



Published in final edited form as:

J Am Stat Assoc. 2014 July ; 109(507): 894–904. doi:10.1080/01621459.2014.899234.

Mechanistic Hierarchical Gaussian Processes

Matthew W. Wheeler¹, David B. Dunson², Sudha P. Pandalai¹, Brent A. Baker^{†,1}, and Amy H. Herring^{†,3}

Matthew W. Wheeler: mwheeler@cdc.gov

¹National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226, MS C-15

²Department of Statistical Science, Duke University

³Department of Biostatistics, University of North Carolina at Chapel Hill

Abstract

The statistics literature on functional data analysis focuses primarily on flexible black-box approaches, which are designed to allow individual curves to have essentially any shape while characterizing variability. Such methods typically cannot incorporate mechanistic information, which is commonly expressed in terms of differential equations. Motivated by studies of muscle activation, we propose a nonparametric Bayesian approach that takes into account mechanistic understanding of muscle physiology. A novel class of hierarchical Gaussian processes is defined that favors curves consistent with differential equations defined on motor, damper, spring systems. A Gibbs sampler is proposed to sample from the posterior distribution and applied to a study of rats exposed to non-injurious muscle activation protocols. Although motivated by muscle force data, a parallel approach can be used to include mechanistic information in broad functional data analysis applications.

Keywords

Functional data analysis; Gaussian process; Muscle force; Ordinary differential equations; Stochastic differential equations

1. INTRODUCTION

Studies of physiologic response to muscle stress are important in developing treatment protocols to combat work-, athletic-, and age-related injury. In order to investigate muscle adaptation and maladaptation following repetitive resistance-type exercises, scientists often obtain a series of functional measures (typically at the beginning and end of a multi-session exercise protocol) on the force produced by the muscle as it moves through its range of motion. These force curves can be compared to determine the benefit/harm of an exercise routine to a population of interest.

[†]These authors share last author Brent Baker is the senior scientist from the lab that motivated the problem, and Amy Herring is the chair of the doctoral research committee that guided this research.

We investigate one such study conducted in rodents. In this study scientists are interested in differences in physiologic response between young (3 month) and old (30 month) rats exposed to the same resistive exercise protocol. Here rats underwent 13 training sessions, described in table 1, on the dorsiflexor (lower leg) muscle group. At the beginning and end of each training session the muscles underwent both isometric (muscle activation without movement) and stretch shortening contractions (muscle activation with joint movement). We are interested in modeling the last stretch shortening contraction (labeled observation 5 in the table) observed in the first (considered pre-exercise) and last (considered post-exercise) exercise protocol. When modeling the stretch shortening contraction there is also an isometric component to force generation, and modeling should estimate both the isometric and stretch shortening components. Our intent is to investigate possible differences in responses between groups (young/old), as well as differences in the response pre- and post-training, for both isometric and stretch shortening components. We are interested in comparing the individual and group level force tracings for isometric as well as stretch shortening contractions.

Each observation was recorded as a muscle force tracing, as illustrated in figure 1. This figure shows the inter-subject variability of 15 muscle force tracings in gray with the mean of these measurements shown in black. In this figure there are five sections separated by a vertical line. The first and last sections represent force generation when the muscles are not contracting. The second and fourth sections represent the force generated during an isometric contraction, with the third section denoting the stretch shortening contraction, which also contains a isometric component.

We have 84 such force tracings, and wish to model the isometric and stretch shortening force generation. The data are defined as follows: for an individual measurement, 565 evenly spaced functional observations were taken. These measurements were taken twice (pre- and post-exercise protocol) resulting in $2 \times 565 = 1130$ functional measurements per animal. All 42 animals (27 old and 15 young) underwent the same resistive exercise protocol resulting in 47, 460 total measurements.

As the observed function is thought to be the product of two unknown functions (the isometric and stretch shortening components) defined by ordinary differential equations (ODEs), the individual components are not directly identifiable using standard functional approaches that allow each component to be essentially any continuous function. Consequently they must be modeled using simplifying assumptions, but such assumptions may not be interpretable in terms of muscle physiology. Our approach allows these assumptions to be based on scientific considerations while retaining the flexibility of the functional approach. Parametric models based upon ODEs do exist but are known to be inadequate for characterizing muscle force tracings (Herzog, 2004). We develop Bayesian nonparametric methods that favor shapes consistent with these parametric models, while being flexible enough to account for deviations from parametric assumptions. As we are primarily interested in mean differences between groups, we further extend these methods to a hierarchical setting to allow functional ANOVA style comparisons.

1.1 Skeletal Muscle Force

The force generated by muscle activation, illustrated in Figure 1, is nonlinear (Maffioletti, 2010; Parsaei and Stashuk, 2011) and is associated with complex physiology, such as motor systems and muscle twitch dynamics. The current lack of accurate statistical models for characterizing force tracings has made effective statistical comparisons challenging.

Models for isometric force measurements date back to Hill (1938). A popular approach uses first order differential equations relating muscle force output to a series of motor, damper, spring systems (Wexler et al., 1997; Ding et al., 1998; Phillips et al., 2004). Such models may reasonably describe areas of observed data across the force activation curve but do not represent important aspects of the response. Other modeling approaches (Geronilla et al., 2006) attempt to characterize the response curve using a time-varying combination of basis functions, leading to improvements in prediction but a lack of interpretability and accommodation of prior mechanistic knowledge.

In an effort to develop better training/rehabilitation protocols tailored to individual needs, recent studies have investigated how age affects muscle adaptation and maladaptation following specific non-injurious, repetitive, resistance-type loading protocols designed to induce increases in performance and muscle mass. Initial investigations (Cutlip et al., 2006; Murlasits et al., 2006) and subsequent validations (Ryan et al., 2008; Baker et al., 2010; Hollander et al., 2010) have supported the use of supramaximal, electrically-evoked stretch-shortening contractions precisely prescribed for inducing adaptation (increases in performance and muscle mass) in young animals following repetitive exposures of resistive muscle contractions. We use such data to study the effects of age on resistive muscle training sessions to better understand the benefits/harm of training across age groups.

Complexities arise when modeling the force tracings of a stretch-shortening contraction. The force output is a product of the isometric force at time t and a function related to joint movement. That is, the total force $h(t)$ measured at time t is thought to be

$$h(t)=Q(t)F(t), \quad (1)$$

where $F:\mathbb{R}^+ \rightarrow \mathbb{R}^+$ is the isometric force at time t and $Q:\mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a function representing the increase ($1 < Q(t)$) or decrease ($0 < Q(t) < 1$) in isometric muscle force generation during a stretch shortening contraction. Scientific interest focuses on differences in $Q(t)$ and $F(t)$ across experimental conditions. Interest in $F(t)$ stems from the fact it is the ‘baseline’ force produced by the muscle. Differences are related to the general health of the muscle. Interest in $Q(t)$ is based upon the fact that it represents the potential force that is released when the muscle moves; differences here relate to the ability of the muscle to adapt. Our focus is on developing nonparametric Bayesian methods that incorporate prior information using ODEs that can estimate both $Q(t)$ and $F(t)$ using minimal assumptions.

1.2 Relevant Literature

We are interested in modeling $F(t)$ and $Q(t)$ using scientific information based upon ODEs. From a Bayesian perspective there is a growing literature on parameter estimation for

models derived from ODEs. We consider parametric and nonparametric approaches that are applicable to our problem.

If one assumes muscle force activation is well described by a known ODE there are several recent contributions on Bayesian parameter estimation in this context. Lunn et al. (2002) develops a framework for pharmacokinetic/pharmacodynamic models. Putter et al. (2002) develop methods based on partial differential equations to estimate HIV infection, and Huang et al. (2006) develop a hierarchical framework to investigate the antiviral response for HIV infection in a population of individuals. To compare different ODEs Vysheirsky and Girolami (2008) investigated Bayes factors between several nonlinear ODE models.

Other parametric work focuses on novel estimation methodologies for non-linear differential equations. For example Calderhead and Girolami (2011) use recent advances in Hamiltonian Monte Carlo methods on Riemann manifolds (Girolami and Calderhead, 2011) to model biochemical processes described by non-linear differential equations. Other authors (Ramsay et al., 2007; Calderhead et al., 2009) develop novel techniques for estimating non-linear differential equations utilizing smoothing based approximations. Though these methodologies explore approaches for modeling data using parametric differential equations, they still assume the data are well characterized by a fully parametric model, which is known to be incorrect for our problem.

From a nonparametric perspective the problem could be approached using a Gaussian process (GP) emulator (Kennedy and O'Hagan, 2000, 2001). In the first stage, a solver is used to obtain the differential equation solution on a finite grid of points. Then, uncertainty and bias are accommodated in the second stage through centering a GP prior on the differential equation solution. Though mechanistic information can be included in the mean function, it is not included in the covariance of the Gaussian process; hence realizations from the GP may be unrealistic and not resemble different ODE solutions. Further the variability in the curves may be inadequately represented, with credible intervals around the estimated curve being quite unrealistic. Our goal is to develop hierarchical Gaussian processes that inherit the behavior of the ODE through the covariance kernel, while also allowing variability among individual trajectories across subjects.

Recent work based upon stochastic differential equations uses differential equations to model data, like the GP emulator approach, but embeds mechanistic information into a stochastic process through stochastic differential equations. In these approaches the forcing function for a differential equation is defined by a stochastic process (the forcing function is the function $r(t)$ such that $Lh(t) = r(t)$ for some differential operator L). This induces a stochastic process over the observed curve that inherits the behavior of the differential equation. Golightly and Wilkinson (2006), Golightly and Wilkinson (2011) and Zhu et al. (2011) are examples of approaches based on diffusions. These methodologies assume a white noise process over the latent forcing function $r(t)$, which is not appropriate as we assume a smooth $r(t)$.

Most similar to our approach is the approach of Lawrence et al. (2007); Álvarez et al. (2009); Honkela et al. (2010); Álvarez et al. (2013), who develop latent force models that

embed mechanistic information into a GP prior while assuming a smooth forcing function. The embedding is accomplished by defining a novel covariance kernel specified as the double integral product of a squared exponential kernel and the Green's function corresponding to the linear ODE of interest. These models can be extended to allow for a limited hierarchical specification by including independent processes for each output. Though similar to our approach, the nature of our problem make it difficult to use directly. Other authors have considered the case in which mechanistic information is used to define stochastic processes over non-linear differential equations (Hartikainen et al., 2012; Titsias et al., 2012). These methods develop approximate sampling algorithms specific to the non-linear case, and do not consider methods for linear ODEs, which is the case we are faced with.

By examining the reproducing kernel Hilbert space (*RKHS*), the latent forcing approach can be shown to be similar to the smoothing approach of Heckman and Ramsay (2000) (see chapters 20 and 21 of Ramsay (2006)). Here a noise process governs the behavior of the latent forcing function, as compared to a squared exponential kernel in the latent forcing approach. In this smoothing approach, the equivalent covariance kernel was extremely ill-conditioned for n as small as 20. In our problem we consider linear ODEs with a smooth forcing function, but have observed similar behavior for large n , which may be caused by the fineness of the temporal sampling between observations.

Alternatively to finding the analytical solution for the convolution one can use approximation methods to solve the ODE. One approach relies on the MAP-Laplace approximation (Gao et al., 2008), which finds the maximum *a posteriori* (MAP) estimate of the latent forcing function and then uses a Taylor approximation at that point. We develop an alternative approach that relies on accurately approximating solutions to the differential equations directly using a Runge-Kutta approximation. We name this process the Mechanistic Hierarchical Gaussian process to differentiate it from the latent force methodology as it can be used on an arbitrary covariance kernel without further derivation; the approximation avoids the quite complex direct analytical solution required when using the latent force approach. We also extend this method using the hierarchical Gaussian process (Behseta et al., 2005) which allows for information sharing between observations. By using the hierarchical Gaussian process we model individual experimental group effects as well as individual subject effects.

2. MECHANISTIC GAUSSIAN PROCESS

Consider modeling an unknown functional response $h: \mathcal{T} \rightarrow \mathbb{R}$, with $\mathcal{T} = [t_0, t_1] \in \mathbb{R}$ and data consisting of error-prone measurements (y_1, \dots, y_n) of h at locations (t_1, \dots, t_n) . A common approach lets

$$y(t_l) = h(t_l) + \epsilon_l, \quad (2)$$

where $h(t) \sim \mathcal{GP}(0, R(\cdot, \cdot))$, a zero mean GP with covariance kernel $R(\cdot, \cdot)$, and $\epsilon_l \stackrel{iid}{\sim} N(0, \tau^{-1})$, with $l = 1, \dots, n$. The covariance kernel $R(\cdot, \cdot)$ is frequently chosen as squared exponential, exponential, Matérn or some default form that leads to flexible

realizations. Although prior information about h can potentially be incorporated through the mean of the Gaussian process and choice of the covariance kernel, it can be difficult to choose appropriate values in practice.

We incorporate prior information by defining a covariance kernel favoring shapes consistent with mechanistic information specified by differential equations. We assume the information is expressible in the form of a linear ODE

$$Lh(t) = \frac{d^m h(t)}{dt^m} + a_{m-1}(t) \frac{d^{m-1} h(t)}{dt^{m-1}} + \dots + a_1(t) \frac{dh(t)}{dt} + a_0(t) h(t) = r(t), \quad (3)$$

where $\{a_0(t), \dots, a_{m-1}(t)\}$ are known non-zero functions on τ . Under these conditions the solution to (3) exists and, given initial values, can be expressed as

$$h(t) = \int_{t_0}^t G(t, \xi) r(\xi) d\xi. \quad (4)$$

Here $G(t, \xi)$ is Green's function that is used to explicitly calculate the covariance kernel for latent force models (Lawrence et al., 2007; Álvarez et al., 2009; Honkela et al., 2010; Álvarez et al., 2013). The integral operator (4), described in shorthand as \mathbf{G} , is a linear operator, and is the inverse of the differential operator L in (3). As \mathbf{G} is linear, if $r(t) \sim \mathcal{G} \mathcal{P}(0, R(\cdot, \cdot))$, then $h(t)$ is also a Gaussian process with a covariance kernel dependent on \mathbf{G} and $R(\cdot, \cdot)$. This defines a GP over h whose covariance kernel favors shapes consistent with (3).

One can alternatively look at (4) from a process-convolution perspective (Higdon, 1998, 2002; Álvarez et al., 2012) Here the covariance kernel can be seen to be derived from the convolution of the Green's function related to the ODE of interest and the covariance kernel of the latent forcing function. From this perspective it can be viewed as developing a kernel specific to the needs of the problem.

In our case, due to the fineness of temporal sampling, the exact solution to the resulting covariance matrix is often extremely ill-conditioned resulting in computational instability. We tried a wide variety of existing methods for addressing ill-conditioning problems in GP regression with no success. The induced covariance of $h(t)$ tends to be substantially more subject to ill-conditioning than even the squared exponential covariance. This is because one is compounding the ill-conditioning of the covariance matrix defined over $r(t)$ with a ODE, which for some initial values may also be extremely ill-conditioned (see chapter 9 of Asaithambi (1995)). We avoid the exact solution and rely on a Runge-Kutta approximation (Asaithambi, 1995) that allows direct modeling of $r(t)$ for an arbitrary covariance kernel $R(\cdot, \cdot)$. In our experience this aids in computation while bypassing the cumbersome calculations necessary to compute the covariance kernel.

2.1 Approximation of the Process

There is a large literature on approximate solutions to differential equations. Given a set of initial conditions corresponding to $h(t_0)$ as well as the first $m-1$ derivatives of h evaluated at the initial point t_0 , Runge-Kutta (RK) methods (see chapter 9 of Asaithambi (1995)) offer

efficient algorithms that approximate the solution to an m^{th} order ODE. When L is linear, RK methods express the numerical solution to the ODE as a linear combination of the forcing function $r(t)$ evaluated at a finite set of points, $\{t_l\}_{l=1}^n$, along with the initial conditions $\mathbf{h}^*=(h_1^*, \dots, h_m^*)'$. We illustrate the approach using the Euler-Cauchy second order approximation, though other RK approximations proceed in much the same way.

The Euler-Cauchy approximation recursively defines a solution to $h(t)$ at $\{t_l\}_{l=1}^n$, by approximating the function as a linear combination of $\mathbf{r}=(r(t_1), \dots, r(t_n))'$ and \mathbf{h}^* . As an example, consider a first order differential equation (i.e., $m = 1$ in (3)) where points are equally spaced with $\Delta = 2(t_j - t_{l-1})$. The approximate Euler-Cauchy solution is formed recursively by:

$$\hat{h}_l = h_{l-1} + \Delta \{g(t_{l-1}, h_{l-1})\} \quad (5)$$

$$h_l = h_{l-1} + \frac{\Delta}{2} \{g(t_{l-1}, h_{l-1}) + g(t_l, \hat{h}_l)\}. \quad (6)$$

Here $g(t_{l-1}, h_{l-1})$ is a function of the derivative evaluated at t_{l-1} and h_{l-1} for $l > 1$ (e.g., for (3) with $m = 1$ one has $g(t_{l-1}, h_{l-1}) = [r(t_{l-1}) + a_0(t_{l-1})h_{l-1}]$ where the function $a_0(\cdot)$ is a known function described in (3)). As long as $g(t, f)$ is linear, the approximation is a linear function of $r(t)$ and the initial conditions \mathbf{h}^* . Consequently the solution can alternatively be expressed as a product of a matrix \mathbf{G} and a vector of elements $\mathbf{r}^*=(\mathbf{h}^*, \mathbf{r})'$. We form the matrix recursively with row l corresponding directly to each function evaluation described above. Continuing with the example, one defines the matrix \mathbf{G} as follows: first set row $l = 1$ to $(1 \ 0 \ \dots \ 0)$, which corresponds to the initial condition specified by h_1^* . For $l > 1$ the approximation proceeds by specifying a row vector

$$\hat{\mathbf{G}}_{\{l,:\}} = [1 + \Delta a_0(t_{l-1})] \mathbf{G}_{\{l-1,:\}} + \hat{K}.$$

Here \hat{K} is a row vector of zeros except at the entry l , which is set to Δ , and $\mathbf{G}_{\{l-1,:\}}$ is the previous row. One then defines row l of \mathbf{G} as

$$\mathbf{G}_{\{l,:\}} = \mathbf{G}_{\{l-1,:\}} + \frac{\Delta}{2} [a_0(t_{l-1})\mathbf{G}_{\{l-1,:\}} + a_0(t_{l-1})\hat{\mathbf{G}}_{\{l,:\}}] + K,$$

where K is a row vector of zeros except at entries l and $l + 1$, which are set to $\frac{\Delta}{2}$. Through this alternate expression one arrives at the approximation $h(t) \approx \mathbf{G}\mathbf{r}^*$. This can be seen as approximating $h(t)$ in the context of a linear regression where \mathbf{r}^* is unknown and \mathbf{G} is a known regression matrix.

Though we describe the method using the Euler-Cauchy approximation (an $O(\Delta^2)$ approximation method), a similar \mathbf{G} matrix can be constructed using higher order RK methods. Higher order methods form better approximations but require more functional

evaluations of $r(t)$. This may require $r(t)$ to be evaluated at points on the index set that have not been observed, often greatly increasing the computational complexity when sampling from the posterior. Before implementation, this trade off between accuracy and computational efficiency should be evaluated, as in many situations a lower order approximation is adequate.

One should also investigate the appropriate step size τ , which also controls the accuracy of the approximation. Too large a step size may result in unacceptable approximation error, and too small step sizes result in unnecessary computational overhead. In our example, with the Euler-Cauchy approximation, our step size was $\tau = (1/260)$. Under these conditions numerical experiments produced results that had a maximum absolute difference on the order of approximately 10^{-3} between the actual and numerical solution across the entire curve.

2.2 Posterior Sampling

For the above approximation, sampling from the mechanistic GP proceeds using a series of conditionally conjugate Gibbs steps. The discussion assumes model (2) with $\mathbf{Y} \sim N(\mathbf{h}, \tau^{-1}\mathbf{I})$ where \mathbf{I} is a $(n \times n)$ identity matrix, $\mathbf{Y} = (y_1(t_1), \dots, y_n(t_n))'$ and $\mathbf{h} = (h(t_1), \dots, h(t_n))'$, with $\tau \sim Ga(a_0, b_0)$. From the above discussion we define $\mathbf{A} = (a_1, \dots, a_z)'$ to be a $z \times 1$ vector of parameters used in the functions $\{a_1(t), \dots, a_{m-1}(t)\}$. These parameters can be specified to be known constants or can be estimated. If the parameters are estimated, in most cases a conjugate prior does not exist and a Metropolis within Gibbs step is necessary. Consequently, in the algorithms that follow we do not specify an exact prior for \mathbf{A} . We assume the initial conditions \mathbf{h}^* are specified as $\mathbf{h}^* \sim N(\mathbf{A}_0, \mathbf{B}_0)$, which are independent of $r(t)$.

Sampling algorithm 1

1. Sample $\mathbf{r}^* \sim N(\mathbf{E}, \mathbf{W})$, where $\mathbf{W} = (\tau\mathbf{G}'\mathbf{G} + \mathbf{\Omega}^{-1})^{-1}$ and $\mathbf{E} = \mathbf{W}(\tau\mathbf{G}'\mathbf{Y} + \mathbf{\Omega}^{-1}\rho)$, where ρ is the prior mean of \mathbf{r}^* . Further $\mathbf{\Omega} = \text{block-diag}(\mathbf{B}_0, \Sigma)$, which is a $(n + m) \times (n + m)$ block diagonal matrix defined using \mathbf{B}_0 and Σ . Here \mathbf{B}_0 is the prior variance of \mathbf{h}^* and Σ is the $n \times n$ covariance matrix for \mathbf{r}^* .
2. Sample τ from $Ga(a_0 + n/2, b_0 + (\mathbf{Y} - \mathbf{G}\mathbf{r}^*(\mathbf{Y} - \mathbf{G}\mathbf{r}^*)/2)$.
3. Marginalizing out \mathbf{r}^* , i.e., $\mathbf{Y} \sim N(0, \mathbf{G}\mathbf{\Omega}\mathbf{G}' + \tau^{-1}\mathbf{I})$ as above \mathbf{I} is a $(n \times n)$ identity matrix, block sample the parameters \mathbf{A} using a Metropolis-Hastings step.

Gibbs sampling proceeds by iterating over the above steps. To keep the algorithm general, we do not provide an explicit description of sampling of the GP hyperparameters in the above algorithm. However, we have observed the best mixing when these hyperparameters are block sampled with \mathbf{A} .

3. ADAPTATION TO MUSCLE FORCE APPLICATION

The mechanistic GP is not directly applicable to the muscle force application, which has the additional complication of decomposing $h(t)$ as

$$h(t) = \begin{cases} F(t) & t \notin [t_a, t_b] \\ F(t)Q(t) & t \in [t_a, t_b] \end{cases} \quad (7)$$

where $F(t)$ and $Q(t)$ are describable through first and second order differential equations, respectively. Additional constraints are needed to separately identify $F(t)$ and $Q(t)$. For $F(t)$, shape constraints are needed that rule out Gaussian processes, so we use restricted splines. As $Q(t)$ is known to equal one at the beginning and end of the stretch shortening contraction, we modify the GP to include this information. In what follows, we describe the individual ODEs governing $F(t)$ and $Q(t)$ and outline an extension of the posterior sampling algorithm of Section 2.2.

3.1 Prior Extended to Muscle Force Data

We define an ODE for $F(t)$ and $Q(t)$ using generalizations of models from the muscle force literature. The isometric force function $F(t)$ is historically related to the first order differential equation (Hill, 1938)

$$\frac{dF(t)}{dt} - BF(t) + p(t) = 0. \quad (8)$$

Here B represents the damping constant of the muscle fibers, and $p(t)$ corresponds to the joint action of muscle at time t . We assume that the form of the motor activation function is unknown but is linear shortly after activation.

Placing a linearity assumption on $p(t)$ only during the SS contraction, we let

$$p(t) = \sum_{s=0}^S \beta_s b_s(t)$$

where $b_0(t) = 1$ and $\beta_s \sim N(a, b)$. Also, we let $b_s(t)$ for $s = 1$ be defined as piecewise polynomial splines on the interval $\mathcal{I}_s = [\tau_{s-1}, \tau_{s+1}]$. For all $s = s'$, which are defined outside of the SS contraction, we use cubic splines defined to be 0 prior to the interval and 1 after the interval. For the interval including the SS contraction we let $b_{s'}$ be a linear spline on the interval, 0 prior to, and 1 after the SS contraction, which defines a linear interpolation of the function on this range. In order to model a flexible curve we use a large number of evenly spaced splines to estimate $p(t)$. This spacing resulted in 35 splines outside of the SS contraction, except during the SS contraction where one linear spline is used to interpolate the area in this region. Equivalent spline spacing was tested on a large number of isometric force tracings (tracings without a SS component) where it was more than adequate in modeling all observations.

When the joint is moved through its range of motion, the force on the joint is related to the angle of the joint and other factors. Angular motion is often described using a second order differential equation, and we follow this approach. As the exact form of the differential equation is unknown (damping constant) we model this function through the fully specified second order differential equation:

$$\frac{d^2 Q(t)}{dt^2} + \lambda \frac{dQ(t)}{dt} - A Q(t) + g(t) = 0, \quad (9)$$

where $g(t) \sim \mathcal{G} \mathcal{P}(0, R(\cdot, \cdot))$, $A > 0$, and the damping constant $\lambda \geq 0$. When $g(t) = 0$ this defines a periodic function having a period of $\pi \sqrt{A}$.

It is further known that the multiplicative effect of $Q(t)$ should be 1 prior to and after joint movement. We add the constraint that at the beginning t_a and end t_b of the stretch shortening contraction, $Q(t_a) = Q(t_b) = 1$. One can easily sample from this using the conditional properties of the multivariate normal distribution.

3.2 Posterior Sampling Extensions

The RK approximation is used to sample both $F(t)$ and $Q(t)$. Analogous to \mathbf{h}^* above, we define $\mathbf{F}^* = (F_0)' \sim N(\mathbf{A}_0, \mathbf{B}_0)$ and $\mathbf{Q}^* = (Q_0, Q_1)' \sim N(\mathbf{A}_1, \mathbf{B}_1)$, initial value vectors for $F(t)$ and $Q(t)$, respectively. Similarly let $\mathbf{p} = (p(t_1), \dots, p(t_n))'$, and $\mathbf{g} = (g(t_1), \dots, g(t_n))'$, which are vectors of the latent forcing functions evaluated at a finite set of points for $F(t)$ and $Q(t)$, respectively. Finally define $\mathbf{P}^* = (\mathbf{F}^{*'}, \mathbf{p}')$ and $\mathbf{g}^* = (\mathbf{Q}^{*'}, \mathbf{g}')$. For convenience we refer to \mathbf{G} as the Euler-Cauchy approximation to either (8) or (9). For all references to $F(t)$, \mathbf{G} is the solution to (8), and for all references to $Q(t)$, \mathbf{G} is the solution to (9).

In sampling $F(t)$ one has $\mathbf{p} = \mathbf{X}\beta$, where \mathbf{X} is the $n \times (S + 1)$ matrix of spline basis functions $\{b_s(t)\}_{s=0}^S$ evaluated at (t_1, \dots, t_n) and $\beta = (\beta_0, \beta_1, \dots, \beta_S)'$. Letting $\beta \sim N(0, \Sigma_\beta)$, step 1 of sampling algorithm 1 is modified as follows:

Sampling algorithm 2

1. Putting the prior $\mathbf{F}^* \sim N(\mathbf{A}_0, \mathbf{B}_0)$ over the initial conditions, define $\mathbf{V} = \mathbf{G}\mathbf{X}$, $\rho = (\mathbf{A}_0' 0)'$, and $\Omega = \text{block-diag}(\mathbf{B}_0, \Sigma_\beta)$. Then sample $\mathbf{P}^* \sim N(\mathbf{E}, \mathbf{W})$, where $\mathbf{W} = (\tau \mathbf{V}'\mathbf{V} + \Omega^{-1})^{-1}$ and $\mathbf{E} = \mathbf{W}(\mathbf{V}'\mathbf{Y} + \Omega^{-1}\rho)$.

We also modify algorithm 1 to sample \mathbf{g}^* given $Q(t_a) = Q(t_b) = 1$. This is done using the conditional properties of the multivariate normal distribution, i.e., for

$$\begin{bmatrix} \mathcal{X}_1 \\ \mathcal{X}_2 \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{bmatrix} \right),$$

one has

$$\mathcal{X}_1 | \mathcal{X}_2 \sim N(\mu_1 - C_{12}C_{22}^{-1}[\mu_2 - \mathcal{X}_2], C_{11} - C_{12}C_{22}^{-1}C_{21}) \quad (10)$$

In the above approximation $Q(t_a) = Q_0$ and $Q(t_b) = \mathbf{G}_{\{n, \cdot\}}\mathbf{g}^*$, where $\mathbf{G}_{\{n, \cdot\}}$ is the last row of \mathbf{G} . We modify step one of sampling algorithm 1 as follows:

Sampling algorithm 3

1. Calculating $\mathbf{g}^* \sim N(\mathbf{E}, \mathbf{W})$ as in algorithm 1, define the following quantities

$$\mathbf{E}^* = \begin{bmatrix} \mathbf{I} \\ \mathbf{G}_{\{n,:\}} \end{bmatrix} \mathbf{E}, \quad \mathbf{W}^* = \begin{bmatrix} \mathbf{I} \\ \mathbf{G}_{\{n,:\}} \end{bmatrix} \mathbf{W} [\mathbf{I} \ \mathbf{G}_{\{n,:\}}].$$

Then sample $g^*|Q(t_a)Q(t_b)$ from a normal distribution whose mean and covariance are derived from \mathbf{E}^* and \mathbf{W}^* as in (10).

On the interval $[t_a, t_b]$, sampling $Q(t)$ and $F(t)$ proceeds conditional on knowledge of the other. To sample $F(t)$ one uses algorithm 2 and multiplies each row of \mathbf{G} by the corresponding value of $Q(t)$ (i.e., for row l one multiplies each element in this row by $Q(t_l)$). Similarly we multiply by $F(t)$ when sampling $Q(t)$ and sample from algorithm 3.

4. HIERARCHICAL MECHANISTIC GAUSSIAN PROCESS

We extend the mechanistic GP to hierarchical modeling (Behseta et al., 2005). This allows modeling of individual curves as well as population means. The extension is described in terms of our application but can be readily used in other settings.

Consider a study in which there is a single factor of interest having I levels. For subject j a functional response $h_{ijk}: \mathcal{T} \rightarrow \mathbb{R}$ is measured K times. In our application the factor is age, $I = 2$, $K = 2$ and represents measurements pre and post exercise routine, and $h_{ijk}(t)$ is the time varying force function. Here, for all i, j, k , the n functional measurements are taken at equally spaced points on the index set \mathcal{T} . Data are modeled as:

$$y_{ijk}(t_l) = h_{ijk}(t_l) + \epsilon_{ijkl},$$

where $\epsilon_{ijkl} \stackrel{iid}{\sim} N(0, \tau_j^{-1})$, and a mechanistic Gaussian process prior is defined over $h_{ijk}(t)$ as in (3).

For subject j , in group i , the pre and post functional measurements are modeled as

$$h_{ijk}(t) = \tilde{h}_{ij1}(t)1(k \geq 1) + \tilde{h}_{ij2}(t)1(k \geq 2), \quad (11)$$

where $1(\cdot)$ is an indicator function that takes the value of 1 if the argument is true, and 0 otherwise. This additive form allows modeling of the muscle force measurements pre- ($k = 1$) and post- ($k = 2$) exercise protocol, with changes in the muscle force output due to the exercise routine modeled through $\tilde{h}_{ij2}(t)$. In terms of the mechanistic GP one models the latent forcing function as

$$r_{ijk}(t) = \tilde{r}_{ij1}(t)1(k \geq 1) + \tilde{r}_{ij2}(t)1(k \geq 2). \quad (12)$$

To get an equivalent expression as in (11) one integrates (12) using (4). For interpretability between observations and groups we use the same integral operator \mathbf{G} across all levels of the hierarchy.

Extending (12) to account for variability between factors we define

$$\tilde{r}_{ijk}(t) \sim \mathcal{G} \mathcal{D}(r_{ik}^{(1)}(t), R_{ik}^{(1)}(\cdot, \cdot))$$

$$r_{ik}^{(1)}(t) \sim \mathcal{G} \mathcal{D}(r_k^{(2)}(t), R_k^{(2)}(\cdot, \cdot)),$$

with $k = 1, 2$ as in (12) and $r_k^{(2)}(t) \sim \mathcal{G} \mathcal{D}(0, R_k^{(3)}(\cdot, \cdot))$. Sampling from this hierarchy proceeds in much the same way as algorithm 1. Analogous to the case of the single curve we define $\tilde{\mathbf{r}}_{ijk}^*$, $\mathbf{r}_{ik}^{(1)*}$, and $\mathbf{r}_{ik}^{(2)*}$ to be vectors representing the latent forcing function and subsequent hierarchies evaluated at sampled points. We define the individual vector of observations $\mathbf{Y}_{ijk} = (y_{ijk}(t_1), \dots, y_{ijk}(t_n))'$. Sampling from the posterior is specified in terms of $\tilde{\mathbf{r}}_{ijk}^*$, $\mathbf{r}_{ik}^{(1)*}$, and $\mathbf{r}_{ik}^{(2)*}$ and proceeds as follows:

Sampling algorithm 4

1. For each i, j, k sample $\tilde{\mathbf{r}}_{ijk}^*$, conditionally on $\tilde{\mathbf{r}}_{ijk}^*$ where $k' = 1$ if $k = 2$ and $k' = 2$ otherwise. Here let $\mathbf{Y}^* = (\mathbf{Y}_{ijk} - \mathbf{G}\tilde{\mathbf{r}}_{ijk}^*)$ and sample $\tilde{\mathbf{r}}_{ijk}^* \sim N(\mathbf{E}, \mathbf{W})$ where $\mathbf{W} = (\tau \mathbf{G}'\mathbf{G} + \Omega_{ijk}^{-1})^{-1}$, $\mathbf{E} = \mathbf{W}(\tau \mathbf{G}'\mathbf{Y}^* + \Omega_{ijk}^{-1}\mathbf{r}_{ik}^{(1)*})$. Here, as in sampling algorithm 1, Ω_{ijk} is a subject specific $(n+m) \times (n+m)$ covariance matrix.
2. For each i, k pair, sample $\mathbf{r}_{ik}^{(1)*} \sim N(\mathbf{E}, \mathbf{W})$, where in this case $\mathbf{W} = (\sum_j \Omega_{ijk}^{-1} + [\Omega_{ik}^{(1)}]^{-1})^{-1}$ and $\mathbf{E} = \mathbf{W} \left(\sum_j \tilde{\Omega}_{ijk}^{-1} \tilde{\mathbf{r}}_{ij}^* + [\Omega_{ik}^{(1)}]^{-1} \mathbf{r}_{ik}^{(2)*} \right)$. Again $\Omega_{ik}^{(1)}$ is an $(n+m) \times (n+m)$ covariance matrix as specified above and $R_{ik}^{(1)}(\cdot, \cdot)$ is used to compute the finite dimensional covariance for the latent forcing function.
3. For $\mathbf{r}_{ik}^{(2)*}$ sample as in step 2 replacing $\tilde{\mathbf{r}}_{ijk}$ with $\mathbf{r}_{ik}^{(1)*}$ etc.
4. For each i, j , sample τ_j from $Ga(a_0 + n, b_0 + \sum_k (\mathbf{Y}_{ijk} - \mathbf{G}\mathbf{r}_{ijk}^*)'(\mathbf{Y}_{ijk} - \mathbf{G}\mathbf{r}_{ijk}^*)/2)$, where $\mathbf{r}_{ijk}^* = \tilde{\mathbf{r}}_{ij1}^* 1(k \geq 1) + \tilde{\mathbf{r}}_{ij2}^* 1(k \geq 2)$.
5. Sample \mathbf{A} similarly to algorithm 1.

Inference on the group average curves $h_{ik}^{(1)}(t)$ and the population average curves $h_k^{(2)}(t)$ proceeds using the approximations $\mathbf{G}\mathbf{r}_{ik}^{(1)*}$ and $\mathbf{G}\mathbf{r}_k^{(2)*}$, and as \mathbf{G} is the same across all i, j, k , the population averages have the same interpretation as other curves in the hierarchy. Extending the above framework by adding more hierarchies, is straightforward. Each additional hierarchy is sampled as in step 2 noting that the previous level is used as the input vector.

4.1 Extensions to the Hierarchical Mechanistic Process

We extend the hierarchical mechanistic process to our application. Here

$$h_{ijk}(t) = \begin{cases} F_{ijk}(t) & t \notin [t_a, t_b] \\ F_{ijk}(t)Q_{ijk}(t) & t \in [t_a, t_b] \end{cases}$$

where the isometric force function $F_{ijk}(t)$ and the stretch shortening multiplier function $Q_{ijk}(t)$ are defined using (8) and (9) respectively. For $F_{ijk}(t)$ and $Q_{ijk}(t)$ we define the hierarchy over the latent forcing function, with the isometric and SS forcing functions $p_{ijk}(t)$ and $g_{ijk}(t)$ specified as in (12). This discussion uses the same notation as above, i.e., the latent functions measured pre and post treatment for the SS contraction and isometric functions are represented as $\tilde{g}_{ijk}(t)$ and $\tilde{p}_{ijk}(t)$ with the corresponding vectors evaluated at sampled points represented as $\tilde{\mathbf{g}}_{ijk}^*$ and $\tilde{\mathbf{p}}_{ijk}^*$.

For $Q_{ijk}(t)$, the forcing functions $\tilde{g}_{ijk}(t)$, $g_{ijk}^{(1)}(t)$, and $g_{ijk}^{(2)}(t)$, are defined such that $Q_{ijk}(t_a) = Q_{ijk}(t_b) = 1$, etc. and these constraints are implemented in exactly the same way as in sampling algorithm 3. To sample $\tilde{\mathbf{g}}_{ijk}^*$, $\mathbf{g}_{ik}^{(1)*}$, $\mathbf{g}_k^{(2)*}$ one proceeds by computing \mathbf{E} and \mathbf{W} as in sampling algorithm 4 and then sampling from the conditionally conjugate distribution specified in sampling algorithm 3.

Hierarchical extensions in modeling $F_{ijk}(t)$ are direct. Using the same hierarchical notation, we place multivariate normal hierarchies over the spline coefficient vector β

$$\tilde{\beta}_{ijk} \sim N(\beta_{ik}^{(1)}, \sum_{\beta, ik}^{(1)})$$

$$\beta_{ik}^{(1)} \sim N(\beta_k^{(2)}, \sum_{\beta, k}^{(2)}),$$

which in turn defines $\tilde{p}_{ijk}(t)$, $p_{ik}^{(1)}(t)$, and $p_k^{(2)}(t)$. Sampling each $\tilde{\mathbf{p}}_{ijk}^*$, $\mathbf{p}_{ik}^{(1)*}$, and $\mathbf{p}_k^{(2)*}$ proceeds by placing the modifications of sampling algorithm 2 into sampling algorithm 4.

5. SIMULATION

We conduct a simulation experiment based upon the model developed in (7). Curves similar to those expected in a muscle force application are generated, and the simulated curves are compared against posterior estimated curves. Mirroring the muscle force application, the hierarchy was generated assuming $I = 2$, $J = 20$ and $K = 2$. The group levels of the hierarchy, $F_{ik}^{(1)}(t)$ and $Q_{ik}^{(1)}(t)$, were simulated to resemble muscle force tracings of isometric and stretch shortening contractions based upon (8) and (9). The individual level data were

generated at 565 equally spaced points, assuming $\Delta = \frac{1}{260}$. Here the first 80 observations represent the force tracing prior to muscle activation. After activation 120 observations were taken of $F_{ijk}(t)$. The next 201 observations were of $F_{ijk}(t)Q_{ijk}(t)$ with the 164 remaining

observations generated from $F_{ijk}(t)$. Similar to the real data, all data were generated assuming little variability between observations; here $\tau_j = 300$ for all observations.

We chose weakly informative priors for all hyper parameters. We place a GP prior over $g_{ijk}(t)$, where the covariance kernel is specified using the squared exponential kernel $K(t, t') = \sigma^2 \exp(-\ell \|t - t'\|^2)$. We set $\sigma^2 \sim Ga(1, 1)$ and let $\ell \sim Ga(100, 0.1)$, which reflects the assumption that $g_{ijk}(t)$ is not expected to be very smooth. The same assumptions are made for all other levels of the hierarchy. For $p_{ijk}(t)$ we put normal priors over the β coefficients, with diffuse priors specified at the topmost level. The precision parameter for all other levels was assigned a $Ga(0.1, 0.1)$ prior. Finally the precision parameter τ_j was specified using a $Ga(100, 0.1)$ prior. This is a vague prior on τ_j centered approximately at the observed error found in muscle force tracings. For the parameters in (8) and (9) we defined discrete uniform priors over a range of plausible values. Here B is defined to be in $[4.1, 7.5]$, based upon analyses of isometric data with a parametric model. Further the parameter A , which defines the period of $Q_{ijk}(t)$, is put in the range of $[-3.5, -0.6]$. This choice corresponds to a range representing a half a period to a period and a half. Finally the damping constant was expected to be negligible, and λ was given a plausible range of $[0.01, 2]$. Note λ can not take on values at 0 due to the identifiability constraints on the ODE.

We collected 50, 000 MCMC samples disregarding the first 5, 000 as a burn-in. Examinations of trace plots for the quantities of interest, the individual curves and curves in the hierarchy, showed adequate mixing. Geweke's diagnostic tests on the chains (Geweke, 1992) indicated convergence. Further the effective sample size was calculated to be 1, 000 or greater for all quantities of interest (hierarchical means had effective sample sizes greater than 30, 000). Hyperparameters for the covariance kernel as well as the parameters specified in (8) and (9) converged as evident in trace plots as well as Geweke's diagnostic, but mixing was slow, with effective sample sizes were typically calculated to be around 400.

For the quantities of interest (i.e., the individual level functions and the first level of the hierarchy $F_{ijk}(t)$, $Q_{ijk}(t)$, $F_{ik}^{(1)}(t)$, and $Q_{ik}^{(1)}(t)$ respectively, which represents 95 total curves), the true curve was within the 95% credible region at the specified level for these curves.

Figure 2 shows the estimates of $Q_{i1}^{(1)}(t)$ and $Q_{i2}^{(1)}(t)$, for one of the groups. Here the true curve is given by the gray line, the estimated curve is shown in solid black, and the 95% credible intervals on the central estimate are given by the dotted lines. One can see that the true curve is estimated within these regions. This figure is representative of the other estimates, where truth is well described by the model.

6. MUSCLE FORCE APPLICATION

With the goal of investigating the effect of non-injurious, repetitive muscle contractions on muscle force generation, we apply our approach to data compiled from Cutlip et al. (2006), Murlasits et al. (2006), and Baker et al. (2010). In these studies, 15 young (3 months), and 27 old (30 months), rats' dorsiflexor muscles were exposed to a resistive muscle contraction protocol that included thirteen sessions. At the end of each session the dorsiflexor muscle group underwent isometric as well as stretch shortening contraction (as described in Figure

1). Individual observations were taken at evenly spaced intervals ($\Delta = \frac{1}{260}$ of a second). The entire measurement lasted just over 2 seconds, resulting in 565 total functional observations as in our above simulation study. Our analysis looks at possible differences between muscle force measurements pre- (after the first resistive muscle contraction protocol) and post- (after the last protocol) study, between young and old animals. Priors for all parameters as well as computational implementation were specified as in the simulation.

Figure 3 shows the individual fits of $h_{ijk}(t)$ for one animal's pre- and post- observations. Here the central posterior estimated curve is shown in black, with the observed data shown using gray hash marks. The credible intervals are not shown, as they are too close to the central estimate to be visible. Figure 4 shows the expected mean isometric contraction for the pre- (dashed line) and post-(solid line) exercise protocol in the old animals (top left) and the young animals (top right). The difference (solid line) between the pre- and the post-training, as well as the 95% pointwise credible interval (dashed line), is shown in the bottom row for the old (bottom left) and young (bottom right) animals. Here it is seen that the young and the old animals, as a group, displayed increased muscle performance related to stretch shortening contractions, and the post exercise increase maximum was approximately 1.5 to 2.2 times an increase in force output. In contrast there was no difference in the group average isometric contraction (i.e., $F_{ik}^{(1)}(t)$) for either young or old animals. Figure 5 shows the estimated posterior curves (top) and corresponding differences (bottom) for the young animals. Here the pre-treatment (dotted line) and post-treatment (solid line) estimated isometric contractions are shown in the top row, and, though the central estimates are different, the bottom row shows that there is not enough evidence to suggest differences between the two groups. Similar results (figure not shown) were observed for the older animals.

The model also allows one to look at individual estimates between curves. Though the animals showed no significant differences at the group level for isometric force generation, individual differences were seen. Figure 6 shows the isometric contraction difference for an individual animal. Here the pre-and post-treatment estimates are shown in the top graph with the estimated differences being shown in the bottom graph. Here individual differences are seen, which is significant as it supports the idea that some older animals vary in their physiology related to the overall muscle health.

Note that there is an additional advantage of modeling the latent forcing function as it may be used to generate or support hypotheses. For example, for $F_{ijk}(t)$, the latent forcing function $p_{ijk}(t)$ represents the muscle motor action at time t . These curves (figure not shown) show a steep increase right after activation, a sharp decrease shortly thereafter, and then a stabilization to a near constant level. This is supportive of the idea that large amounts of calcium influx the cytoplasm and bind rapidly to troponin upon muscle activation (steep increase in force tracing), until finally calcium is sequestered from the cytosol upon deactivation (return to baseline in force tracing).

7. DISCUSSION

This article proposes a flexible nonparametric Bayesian method that takes into account prior scientific information based upon an ODE. This method relies on a sampling algorithm using a Runge-Kutta approximation to the ODE. The nature of the approximation allows for a hierarchical specification at the population level estimates by modeling the latent forcing function directly. The goal was to develop an inferential framework for muscle force tracings, and investigate the effect of non-injurious resistive exercise protocols on muscles of different ages (i.e., young and old muscles). It was important to accurately model both the overall functional response and the two constituent functions, which themselves have scientific interest.

For our specific problem the model is able to show that the dynamic force generated through the stretch shortening contraction may be more informative and specific in showing adaptation and maladaptation following non-injurious mechanical loading. Both groups of rats have an adaptive response in dynamic muscle force produced. Force generated is, at the population level, greater after the exercise protocol, but both groups have statistically non-significant responses in isometric force generation. As maximal isometric force generation, which is represented by a single value, has been seen as the gold standard response of interest in the muscle force literature in terms of measuring adaptation/maladaptation. The models presented here may allow researchers to begin to model more sensitive endpoints and look at their data from a functional data perspective. Further, such analyses may lead to new insights. For example, the maximal force for older rats appears to occur at a different joint angle, suggesting impacts affecting the length=tension curve dynamics as the animals age.

The proposed approach can be extended to human muscle force tracings, and may allow for in-depth study of human physiologic responses to exercise routine post training. Such an analysis, on the entire force output, has previously not been attempted. For such a study, age, as well as other variables, can be included in the hierarchical framework. In this manner the efficacy of the current 'one size fits all' approach across the spectrum of prevention/intervention including occupational, athletic, physical therapy, strength and conditioning, and wellness programs can be studied. If similar results are seen in human populations, this might suggest that different treatment protocols are appropriate depending on age or other variables of interest.

In domains outside of the muscle force literature, the approximation method allows researchers to easily incorporate mechanistic information into analyses, while bypassing the cumbersome integrals that are required for direct calculation of the covariance kernel. In situations where there are fewer observations that are not evenly spaced, one can still use this approach. This can be done either by using our regression based approach directly, or, when inference is desired on less closely spaced observations, one can approximate the covariance matrix by computing $\mathbf{G}\Sigma\mathbf{G}'$. Here, as above, Σ is the finite dimensional covariance defined on the latent forcing function. As this matrix will represent a fine grid of points across the domain of interest, one needs only to extract the rows and columns that correspond to locations where inference is desired. Finally, this method can be used where

there is prior knowledge on the approximate shape of the response, but no direct mechanistic information is available. Here, one would tailor a covariance kernel to the problem at hand, which may result in some gains in efficiency for problems where points are sampled at sparse intervals.

Acknowledgments

The authors would like to thank John Bailer and Rich Lovering as well as three anonymous referees for their comments and suggestions on an earlier version of this manuscript.

References

- Álvarez M, Luengo D, Lawrence N. Linear latent force models using Gaussian processes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 2013; 35(11):2693–2705.
- Álvarez M, Luengo-García D, Lawrence N. Latent force models. *Proceedings of the Twelfth International Workshop on Artificial Intelligence and Statistics*. 2009; 5:9–16.
- Álvarez MA, Rosasco L, Lawrence ND. Kernels for vectored-valued functions: a review. *Foundations and Trends in Machine Learning*. 2012; 4(3):195–266.
- Asaithambi, N. *Numerical Analysis: Theory and Practice*. Saunders; College Pub: 1995.
- Baker B, Hollander M, Kashon M, Cutlip R. Effects of glutathione depletion and age on skeletal muscle performance and morphology following chronic stretch-shortening contraction exposure. *European Journal of Applied Physiology*. 2010; 108(3):619–630. [PubMed: 19882165]
- Behseta S, Kass R, Wallstrom G. Hierarchical models for assessing variability among functions. *Biometrika*. 2005; 92(2):419–434.
- Calderhead B, Girolami M. Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods. *Interface focus*. 2011; 1(6):821–835. [PubMed: 23226584]
- Calderhead B, Girolami M, Lawrence ND. Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes. *Advances in Neural Information Processing Systems*. 2009; 21:217–224.
- Cutlip R, Baker B, Geronilla K, Mercer R, Kashon M, Miller G, Murlasits Z, Alway S. Chronic exposure to stretch-shortening contractions results in skeletal muscle adaptation in young rats and maladaptation in old rats. *Applied Physiology, Nutrition, and Metabolism*. 2006; 31(5):573–587.
- Cutlip RG, Baker BA, Geronilla KB, Kashon ML, Wu JZ. The influence of velocity of stretch-shortening contractions on muscle performance during chronic exposure: age effects. *Applied Physiology, Nutrition, and Metabolism*. 2007; 32(3):443–453.
- Ding J, Binder-Macleod S, Wexler A. Two-step, predictive, isometric force model tested on data from human and rat muscles. *Journal of Applied Physiology*. 1998; 85(6):2176–2189. [PubMed: 9843541]
- Gao P, Honkela A, Rattray M, Lawrence ND. Gaussian process modelling of latent chemical species: applications to inferring transcription factor activities. *Bioinformatics*. 2008; 24(16):i70–i75. [PubMed: 18689843]
- Geronilla K, Wu J, Baker B, Cutlip R. Characterization of isometric contractions of rat skeletal muscle in vivo: Duty cycle effects. *Bio-Medical Materials and Engineering*. 2006; 16(6):369–380. [PubMed: 17119276]
- Geweke, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: B, JM.; B, A.; D, AP.; S, AFM., editors. *Bayesian Statistics 4*. 1992.
- Girolami M, Calderhead B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2011; 73(2):123–214.
- Golightly A, Wilkinson DJ. Bayesian sequential inference for stochastic kinetic biochemical network models. *Journal of Computational Biology*. 2006; 13(3):838–851. [PubMed: 16706729]
- Golightly A, Wilkinson DJ. Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo. *Interface Focus*. 2011; 1(6):807–820. [PubMed: 23226583]

- Hartikainen J, Seppanen M, Sarkka S. State-space inference for non-linear latent force models with application to satellite orbit prediction. arXiv preprint arXiv:1206.4670. 2012
- Heckman NE, Ramsay JO. Penalized regression with model-based penalties. *Canadian Journal of Statistics*. 2000; 28(2):241–258.
- Herzog W. History dependence of skeletal muscle force production: Implications for movement control. *Human Movement Science*. 2004; 23(5):591–604. [PubMed: 15589623]
- Higdon D. A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*. 1998; 5(2):173–190.
- Higdon, D. Quantitative methods for current environmental issues. Springer; 2002. Space and space-time modeling using process convolutions; p. 37-56.
- Hill A. The heat of shortening and the dynamic constants of muscle. *Proceedings of the Royal Society of London. Series B, Biological Sciences*. 1938; 126(843):136–195.
- Hollander M, Baker B, Ensey J, Kashon M, Cutlip R. Effects of age and glutathione levels on oxidative stress in rats after chronic exposure to stretch-shortening contractions. *European Journal of Applied Physiology*. 2010; 108(3):589–597. [PubMed: 19882168]
- Honkela A, Girardot C, Gustafson E, Liu Y, Furlong E, Lawrence N, Rattray M. Model-based method for transcription factor target identification with limited data. *Proceedings of the National Academy of Sciences*. 2010; 107(17):7793–7798.
- Huang Y, Liu D, Wu H. Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics*. 2006; 62(2):413–423. [PubMed: 16918905]
- Kennedy M, O'Hagan A. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*. 2000; 87(1):1–13.
- Kennedy M, O'Hagan A. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*. 2001; 63(3):425–464.
- Lawrence ND, Sanguinetti G, Rattray M. Modelling transcriptional regulation using Gaussian processes. *Advances in Neural Information Processing Systems*. 2007; 19:785–792.
- Lunn D, Best N, Thomas A, Wakefield J, Spiegelhalter D. Bayesian analysis of population pk/pd models: General concepts and software. *Journal of Pharmacokinetics and Pharmacodynamics*. 2002; 29(3):271–307. [PubMed: 12449499]
- Maffiuletti N. Physiological and methodological considerations for the use of neuromuscular electrical stimulation. *European Journal of Applied Physiology*. 2010; 110(2):223–234. [PubMed: 20473619]
- Murlasits Z, Cutlip R, Geronilla K, Rao K, Wonderlin W, Alway S. Resistance training increases heat shock protein levels in skeletal muscle of young and old rats. *Experimental Gerontology*. 2006; 41(4):398–406. [PubMed: 16524679]
- Parsaei H, Stashuk D. Adaptive motor unit potential train validation using mup shape information. *Medical Engineering and Physics*. 2011; 33(5):581–589. [PubMed: 21269867]
- Phillips C, Repperger D, Neidhard-Doll A, Reynolds D. Biomimetic model of skeletal muscle isometric contraction: I. an energetic-viscoelastic model for the skeletal muscle isometric force twitch. *Computers in Biology and Medicine*. 2004; 34(4):307–322. [PubMed: 15121002]
- Putter H, Heisterkamp S, Lange J, De Wolf F. A Bayesian approach to parameter estimation in HIV dynamical models. *Statistics in Medicine*. 2002; 21(15):2199–2214. [PubMed: 12210633]
- Ramsay, JO. *Functional data analysis*. Wiley; 2006.
- Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69(5):741–796.
- Ryan M, Dudash H, Docherty M, Geronilla K, Baker B, Haff G, Cutlip R, Alway S. Aging-dependent regulation of antioxidant enzymes and redox status in chronically loaded rat dorsiflexor muscles. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. 2008; 63(10):1015–1026.
- Titsias MK, Honkela A, Lawrence ND, Rattray M. Identifying targets of multiple co-regulating transcription factors from expression time-series by Bayesian model comparison. *BMC systems biology*. 2012; 6(1):53. [PubMed: 22647244]

- Vyshemirsky V, Girolami MA. Bayesian ranking of biochemical system models. *Bioinformatics*. 2008; 24(6):833–839. [PubMed: 18057018]
- Wexler A, Ding J, Binder-Macleod S. A mathematical model that predicts skeletal muscle force. *IEEE Transactions on Biomedical Engineering*. 1997; 44(5):337–348. [PubMed: 9125818]
- Zhu B, Taylor JMG, Song PXX. Semiparametric stochastic modeling of the rate function in longitudinal studies. *Journal of the American Statistical Association*. 2011; 106(496):1485–1495. [PubMed: 22423170]

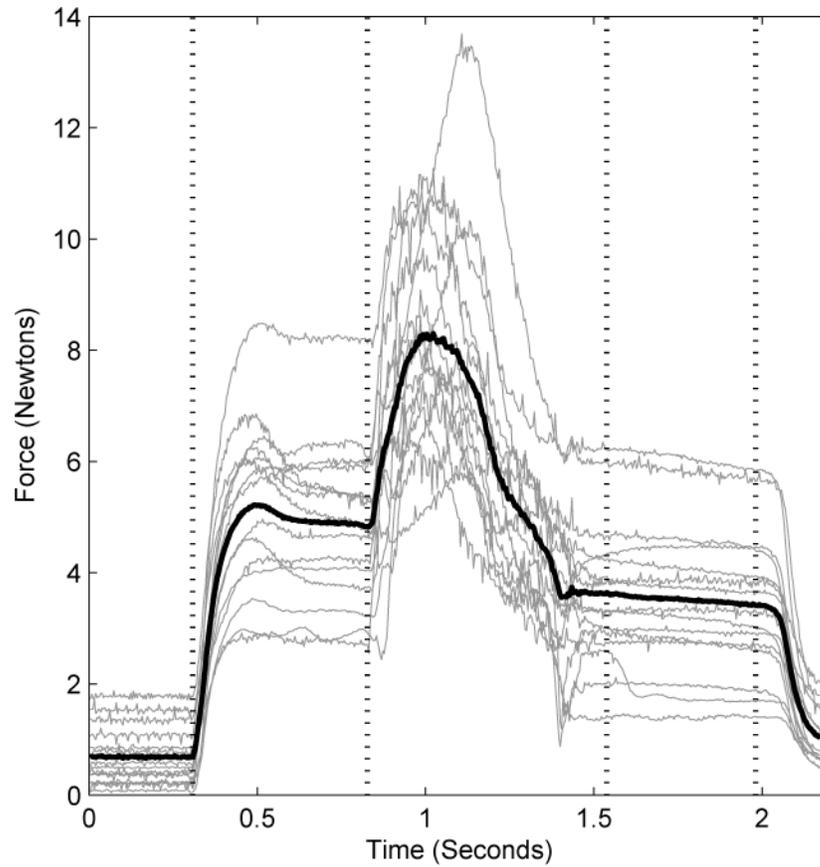


Figure 1.

Fifteen muscle force measurements (gray lines) made on young rats before training, with the mean of these measurements shown in black. The five sections separated by a vertical line describe different portions of the muscle activation. The first and last sections represent force generation when the muscles are not contracting. The second and fourth sections represent the force generated during an isometric contraction; with the third section denoting the stretch shortening contraction, which takes place during an isometric contraction.

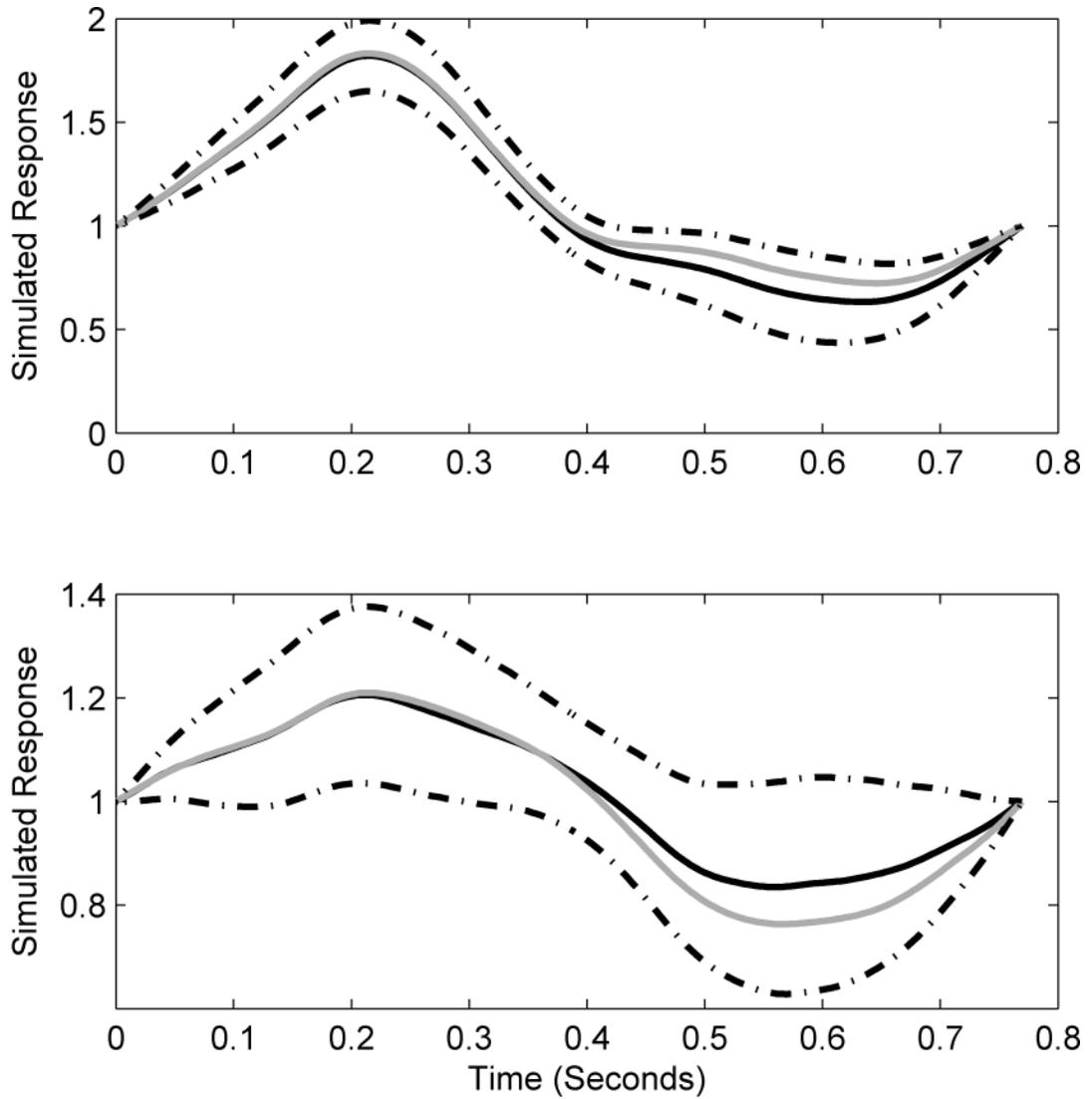


Figure 2. Estimated group level curves for the dynamic force in a stretch shortening contraction for a simulated data set. The black line and corresponding 95% credible region (dotted lines) represent the mean estimated curve and credible region. The true curve is given by the gray line and can be seen to be well represented by the model.

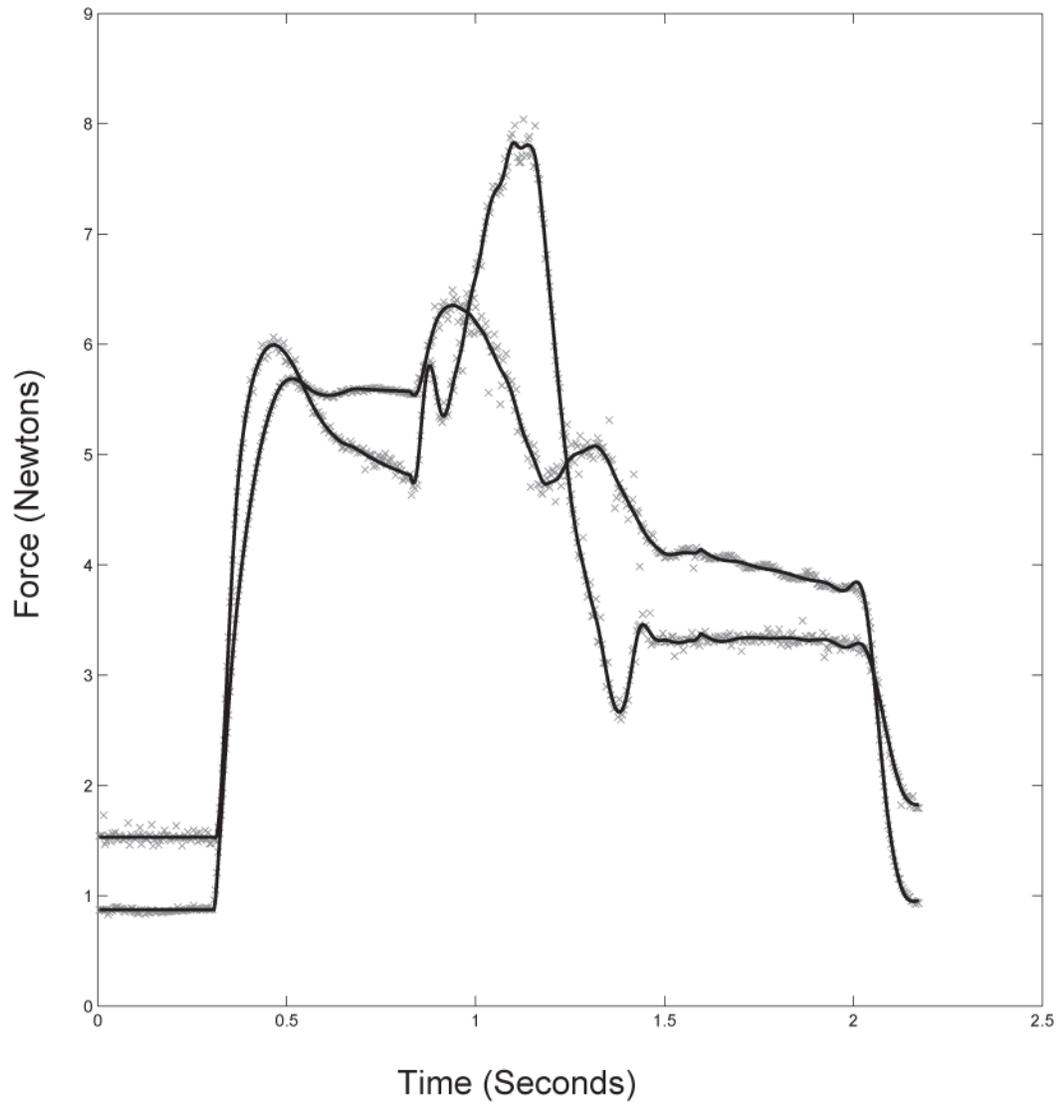


Figure 3. The estimated mean isometric force generated for a single animal pre- and post-treatment. The dark black line represents central estimates of the function $h(t) = Q(t)F(t)$, with the dark gray hash marks representing the observed data. Here credible interval estimates are not shown as they are very narrow.

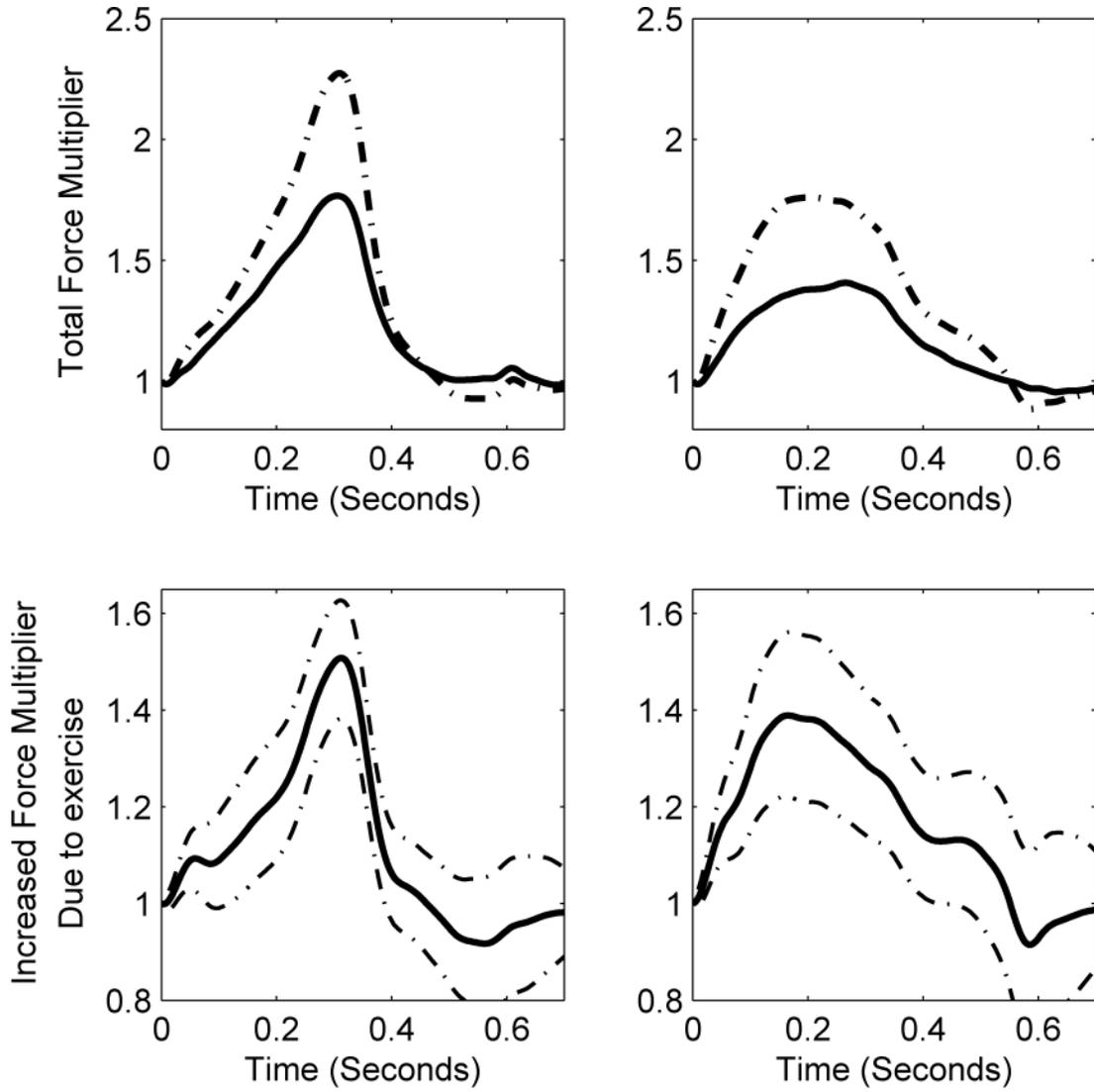


Figure 4. The estimated group level dynamic force multiplier generated by old (left column) and young (right column) animals. Here the top row represents the average force increase during a stretch shortening contraction pre- (black line) and post- (dotted line) exercise protocol. The bottom row represents the observed increase (solid black lines) thought to be due to exercise along with the 95% pointwise credible interval (dotted line) for this increase due to exercise.

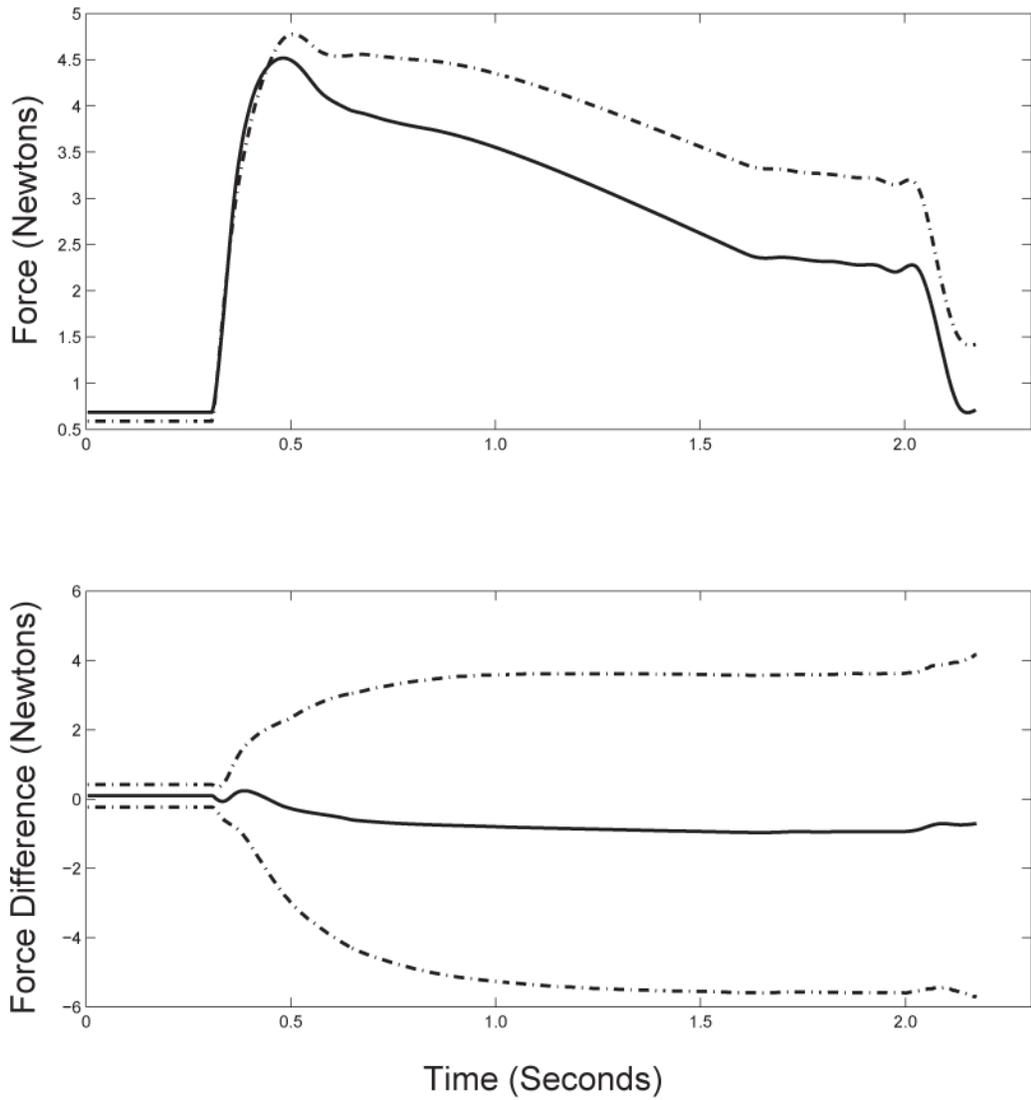


Figure 5. Estimated mean isometric muscle force generated for the young animals pre (dashed line) and post (solid line). The bottom row gives the estimates, and 95% pointwise credible intervals of the difference between the two.

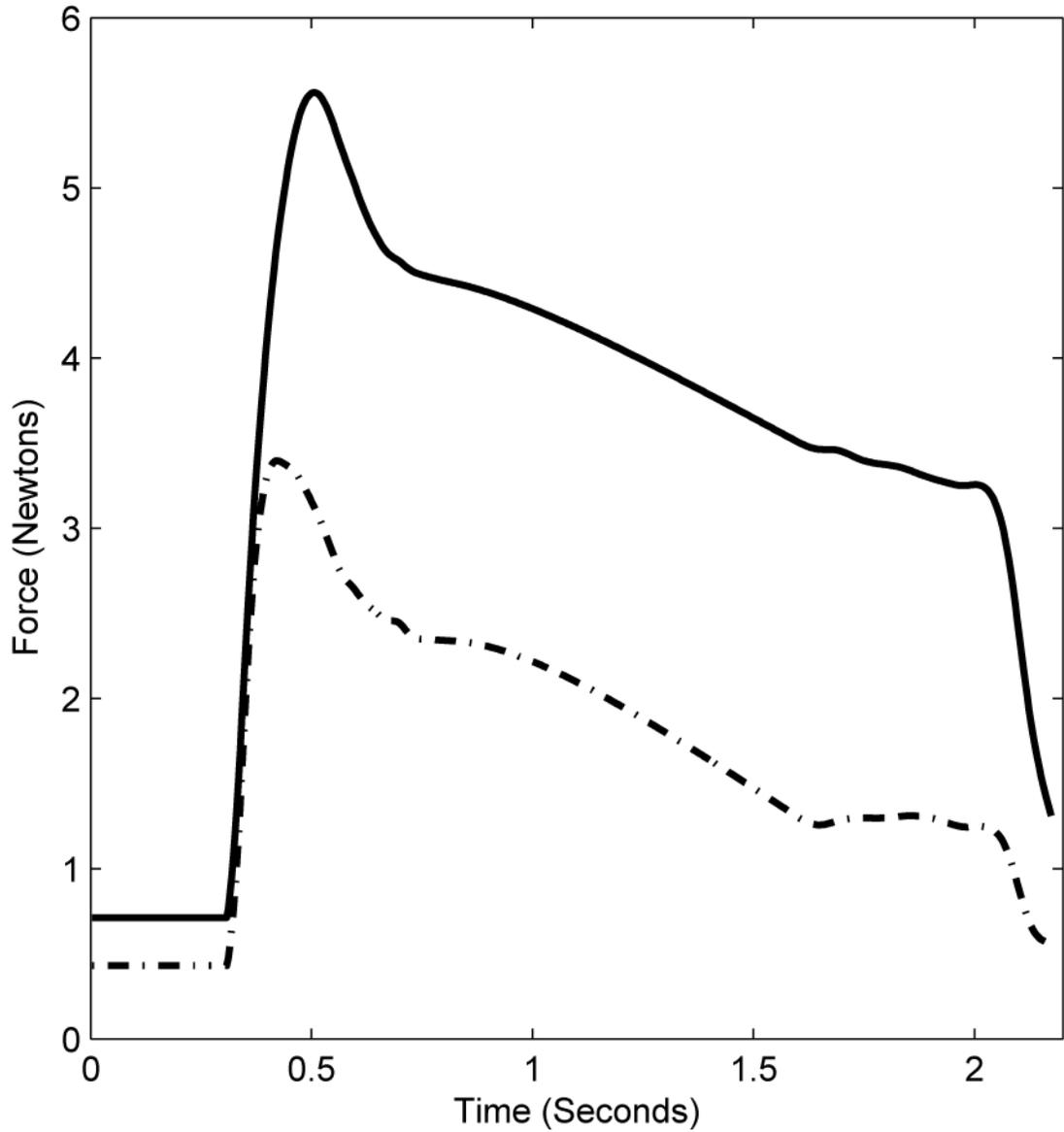


Figure 6. Estimated dynamic muscle force for an old animal. This figure shows the central estimate for the pre- (dashed line) and post- (solid line) for the isometric contraction of an individual old animal. Credible intervals are not shown as they are indistinguishable from the central estimate.

Table 1

This table presents the steps of a resistive-type exercise protocol administered in a rat model of muscle function. Degrees indicate the angle of the joint during application of force. We investigate differences in observation 5 from the first and last experiment. Adapted from Cutlip et al. (2007).

Observation	Experimental Test Protocol (3 Times/wk for 4.5 weeks)
1	Single Isometric force test conducted at 90°
2	1 Stretch-Shortening Contraction at 70° – 140° – 70°
3	8 sets of 10 Stretch-Shortening Contractions 90° – 140° – 90°
4	Single Isometric test conducted at 90°
5	1 Stretch-Shortening Contraction at 70° – 140° – 70°