2009-07-02

Modeling Temperature Reduction in Tendons Using Gaussian Processes Within a Dynamic Linear Model

Richard David Wyss
Brigham Young University - Provo

MODELING TEMPERATURE REDUCTION IN TENDONS USING GAUSSIAN PROCESSES WITHIN A DYNAMIC LINEAR MODEL

by

Richie Wyss

A project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Statistics

Brigham Young University

August 2009

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a project submitted by

Richie Wyss

This project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Gilbert W. Fellingham, Chair
Date	C. Shane Reese
Date	Bruce J. Collings

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the project of Richie Wyss in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Gilbert W. Fellingham
Chair, Graduate Committee

Accepted for the Department

Scott D. Grimshaw
Graduate Coordinator

Accepted for the College

Thomas W. Sederberg
Associate Dean, College of Physical and Mathematical Sciences

ABSTRACT

MODELING TEMPERATURE REDUCTION IN TENDONS USING GAUSSIAN PROCESSES WITHIN A DYNAMIC LINEAR MODEL

Richie Wyss

Department of Statistics

Master of Science

The time it takes an athlete to recover from an injury can be highly influenced by training procedures as well as the medical care and physical therapy received. When an injury occurs to the muscles or tendons of an athlete, it is desirable to cool the muscles and tendons within the body to reduce inflammation, thereby reducing the recovery time. Consequently, finding a method of treatment that is effective in reducing tendon temperatures is beneficial to increasing the speed at which the athlete is able to recover. In this project, Bayesian inference with Gaussian processes will be used to model the effect that different treatments have in reducing tendon temperature within the ankle. Gaussian processes provide a powerful methodology for modeling data that exhibit complex characteristics such as nonlinear behavior while retaining mathematical simplicity.

ACKNOWLEDGEMENTS

I am very thankful to my family for their support throughout my education. I also wish to extend my deep appreciation to my advisor Dr. Fellingham for his patience and mentorship, as well as my graduate committee members Dr. Reese and Dr. Collings.

CONTENTS

FIGURES

Figure

1. INTRODUCTION

In order to increase the likelihood of success in athletic competition, it is important to be able to ascertain the most effective ways to treat sports injuries and to train athletes to help them perform at their optimum level. When an athlete is injured in an athletic event, the muscles and joints in the body tend to swell, which slows the recovery time and increases the chance for injury in future athletic events. To reduce the swelling and speed the recovery process, various treatments are typically administered to the athletes in an effort to cool the temperature in the muscles and joints. Studies have shown that the more quickly the temperature in the muscles and tendons can be reduced, the more quickly the athlete can recover, thus lowering the chances for injury in future athletic events (Safran, McKeag, and Camp 2002). Therefore, it is essential in athletic competition to find the most effective methods of reducing muscle and tendon temperature and to utilize those methods in both the training and treatment of the athletes.

In this study, two sets of data will be analyzed to determine whether ice-cooled whirlpool baths or ice packs are more effective for cooling the temperature of an athlete's tendons after exercise. In the first data set, the temperatures of the tendons of 15 noninjured subjects were measured in degrees Celsius. The temperatures were measured every 30 seconds at room temperature for 5 minutes, every 30 seconds in an ice-filled whirlpool bath for 20 minutes, and again at room temperature every 30 seconds for 23 minutes. The second data set consisted of the same sequence of measurements on the same 15 subjects. The only difference was the treatment applied. In the second data set, measurements on the temperatures were taken every 30 seconds at room temperature for 5 minutes, every 30 seconds with an ice pack applied to the tendon for 20 minutes, and again at room temperature with the ice

pack removed every 30 seconds for 23 minutes. A plot of the data for the 15 subjects in each set of data can be found in Figure 1.1. Figure 1.1 reveals clear breaks in both data sets when the treatments are applied and removed.

The purpose of this study is to demonstrate the use of Gaussian processes within a dynamic linear model setting in describing the behavior of the temperatures in the tendons as they cool due to treatment. The models will determine if one method of treatment is significantly more effective than the other in cooling the temperatures and keeping the temperatures down once the treatment period is over. More specifically, a model will be fit to each of the first two sets of data described above using a Bayesian implementation of Gaussian processes. This methodology models the average behavior of the temperature in the tendons for each set of data as a function of time and treatment, or environment. To further assess the differences between the two treatments, the posterior distributions for the difference in the mean temperature at each time period will be computed. These posterior distributions for the differences in the mean tendon temperatures will be constructed two separate ways and the results will be compared to determine the most effective treatment.

As mentioned previously, the main purpose of this project is the incorporation of Gaussian processes into the methodology used for the analysis. A dynamic model-building approach will then be taken. Dynamic linear models allow us to continually update the study with information as it becomes available, thereby increasing the accuracy of the predictions and models.

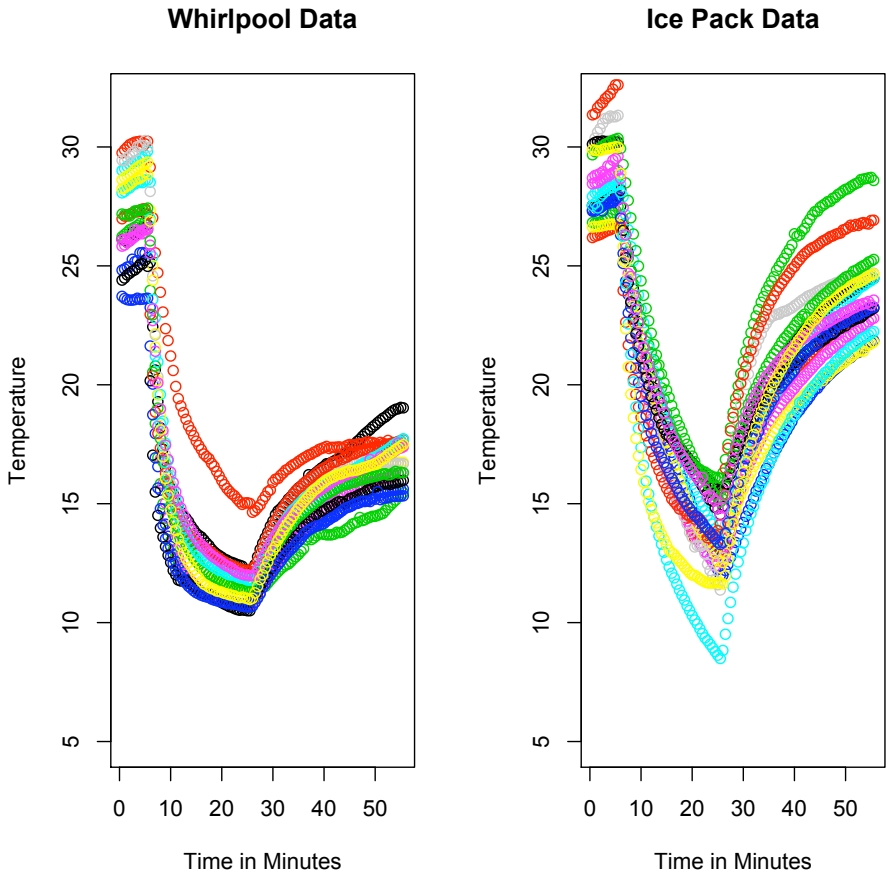

Figure 1.1: Tendon Temperatures of 15 Subjects. The colors represent different individuals.

2. LITERATURE REVIEW

Bayesian inference is a powerful tool for performing statistical modeling. The primary focus of this project is the use of a technique known as Gaussian processes to model the complex behavior of data. Before we proceed to a detailed description of the application of these methods, we first describe the concepts and previous use of the methods.

2.1 Bayesian Methods

In 1764 an essay was published by Reverend Thomas Bayes (Bayes 1764). He proposed a theorem that has had a significant effect on the methods used in statistics. Bayes proposed a theorem (later known as Bayes Theorem) that became the foundation for an entire branch of statistics known as Bayesian methods.

The fundamental idea behind Bayesian inference, which differentiates it from frequentist methodology, is to consider parameters as random quantities. Ashby (2006) asserts that doing so requires specification of distributions, known as the prior distributions, for each of the parameters of interest. Ashby further states that these prior distributions describe the behavior of the parameters and not the data. Using the prior distributions along with the likelihood function of the data, a posterior distribution for the parameter or parameters is calculated using Bayes rule (Gelman, Carlin, Stern, and Rubin 2004). The posterior distributions are then used to make inference about the parameters of interest.

Formally, given a model or hypothesis, H, the associated set of parameters, θ, is used to make inference about the behavior of a given set of random variables or data set $\mathbf{X}$. Inference about the model parameters is made by first specifying prior knowledge or beliefs about the parameter vector. This specification is made in

terms of prior probability distributions on the parameters $P(\theta)$. If $P(X|\theta)$ represents the likelihood of the data given the parameters and $P(\mathbf{X})$ represents the marginal distribution of the data, then the posterior distribution from which inference is made is produced using Bayes rule as follows:

$$P(\theta|\mathbf{X}) = \frac{\mathbf{P}(\mathbf{X}|\theta)\mathbf{P}(\theta)}{\mathbf{P}(\mathbf{X})}. \tag{2.1}$$

In practice, the commonly used notation for the above equation is $\pi(\theta|\mathbf{y})$ to represent the posterior distribution, $\pi(\theta)$ to represent the prior distribution on the parameters, and $f(\mathbf{y}|\theta)$ to represent the likelihood of the data. The marginal distribution of the data, shown as $P(\mathbf{X})$ in the above equation, is equal to $\int P(X|\theta)\mathbf{P}(\theta)\mathbf{d}\theta$, which is more commonly seen as $\int \pi(\theta)\mathbf{f}(\mathbf{y}|\theta)\mathbf{d}\theta$.

When the prior distribution $P(\theta)$ on a parameter θ is conjugate, the derivation of the posterior distribution is greatly simplified and is numerically tractable. Formally, if F represents a class of sampling distributions $p(y|\theta)$, then a class of prior distributions, P, for θ is defined as conjugate for F if $p(\theta|y) \in P$ for all $p(.|\theta) \in F$ and $p(.) \in P$ (Gelman et al. 2004). The calculated posterior distribution will always have the same distributional form as the prior distribution (Gelman et al. 2004). Conjugate priors imply the posterior distribution is a known closed form. The benefit of this is that samples can be taken directly from the posterior, thereby simplifying the computational procedure of obtaining draws from the posterior.

2.2 Markov Chain Monte Carlo

The derivation of the posterior distribution through the use of Bayes theorem is not always possible through analytical methods. Consequently, numerical methods are often employed, and Bayesian analysis usually requires the implementation of a technique known as Markov chain Monte Carlo. Monte Carlo and Markov chains are two separate ideas that are used in conjunction with one another: Monte Carlo

methods and Markov chains.

Physicists working at the Los Alamos Laboratory showed that complex mathematical problems can be solved through statistical sampling (Hammersley and Handscomb 1964). Stanislaw Ulam, a physicist at Los Alamos laboratory who is considered the primary inventor of Monte Carlo methods, explained that the central idea behind Monte Carlo methods is using draws of random numbers or random sampling to simulate mathematical systems (Metropolis and Ulam 1949). Monte Carlo methods are useful when modeling systems that are too complex to use direct deterministic algorithms or procedures.

Andrey Markov developed a theory in stochastic processes which came to be known as Markov chains. Markov chains have the property that the future state of a system or variable is independent of past states and is dependent solely on the present state (Marcus and Rosen 2006). Formally, if $X_1, X_2, X_3, ...$ represents a sequence of random variables, then $P(X_{n+1} = x|X_n = x_n, ..., X_1 = x_1) = P(X_{n+1} = x|X_n = x_n)$. $P(X_{n+1} = x|X_n = x_n, ..., X_1 = x_1)$ represents the probability that the stochastic process is in state x at time $n + 1$ given the present state along with all previous states at each time period. $P(X_{n+1} = x|X_n = x_n)$ represents the probability that the stochastic process is in state x at time $n + 1$ given the state of the process at the current time period (Ross 2007).

Combining these concepts, Nicholas Metropolis developed an algorithm later generalized by Hastings which uses Markov chain Monte Carlo simulation to obtain draws of random numbers from posterior distributions (Hastings 1970). A special case of Metropolis-Hastings is Gibbs sampling, which was developed by Geman and Geman (1984). Gibbs sampling is less general than Metropolis-Hastings, but still provides a way to obtain samples from marginal densities using conditional densities. If the complete conditional distribution of each parameter is available in a closed form, Gibbs sampling can be utilized in an iterative process to sample from each of

the conditional posterior distributions and successively substitute the drawn value into the other complete conditionals (Prabhu and Basawa 1990). When the complete conditionals are not available in closed form, a more general iterative sampling procedure known as Metropolis-Hastings can be used (Prabhu and Basawa 1990). Because of the relative ease with which samples can be obtained from known distributional forms, when possible, priors are chosen in such a way as to result in closed-form solutions for the complete conditionals. As a result, Gibbs sampling is usually considered the primary computational method when performing Bayesian inference.

The above results provide a way to iteratively sample from the complete conditional density of each parameter for which inference is made. The distribution of the resulting draws converge to the marginal posterior density for each parameter (Gelman et al. 2004). The resulting posterior densities can then be used to make inference about the parameters of interest. These results are important in that they allow for inference about the posterior distributions to be made when the derivation for the posterior distributions cannot be obtained analytically. The universal applicability of Metropolis-Hastings and Gibbs sampling is what gives modern Bayesian methods its power when conducting statistical inference.

2.3 Gaussian Processes

Formally, a Gaussian process is a system of random variables $X = X_\lambda : \lambda \in \Lambda$ such that any finite linear combination $\sum a_k X_\lambda$ is a Gaussian random variable, with X being a stochastic process (Rasmussen and Williams 2006). Hida and Hitsuda (1976) explain that Gaussian Processes (GPs) are generalizations of multivariate Gaussian random variables extending to infinite dimensionality.

Rasmussen and Williams (2006) further explain that a function can be thought of as an infinitely long vector. Where a Gaussian distribution is fully specified by a mean vector μ and covariance matrix $\mathbf{\Sigma}$, a Gaussian process is completely specified

by a mean function m($\mathbf{x}$)=E[f($\mathbf{x}$]) and covariance function $\mathbf{k}(\mathbf{x}, \mathbf{x}')$ and is written as $f(x) \sim GP(m(x), k(x, x'))$. Using the notation above, the random variables that are used to define a Gaussian process are the values of the function $f(x)$. Usually, a Gaussian process is defined over time, which is represented or indexed by x.

Thinking of a GP as a Gaussian distribution with an infinitely long vector and an infinite by infinite covariance matrix may seem impractical when performing statistical inference. However, the marginalizing property allows inference to be performed on any subset of the Gaussian variables with the variables still retaining their Gaussian properties. The marginalizing property simply states that if $p(\mathbf{x}, \mathbf{y}) = N\left(\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B^T} & \mathbf{C} \end{pmatrix}\right)$, then $p(\mathbf{x}) = N(\mathbf{a}, \mathbf{A})$.

The marginalizing property makes it possible for inference to be made for a large set of parameters while maintaining mathematical simplicity. This ability to perform inference on large numbers of parameters is one of the primary advantages of incorporating a Gaussian process approach when constructing a statistical model (Mackay 1999). This allows for estimation to be made about complex data behavior such as nonlinear regression without having to define a functional form of the data and perform inference on the functional parameters (Mackay 1999). Ebden (2008) further explains this idea by stating that when modeling any arbitrary data set $y = y_1, ..., y_n$, the n observations can always be thought of as a sample from some n dimensional multivariate Gaussian distribution and the data can be modeled using a Gaussian process. Thus Gaussian processes are as universal as they are simple.

Ebden (2008) illustrates this concept in a typical prediction problem using Gaussian process regression (GPR). Given a set of random variables $\mathbf{Y}$, Ebden first explains that using traditional methods, the behavior of $\mathbf{Y}$ can be described by an underlying function $f(\mathbf{x})$ through the relation $\mathbf{Y} = f(\mathbf{x}) + N(\mathbf{0}, \mathbf{\Sigma})$. Ebden explains that statistical methods can be used to approximate $E(\mathbf{Y}|\mathbf{x^*})$ by estimating $f(\mathbf{x})$ from the given set $\mathbf{Y}$. However, he further explains that Gaussian process regression

is a finer approach than the traditional methods which specify models. A Gaussian process regression approach allows for $f(\mathbf{x})$ to be represented obliquely, thereby letting the data "speak" more freely. With Gaussian processes, given the set $\mathbf{Y}$, the objective is to predict y_*, not the actual f_*. Theoretically their expected values are identical, but the variances of the two methods differ owing to the observational noise in the data, as well as the covariance function specification. Utilizing the assumption that $\mathbf{Y}$ is a sample from a multivariate Gaussian distribution, the prediction or estimation of y_* is relatively simple according to Ebden. With $P(y_*|y)$ following a Gaussian distribution, the best estimate for y_* is simply the mean of this distribution, where the distribution is obtained after the specification of the mean and covariance functions of the Gaussian process (Ebden 2008).

Schervish (2004) further demonstrates the use of Gaussian processes in terms of a simple two-dimensional regression problem. He explains that the regression problem consists of a set of data $\mathbf{y}$ where each observation is assumed to satisfy $y = f + e$ where $e \sim N(0, \sigma^2)$ and f is an underlying function. Placing a zero mean Gaussian process prior on f, $f \sim N(\mathbf{0}, \mathbf{K})$, the Gaussian marginal likelihood, $p(\mathbf{y}|\theta) = N(\mathbf{0}, \mathbf{K} = \mathbf{I}\sigma^2)$, is obtained where $\mathbf{K}$ is the covariance matrix with parameter vector θ. Placing priors on the hyperparameters, where hyperparameters are parameters on prior distributions, and using Markov chain Monte Carlo methods, Schervish obtains estimates $\hat{\theta}$ which are used to make predictions or estimations about the underlying behavior of the data.

Gaussian processes also allow for parameters within the model to be related to the other parameters in some way, either through the specification of the covariance matrix or in the prior distribution specifications (Mackay 1999). This relational idea between the parameters, or the idea of borrowing strength from previous inference or data, is a major advantage of the Gaussian process approach. The concept of borrowing strength allows for an adaptive approach when estimating parameters,

creating a dynamic model that can continually be updated as information becomes available.

Gaussian processes are usually used within a Bayesian framework. Within the Bayesian framework, the Gaussian model is based on a prior distribution specified over a possibly infinite dimensional space of functions (Ashby 2006). When performing inference on the mean function of a given set of data, if the likelihood function as well as the prior specifications follow a Gaussian process, the resulting posterior distribution will also be a Gaussian process (Ashby 2006).

These concepts are illustrated by Gray, Murray-Smith, and Thompson (2003) in modeling twin tank systems. In this report, Gaussian process priors are used in the context of this dynamic modeling. The authors demonstrate how the Gaussian process modeling predicts the behavior of a dynamic system by predicting the distribution of the next data point based on the system input and the predicted distribution of the current point (Gray et al. 2003).

Berry and Ruppert (2002) also illustrate the ability of Gaussian processes to borrow strength in the context of fitting a smoothing spline in a nonparametric regression setting. They define their cubic estimator g of the true spline m as the minimizer of the sum of squared errors $\sum[m(X_i) - g(X - i)]^2$. Their g estimator at each knot is a function of the estimator at all other knots, thereby borrowing strength or information from the other knots.

Gaussian processes provide a very powerful, yet tractable, way to model complex nonlinear behavior without having to specify complex functional forms. In this project, we assume that the temperatures at different time periods follow a Gaussian distribution. By setting Gaussian priors on the temperature means, a Gaussian process approach can be employed in performing the analysis on the data sets containing the temperature values.

3. METHODS AND APPLICATION

3.1 Distribution of Data and Likelihood

The two data sets that will be analyzed in this project consist of the Achilles tendon temperatures of individual subjects measured at different time periods. There are 15 subjects, each being measured at 111 different time periods for a total of 1665 observations. It is reasonable to assume that the expected value of the temperatures at each time period will differ since the temperatures change as time elapses and as the treatments are applied. Further, since physical measurements on individuals are usually normally distributed, it is reasonable to assume that the temperature measurements on these observations are normally distributed as well. It is assumed that the variance at each time measurement is the same. Although other assumptions could be used allowing for heterogeneity among the variance components, for the purposes of this project, constant variance across time measurements is fairly reasonable and, for simplicity, will be implemented. Consequently, it is assumed that the observations $y_{ij} \sim Normal(\theta_i, \sigma^2)$, where i represents the i^{th} time period and j represents the j^{th} individual measured at each time period.

With the data distributed as indicated above and assuming independence between subjects and conditional independence within the measurements of each subject, the likelihood function is as follows:

$$f(\mathbf{y_{ij}}|\theta_i, \sigma^2) = (2\pi\sigma^2)^{\frac{-N}{2}} exp\left\{\frac{-\sum_{i=1}^{t}\sum_{j=1}^{n}(y_{ij}-\theta_i)^2}{2\sigma^2}\right\},$$

where:

- n = the number or subjects measured at each time period (15 subjects),
- t = the number of time periods (111 in this case), and

- N = the total number of responses (1665 observations).

3.2 Prior Distribution for $\mathbf{\Theta}$ and θ_i: A Gaussian Process Approach

A Gaussian process distribution will be used for the prior distribution on $\mathbf{\Theta}$, which is the vector of the θ_is. Recall that a Gaussian process is simply a stochastic process in which the collection of random variables have a joint Gaussian distribution. Also recall that a Gaussian process is fully specified by its mean function and covariance function and is written as follows:

$$\mathbf{\Theta} \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x'}). \tag{3.1}$$

According to probability theory, the joint distribution of a set of random variables or events can be written as the product of the conditional distributions. More specifically, the joint distribution of $\mathbf{\Theta}$ can be written as follows:

$$P(\mathbf{\Theta}) = P(\theta_{111}|\theta_{110}...\theta_1)P(\theta_{110}|\theta_{109}...\theta_1)...P(\theta_1). \tag{3.2}$$

If we apply the marginalizing property of Gaussian processes or multivariate normal distributions, it follows that the marginal distribution of each θ_i, conditioned on the previous θ_is, also follows a normal or Gaussian distribution. Because the time lengths between measurements are equal, it is reasonable to assume that the variance is constant across time. Further, it is also reasonable to assume that the mean of the conditional distribution of θ_i will be closely related to the previous θ or θ_{i-1}. Consequently, it is reasonable to assume that the conditional distribution of θ_i will be normally distributed centered around θ_{i-1} with variance τ^2. Formally, the conditional distribution of θ_i is written as

$$(\theta_i|\theta_{i-1}\theta_{i-2}...\theta_1) \sim N(\theta_{i-1}, \tau^2). \tag{3.3}$$

If we describe the behavior of the conditional distribution of θ_i as $N(\theta_{i-1}, \tau^2)$, the conditional distribution of the current θ_i is dependent on the previous θ, or θ_{i-1}.

This prior specification applies to each of the θ_is except for the first θ or θ_1. The prior set on θ_1 is called the initial information, and is given specific values. By setting up the model such that the current information is affected or updated by previous information (θ_i is influenced by the previous θ or θ_{i-1}), the model is considered a dynamic model where initial information or a starting point is necessary so that subsequent θs can be updated. In choosing values for the distribution of θ_1, it was assumed that since the measurements were taken at room temperature (degrees Celsius) before the treatments were applied, the temperatures within the tendons should be between room temperature and core temperature with mild variation. Consequently, it was assumed that $\theta_1 \sim N(25, 2)$.

With the distribution of $\theta_i|\theta_{i-1}\theta_{i-2}...\theta_1$ solely dependent on θ_{i-1}, we can apply the Markovian property, which states that the conditional distribution of the future state or states of a stochastic process is independent of past states and solely dependent on the current state. In other words, applying the Markovian property, it is reasonable to assume that the distribution of $\theta_i|\theta_{i-1}\theta_{i-2}...\theta_1$ is equal to the distribution of $\theta_i|\theta_{i-1}$, and therefore $P(\theta_i|\theta_{i-1}\theta_{i-2}...\theta_1) = P(\theta_i|\theta_{i-1}) = N(\theta_{i-1}, \tau^2)$ and the joint distribution of $\mathbf{\Theta}$ can be written as follows:

$$P(\mathbf{\Theta}) = \prod_{i=2}^{111} P(\theta_i|\theta_{i-1})P(\theta_1) = \prod_{i=2}^{111} N(\theta_{i-1}, \tau^2)N(25, 2). \tag{3.4}$$

One of the most advantageous aspects of using a Gaussian process approach is incorporated by describing the behavior of $\theta_i|\theta_{i-1}$ as normal or Gaussian and basing its mean value solely on the value of the previous θ or θ_{i-1}. Setting up $\theta_i|\theta_{i-1}$ in this fashion allows for strength, or information, to be borrowed from the information that has already been gathered. As mentioned previously, this idea of borrowing strength or information creates a dynamic model by updating the information as it becomes available, thereby increasing the accuracy of the model.

3.3 Prior Distributions for σ^2 and τ^2

With σ^2 representing a variance component, the support set is nonnegative. Therefore, to preserve the parameter space and allow a closed-form solution to the complete conditionals, an Inverse Gamma prior distribution was used. In choosing the values for a_σ and b_σ, it was determined from discussions with professors and medical doctors that the temperature in the Achilles tendon should not have a range of more than three or four degrees at each of the time periods. Therefore, to take the conservative approach from these suggestions, values were chosen for a range of approximately four degrees. The conservative approach was chosen to provide a slightly more diffuse prior distribution, thereby allowing the data to carry more weight in the analysis. It was determined that the average variance of the data should be approximately one degree. With this reasoning, a prior distribution is placed on σ^2 so that it will be centered around one. Values were also chosen so that three standard deviations from the mean would be approximately two degrees, thereby allowing most of the data to be contained within a range of four degrees. Using this reasoning, the values of 5 and 0.2 were chosen for a_σ and b_σ and the distribution for σ^2 can be written as follows:

$$\sigma^2 \sim InverseGamma(5, 0.2).$$

By introducing a prior distribution on the variance component of θ_i, a hierarchical component to the model is introduced. Again, because τ^2 represents a variance term, the support set is nonnegative. Similar to the reasoning used in choosing the prior distribution for σ^2, to preserve the parameter space and to allow the complete conditionals to be solved in closed form, an Inverse Gamma distribution was chosen. In choosing the prior values for a_τ and b_τ, since τ^2 represents the variance of the means of the observations, it is reasonable to assume that the values for τ^2 will be slightly smaller and tend to vary less than the values for σ^2. Using these assumptions,

along with the opinions of experts in the field, it was determined that the variance of the means should be minimal and that the range of the means should not be much larger than two. It was also determined that the mean variance of τ^2 would be slightly less than the mean variance of σ^2 but would still be close to one degree. Using this reasoning, the following values were chosen for the prior distribution of τ^2: $a_\tau = 10$ and $b_\tau = 0.1$, and the distribution for τ^2 is written as follows:

$$\tau^2 \sim InverseGamma(10, 0.1).$$

The plotted prior distributions for σ^2 and τ^2 can be seen in Figure 3.1.

3.4 The Complete Conditionals

As noted earlier in this discussion, one advantage to this choice of prior distributions is that the complete conditionals have closed-form solutions. The joint posterior distribution is proportional to the product of each of the relevant prior distributions and the likelihood function, so that

$$\pi(\Theta, \sigma^2, \tau^2|\mathbf{y}) \propto \mathbf{f(y_{ij}}|\theta_\mathbf{i}, \sigma^\mathbf{2})\pi(\sigma^\mathbf{2})\pi(\tau^\mathbf{2}) \prod_{\mathbf{i=1}}^{\mathbf{t}} \pi(\theta_\mathbf{i}|\theta_\mathbf{i-1})\pi(\theta_\mathbf{1}). \tag{3.5}$$

Simplifying the above equation, the distributions for the complete conditionals were calculated.

To obtain the complete conditional for θ_i, where the complete conditional is represented as $[\theta_i]$, it is important to realize that all of the parts in the joint posterior distribution that do not contain θ_i can be considered constants. The only distributions in the joint posterior that contain the value of θ_i in their distribution functions are the prior distributions, $\pi(\theta_i|\theta_{i-1})$ and $\pi(\theta_{i+1}|\theta_i)$, and the likelihood function, $f(y_{ij}|\theta_i, \sigma^2)$. Consequently, the complete conditional for θ_i is proportional to the simplified product: $\pi(\theta_i|\theta_{i-1})\pi(\theta_{i+1}|\theta_i)f(y_{ij}|\theta_i, \sigma^2)$.

Treating all terms in this product that do not contain θ_i as constants and

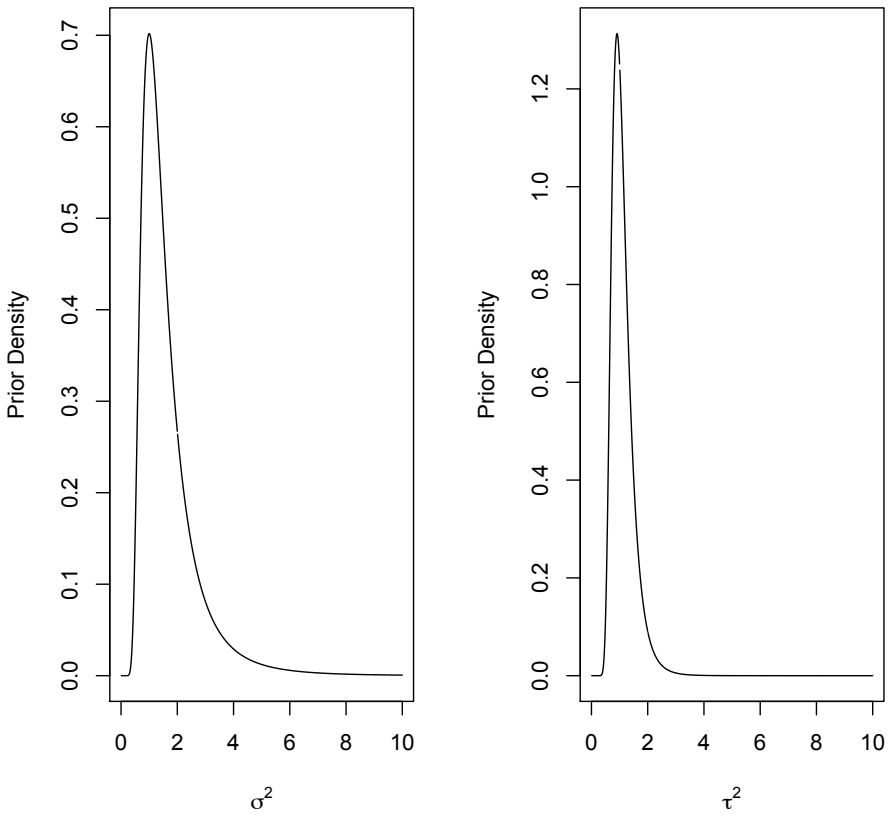

Figure 3.1: Prior Densities for σ^2 and τ^2

simplifying, the following results were obtained:

$$
\begin{aligned}
[\theta_i] &\propto \pi(\theta_i|\theta_{i-1})\pi(\theta_{i+1}|\theta_i)f(y_{ij}|\theta_i,\sigma^2) \\
&= (2\pi\tau^2)^{-\frac{1}{2}} exp\left[-\frac{(\theta_i-\theta_{i-1})^2}{2\tau^2}\right](2\pi\tau^2)^{-\frac{1}{2}} exp\left[-\frac{(\theta_{i+1}-\theta_i)^2}{2\tau^2}\right] \\
&\quad \times \prod_{i=1}^{t}\prod_{j=1}^{n_i}(2\pi\sigma^2)^{-\frac{1}{2}} exp\left[-\frac{(y_{ij}-\theta_i)^2}{2\sigma^2}\right] \\
&\propto exp\left[-\frac{(\theta_i-\theta_{i-1})^2}{2\tau^2}\right] exp\left[-\frac{(\theta_{i+1}-\theta_i)^2}{2\tau^2}\right] exp\left[-\frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2}{2\sigma^2}\right] \\
&= exp\left[-\frac{1}{2\sigma^2\tau^2}(\sigma^2(\theta_i-\theta_{i-1})^2+\sigma^2(\theta_{i+1}-\theta_i)^2+\tau^2\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2)\right].
\end{aligned}
$$

Multiplying this out and removing all terms that do not contain θ_i, the above equation is proportional to

$$
\begin{aligned}
&exp\left[-\frac{1}{2\sigma^2\tau^2}(\sigma^2(\theta_i^2-2\theta_i\theta_{i-1})+\sigma^2(-2\theta_{i+1}\theta_i+\theta_i^2)+\tau^2(-2\sum_{i=1}^{t}\sum_{j=1}^{n_i}y_{ij}\theta_i+\sum_{i=1}^{t}\sum_{j=1}^{n_i}\theta_i^2))\right] \\
&\propto exp\left[-\frac{1}{2\sigma^2\tau^2}(\sigma^2(\theta_i^2-2\theta_i\theta_{i-1})+\sigma^2(-2\theta_{i+1}\theta_i+\theta_i^2)+\tau^2(-2\theta_i n_i\bar{y}_i+n_i\theta_i^2))\right] \\
&= exp\left[-\frac{1}{2\sigma^2\tau^2}(\theta_i^2\sigma^2-2\theta_i\theta_{i-1}\sigma^2-2\theta_{i+1}\theta_i\sigma^2+\theta_i^2\sigma^2-2\theta_i n_i\bar{y}_i\tau^2+n_i\theta_i^2\tau^2)\right] \\
&= exp\left[-\frac{1}{2\sigma^2\tau^2}(\theta_i^2(2\sigma^2+n_i\tau^2)-2\theta_i(\theta_{i-1}\sigma^2+\theta_{i+1}\sigma^2+n_i\bar{y}_i\tau^2))\right].
\end{aligned}
$$

Factoring out $2\sigma^2+n_i\tau^2$,

$$
[\theta_i] \propto exp\left[-\frac{2\sigma^2+n_i\tau^2}{2\sigma^2\tau^2}\left(\theta_i^2-2\theta_i\left(\frac{\theta_{i-1}\sigma^2+\theta_{i+1}\sigma^2+n_i\bar{y}_i\tau^2}{2\sigma^2+n_i\tau^2}\right)\right)\right].
$$

The above expression is the kernel of a Normal, implying that the complete conditional for θ_i is distributed as follows:

$$
[\theta_i] \sim N\left(\frac{\theta_{i-1}\sigma^2+\theta_{i+1}\sigma^2+n_i\bar{y}_i\tau^2}{2\sigma^2+n_i\tau^2}, \frac{\sigma^2\tau^2}{2\sigma^2+n_i\tau^2}\right). \tag{3.6}
$$

Similar to obtaining the complete conditional for θ_i, the complete conditional for σ^2 is obtained by simplifying the joint posterior distribution while treating all

terms that do not contain σ^2 as constants. Doing this results in the following:

$$\begin{aligned}
[\sigma^2] &\propto \pi(\sigma^2) f(y_{ij}|\theta_i, \sigma^2) \\
&= \frac{\beta_\sigma^{-\alpha_\sigma}}{\Gamma(\alpha_\sigma)} (\sigma^2)^{-\alpha_\sigma - 1} exp\left[\frac{-1}{\beta_\sigma \sigma^2}\right] (2\pi\sigma^2)^{-N/2} exp\left[\frac{-\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2}{2\sigma^2}\right] \\
&\propto (\sigma^2)^{-\alpha_\sigma - 1 - \frac{N}{2}} exp\left[\frac{-1}{\sigma^2}\left(\frac{1}{\beta_\sigma} + \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2}{2}\right)\right].
\end{aligned}$$

The above expression implies that the distribution of the complete conditional for σ^2 is as follows:

$$[\sigma^2] \sim IG\left(a_\sigma + \frac{N}{2}, \left[\frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2}{2} + \frac{1}{\beta_\sigma}\right]^{-1}\right). \tag{3.7}$$

The complete conditional for τ^2 is obtained by simplifying the following:

$$\begin{aligned}
[\tau^2] &\propto \pi(\tau^2) \prod_{i=1}^{t} \pi(\theta_i|\theta_{i-1}) \\
&= \frac{\beta_\tau^{-\alpha_\tau}}{\Gamma(\alpha_\tau)} (\tau^2)^{-\alpha_\tau - 1} exp\left[\frac{-1}{\beta_\tau \tau^2}\right] \prod_{i=1}^{t} (2\pi\tau^2)^{\frac{-1}{2}} exp\left[\frac{-(\theta_i - \theta_{i-1})^2}{2\tau^2}\right] \\
&\propto (\tau^2)^{-\alpha_\tau - 1 - \frac{t}{2}} exp\left[\frac{-1}{\tau^2}\left(\frac{1}{\beta_\tau} + \frac{\sum_{i=1}^{t}(\theta_i - \theta_{i-1})^2}{2}\right)\right].
\end{aligned}$$

The above expression implies that the distribution for the complete conditional of τ^2 is

$$[\tau^2] \sim IG\left(a_\tau + \frac{t}{2}, \left[\frac{\sum_{i=1}^{t}(\theta_i - \theta_{i-1})^2}{2} + \frac{1}{\beta_\tau}\right]^{-1}\right). \tag{3.8}$$

Using the specified complete conditionals along with the statistical methods described above, Gaussian process models representing the behavior of the temperature for both treatments will be produced. By placing a Gaussian process prior on the parameter vector $\mathbf{\Theta}$, the posterior distributions for a potentially infinite vector of θ_is can be analyzed without increased complexity. In this project, since we only have data collected at 111 different time periods, the posterior distributions at each of the 111 periods will be used to model the behavior of the temperatures in each of the data sets.

3.5 How to Deal with First and Last θ_i

The model has been set up in such a way that the complete conditional of θ_i depends upon θ_{i-1} and θ_{i+1}. This framework presents problems for the first and last means of interest. For θ_1 there is no previous mean temperature or θ_0, and for θ_{111} there is no subsequent mean temperature or θ_{112} on which to base updates.

Since the complete conditionals are computed conditioning on the assumed known information, the complete conditional for θ_{111} is computed without conditioning on the prior conditional distribution for $\pi(\theta_{i+1}|\theta_i)$. Although the prior for θ_{i+1} incorporates information contained in θ_i, since θ_{i+1} does not exist at time 111, that information is simply removed and the complete conditional for θ_{111} is calculated by simplifying the product $\pi(\theta_{111}|\theta_{110})f(y_{ij}|\theta_i, \sigma^2)$ to obtain

$$
\begin{aligned}
[\theta_{111}] &\propto \pi(\theta_{111}|\theta_{110})f(y_{ij}|\theta_i,\sigma^2) \\
&= (2\pi\tau^2)^{-\frac{1}{2}} exp\left[-\frac{(\theta_{111}-\theta_{110})^2}{2\tau^2}\right]\prod_{i=1}^{t}\prod_{j=1}^{n_i}(2\pi\sigma^2)^{-\frac{1}{2}}exp\left[-\frac{(y_{ij}-\theta_i)^2}{2\sigma^2}\right] \\
&\propto exp\left[-\frac{(\theta_{111}-\theta_{110})^2}{2\tau^2}\right] exp\left[-\frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2}{2\sigma^2}\right] \\
&= exp\left[-\frac{1}{2\sigma^2\tau^2}(\sigma^2(\theta_{111}-\theta_{110})^2+\tau^2\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2)\right] \\
&\propto exp\left[-\frac{1}{2\sigma^2\tau^2}(\sigma^2(\theta_{111}^2-2\theta_{111}\theta_{110})+\tau^2(-2\sum_{i=1}^{t}\sum_{j=1}^{n_i}y_{ij}\theta_i+\sum_{i=1}^{t}\sum_{j=1}^{n_i}\theta_i^2))\right] \\
&\propto exp\left[-\frac{1}{2\sigma^2\tau^2}(\sigma^2(\theta_{111}^2-2\theta_{111}\theta_{110})+\tau^2(-2\theta_{111}n\bar{y}_{111}+n_{111}\theta_{111}^2))\right] \\
&= exp\left[-\frac{1}{2\sigma^2\tau^2}(\theta_{111}^2\sigma^2-2\theta_{111}\theta_{110}\sigma^2-2\theta_{111}n\bar{y}_{111}\tau^2+n\theta_{111}^2\tau^2)\right] \\
&= exp\left[-\frac{1}{2\sigma^2\tau^2}(\theta_{111}^2(\sigma^2+n\tau^2)-2\theta_{111}(\theta_{110}\sigma^2+n\bar{y}_{111}\tau^2))\right] \\
&\propto exp\left[-\frac{\sigma^2+n\tau^2}{2\sigma^2\tau^2}\left(\theta_{111}^2-2\theta_{111}\left(\frac{\theta_{110}\sigma^2+n\bar{y}_{111}\tau^2}{\sigma^2+n\tau^2}\right)\right)\right].
\end{aligned}
$$

The above expression is the kernel of a Normal distribution, implying that the com-

plete conditional distribution for θ_{111} is as follows:

$$[\theta_{111}] \sim N\left(\frac{n\tau^2\bar{y}_{111} + \theta_{110}\sigma^2}{n\tau^2 + \sigma^2}, \frac{\sigma^2\tau^2}{n\tau^2 + \sigma^2}\right). \tag{3.9}$$

The product $\pi(\theta_1)\pi(\theta_2|\theta_1)f(y_{ij}|\theta_i, \sigma^2)$ is simplified to obtain the complete conditional distribution for θ_1. The simplification process to obtain the complete conditional for θ_1 is similar to the process used to obtain the complete conditional for θ_i. The only difference is that prior values are placed on θ_1 and the distribution on θ_1 is called the initial information as mentioned above. (See above for prior value specification). Simplifying the above product of distributions, the following results were obtained:

$$\begin{aligned}
[\theta_1] &\propto \pi(\theta_1)\pi(\theta_2|\theta_1)f(y_{ij}|\theta_i, \sigma^2) \\
&= (2\pi(2))^{-\frac{1}{2}} exp\left[-\frac{(\theta_1 - 25)^2}{2(2)}\right](2\pi\tau^2)^{-\frac{1}{2}} exp\left[-\frac{(\theta_2 - \theta_1)^2}{2\tau^2}\right] \\
&\quad \times \prod_{i=1}^{t}\prod_{j=1}^{n_i}(2\pi\sigma^2)^{-\frac{1}{2}} exp\left[-\frac{(y_{ij} - \theta_i)^2}{2\sigma^2}\right] \\
&\propto exp\left[-\frac{(\theta_1 - 25)^2}{2(2)}\right] exp\left[-\frac{(\theta_2 - \theta_1)^2}{2\tau^2}\right] exp\left[-\frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij} - \theta_i)^2}{2\sigma^2}\right] \\
&= exp\left[-\frac{1}{2(2)\sigma^2\tau^2}(\sigma^2\tau^2(\theta_1 - 25)^2 + 2\sigma^2(\theta_2 - \theta_1)^2 + 2\tau^2\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij} - \theta_i)^2)\right]. \\
&\propto exp\left[-\frac{1}{2(2)\sigma^2\tau^2}(\sigma^2\tau^2(\theta_1^2 - 2(25)\theta_1) + 2\sigma^2(-2\theta_2\theta_1 + \theta_1^2) + 2\tau^2(-2\theta_1 n_1\bar{y}_1 + n_1\theta_1^2))\right] \\
&= exp\left[-\frac{1}{2(2)\sigma^2\tau^2}(\theta_1^2\sigma^2\tau^2 - 2(25)\theta_1\sigma^2\tau^2 - 2(2)\theta_2\theta_1\sigma^2 + 2\theta_1^2\sigma^2 - 2(2)\theta_1 n_1\bar{y}_1\tau^2 + 2n_1\theta_1^2\tau^2)\right] \\
&= exp\left[-\frac{1}{2(2)\sigma^2\tau^2}(\theta_1^2(\sigma^2\tau^2 + 2\sigma^2 + 2n_1\tau^2) - 2\theta_1(25\sigma^2\tau^2 + 2\theta_2\sigma^2 + 2n_1\bar{y}_1\tau^2))\right].
\end{aligned}$$

Factoring out $\sigma^2\tau^2 + 2\sigma^2 + 2n_1\tau^2$,

$$[\theta_1] \propto exp\left[-\frac{\sigma^2\tau^2 + 2\sigma^2 + 2n_1\tau^2}{2(2)\sigma^2\tau^2}\left(\theta_1^2 - 2\theta_1\frac{(25\sigma^2\tau^2 + 2\theta_2\sigma^2 + 2n_1\bar{y}_1\tau^2)}{\sigma^2\tau^2 + 2\sigma^2 + 2n_1\tau^2}\right)\right],$$

which is the kernel of a Normal distribution, implying that the complete conditional

distribution for θ_1 is as follows:

$$[\theta_1] \sim N\left(\frac{(25\sigma^2\tau^2 + 2\theta_2\sigma^2 + 2n_1\bar{y}_1\tau^2)}{\sigma^2\tau^2 + 2\sigma^2 + 2n_1\tau^2}, \frac{2\sigma^2\tau^2}{\sigma^2\tau^2 + 2\sigma^2 + 2n_1\tau^2}\right). \tag{3.10}$$

3.6 Comparison of Treatments

To assess whether there is a difference between the ice pack and whirlpool bath in reducing tendon temperature, the posterior distributions of the differences between the means of the data sets at each time period were used to model the behavior of the differences between the treatments. This is similar to how the posterior distributions of the mean of each of the treatments was used to model the behavior of the temperatures, and this technique uses the same idea as the Gaussian process analysis that was performed on the original data sets in that the inference is performed on the parameter space directly.

To obtain the posterior distributions of the differences, two different methods which should yield similar results were applied. First, a third data set was created by pairing up the individuals for each of the two treatments and taking the difference in their respective temperatures at each time period. The same Gaussian process methods were used in modeling this set of data as were used in modeling the previous two sets of data. Also, the same likelihood and prior distributions were used as previously described, resulting in the same complete conditionals, with the exception of the complete conditional for θ_{diff_1}. With θ_{diff_1} representing the mean difference in the temperatures between the two data sets at time 1, a prior mean of zero was placed on θ_{diff_1} instead of a value of 25, which was used for the original data sets, and $\theta_{diff_1} \sim N(0, 2)$.

In addition, the posterior distributions for the differences at each time period were obtained by using draws from the posterior distributions of the individual treatment means and subtracting the difference in the temperature value. This process was repeated for each time period and the resulting mean differences produced values

that model the behavior of the posterior distribution for the differences in temperature means between the two treatments at that time period. Continuing this process for each time period resulted in a posterior distribution for the difference in the mean temperature for each time period.

With the posterior distributions of the differences at each time period computed as described above, a 95% credible interval at each time period was calculated to determine if there was a significant difference between the treatment means at each of the time periods. By noting if zero was contained within the interval, it was determined if there is a significant difference between the two treatments at any time period.

4. RESULTS

4.1 Modeling Temperature Means

The posterior distribution at each of the time periods for both the whirlpool and ice pack data sets were obtained. Figure 4.1 shows all 111 posterior distributions for the whirlpool treatment data set plotted on the same plot. The marginal posterior distributions for the θ_is for the ice pack treatment data set are similar.

As mentioned previously, it is desirable to understand the behavior of the temperatures as they change throughout time. To do this, the mean of each density was used as an estimate of the temperature at each time period. Figure 4.2 shows the mean of the posterior distributions for each data set plotted against time along with the 95% credible interval for the mean of each distribution at each time.

The behavior in the temperatures as modeled through time is similar in both sets of data. It is apparent from Figure 4.2 that the temperatures of the tendons are drastically reduced in both sets of data when the treatments are applied. The plots also indicate a gradual increase in the temperature in both the data sets when the treatments are removed.

4.2 Modeling Temperature Differences

To better assess which treatment is more effective in decreasing the temperature of the tendons as well as keeping the temperatures low after the treatment is removed, the posterior distributions of the differences of the temperatures between the two data sets were computed.

The individuals used when measuring tendon temperatures in the whirlpool data set were the same individuals used when taking measurements on tendon tem-

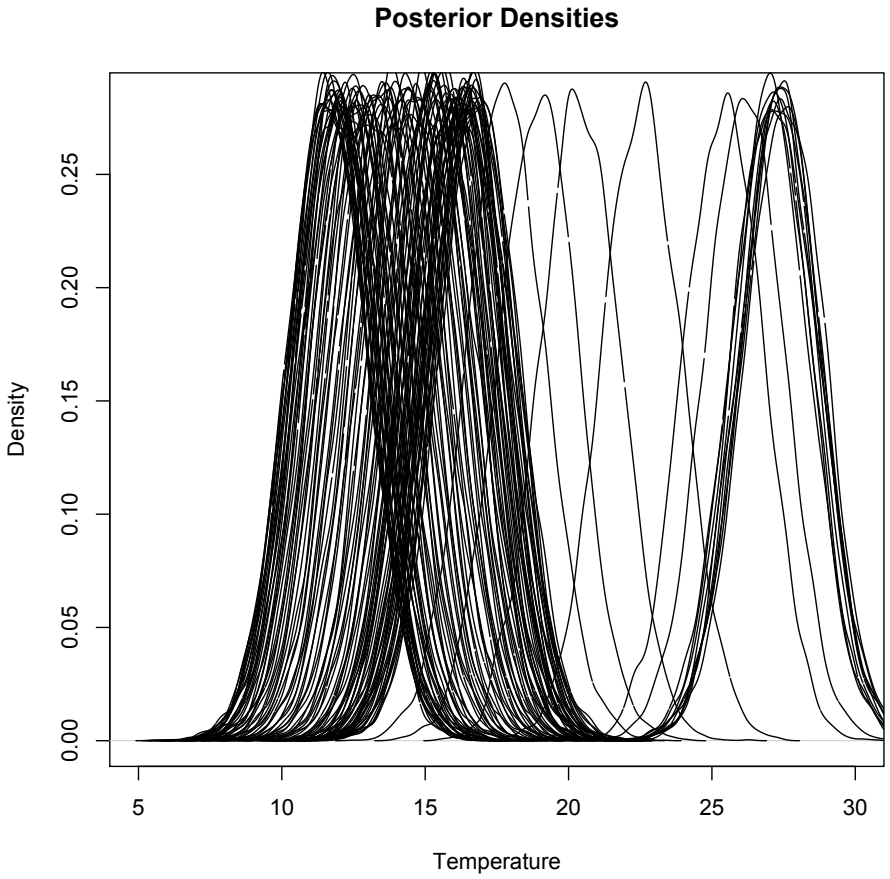

Figure 4.1: Posterior Densities for θ_i i=1,...,111

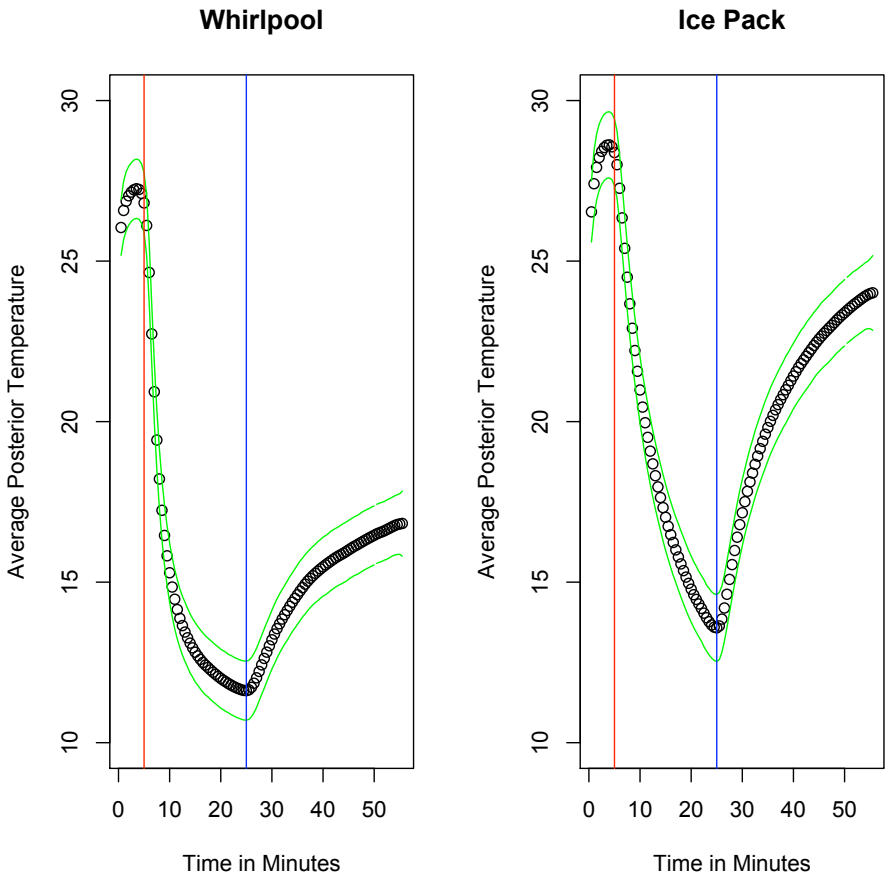

Figure 4.2: Posterior Densities for θ_i i=1,...,111

peratures in the ice pack data set. Consequently, the differences in the treatment temperatures for each individual can be modeled.

Using the two methods described previously, the posterior distributions of the differences in mean temperature at each time period were obtained (the whirlpool data set was subtracted from the ice pack data set). The posterior means along with the 95% credible intervals for the differences in temperatures between the two treatments for both methods can be seen in Figure 4.3. The posterior distributions of the differences in temperatures that were obtained are similar for both methods.

4.3 Conclusions and Comparisons

Its characteristics of general applicability, mathematical simplicity, and the ability to update the model as information becomes available make the method of Gaussian processes implemented within a dynamic linear model setting arguably more powerful in modeling complex nonlinear behavior than other methods. In 2007 John Howell performed a similar analysis on the tendon temperature data sets using a technique known as smoothing splines. Howell (2007) shows that, similar to Gaussian processes, smoothing splines offer more flexibility than traditional polynomial regression when fitting nonlinear behavior. However, Howell points out that some challenges and limitations arise due to assumptions about parameters and distributional forms that need to be made when using smoothing splines. Although Gaussian processes offer more flexibility when modeling complex behavior, the results obtained by Howell are very similar to those obtained using Gaussian processes.

Figure 4.3 shows the 95% credible intervals for each of the posterior distributions at each time period. Because none of the credible intervals contain the value 0 after the treatments are applied, we believe that the treatments are significantly different at each time period when the treatment is applied as well as when the treatments are removed. Again, because Figure 4.3 displays the differences in temperatures where the

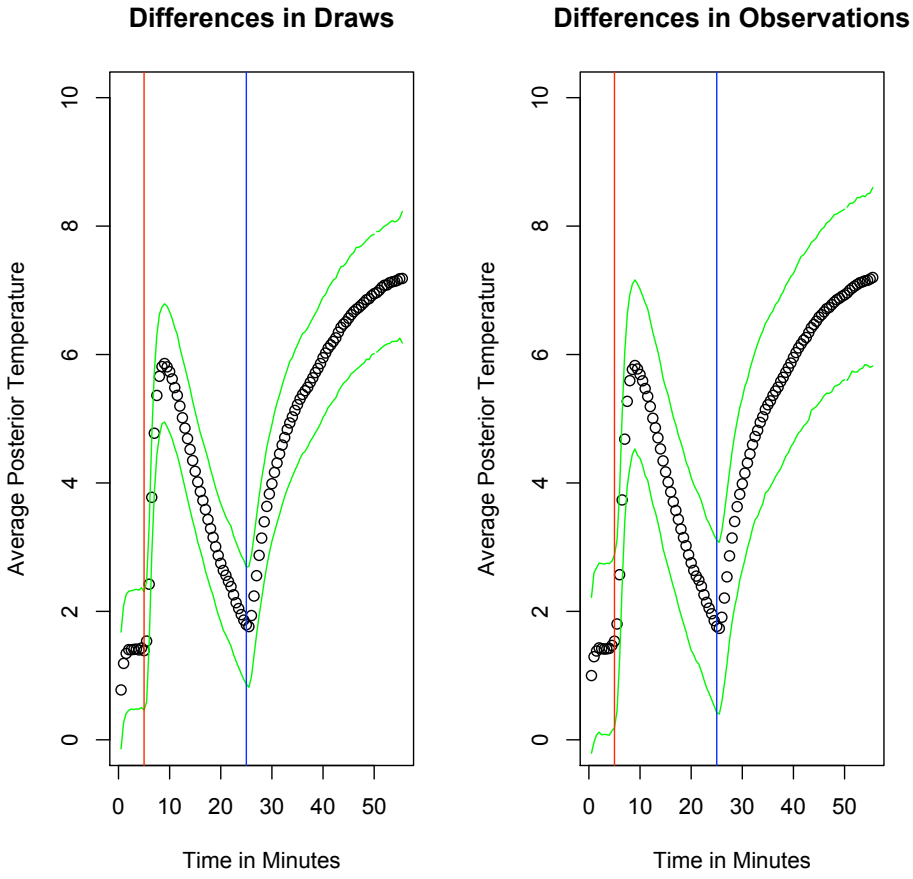

Figure 4.3: Posterior Means of Differences (Ice Pack Minus Whirlpool)

observations from the whirlpool treatment are subtracted from the ice pack treatment, we conclude that the whirlpool treatment is significantly more effective in reducing the Achilles tendon temperatures as well as keeping the temperatures down once the treatments are removed.

By examining Figure 4.3, we also conclude that the temperature differences are the greatest between the two treatments during the first few time periods when the treatments are applied as well as the later time periods when the treatments are removed. Figure 4.3 indicates that the whirlpool treatment decreases the temperatures faster in the tendons, but that the temperatures eventually become more similar the longer the treatments are applied. Once the treatments are removed, the temperatures in the tendons increase faster in the means where the ice pack treatment was applied. This further indicates that the whirlpool treatment is preferable to the ice pack treatment. Incorporating a Gaussian process approach within a dynamic linear model setting to model the behavior of the temperatures makes it clear that the more effective treatment is the whirlpool bath. The whirlpool treatment both reduces the temperatures more quickly and keeps the temperatures lower for a longer period of time after the treatments are removed.

BIBLIOGRAPHY

Ashby, D. (2006), "Bayesian Statistics in Medicine: A 25 Year Review," *Statistics in Medicine*, 25, 3589–3631.

Bayes, T. (1764), "An Essay Towards Solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F.R.S. communicated by Mr. Price, in a letter to John Canton, A.M.F.R.S." *Philosophical Transactions of the Royal Society of London*, 53, 370–418.

Berry, S. M. and Ruppert, D. (2002), "Bayesian Smoothing and Regression Splines for Measurement Error Problems," *Journal of the American Statistical Association*, 97, 160–169.

Dym, H. and McKean, H. (1976), *Gaussian Processes, Function Theory and Inverse Spectral Problem*, New York: Academic Press.

Ebden, M. (2008), "Gaussian Processes for Regression," unpublished lecture notes, Oxford University, 1–6.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis (2nd Ed.)*, New York: Chapman and Hall.

Geman, S. and Geman, D. (1984), "Stochastic Relaxation Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Gray, G., Murray-Smith, R., and Thompson, K. (2003), "Tutorial Example of Gaussian Process Prior Modelling Applied to Twin-Tank System," unpublished report, Glasgow University, 1–12.

Hammersley, J. M. and Handscomb, D. C. (1964), *Monte Carlo Methods*, London: Taylor and Francis.

Hastings, W. (1970), "Monte Carlo Sampling Methods Using Markov Chains and their Applications," *Biometrika*, 57, 97–109.

Hida, T. and Hitsuda, M. (1976), "Gaussian Processes," *Translations of Mathematical Monographs*, 120, 21–67.

Howell, J. (2007), "Analysis Using Smoothing Splines as Implemented in LME() in R," unpublished master's project, Brigham Young University, Dept. of Statistics.

Ibragimov, I. and Rozanov, Y. (1970), *Gaussian Random Processes*, New York: Springer-Verlag.

Mackay, D. (1999), "Introduction to Gaussian Processes," unpublished lecture notes, Cambridge University, 1–20.

Marcus, M. B. and Rosen, J. (2006), *Markov Processes, Gaussian Processes, and Local Times*, Cambridge: University Press.

Metropolis, N. and Ulam, S. (1949), "The Monte Carlo Method," *Journal of the American Statistical Association*, 44, 335–341.

Murray-Smith, R. and Sbarbaro, D. (2006), "Nonlinear Adaptive Control Using Non-Parametric Gaussian Prior Models," International Federation of Automatic Control, Conference Paper, 1–6.

Prabhu, N. and Basawa, I. (eds.) (1990), *Statistical Inference in Stochastic Processes*, New York: Marcel Dekker Inc.

Rasmussen, C. E. and Williams, C. (2006), *Gaussian Processes for Machine Learning*, Cambridge, MA: The MIT Press Cambridge.

Ross, S. M. (2007), *Introduction to Probability Models*, New York: Academic Press.

Safran, M. R., McKeag, D. B., and Camp, S. P. V. (2002), *Manual of Sports Medicine*, Philadelphia, PA: Lippincott-Raven.

Schervish, M. J. (2004), "Nonstationary Covariance Functions for Gaussian Process Regression," *Proc. of the Conf. on Neural Information Processing Systems (NIPS)*, 6, 1–5.

```
library(MASS)
library(MCMCpack)
library(msm)
library(rgl)

#PLOTTING THE DATA
data<-read.table('whirlpool.txt',header=TRUE)
data<-data[,c(-1,-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25,-27,-29)]
data<-as.matrix(data)
data.Ice<-read.table('icepack.txt',header=TRUE)
data.Ice<-data.Ice[,c(-1,-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25,-27,-29)]
data.Ice<-as.matrix(data.Ice)
Time<-seq(.5,55.5,.5)
par(mfrow=c(1,2))
plot(Time,data[,1],main='Whirlpool Data',ylim=c(5,32),pch=1,ylab='Temperature'
   ,xlab='Time in Minutes')
for(i in 2:15){
points(Time,data[,i],pch=i,col=i)
}
plot(Time,data.Ice[,1],main='Ice Pack Data',ylim=c(5,32),pch=1,ylab='Temperature'
   ,xlab='Time in Minutes')
for(i in 2:15){
points(Time,data.Ice[,i],pch=i,col=i)
}

###PLOTTING PRIORS#####
mean(rgamma(10000,shape=5,scale=.2))
mean(1/rgamma(10000,shape=5,scale=.2))
var(1/rgamma(10000,shape=5,scale=.2))
xxx<-seq(0,10,.01)
par(mfrow=c(1,2))
plot(xxx,dinvgamma(xxx,shape=4,scale=1/.2),type='l',main='',xlab=expression(sigma^2)
   ,ylab='Prior Density')
plot(xxx,dinvgamma(xxx,shape=10,scale=1/.1),type='l',main='',xlab=expression(tau^2)
   ,ylab='Prior Density')
alpha<-10
beta<-.1

data<-t(data)
data.Ice<-t(data.Ice)
nobs<-10000

#whirlpooldata
meany<-apply(data,2,mean)
n<-dim(data)[1]
N<-dim(data)[1]*dim(data)[2]
T<-dim(data)[2]
theta<-matrix(1,nrow=nobs,ncol=111)
tau2<-rep(5,nobs)
sig2<-rep(7.5,nobs)
```

```
#priors for whirlpool data
asig<-5
bsig<-.2
atau<-100
btau<-100

#icepack data
meany.Ice<-apply(data.Ice,2,mean)
theta.Ice<-matrix(1,nrow=nobs,ncol=111)
tau2.Ice<-rep(1,nobs)
sig2.Ice<-rep(1,nobs)

#priors for icepack data
asig.Ice<-5
bsig.Ice<-2
atau.Ice<-50
btau.Ice<-30

#here theta1 and theta111 are different
#start MCMC loop
for (j in 2:nobs)
{
#WHIRL POOL DATA-update theta1
   alpha<-((2*n*meany[1]*tau2[j-1])+(sig2[j-1]*tau2[j-1]*25)+
              (2*theta[j-1,2]*sig2[j-1]))/((2*n*tau2[j-1])+
              (2*sig2[j-1])+(sig2[j-1]*tau2[j-1]))
   gamma<-(2*sig2[j-1]*tau2[j-1])/((2*n*tau2[j-1])+(2*sig2[j-1])
               +(sig2[j-1]*tau2[j-1]))
   theta[j,1]<-rnorm(1,alpha,sqrt(gamma))

   #ICE PACK DATA-update theta1.Ice
   alpha.Ice<-((2*n* meany.Ice[1]*tau2.Ice[j-1])+(sig2.Ice[j-1]*tau2.Ice[1]*25)
                   +(2*theta.Ice[j-1,2]*sig2.Ice[j-1]))/((2*n*tau2.Ice[j-1])
                   +(2*sig2.Ice[j-1])+(sig2.Ice[j-1]*tau2.Ice[j-1]))
   gamma.Ice<-(2*sig2.Ice[j-1]*tau2.Ice[j-1])/((2*n*tau2.Ice[j-1])
                   +(2*sig2.Ice[j-1])+(sig2.Ice[j-1]*tau2.Ice[j-1]))
   theta.Ice[j,1]<-rnorm(1,alpha.Ice,sqrt(gamma.Ice))

#WHIRL POOL DATA-update theta2 - theta110
   for (i in 2:110)
   {
     alpha<-((n*meany[i]*tau2[j-1])+(theta[j,i-1]*sig2[j-1] )+
                (theta[j-1,i+1]*sig2[j-1]))
                 /((n*tau2[j-1])+(2*sig2[j-1]))
    gamma<-(sig2[j-1] *tau2[j-1]) / ((n*tau2[j-1])+(2*sig2[j-1]))
     theta[j,i]<- rnorm(1,alpha,sqrt(gamma))
   }

   #ICE PACK DATA-update theta2 - theta110
   for (i in 2:110)
   {
     alpha.Ice<-((n*meany.Ice[i]*tau2.Ice[j-1])+(theta.Ice[j,i-1]
                   *sig2.Ice[j-1])+(theta.Ice[j-1,i+1]*sig2.Ice[j-1]))
```

```
                  /((n*tau2.Ice[j-1])+(2*sig2.Ice[j-1]))
    gamma.Ice<-(sig2.Ice[j-1]*tau2.Ice[j-1])/((n*tau2.Ice[j-1])
                  +(2*sig2.Ice[j-1]))
     theta.Ice[j,i]<- rnorm(1,alpha.Ice,sqrt(gamma.Ice))
   }

#WHIRL POOL DATA-update theta111
   alpha<-((tau2[j-1]*n*meany[111]+sig2[j-1]*theta[j,110])
              /(n*tau2[j-1]+sig2[j-1]))
   gamma<-(sig2[j-1]*tau2[j-1])/(n*tau2[j-1]+sig2[j-1])
   theta[j,111]<-rnorm(1,alpha,sqrt(gamma))

    #ICE PACK DATA-update theta111
    alpha.Ice<-((tau2.Ice[j-1]*n*meany.Ice[111]+sig2.Ice[j-1]
              *theta.Ice[j,110])/(n*tau2.Ice[j-1]+sig2.Ice[j-1]))
   gamma.Ice<-(sig2.Ice[j-1]*tau2.Ice[j-1])/(n*tau2.Ice[j-1]+sig2.Ice[j-1])
   theta.Ice[j,111]<-rnorm(1,alpha.Ice,sqrt(gamma.Ice))

   #WHIRL POOL DATA-update tau2
    astartau<-atau +T/2
    tausum<-0
    for (i in 2:111)
    {
       tausum<- tausum + ((theta[j,i] - theta[j,i-1])^2)/(2)
    }
    bstartau<-(1/btau + tausum)^(-1)
    tau2[j]<-1/rgamma(1,shape=astartau,scale=bstartau)

    #ICE PACK DATA-update tau2
    astartau.Ice<-atau.Ice+T/2
    tausum.Ice<-0
    for (i in 2:111)
    {
       tausum.Ice<- tausum.Ice + ((theta.Ice[j,i] - theta.Ice[j,i-1])^2)/(2)
    }
    bstartau.Ice<-(1/btau.Ice + tausum.Ice)^(-1)
    tau2.Ice[j]<-1/rgamma(1,shape=astartau.Ice,scale=bstartau.Ice)

#WHIRL POOL DATA-update sigma2
    astarsig<-asig  + N/2
    tempthetamat<-rep(theta[j,1],n)
    for(i in 2:111)
    {
     tempthetamat<-cbind(tempthetamat,rep(theta[j,i],n))
    }
    sigsum <- sum((data-tempthetamat)^2)/2
    bstarsig<-(sigsum+1/bsig)^(-1)
    sig2[j]<-1/rgamma(1,shape=astarsig,scale=bstarsig)

    #ICE PACK DATA-update sigma2
    astarsig.Ice<- asig.Ice  + N/2
    tempthetamat.Ice<-rep(theta.Ice[j,1],n)
    for(i in 2:111)
    {
```

```
     tempthetamat.Ice<-cbind(tempthetamat.Ice,rep(theta.Ice[j,i],n))
    }
    #tempthetamat<-cbind(rep(theta[j,1],n),rep(theta[j,2],n),rep(theta[j,3],n)
                  ,rep(theta[j,4],n),rep(theta[j,5],n),rep(theta[j,6],n))
    sigsum.Ice <-sum((data.Ice-tempthetamat.Ice)^2)/2
    bstarsig.Ice<-(sigsum.Ice+1/bsig.Ice)^(-1)
    sig2.Ice[j]<-1/rgamma(1,shape=astarsig.Ice,scale=bstarsig.Ice)
}


PLOTTING MARGINAL POSTERIOR DENSITIES FOR WHIRLPOOL DATA
par(mfrow=c(1,3))
plot(density(theta[500:nobs,1]),xlim=c(5,30),main='Posterior
   Densities',xlab='Temperature')
abline(v=meany[1],col="red")
for(i in 2:111)
{
lines(density(theta[500:nobs,i]))
abline(v=meany[i],col="red")
}

#WHIRL POOL DATA-average thetas and plots of average thetas
average.thetas<-NULL
HPD1<-NULL
HPD2<-NULL
for(i in 1:111)
{
average.thetas[i]<-mean(theta[500:nobs,i])
HPD1[i]<-quantile(theta[500:nobs,i],.025)
HPD2[i]<-quantile(theta[500:nobs,i],.975)
}
time<-seq(.5,55.5,.5)
par(mfrow=c(1,2))
plot(time,average.thetas,xlab='Time in Minutes',
   ylab='AveragePosterior Temperature',type='p',
   ylim=c(10,30),main='Whirlpool')
lines(time,HPD1,type='l',col='green')
lines(time,HPD2,type='l',col='green')
abline(v=5,col='red')
abline(v=25,col='blue')

#ICE PACK DATA-average thetas and plots of average thetas
average.thetas.Ice<-NULL
HPD1.Ice<-NULL
HPD2.Ice<-NULL
for(i in 1:111)
{
average.thetas.Ice[i]<-mean(theta.Ice[500:nobs,i])
HPD1.Ice[i]<-quantile(theta.Ice[500:nobs,i],.025)
HPD2.Ice[i]<-quantile(theta.Ice[500:nobs,i],.975)
}
time<-seq(.5,55.5,.5)
plot(time,average.thetas.Ice,xlab='Time in Minutes',
   ylab='AveragePosterior Temperature',type='p',
```

```
   ylim=c(10,30),main='Ice Pack')
lines(time,HPD1.Ice,type='l',col='green')
lines(time,HPD2.Ice,type='l',col='green')
abline(v=5,col='red')
abline(v=25,col='blue')

###PART 2 POST DISTRIBUTION OF DIFFERENCES IN DRAWS FROM POSTERIOR####
theta.diff<-theta.Ice-theta
#DIFFERENCES IN THETAS
average.thetas.diff<-NULL
HPD1.diff<-NULL
HPD2.diff<-NULL
for(i in 1:111)
{
average.thetas.diff[i]<-mean(theta.diff[500:nobs,i])
HPD1.diff[i]<-quantile(theta.diff[500:nobs,i],.025)
HPD2.diff[i]<-quantile(theta.diff[500:nobs,i],.975)
}
time<-seq(.5,55.5,.5)
par(mfrow=c(1,2))
plot(time,average.thetas.diff,xlab='Time',ylab='Average
   PosteriorTemperature',type='p',main='Differences in Draws'
   ,ylim=c(0,10))
lines(time,HPD1.diff,type='l',col='green')
lines(time,HPD2.diff,type='l',col='green')
abline(v=5,col='red')
abline(v=25,col='blue')

###PART 3 DATA SET 3 (DIFFERENCES IN TREATMENTS)####
#Creating data
data<-read.table('whirlpool.txt',header=TRUE)
data.Ice<-read.table('icepack.txt',header=TRUE)
data<-as.matrix(data)
data.Ice<-as.matrix(data.Ice)
for(i in 1:15)
{
data<-data[,-i]
data.Ice<-data.Ice[,-i]
}
data<-t(data)
data.Ice<-t(data.Ice)
data.minus<-data.Ice-data
nobs<-10000

#data.three
meany.minus<-apply(data.minus,2,mean)
n<-dim(data.minus)[1]
N<-dim(data.minus)[1]*dim(data.minus)[2]
T<-dim(data.minus)[2]
theta.minus<-matrix(1,nrow=nobs,ncol=111)
tau2.minus<-rep(1,nobs)
sig2.minus<-rep(1,nobs)

#priors for data.minus
```

```
asig.minus<-5
bsig.minus<-.2
atau.minus<-10
btau.minus<-.1
j<-1000
#here theta1 and theta111 are different
#start MCMC loop
for (j in 2:nobs)
{
     #MINUS DATA-update theta1 ####
     alpha.minus<-((2*n*meany.minus[1]*tau2.minus[j-1])+
                  (sig2.minus[j-1]*tau2.minus[j-1]*0)+
                  (2*theta.minus[j-1,2]*sig2.minus[j-1]))
                  /((2*n*tau2.minus[j-1])
                  +(2*sig2.minus[j-1])+(sig2.minus[j-1]*
                  tau2.minus[j-1]))
     gamma.minus<-(2*sig2.minus[j-1]*tau2.minus[j-1])
                  /((2*n*tau2.minus[j-1])+(2*sig2.minus[j-1])
                  +(sig2[j-1]*tau2[j-1]))
     theta.minus[j,1]<-rnorm(1,alpha.minus,sqrt(gamma.minus))

     #MINUS DATA-update theta2 - theta110
     for (i in 2:110)
   {
        alpha.minus<-((n*meany.minus[i]*tau2.minus[j-1])
                        +(theta.minus[j,i-1]*sig2.minus[j-1])
                        +(theta.minus[j-1,i+1]*sig2.minus[j-1]))
                        /((n*tau2.minus[j-1])+(2*sig2.minus[j-1]))
      gamma.minus<-(sig2.minus[j-1] *tau2.minus[j-1])/
                        ((n*tau2.minus[j-1])+(2*sig2.minus[j-1]))
       theta.minus[j,i]<-rnorm(1,alpha.minus,sqrt(gamma.minus))
   }

     #MINUS DATA-update theta111
      alpha.minus<-((tau2.minus[j-1]*n*meany.minus[111]+
                   sig2.minus[j-1]*theta.minus[j,110])
                   /(n*tau2.minus[j-1]+sig2.minus[j-1]))
      gamma.minus<-(sig2.minus[j-1]*tau2.minus[j-1])/
                   (n*tau2.minus[j-1]+sig2.minus[j-1])
      theta.minus[j,111]<-rnorm(1,alpha.minus,sqrt(gamma.minus))

      #MINUS-update tau2
      astartau.minus<-atau.minus +T/2
      tausum.minus<-0
      for (i in 2:111)
       {
        tausum.minus<-tausum.minus+((theta.minus[j,i]-theta.minus[j,i-1])^2)/(2)
       }
       bstartau.minus<-(1/btau.minus + tausum.minus)^(-1)
       tau2.minus[j]<-1/rgamma(1,shape=astartau.minus,scale=bstartau.minus)

#MINUS-update sigma2
    astarsig.minus<-asig.minus+N/2
    tempthetamat.minus<-rep(theta.minus[j,1],n)
```

```
    for(i in 2:111)
    {
     tempthetamat.minus<-cbind(tempthetamat.minus,rep(theta.minus[j,i],n))
    }
    sigsum.minus <- sum((data.minus-tempthetamat.minus)^2)/2
    bstarsig.minus<-(sigsum.minus+1/bsig.minus)^(-1)
    sig2.minus[j]<-1/rgamma(1,shape=astarsig.minus,scale=bstarsig.minus)

}

#WHIRL POOL DATA-average thetas and plots of average thetas
average.thetas.minus<-NULL
HPD1.minus<-NULL
HPD2.minus<-NULL
for(i in 1:111)
{
average.thetas.minus[i]<-mean(theta.minus[500:nobs,i])
HPD1.minus[i]<-quantile(theta.minus[500:nobs,i],.025)
HPD2.minus[i]<-quantile(theta.minus[500:nobs,i],.975)
}
time<-seq(.5,55.5,.5)
plot(time,average.thetas.minus,xlab='Time',ylab='Average Posterior Temperature'
   ,type='p',ylim=c(0,10),main='Differences in Observations')
lines(time,HPD1.minus,type='l',col='green')
lines(time,HPD2.minus,type='l',col='green')
abline(v=5,col='red')
abline(v=25,col='blue')
```