Calibration

After completing the initial Griddy Gibbs verification process, we applied the algorithm to a vehicle dynamics problem using time series data. We aimed to estimate the inputs to the Fiala tire model used to generate a set of experimental data, specifically the normal force generated in the contact patch of a tire as it runs over a speed bump. In the absence of experimental data, we are calibrating to simulation data with an added Gaussian discrepancy. This data point was held back from the simulation data the algorithm was able to sample from. In the first study, the Fiala tire model was used to generate both the "experimental" data (in this case model data with a discrepancy added) and the model simulation data used to inform the calibration. In the second study, we calibrate Fiala tire parameters to PAC2002 simulation data. This simulation data was generated at a large number (6,000) of simulations, for 2,000 unique parameter combinations at 3 settings of the independent x variable, speed bump height. 50,000 MCMC iterations were used for calibration.

5.0.1 Data Model

We again assume our unknown parameters follow the multivariate normal distribution. We take the same relationship between the data model and experimental data as in equation 3.1, with unknown parameters vertical stiffness (θ_1) and vertical damping (θ_2) and independent variables speed bump height and time. Again, discrepancy was not included in the formulation of the data model, but a discrepancy was added to the results to which we are calibrating in order to more closely resemble true experimental data.

5.1 Simulation and Tire Models

The Tire Testrig within ADAMS/Car was used to perform this simulation and isolate its effects on the tire. A "flat road with cleat" simulation was performed at a speed of 20 m/s. The cleat was 100mm long and occurred immediately after the simulation began. The cleat height was taken as a controllable independent variable in the calibration. Three height settings were used: 412 mm, 274.4 mm, and 121.3 mm. A constant static load of 3 kN was applied to the tire.

The Fiala tire model assumes the tire acts as a beam on an elastic foundation. It assumes the contact patch is rectangular and pressure is uniform across it. It neglects the effects of the tire's camber angle. model calculates the normal force as

$$F_z = min(0, F_{zk} + F_{zc}) \tag{5.1}$$

where F_{zk} is the normal force due to the vertical stiffness of the tire, and F_{zk} is the normal force due to the vertical damping of the tire. These quantities can be

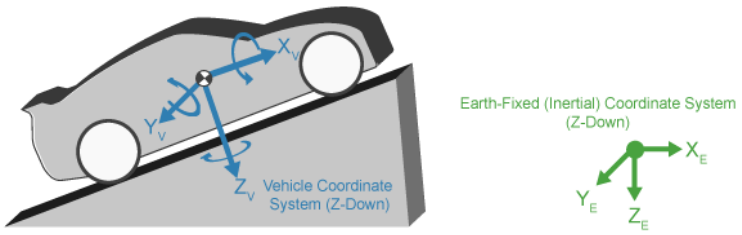

Figure 5.1: SAE coordinate system [14]

calculated per the following:

$$F_{zk} = -k * \rho \tag{5.2}$$

$$F_{zc} = -c * \frac{d\rho}{dt} \tag{5.3}$$

where ρ is vertical deflection or penetration of the tire. These equations use SAE coordinates, in which the positive z-axis is directed downward [9].

In the PAC2002 model, the vertical tire response is modeled as a spring and damper in parallel. The contact patch is found by assuming the tire acts as a rigid disc. The normal force in the tire is calculated as follows:

$$\begin{aligned} F_z = K_z\dot{\rho} + ([1 + q_{pFz1}dp_i]\gamma_{Cz}F_{zo}) \times \left(\frac{q_{Fz1}\rho}{R_o} + \frac{q_{Fz2}\rho^2}{R_o^2} + \frac{q_{Fz3}\gamma^2\rho}{R_o}\right) \times \dots \\ \left(q_{reo} + \frac{q_{v2}|\Omega|R_o}{V_o} - \frac{q_{Fcx1}F_x}{F_{zo}} - \frac{q_{Fcy1}F_y}{F_{zo}} + q_{Fc\gamma 1}\gamma^2\right) \end{aligned} \tag{5.4}$$

where

F_x = Longitudinal force

F_{zo} = Nominal wheel load

Ω = Wheel rotational velocity

R_o = Unloaded radius

V_o = Measurement speed at test bench

γ = Camber (inclination) angle

ρ = Tire deflection

λ_{Cz} = Scale factor for vertical tire stiffness

K_z = Tire vertical damping

q_{REO} = Correction factor for measured unloaded radius

q_{V2} = Tire stiffness variation coefficient with speed

q_{Fcx1} = Tire stiffness interaction with F_x

$q_{Fc\gamma 1}$ = Tire stiffness interaction with camber

q_{Fz1} = Tire vertical stiffness coefficient (linear)

q_{Fz2} = Tire vertical stiffness coefficient (quadratic)

q_{Fz3} = Camber dependency of the tire vertical stiffness

q_{pFz1} = Factor to include effect of inflation pressure

dp_i = Normalized inflation pressure

The PAC2002 tire model is more complex than the Fiala model, and it considers much more information in its calculation.

5.2 Sensitivity Study

A sensitivity study was performed to determine parameters and a response metric with which to perform calibration. The simulation was the same as that de-

scribed previously, except it only involved one speed bump of height 100 mm. This simulation was repeated at 10% intervals across the expected ranges for each of the nine input parameters in the Fiala tire model: vertical stiffness, vertical damping, rolling resistance, "cslip" (the partial derivative of longitudinal force with respect to longitudinal slip ratio at zero longitudinal slip), "calpha" (the partial derivative of lateral force with respect to slip angle at zero slip angle, also known as cornering stiffness), "umin" (the coefficient of friction with full slip, or slip ratio = 1), "umax" (the coefficient of friction with zero slip), and relaxation length in the x and y directions. It was found that changing vertical stiffness and vertical damping at these increments generated pronounced effects on the normal force in the tire. As a result of this study, vertical stiffness and vertical damping were chosen as the two parameters to calibrate to normal force data. For simplicity, and because they were not sensitive to the normal force metric chosen, the remaining seven Fiala input parameters were held constant in this study. Further work into calibrating all Fiala parameters to PAC2002 results is needed in order to apply this calibration methodology to a model selection tool.

The ranges for vertical stiffness and damping can be found in table 5.1. The results of this sensitivity study for vertical stiffness and damping can be found in figures 5.2 and 5.3. A higher vertical stiffness corresponded to higher peak force values which dissipated slower than lower stiffness values. A higher vertical damping corresponded to a much quicker dissipation of force in the tire.

	Vertical stiffness (N/mm)	**Vertical damping (Ns/mm)**
Minimum	50	0
Maximum	800	5
Range	750	5

Table 5.1: Parameter ranges for tire model calibration

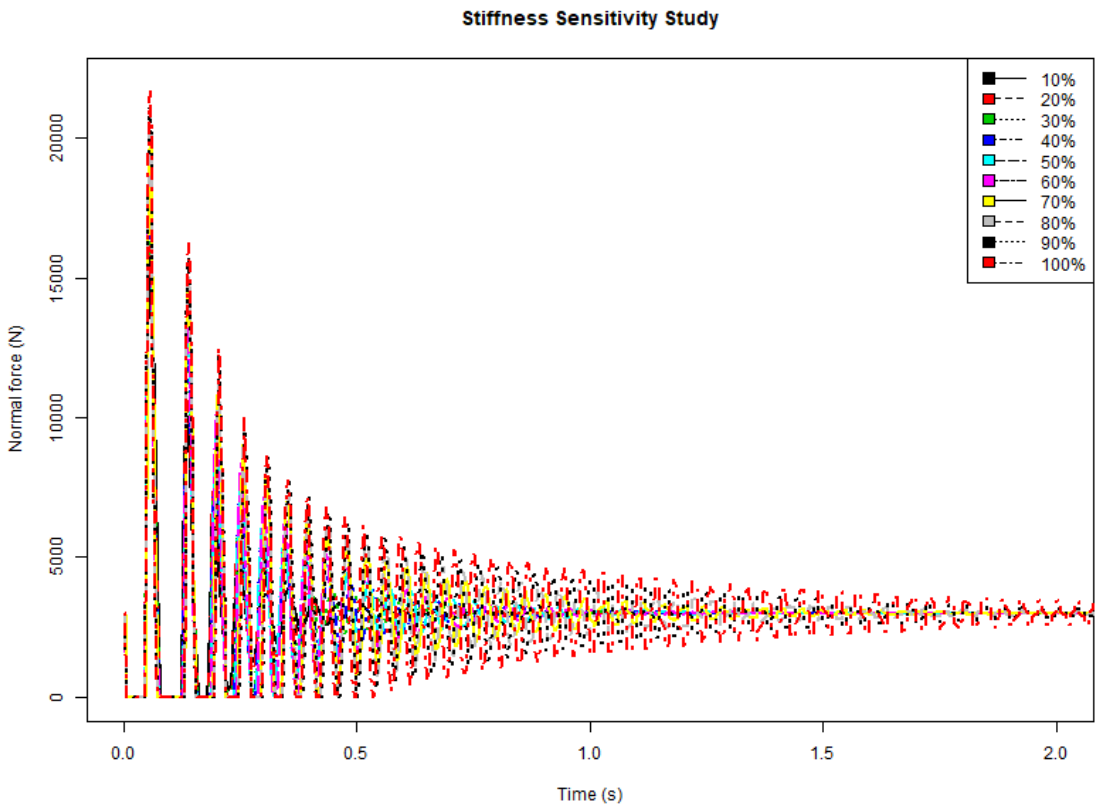

Figure 5.2: Change in normal force subject at 10% increments of vertical stiffness range

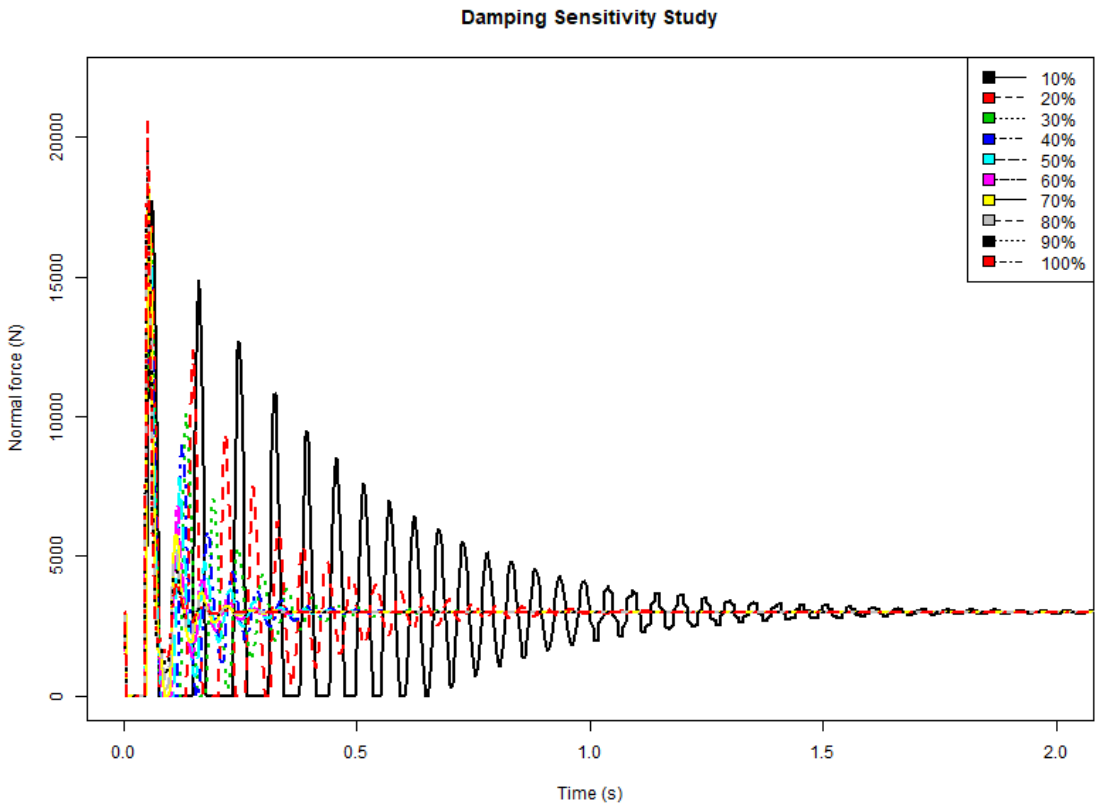

Figure 5.3: Change in normal force subject at 10% increments of vertical damping range

5.3 Fiala to Fiala Calibration

The results for this calibration are summarized in table 5.2.

	95% Lower bound	**θ estimate**	**True value**	**95% Upper bound**
θ_1	460.9	460.9	450	460.9
θ_2	1.53	1.53	1.5	1.53

Table 5.2: Fiala to Fiala calibration results

The 95% confidence interval is reported here, but note that this metric may be misleading for our case where the posterior distribution is composed of a discrete number of bins. For example, if the expected value in a calibration is 2000, and the sampler chooses a value of 2100 with a very narrow confidence interval that does

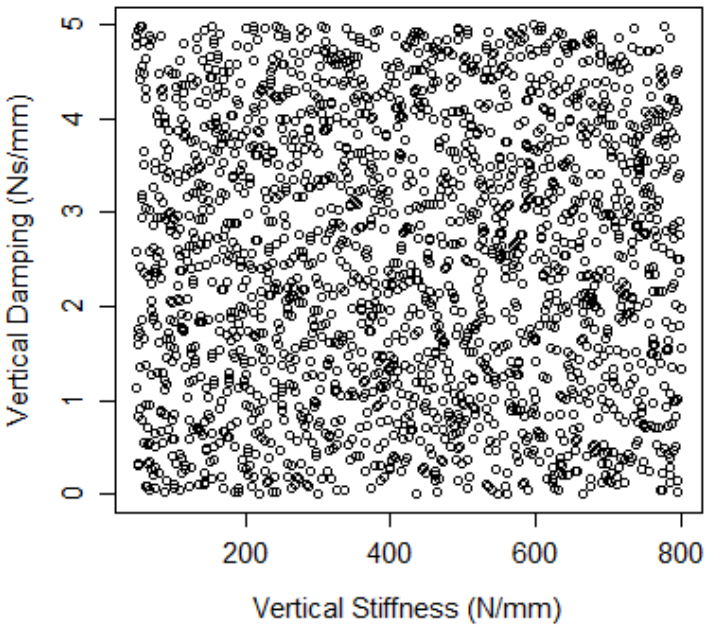

Figure 5.4: Distribution of parameters in design of experiments

not include 2000, this would still represent an accurate calibration if the preceding data point was 1800. Regardless, the true values for each parameter is not captured in this interval. This interval is extremely narrow due to the large number of data points (500) used in calibration, which misleadingly decreases the estimated standard deviation and causes the sampler to select a single data point.

The posterior distributions (figure 5.5) for both parameters are centered around the expected mean values. These distributions are very narrow, likely due to the large amount of data (500 points) used in calibration. As can be seen in the trace plots from figure 5.6, only one value for each parameter is sampled in the 50,000 MCMC iterations. As seen in figure 5.7, simulation results generated with the median estimate (shown in red) closely match the "experimental" data (shown in black). This posterior predictive check is also reproduced with the random noise included in calibration

in figure 5.8.

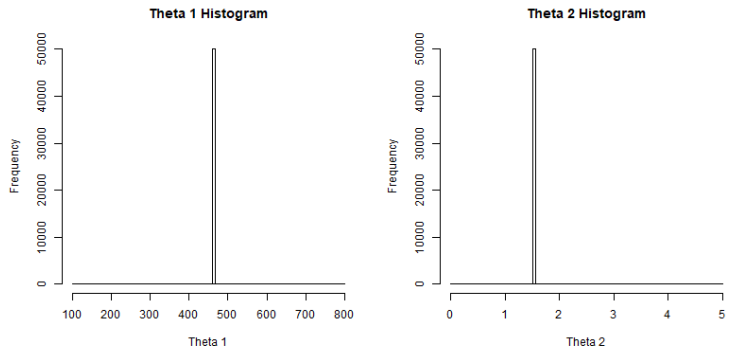

Figure 5.5: Posterior distributions for Fiala-to-Fiala calibration

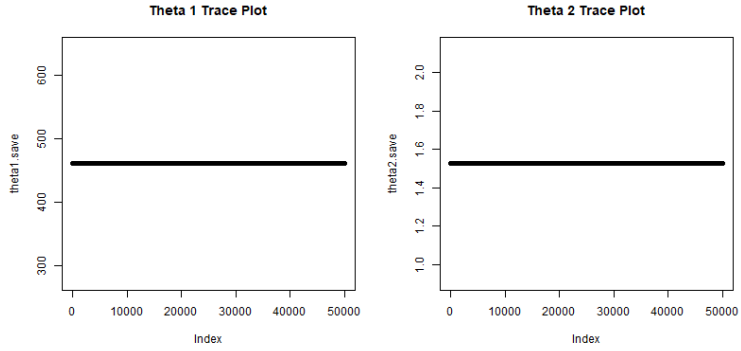

Figure 5.6: Trace plots for Fiala-to-Fiala calibration

5.4 Fiala to PAC2002 Calibration

The results for this calibration are summarized in table 5.3. Unlike in the Fiala-to-Fiala calibration, the true values for the parameter estimates are unknown, as the PAC2002 model includes many more parameters which influence the normal force generated in the tire.

The posterior distributions for each parameter are represented in figure 5.9. As in the Fiala-to-Fiala calibration, they are extremely narrow due to the large amounts

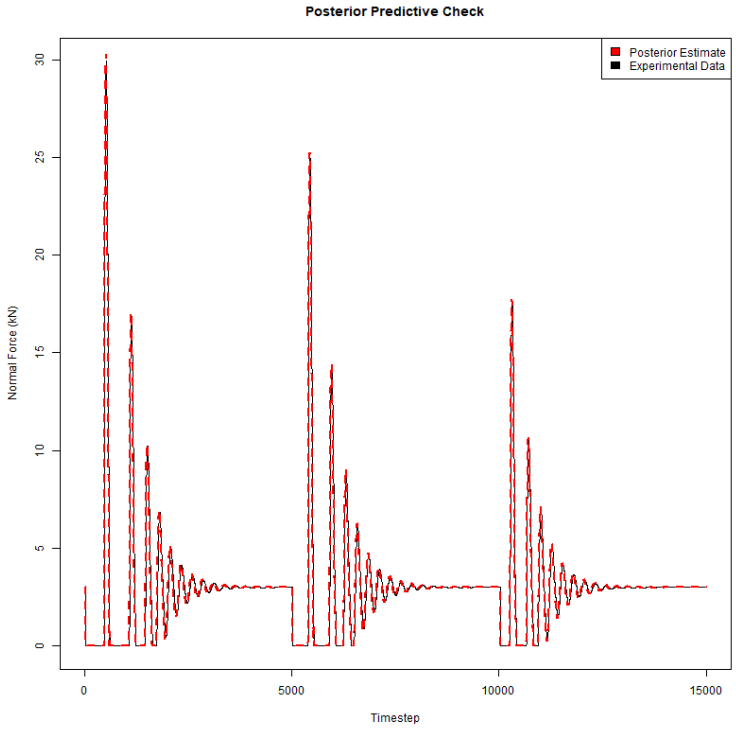

Figure 5.7: Posterior predictive check without noise for Fiala-to-Fiala calibration

	95% Lower bound	**θ estimate**	**95% Upper bound**
θ_1	313.2	313.2	313.2
θ_2	1.37	1.37	1.37

Table 5.3: Fiala to PAC2002 calibration results

of data points (500) used in calibration. Again, the sampler heavily favors one value for each parameter, but more are visited, likely due to the greater uncertainty involved in calibrating to a different model type (figure 5.10). The posterior predictive check shows a close match between the data which was calibrated to and the selected simulation observations (figure 5.11), though the peaks are generally underestimated. This is likely due to the inherent differences between the PAC2002 and Fiala models; no Fiala parameter combination will accurately represent those peaks. Again, this figure is reproduced including random noise which was included during the calibration (figure 5.12).

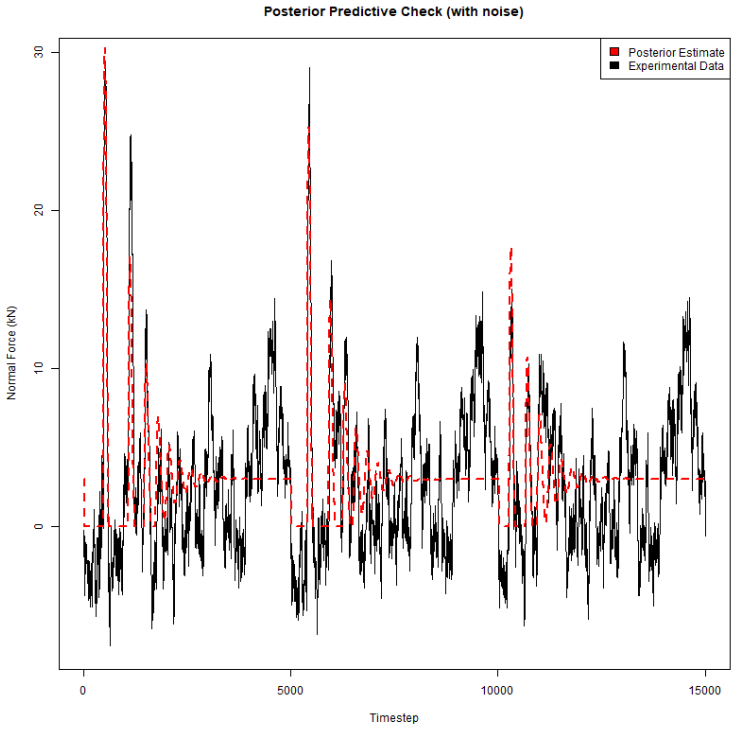

Figure 5.8: Posterior predictive check with noise for Fiala-to-Fiala calibration

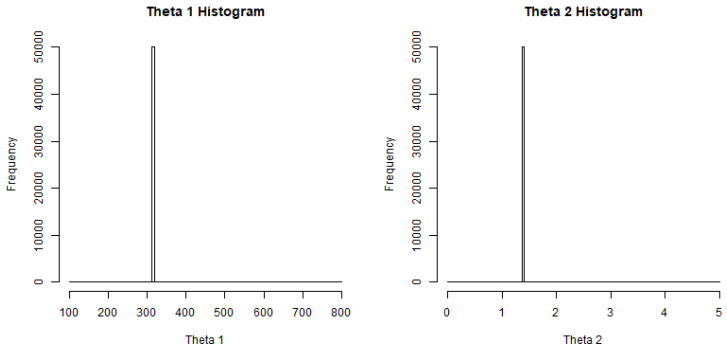

Figure 5.9: Posterior distributions for Fiala-to-Pac2002 calibration

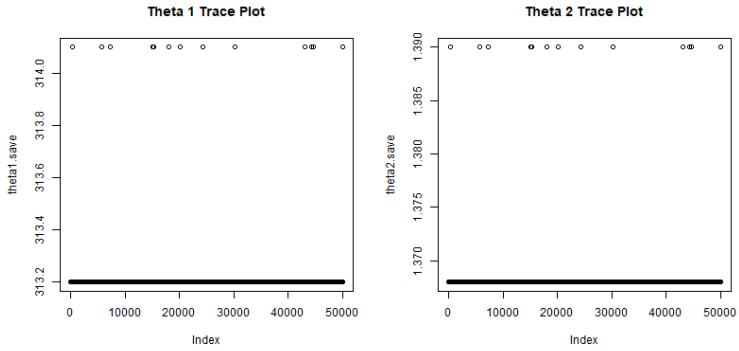

Figure 5.10: Trace plots for Fiala-to-Pac calibration

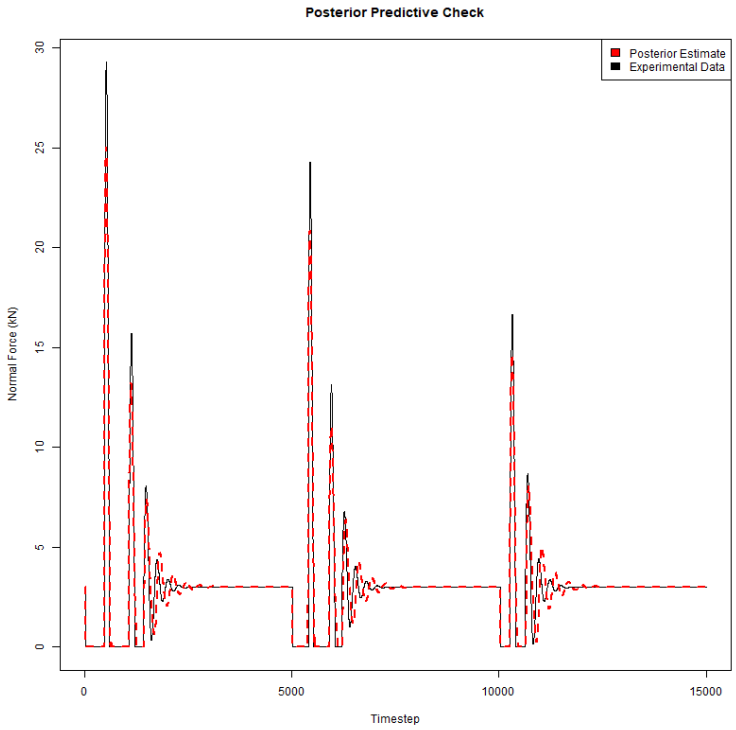

Figure 5.11: Posterior predictive check without noise for Fiala-to-Pac2002 calibration

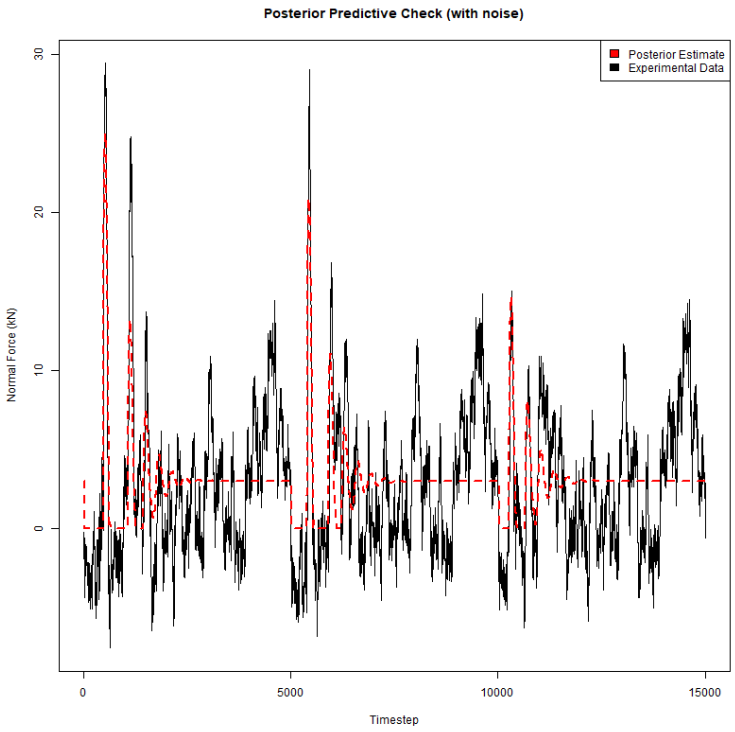

Figure 5.12: Posterior predictive check with noise for Fiala-to-Pac2002 calibration

Chapter 6

Conclusions

This thesis describes the novel application of the Griddy Gibbs sampler to a Bayesian model calibration process. This technique is directly compared to a traditional Bayesian calibration method, which used the Metropolis-Hastings algorithm, in the calibration of boundary conditions in a piston heat transfer analysis. The Griddy Gibbs method was found to generate nearly identical results at a high number of simulations. When the number of data points was large, the Griddy Gibbs method presented significant time savings compared to the traditional method. This is because it is a discrete method and samples only from a library of existing data; however, because of this fact, a robust library of existing data must exist to gain accurate calibration results with the Griddy Gibbs method.

The Griddy Gibbs calibration method was then used to calibrate time dependent data in a vehicle dynamics problem. This calibration utilized large amounts of data, which would be extremely time intensive using traditional sampling methods. In this application, the Griddy Gibbs method was able to calibrate tire parameters with very small error, but the large amount of data caused an underestimation of the variance of the data. This caused the posterior distributions to be extremely

narrow. More work is needed in this area to accurately describe the uncertainty in the Griddy Gibbs calibration, as well as to study the method's behavior subject to varying numbers of unknown parameters.

6.1 Future Work

In addition to this uncertainty problem, more work is needed in assessing the importance of including the correlation parameter in calibration. This is especially relevant because in its absence, the Griddy Gibbs calibration process can operate in real time. This presents the opportunity for a powerful tool which could aid in the navigation of autonomous vehicles. More work is planned for this application, specifically the development of a tool to select optimal vehicle subsystem models to use in varying environmental conditions.

Appendices

Appendix A R Code Steps

A.1 Inputs

The Griddy Gibbs code takes three input files: design, expobs, and simobs. The design file contains the values of the parameter settings for each simulation and should be organized with independent x settings first, followed by θ parameters. The "expobs" file holds the experimental data to calibrate to, and the "simobs" file holds the simulation observations used in calibration. The expobs file should include the x settings used to generate the data, and its x settings should match those in the design file.

These inputs should be arranged as follows, for n_x x settings, n_y responses, per simulation i:

$$Y(i) = \begin{bmatrix} Y_1(i) & Y_2(i) & \dots & Y_{n_x}(i) \end{bmatrix} \tag{1}$$

where

$$Y_1(i) = \begin{bmatrix} Y_1(x_1) & Y_2(x_1) & \dots & Y_{n_y}(x_1) \end{bmatrix} \tag{2}$$

$$Y_2(i) = \begin{bmatrix} Y_1(x_2) & Y_2(x_2) & \dots & Y_{n_y}(x_2) \end{bmatrix} \tag{3}$$

$$Y_{n_x}(i) = \begin{bmatrix} Y_1(x_{n_x}) & Y_2(x_{n_x}) & \dots & Y_{n_y}(x_{n_x}) \end{bmatrix} \tag{4}$$

This is not the standard format for input files generated using the Los Alamos

"simobs" Matlab code, which considers each x setting a separate simulation, hence the "rearrange truth & simobs" section in lines 36-64.

A.2 Precalculation

In this section, the correlation matrices and portions of the logarithmic likelihood function are precalculated. The correlation matrices are calculated in lines 68-83 according to the following equation:

$$\sum_{i=1}^{n_x} \rho_{x_n}^{\left|x_{n_i} - x_{n_j}\right|} \tag{5}$$

For the term in the i^{th} row and j^{th} column of R describing the correlation between n independent variables x. If it is not desired to include the correlation parameter in calibration, it should be set to zero.

The variable $C_\theta = (Y(\theta, x, z) - g(\theta))^T R^{-1} (Y(\theta, x, z) - g(\theta))$ is calculated for each simulation and rho setting specified. This is taken from the equation for the logarithmic likelihood

$$ln(L(\theta)) = \frac{n}{2} ln\left(\frac{\tau}{2\pi}\right) - \frac{1}{2} ln(|R|) - \frac{\tau}{2} C_\theta \tag{6}$$

In lines 114-118, each combination of the correlation parameter ρ per x value is assembled. If more or fewer than three x variables are used, these lines should be edited.

A.3 Gibbs Sampling

The Griddy Gibbs sampling algorithm includes the following steps, for each i iteration:

1. Sample $\theta^{(i)}$

   (a) Calculate the full conditional distribution for each simulation: $p(\theta|\tau^{(i)}, \rho^{(i)}) = L(\theta)\pi(\theta)$. On the first iteration $i = 1$, assume arbitrary values for $\tau^{(1)}$ and $\rho^{(1)}$. On each successive iteration i, use the previous $i - 1$ simulation's value.

   (b) Normalize the full conditional distribution: $p_N(\theta|else)^{(i)} = \frac{p^{(i)}(\theta|\tau^{(i)},\rho^{(i)})}{\sum_{i=1}^{n} \pi^{(i)}(\theta|\tau^{(i)},\rho^{(i)})}$. Because this function is proportional to the joint posterior distribution, this normalization process in effect removes that proportionality constant.

   (c) Take the cumulative logarithmic likelihood: $cp^{(i)}(\theta|\tau^{(i)}, \rho^{(i)})$. Draw a random number $k^{(i)}$ between 0 and 1. The i^{th} estimate for θ, $\theta^{(i)}$, will be the combination of parameters used in the simulation run for which $p^{(i)}(\theta|\tau^{(i)}, \rho^{(i)}) > k^{(i)}$.

2. Sample $\tau^{(i)}$

   (a) Because the sampling distribution for τ follows the multivariate normal distribution and the prior distribution follows the gamma distribution, τ can be sampled directly using a value for C_θ calculated using the i^{th} estimate for θ, $\theta^{(i)}$. The parameters α and β should be assigned values 1 and 0, respectively, to be uninformative.

$$\pi(\tau|\theta) \sim Gamma(\alpha + n/2, \beta + C_\theta^{(i)}/2) \tag{7}$$

3. Sample $\rho^{(i)}$

   (a) Calculate the full conditional distribution, $p(\rho|\theta^{(i)}, \tau^{(i)}) = L(\rho)\pi(\rho)$, using the i^{th} calculated value for $\tau^{(i)}$, $\theta^{(i)}$.

(b) Normalize the full conditional distributions:

$p_N(\rho|\theta^{(i)}\tau^{(i)})^{(i)} = \frac{p^{(i)}(\rho|\theta^{(i)},\tau^{(i)})}{\sum_{i=1}^{n} p^{(i)}(\rho|\theta^{(i)},\tau^{(i)})}$.

(c) Take the cumulative log likelihood, $cp^{(i)}(\rho|\theta^{(i)}\tau^{(i)})$. Draw a random number $k^{(i)}$ between 0 and 1. The i^{th} estimate for ρ, $\rho^{(i)}$, will be the combination of ρ for which $p^{(i)}(\rho|\theta^{(i)}\tau^{(i)}) > k^{(i)}$.

The mean of the i estimates for θ, τ, ρ_x, and ρ_t are considered the sampler's estimate for their true values.

Appendix B Tire Model Calibration

The application of the Griddy Gibbs code to the vehicle dynamics calibration generally follows the same process as listed in Appendix A, except a different method for calculation the correlation matrices is used.

In this application, we observe $Y(t, x)$ at a discrete number of x and t values. For one case, we observe the normal force at $n_t = 500$ and $n_x = 3$. Because we observe each of the time points at each x setting, we can take the general correlation matrix R to be the Kronecker product of the two separate correlation matrices, R_x and R_t.

$$R = R_x \otimes R_t \tag{8}$$

The Kronecker product multiplies the entire R_t matrix by each value in R_x.

$$R_x \otimes R_t = \begin{bmatrix} R_{x_{11}} \times R_t & R_{x_{12}} \times R_t & \cdots & R_{x_{1n_x}} \times R_t \\ R_{x_{21}} \times R_t & R_{x_{22}} \times R_t & \cdots & R_{x_{2n_x}} \times R_t \\ \vdots & \vdots & \ddots & \vdots \\ R_{x_{n_x1}} \times R_t & R_{x_{n_x2}} \times R_t & \cdots & R_{x_{n_xn_x}} \times R_t \end{bmatrix} \tag{9}$$

The determinant of the Kronecker product of two matrices is given by

$$det(R_x \otimes R_t) = det(R_x)^{n_t} \times det(R_t)^{n_x} \tag{10}$$

For large values of ρ, $det(R)$ tends towards 0, and for large values of n_y, $det(R_t)^{n_x}$ tends towards infinity. For numerical stabilization, we can use the natural log of the determinant, and use the fact that R_t is composed of an evenly spaced grid of evaluation points with step size ϵ, to alternatively calculate the determinant of R:

$$det(R_t) = (1 - (\rho^{\epsilon})^2)^{(ny-1)} \tag{11}$$

$$ln(det(R_t)) = n_y \times ln(det(R_x)) + n_x \times (n_y - 1) \times ln(1 - (\rho^{\epsilon})^2) \tag{12}$$

In the preceding equation, the determinant of R_x can be directly calculated as its exponent is sufficiently small that it does not tend towards infinity.

The inverse of the Kronecker product of two matrices is given by

$$(R_x \otimes R_t)^{-1} = R_x^{-1} \otimes R_t^{-1} \tag{13}$$

Bibliography

[1] D. M. Allen. Mean square error of prediction as a criterion for selecting variables. *Technometrics*, 13(3):469–475, 1971.

[2] S. P. Brooks. Markov chain monte carlo method and its application. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):69–100, 2021.

[3] T. Chai and R. R. Draxler. Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding rmse in the literature. *Geoscientific Model Development Discussions*, 7(1):1525–1534, 2014.

[4] S. Chib and E. Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[5] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London.*, 222:309–368, 1922.

[6] R. A. Fisher. A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Monthly Notices of the Royal Astronomical Society*, 80:758–770, 1930.

[7] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[8] M. Gu and L. Wang. Scaled gaussian stochastic process for computer model calibration and prediction. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1555–1583, 2018.

[9] Hexagon. Adams tire user's guide. `https://help.hexagonmi.com/bundle/Adams_2020_Adams_Tire_User_Guide/resource/Adams_2020_Adams_Tire_User_Guide.pdf`, 2020.

[10] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne. Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2), 2004.

[11] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society*, 63(3):425–464, 2001.

[12] Z. Li and M. Hwai Yong Tan. A gaussian process emulator based approach for bayesian calibration of a functional input. *Technometrics*, 64(3):769–784, 2022.

[13] A. Lye, A. Cicirello, and E. Patelli. Sampling methods for solving bayesian model updating problems: A tutorial. *Mechanical Systems and Signal Processing*, 159, 2021.

[14] Mathworks. Coordinate systems in vehicle dynamics blockset. `https://es.mathworks.com/help/vdynblks/ug/coordinate-systems-in-vehicle-dynamics-blockset.html`, 2023.

[15] J. Oakley and A. O'Hagan. Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4):769–784, 2002.

[16] J. Orloff and J. Bloom. Maximum likelihood estimates. , 2014.

[17] K. Pearson. Method of moments and method of maximum likelihood. *Biometrika*, 28(1/2):34–59, 1936.

[18] C. Ritter and M. A. Tanner. Facilitating the gibbs sampler: the gibbs stopper and the griddy-gibbs sampler. *Journal of the American Statistical Association*, 87(419):861–868, 1992.

[19] C. J. Willmott and K. Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.

[20] R. K. W. Wong, C. B. Storlie, and T. C. M. Lee. A frequentist approach to computer model calibration. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(2):635–648, 2017.

[21] J. M. Wooldridge. Applications of generalized method of moments estimation. *Journal of Economic Perspectives*, 15(4):87–100, 2001.

[22] S. Wright, A. Ravikumar, L. Redmond, C. McMahan, B. Lawler, M. Castanier, E. Gingrich, and M. Tess. Benchmarking of machine learning methodologies for piston thermal model calibration. In *SAE WCX*. SAE International, 2024.