Improving the Fitness of High-dimensional Biomechanical Models via Data-driven Stochastic Exploration

Veronica J. Santos, *Member, IEEE,* Carlos D. Bustamante, and Francisco J. Valero-Cuevas, *Member, IEEE*

***Abstract*—The field of complex biomechanical modeling has begun to rely on Monte Carlo techniques to investigate the effects of parameter variability and measurement uncertainty on model outputs, search for optimal parameter combinations, and define model limitations. However, advanced stochastic methods to perform data-driven explorations, such as Markov chain Monte Carlo (MCMC), become necessary as the number of model parameters increases. Here we demonstrate the feasibility and, what to our knowledge is, the first use of an MCMC approach to improve the fitness of realistically large biomechanical models. We used a Metropolis-Hastings algorithm to search increasingly complex parameter landscapes (3, 8, 24, and 36 dimensions) to uncover underlying distributions of anatomical parameters of a "truth model" of the human thumb on the basis of simulated kinematic data (thumbnail location, orientation, linear and angular velocities) polluted by zero-mean, uncorrelated multivariate Gaussian "measurement noise." Driven by these data, ten Markov chains searched each model parameter space for the subspace that best fit the data (posterior distribution). As expected, the convergence time increased, more local minima were found, and marginal distributions broadened as the parameter space complexity increased. In the 36-dimensional scenario, some chains found local minima but the majority of chains converged to the true posterior distribution (confirmed using a cross-validation dataset), thus demonstrating the feasibility and utility of these methods for realistically large biomechanical problems.**

***Index Terms*—Bayesian statistics, biomechanical model, Markov chain Monte Carlo, Metropolis-Hastings, parameter estimation, thumb.**

Manuscript received February 22, 2008; revised June 20, 2008. This material is based upon work supported under a National Science Foundation (NSF) Graduate Research Fellowship (to V. J. Santos), a CAREER Award BES-0237258 from NSF's Biomedical Engineering/Research to Aid Persons with Disabilities Program, and Grant Nos. AR050520 and AR052345 from the National Institutes of Health (NIH) (to F. J. Valero-Cuevas). Any opinions, findings, conclusions or recommendations expressed in this publication are ours and do not necessarily reflect the views of the NSF, the National Institute of Arthritis and Musculoskeletal and Skin Diseases, or the NIH.

V. J. Santos is with the Dept. of Mechanical and Aerospace Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: Veronica.Santos@asu.edu). This work was performed while Dr. Santos was with the Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853 USA.

C. D. Bustamante is with the Dept. of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853 USA.

F. J. Valero-Cuevas is with the Dept. of Biomedical Engineering and the Div. of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA 90089 USA. This work was performed while Dr. Valero-Cuevas was with the Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853 USA.

I. Introduction

THE field of complex biomechanical modeling has begun to rely on Bayesian techniques to explore the effects of anatomical variability and measurement uncertainty on model predictions in an open-loop manner. Prior studies have used Monte Carlo methods on biomechanical models of the *thumb* [1], [2], *arm* [3], *elbow* [4], *shoulder* [5]–[7], *knee* [8]–[10], *spine* [11], *tongue* [12], *motor-unit populations* [13], and *bone mechanics* [14]. Recently, response-surface mapping has also been used to establish distributions of biomechanical model performance resulting from uncertainty in input parameters [15]. A key feature of Monte Carlo methods, the most commonly used subset of Bayesian techniques, is the explicit exploration of the parameter space by the dense sampling of each parameter as per assumed probability distribution functions. Monte Carlo simulations are "open-loop" in that independent sets of model parameters are drawn randomly from assumed *prior* distributions $p(\underline{\theta})$, which reflect all prior knowledge about the model parameters, and fed through deterministic equations $f(\underline{\theta})$ to predict distributions of system behavior (Fig. 1).

Unfortunately, a severe limitation of Monte Carlo methods is that their computational demands grow exponentially with the number of model parameters [16]. When the number of parameters is large and/or each iteration is computationally expensive, such open-loop methods become computationally prohibitive. We can, however, improve the efficiency of the simulations using advanced stochastic methods. One approach is to make the stochastic search data-driven (or "closed-loop") by drawing dependent sets of parameters randomly from proposal distributions, feeding them through deterministic equations to predict system behavior, and comparing the predictions to experimental observations in a feedback manner (Fig. 1). As a result, such *data-driven* Markov chain Monte Carlo (MCMC) sampling is biased toward parameter values that lead to model predictions that best agree with experimental data in a least-squares sense [17] (see Appendix A).

While Monte Carlo methods have been applied to biomechanical model parameter spaces ranging from four [14] to 50 dimensions [1], the use of MCMC methods has not been demonstrated for realistically large biomechanical models. To our knowledge, MCMC methods have only been used to match 3D models of the human body to markerless motion capture

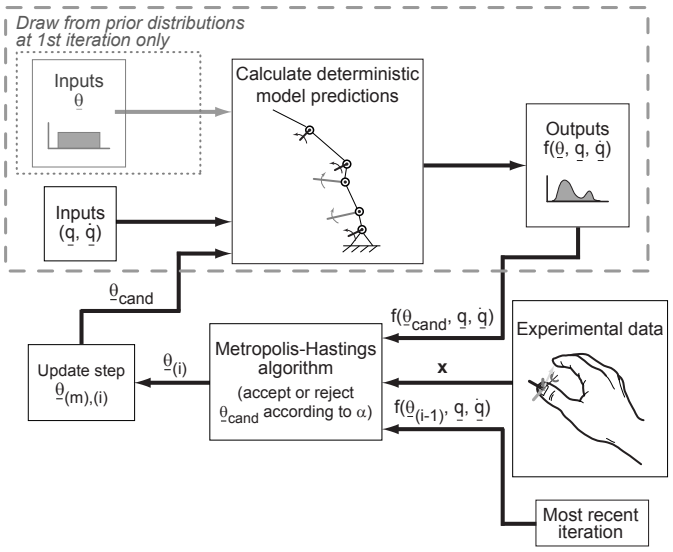


Fig. 1. The open-loop structure of Monte Carlo simulations is enclosed by the gray dashed line. Note that Monte Carlo simulations always draw from the prior distributions while MCMC simulations only draw from the prior distributions at the first iteration. All subsequent MCMC candidate parameter sets are perturbations of the most recent parameter sets. MCMC simulations directly compare model predictions and experimental data in a closed-loop manner. We built our Markov chains using a Metropolis-Hastings algorithm and single-site updating (see section II-A for details on the notation).

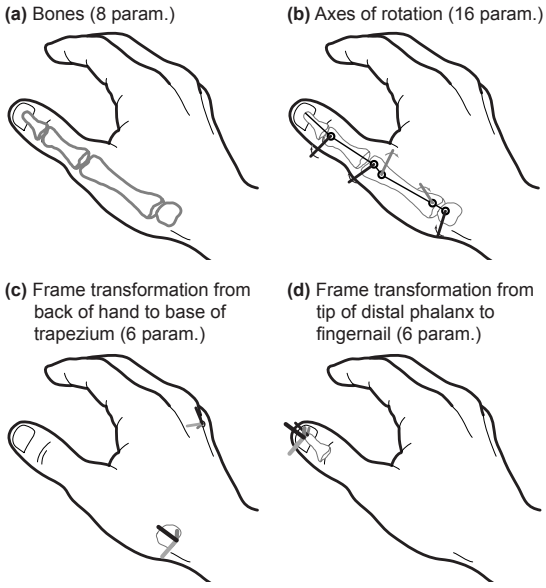


Fig. 2. The model structure consists of 36 independent model parameters: 8 bone dimensions [30] (Subplot *a*), 16 axis of rotation parameters [27]–[29] (Subplot *b*), and 12 coordinate transformation parameters [31] (Subplots *c*, *d*). One transformation relates the proximal end of the thumb at the base of the trapezium to a space-fixed coordinate system on the back of the hand while the other relates the distal end of the thumb at the tip of the distal phalanx to a body-fixed coordinate system on the thumbnail.

data [18] and to estimate neural activity from biosignals [19]–[21]. In this work, we use synthetic data from a model of the human thumb as a proof-of-concept of an MCMC Metropolis-Hastings algorithm to explore a complex 36-dimensional (36D) model parameter space and improve the fitness of the model. In addition, we explicitly investigate the effects of parameter variability and uncertainty on biomechanical model performance.

II. Methods

Using a truth model approach, we arbitrarily defined model parameters for a five degree of freedom kinematic model of the thumb. We then generated synthetic, noisy kinematic output data for this model. Ten independent Markov chains [22] driven by these noisy data searched the model parameter space, which we systematically expanded from 3 to 8, 24, and 36 dimensions, for the subspace that best fit the data. We constructed chains with a Metropolis-Hastings sampling algorithm [23]–[25] on dual Intel® Xeon®, 2.4GHz machines. We developed custom C code using the GNU Scientific Library [26], the Microsoft® Visual Studio .NET Development Environment 2003, and the Intel® C++ Compiler 9.0. We performed the post-analysis in MATLAB® 7.1 for visualization purposes.

Our prior work on Monte Carlo simulations of an anatomy-based model of the thumb [27]–[29] employed open-loop simulations and used experimental data to establish prior distributions, but did not use experimental data as a means to effectively explore the parameter space [2]. In this work, we apply closed-loop MCMC simulations to the same model structure with non-orthogonal and non-intersecting axes of rotation and add model parameters to relate the model to surface marker experimental data. Thus, the model has 36 independent parameters (8 bone dimensions [30], 16 axis of rotation parameters [27]–[29], 12 coordinate transformation parameters [31]) summarized by $\underline{\theta}$ (Fig. 2). The input variables to the five degree of freedom serial linkage model are joint angles $\underline{q}$ and joint angular velocities $\underline{\dot{q}}$ representative of unimpaired kinematic behavior [31]. The output variables obtained deterministically from the serial linkage model are 3D thumbnail location, orientation, linear velocities, and angular velocities.

The experimental data matrix $\mathbf{X}$ is comprised of n rows of observations by $k = 13$ columns of kinematic outputs. The j^{th} observation makes up the j^{th} row of data matrix $\mathbf{X}$, whose individual elements are represented by $x_{j,k}$. Three-dimensional thumbnail location (x, y, z), orientation (expressed by four quaternion parameters q_0, q_1, q_2, q_3 [32], [33]), and linear velocity $(\dot{x}, \dot{y}, \dot{z})$ are expressed in F_{fixed}, a fixed global reference frame whose origin is located on the dorsum of the hand just proximal to the distal tip of the third metacarpal (Fig. 2, Subplot *c*). Angular velocity $(\omega_x', \omega_y', \omega_z')$ is expressed in $F_{thumbnail}$, a local reference frame whose origin is located in the center of the thumbnail (Fig. 2, Subplot *d*). For both F_{fixed} and $F_{thumbnail}$, the $(+)x$, y, and z axes point palmarly, radially, and distally, respectively.

$$\underline{\mathbf{X}}_j = \begin{bmatrix} x & y & z & q_0 & q_1 & q_2 & q_3 & \dot{x} & \dot{y} & \dot{z} & \omega_x' & \omega_y' & \omega_z' \end{bmatrix}_j \quad (1)$$

A. Constructing Markov chains using a Metropolis-Hastings sampling scheme

Ideally, we would know our target distribution $\pi(\underline{\theta})$ (the *posterior* distribution $p(\underline{\theta}|\mathbf{X})$, or the probability of the model

parameter set $\underline{\theta}$ given observed data $\mathbf{X}$) and it would be a simple analytical expression or easy to sample from directly. When this is not the case, we can still estimate the target distribution by sampling from a *stationary* distribution $\pi^*(\underline{\theta})$ that is *proportional* to the target distribution $\pi(\underline{\theta})$ and draw inferences on those results. We can estimate the expectation of any function of $\underline{\theta}$ using averages of samples drawn after convergence to the stationary distribution.

The key to the MCMC approach is the use of an appropriate sampling scheme to construct *ergodic Markov chains* that converge to such a stationary distribution. If run long enough, these chains will locate and forever sample from a stationary distribution whose existence and uniqueness is guaranteed [34] and is our posterior distribution of interest by design. With the Metropolis-Hastings algorithm, we build ergodic chains by drawing each candidate parameter set $\underline{\theta}_{cand}$ from a *proposal* distribution $q(\underline{\theta}_{cand}|\underline{\theta}_{(i-1)})$ and accepting $\underline{\theta}_{cand}$ with an acceptance probability α as per [23]–[25].

$$\alpha(\underline{\theta}_{cand}, \underline{\theta}_{(i-1)}) = \min\left\{\frac{\pi(\underline{\theta}_{cand})q(\underline{\theta}_{(i-1)}|\underline{\theta}_{cand})}{\pi(\underline{\theta}_{(i-1)})q(\underline{\theta}_{cand}|\underline{\theta}_{(i-1)})}, 1\right\} \quad (2)$$

At each iteration, if a randomly drawn uniform variable $U(0,1)$ lies in the $[0,\alpha]$ range, we accept $\underline{\theta}_{cand}$ and set $\underline{\theta}_{(i)} = \underline{\theta}_{cand}$. Otherwise, we reject $\underline{\theta}_{cand}$ and set $\underline{\theta}_{(i)} = \underline{\theta}_{(i-1)}$.

1) Prior distributions: Boundaries on the prior distributions $p(\underline{\theta})$ can be set to known physiologically-limited ranges (e.g., min/max bone length). Certainty or uncertainty can be represented by using distributions having small or large variances, respectively. We created bounds (mean $\pm$ std) for our conservative, noninformative, uniform prior distributions using sparse published [27]–[29] and measured data [30], [31]. Inappropriate selection of prior distribution bounds may exclude regions of the model parameter space where the posterior distribution actually resides. However, the shape of the prior distributions is not that critical due to the memoryless nature of the Markov chains. If run long enough, the chains will eventually converge to the distribution of interest regardless of "poor" initial conditions.

2) Proposal distributions: The initial candidate parameter set $\underline{\theta}_{cand}$, the starting point of the chain, is drawn randomly from the prior distributions (Fig. 1). The remaining $\underline{\theta}_{cand}$ are perturbations of the most recent parameter set $\underline{\theta}_{(i-1)}$ by a "jump-width" vector $\underline{c}$ whose elements are independently scaled by a uniform random variable $u \sim i.i.d.$[1] $U(0,1)$ [35] for $p = 1 \ldots d_{param}$, the dimensionality of the model parameter space.

$$\theta_{p,cand,temp} = \theta_{p,(i-1)} \pm (c_p u_p) \quad (3)$$

Each $\underline{c}$ element is first arbitrarily set to 0.1% of each parameter's allowable range. To gauge search efficiency, we calculate the *acceptance rate* as the percent of the total iterations that we accept $\underline{\theta}_{cand}$. High and low rates mean that the chain is "wandering" excessively (need to increase $\underline{c}$) or "stuck" (need to decrease $\underline{c}$), respectively. While convergence is guaranteed by ergodicity regardless of acceptance rate, we adjust $\underline{c}$ automatically to achieve rates between 30-70% for each model parameter to minimize computing time [36]. We use a reflection technique [37] to ensure that all proposed model parameters remain within the prior distribution boundaries (LB_p, UB_p).

$$\theta_{p,cand} := \begin{cases} 2LB_p + c_p u_p - \theta_{p,(i-1)} & \text{if } \theta_{p,cand,temp} < LB_p \\ 2UB_p - c_p u_p - \theta_{p,(i-1)} & \text{if } \theta_{p,cand,temp} > UB_p \\ \theta_{p,cand,temp} & \text{otherwise} \end{cases} \quad (4)$$

3) Likelihood function: The *likelihood function* $p(\mathbf{X}|\underline{\theta})$ is the probability of observing data $\mathbf{X}$ given the model parameter set $\underline{\theta}$. For our likelihood function, we assume that the experimental data matrix $\mathbf{X}$ is composed of n i.i.d. observations. We assume each observation $\underline{\mathbf{X}}_j$ (the j^{th} row of data matrix $\mathbf{X}$, see eqn. 1) is i.i.d. from a multivariate Gaussian distribution $MVN(f(\underline{\theta}), \hat{\Sigma})$ with a vector of means $f(\underline{\theta})$ that is a deterministic function of $\underline{\theta}$ and a covariance matrix $\hat{\Sigma} = diag(\hat{\sigma}_k^2)$, where $k = 1 \ldots d_{expt}$, the dimensionality of the experimental measurements (number of columns in $\mathbf{X}$).

At each iteration, we estimate $\hat{\sigma}_k^2$ as the sum of the squared residuals divided by a random variable from a χ^2 distribution with $(n-1)$ degrees of freedom. The variance estimates are affected by both measurement uncertainty in $\underline{\mathbf{X}}_j$ and modeling errors in $f_{pred}(\underline{\theta}_{(i)}, \underline{q}_j, \underline{\dot{q}}_j)$, where $\underline{q}_j$ and $\underline{\dot{q}}_j$ represent the constant input joint angles and joint angular velocities for the j^{th} observation.

$$\hat{\sigma}_k^2 = \frac{\sum_{j=1}^{n}[x_{j,k} - f_k(\underline{\theta}, \underline{q}_j, \underline{\dot{q}}_j)]^2}{\chi^2_{(n-1)}} \quad (5)$$

After applying these assumptions and simplifying (App. A), we are left with the following acceptance probability equation.

$$\alpha(\underline{\theta}_{cand}, \underline{\theta}_{(i-1)}) = \min\left\{\frac{\exp\left\{-\frac{1}{2}\sum_{k=1}^{d_{expt}}\sum_{j=1}^{n}\frac{[x_{j,k} - f_k(\underline{\theta}_{cand}, \underline{q}_j, \underline{\dot{q}}_j)]^2}{\hat{\sigma}^2_{k,(i-1)}}\right\}}{\exp\left\{-\frac{1}{2}\sum_{k=1}^{d_{expt}}\sum_{j=1}^{n}\frac{[x_{j,k} - f_k(\underline{\theta}_{(i-1)}, \underline{q}_j, \underline{\dot{q}}_j)]^2}{\hat{\sigma}^2_{k,(i-1)}}\right\}}, 1\right\} \quad (6)$$

If $\underline{\theta}_{cand}$ does not predict the experimental data matrix $\mathbf{X}$ well, α is small and $\underline{\theta}_{cand}$ is less likely to be accepted.

4) Updating scheme: The probability of proposing *and* accepting a "jump" from one point in our high-dimensional parameter space to an entirely unrelated point in the space is very low. Thus, we adopt the "single-site" updating scheme, or "single-component Metropolis-Hastings," in which one element of the model parameter set is varied at a time [37]. For each iteration i we perform an update step m for each dimension in the model parameter space (e.g., $m = 36$ update steps for the 36D space of test case 6). At the first update step, we perturb the first element of $\underline{\theta}_{cand}$ (via eqn. 3), shown here for the 36D test case.

$$\underline{\theta}_{cand,(m=1),(i)} = \begin{bmatrix} \theta_{1,cand} \\ \theta_{2,(i-1)} \\ \vdots \\ \theta_{36,(i-1)} \end{bmatrix} \quad (7)$$

We calculate the acceptance probability and either accept $(\theta_{1,(i)} = \theta_{1,cand})$ or reject $(\theta_{1,(i)} = \theta_{1,(i-1)})$ the element

[1] i.i.d. = independent and identically-distributed

$\theta_{1,cand}$. At the second update step, we perturb the second element $\theta_{2,cand}$ and so on.

$$\underline{\theta}_{cand,(m=2),(i)} = \begin{bmatrix} \theta_{1,(i)} \\ \theta_{2,cand} \\ \theta_{3,(i-1)} \\ \vdots \\ \theta_{36,(i-1)} \end{bmatrix} \quad (8)$$

B. Post hoc analysis

1) Assessing convergence of the MCMC simulations: MCMC being a stochastic process, there is no analytical expression for predicting a sufficient number of iterations prior to running the simulations. Thus, we run diagnostics on MCMC simulation outputs in a *post hoc* manner. We use the Gelman-Rubin convergence diagnostic $\sqrt{\hat{R}}$ [38], [39] to assess convergence because it is an empirical metric that is recommended for use with multiple chains. Briefly, $\sqrt{\hat{R}}$ compares the variance among independent Markov chains to the variance within each chain [22], [36], [37] and approaches one as the independent chains converge to and sample from the same stationary distribution. In practice, we want $\sqrt{\hat{R}}$ to permanently fall below a threshold (1.2, per [39]) for all input and output scalars of interest, as this indicates convergence. The iterations preceding the crossing of the $\sqrt{\hat{R}}$ threshold are discarded as part of the *burn-in* period.

The iterations after the burn-in period are often *thinned* according to a user-specified thinning interval p to reduce costs such as computational storage and post-analysis time [22], [39]. Every p^{th} iteration of each chain is selected and pooled. The model parameter set at each of these pooled iterations is a draw from the joint posteriors and can be used to draw inferences about the model. We know of no widely-accepted number of converged iterations necessary for adequate sampling of the posterior distribution, but due to the Law of Large Numbers [40], many iterations beyond convergence will only serve to refine the distribution. We arbitrarily extended chains to obtain $25,000$ converged iterations per chain, thus ensuring $1,000$ thinned iterations with $p = 25$.

2) Posterior predictive sampling: To obtain *posterior predictive samples* $f_{pred,noisy}(\underline{\theta}_{(i)}, \underline{q}_j, \underline{\dot{q}}_j)$ that account for measurement noise, we add zero-mean, uncorrelated multivariate Gaussian noise ($\hat{\Sigma}_{(i)} = diag(\hat{\sigma}^2_{(i)})$; eqn. 5) to the deterministic model predictions $f_{pred}(\underline{\theta}_{(i)}, \underline{q}_j, \underline{\dot{q}}_j)$.

$$f_{pred,noisy}(\underline{\theta}_{(i)}, \underline{q}_j, \underline{\dot{q}}_j) = f_{pred}(\underline{\theta}_{(i)}, \underline{q}_j, \underline{\dot{q}}_j) + MVN(\underline{0}, \hat{\Sigma}_{(i)}) \quad (9)$$

The distributions of the posterior predictive samples (over all thinned, converged iterations) are then compared to each inherently noisy experimental observation $\underline{\mathbf{X}}_j$.

3) Cross-validating the model: Posterior predictive sampling indicates how well the joint posteriors can predict the experimental data provided to the Metropolis-Hastings algorithm. However, model cross-validation must be performed using a separate set of experimental observations $\mathbf{X}_{val}$ not provided to the algorithm. Cross-validation is identical to posterior predictive sampling except that each joint posterior draw $\underline{\theta}_{(i)}$ is used to predict model outputs under different conditions (i.e., different joint angles $\underline{q}$ and joint angular velocities $\underline{\dot{q}}$). The variance estimates are recalculated (eqn. 5) and noise is added (eqn. 9) prior to comparison with $\mathbf{X}_{val}$.

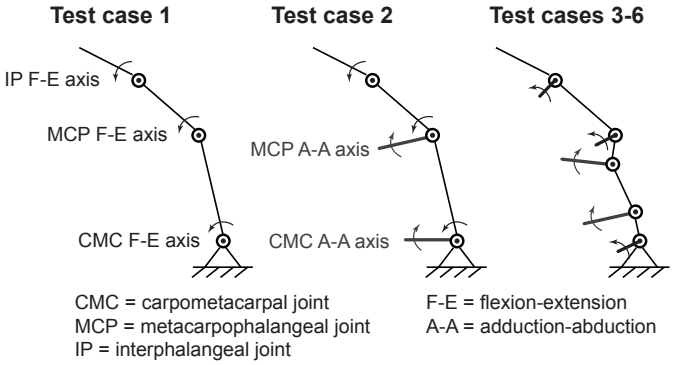


Fig. 3. Truth model test cases used for the validation of our Metropolis-Hastings sampling algorithm are shown. From left to right, the models were assigned 3, 3, and (3, 8, 24, and 36) free model parameter values, respectively (see Table I).

C. Validating the sampling algorithm using truth model "test cases"

Truth models are used to validate algorithms because the researcher defines the model structure, its parameter values, and its deterministic behavior and focuses on the algorithm's performance itself. In this work, we defined a "virtual subject" having a true model parameter set $\underline{\theta}_{true}$ and model inputs (joint angles $\underline{q}$ and joint angular velocities $\underline{\dot{q}}$). We calculated the deterministic model outputs $f(\underline{\theta}_{true}, \underline{q}, \underline{\dot{q}})$, which were necessarily free of measurement noise. We added simulated zero-mean, uncorrelated multivariate Gaussian "measurement noise" $MVN(\underline{0}, \hat{\Sigma}_{true})$ to our deterministic model outputs $f(\underline{\theta}_{true}, \underline{q}, \underline{\dot{q}})$ and treated the noisy results $f(\underline{\theta}_{true,noisy}, \underline{q}, \underline{\dot{q}})$ as experimental data, similar to our posterior predictive methods (eqn. 9). For convenience, we defined $\hat{\Sigma}_{true} = diag(\hat{\sigma}^2_{k,true})$. We used these simulated "noisy" experimental data to drive our MCMC simulations.

$$f(\underline{\theta}_{true,noisy}, \underline{q}, \underline{\dot{q}}) = f(\underline{\theta}_{true}, \underline{q}, \underline{\dot{q}}) + MVN(\underline{0}, \hat{\Sigma}_{true}) \quad (10)$$

Due to the high-dimensional nature of the parameter space and the fact that we were developing custom C code, we worked through a series of truth model "test cases" to validate our algorithm (Fig. 3 and Table I). Test cases 1-3 used three free parameters only, requiring a search of a 3D parameter space. Test cases 1 and 2 were planar and 3D kinematic models of the thumb, respectively, with orthogonal and intersecting axes of rotation. Test case 3 was a trivial upgrade to test case 2 that used different constants for the axes of rotation model parameters to simulate a kinematic model of the thumb based on the "virtual five-link model" with non-orthogonal and non-intersecting axes of rotation (Table I). Once the sampling scheme worked for these three test cases, we systematically increased the dimensionality of the parameter space from 3 to 8, 24, and 36. Test case 6, with the 36D parameter space, is

TABLE I

SIX TRUTH MODEL TEST CASES WERE USED TO VALIDATE OUR METROPOLIS-HASTINGS SAMPLING ALGORITHM (SEE FIG. 3). BOXES WITH DASHES INDICATE FREE VARIABLES THAT MAKE UP THE MODEL PARAMETER SPACE.

		Test case		**1**	**2**	**3**	**4**	**5**	**6**
		Dimensionality of model parameter space		**3**	**3**	**3**	**8**	**24**	**36**
		Rotational degrees of freedom		3	5				
		Axes of rotation		orthogonal, intersecting	non-orthogonal, non-intersecting				
Parameters for bone dimensions	1	metaz	Distal length of metacarpal	-	-	-	-	-	-
	2	metaratx	Palmar/distal length for metacarpal	0.3402	0.3402	0.3402	-	-	-
	3	metaraty	Radial/distal length for metacarpal	0.3602	0.3602	0.3602	-	-	-
	4	proxphalz	Distal length of proximal phalanx	-	-	-	-	-	-
	5	proxratx	Palmar/distal length for proximal phal.	0.4176	0.4176	0.4176	-	-	-
	6	distphalz	Distal length of distal phalanx	-	-	-	-	-	-
	7	trapratz	Trapezium/proximal phalanx distal lengths	0	0	0	-	-	-
	8	trapratx	Palmar/distal length for trapezium	1.1855	1.1855	1.1855	-	-	-
Parameters specifying the location and orientation of the axes of rotation (orientation param. in rad.)	9	BETAcmcFE	Orientation parameter for CMC FE axis	0	0	0.1151	0.1151	-	-
	10	bpercCMCFE	Location parameter for CMC FE axis	0.5	0.5	0.347	0.347	-	-
	11	ALPHAcmcAA	Orientation parameter for CMC AA axis	1.5708	1.5708	1.4824	1.4824	-	-
	12	BETAcmcAA	Orientation parameter for CMC AA axis	1.5708	1.5708	1.2027	1.2027	-	-
	13	tpercCMCAA	Location parameter for CMC AA axis	0.5	0.5	0.5573	0.5573	-	-
	14	lpercCMCAA	Location parameter for CMC AA axis	0	0	0.1586	0.1586	-	-
	15	ALPHAmpAA	Orientation parameter for MP AA axis	1.5708	1.5708	1.4099	1.4099	-	-
	16	BETAmpAA	Orientation parameter for MP AA axis	1.5708	1.5708	1.3123	1.3123	-	-
	17	tpercMPAA	Location parameter for MP AA axis	0.5	0.5	0.4792	0.4792	-	-
	18	lpercMPAA	Location parameter for MP AA axis	1	1	0.8617	0.8617	-	-
	19	ALPHAmpFE	Orientation parameter for MP FE axis	1.5708	1.5708	1.8289	1.8289	-	-
	20	tpercMPFE	Location parameter for MP FE axis	0.5	0.5	0.6425	0.6425	-	-
	21	lpercMPFE	Location parameter for MP FE axis	1	1	0.835	0.835	-	-
	22	BETAipFE	Orientation parameter for IP FE axis	1.5708	1.5708	1.4265	1.4265	-	-
	23	tpercIPFE	Location parameter for IP FE axis	0.5	0.5	0.5918	0.5918	-	-
	24	lpercIPFE	Location parameter for IP FE axis	1	1	0.9036	0.9036	-	-
Parameters for 3D frame transform. at the proximal base (FF) and distal tip (FN) of the thumb	25	FF_rotz	Euler rot. ang. about z-axis for FF frame	0	0	0	0	0	-
	26	FF_roty	Euler rot. ang. about y-axis for FF frame	0	0	0	0	0	-
	27	FF_rotx	Euler rot. ang. about x-axis for FF frame	0	0	0	0	0	-
	28	FF_trx	Translation along x-axis for FF frame	0	0	0	0	0	-
	29	FF_try	Translation along y-axis for FF frame	0	0	0	0	0	-
	30	FF_trz	Translation along z-axis for FF frame	0	0	0	0	0	-
	31	FN_rotz	Euler rot. ang. about z-axis for FN frame	0	0	0	0	0	-
	32	FN_roty	Euler rot. ang. about y-axis for FN frame	0	0	0	0	0	-
	33	FN_rotx	Euler rot. ang. about x-axis for FN frame	0	0	0	0	0	-
	34	FN_trx	Translation along x-axis for FN frame	0	0	0	0	0	-
	35	FN_try	Translation along y-axis for FN frame	0	0	0	0	0	-
	36	FN_trz	Translation along z-axis for FN frame	0	0	0	0	0	-

the level of complexity we are truly interested in for practical application to experimental data.

III. RESULTS

Despite the high-dimensional model parameter space, the MCMC simulations converged, and the joint posterior distribution for all model parameters was obtained for test cases ranging from 3 to 36 dimensions. As expected, the burn-in period increased as the dimensionality of the model parameter space increased (Table II). The true posterior distribution for the 3D test case 3 was found quickly while the 36D test case 6 took much longer (burn-in iterations of 919 and 56,400, respectively). The Markov chains also found more local minima as the dimensionality of the model parameter space increased. All ten chains settled on a single minima (true posterior) for the 3D and 8D test cases. The ten chains located two and four local minima (including the true posterior) for the 24D and 36D test cases, respectively.

Not surprisingly, our search of the 36D test case 6 took the longest to converge likely due to the "curse of dimensionality" [16] as well as the potential for an increased number of local minima in the fitness landscape. For test case 6, we ran ten

TABLE II
A SUMMARY OF THE MCMC CONVERGENCE STATISTICS IS SHOWN FOR THE SIX TRUTH MODEL TEST CASES. WE USED A THINNING INTERVAL OF 25 ITERATIONS.

Test case	Dim. of param. space	Total iter. per chain	Burn-in iter.	# of chains to find true posterior (out of 10)	Converged iter. after thinning [a]	# of minima found thus far
1	3	4,005	455	10	1,420	1 (true)
2	3	3,005	480	10	1,010	1 (true)
3	3	3,020	919	10	840	1 (true)
4	8	4,025	1,600	10	970	1 (true)
5	24	25,000	13,841	8	3,570	2 (inc. true)
6	36	105,000	56,400	6	11,664	4 (inc. true)

[a] # converged iter. after thinning = [(# total - # burn-in) iter. per chain] * (# converged chains) / (# iterations for thinning interval)

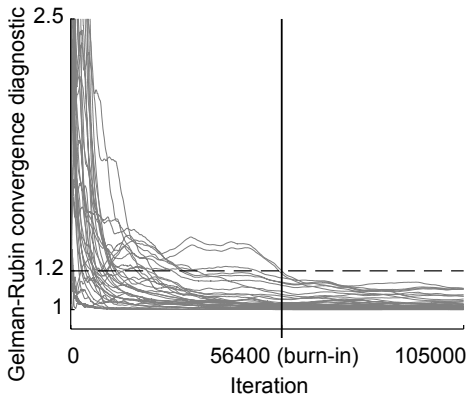


Fig. 5. For the best-performing six-chain subset from the 36D test case 6, the Gelman-Rubin convergence diagnostic values $\sqrt{\hat{R}}$ satisfied our threshold of 1.2 (dashed line) by a burn-in iteration of 56400 (vertical line).

independent Markov chains of $105,000$ iterations each on 2.4GHz dual Intel® Xeon® machines, which took 255.5 hours (49 hours for MCMC simulations in C and 206.5 hours for post-analysis in MATLAB® 7.1), or approximately 10.6 days of wall-clock time.

For the 36D model parameter space, the time history of the sum of the squared residuals between the model predictions and the simulated experimental data (numerator of eqn. 5) showed that a six-chain subset (majority of the ten independent chains) had sums of squared residuals that were two orders of magnitude smaller than the next best-performing chain. Upon further inspection, it was found that this six-chain subset had converged to the true posterior distribution while the remaining four chains (a single two-chain subset and two individual chains) found other local minima in the 36D landscape (Fig. 4). In all, four distinct types of behavior were observed after convergence.

When including all ten Markov chains for the 36D test case 6, the Gelman-Rubin convergence diagnostic failed to converge to values below our threshold of 1.2. By restricting the $\sqrt{\hat{R}}$ calculations to the six-chain subset, we determined a burn-in period of 56,400 iterations (Fig. 5). At this point, each of the six independent chains had located the true posterior distribution and were sampling from the same region of the model parameter space despite their overdispersed initial conditions.

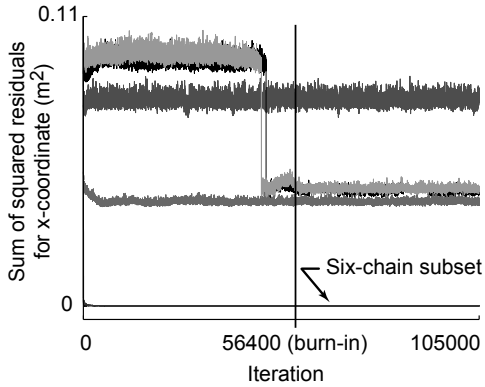


Fig. 4. The time history of the sum of the squared residuals (summed across all data points, see numerator of eqn. 5) is shown for a representative model output (thumbnail x-coordinate) for the 36D test case 6. Each trace corresponds to the performance of a single Markov chain. The burn-in iteration is represented by the vertical line at iteration 56400. The six-chain subset that located the true posterior distribution had values on the order of 2E-$4m^2$. Four other chains had larger residuals ranging from 4E-2 to 8E-$2m^2$, suggesting that they had found local minima.

For the 36D test case 6, the first 7500 iterations are shown for two representative model parameters in Figure 6, Subplot *a*. This subplot illustrates the overdispersed initial conditions of the ten Markov chains, exploration of the parameter space, and the movement of some of the chains towards the true posterior distribution. Note that an open-loop Monte Carlo integrator must randomly sample from the entire volume to achieve convergence. Figure 6, Subplot *b* shows the burn-in iterations as well as samples from the true posterior for three representative model parameters. While only 3D results are shown, convergence occurred in 36D for this test case 6, making the efficiency of the MCMC approach all the more striking. It is clear that independent chains are able to converge to the appropriate region of the high-dimensional parameter space without the computationally prohibitive exploration of every point in the 36D space.

From our thinned, converged posterior distribution draws, we immediately have the probability density function for each element of $\underline{\theta}$, also called the *marginal* distribution. The marginal distributions for the 3D test case 3 are sharp and have tight ranges (Figure 7, Subplots *a*, *d*, and *f*). With only three free model parameters, there are relatively few parameter combinations that lead to model predictions that agree well with the simulated experimental observations. In contrast, the marginal distributions for the 36D test case 6 varied in breadth (Figures 7, 8, and 9). Of the eight bone parameters, the longitudinal lengths of the thumb bones appeared to be the most important for predicting the simulated kinematic experimental data. Of the 16 axis of rotation model parameters, those specifying the orientation of the most proximal joint axes (e.g., BETAcmcAA) had the most narrow marginal distributions, suggesting their importance in the model (Figure 8). Almost all model parameters associated with the 3D frame

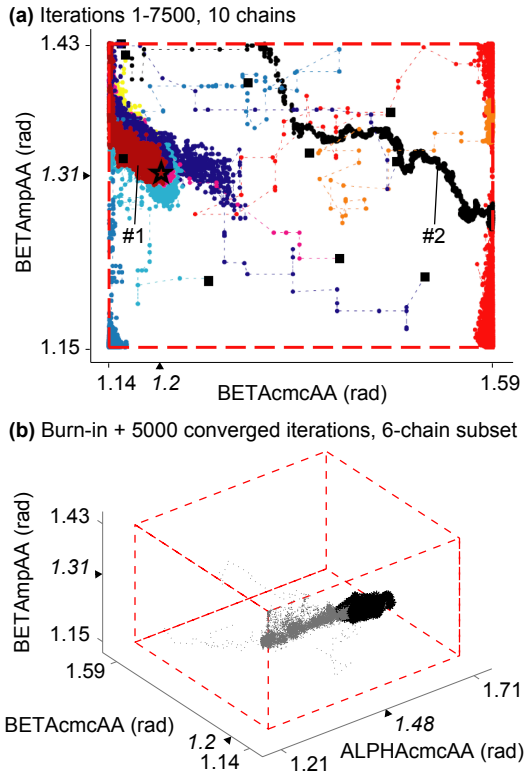


Fig. 6. **(a)** The overdispersed initial conditions (solid squares) for ten independent Markov chains are shown for two representative parameters for the 36D test case 6. Some chains (e.g., #1) located the true posterior (true parameter values marked by a star) while others (e.g., #2) found local minima. **(b)** The history of the six-chain subset is shown for three representative parameters for the 36D test case 6. Burn-in and pooled converged iterations are shown in gray and black, respectively. For both Subplots *a* and *b*, prior distribution bounds are represented by dashed lines, and tick marks and triangles on the axes indicate the parameter bounds and true values, respectively.

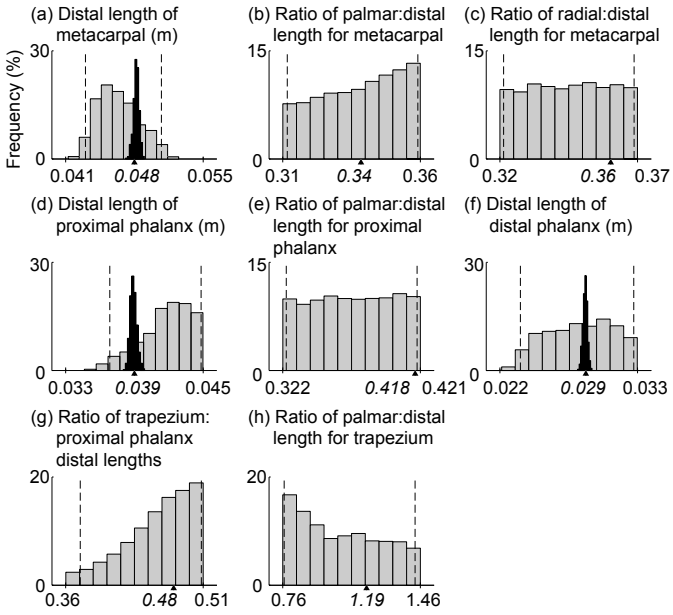


Fig. 7. Marginal distributions are shown for the eight bone parameters [30] for the 36D test case 6 (gray bars) and the 3D test case 3 (black bars in Subplots *a*, *d*, and *f*). Dashed lines indicate the 95% confidence intervals (for the 36D test case 6) while tick marks and triangles on the x-axes indicate the parameter bounds and true values, respectively.

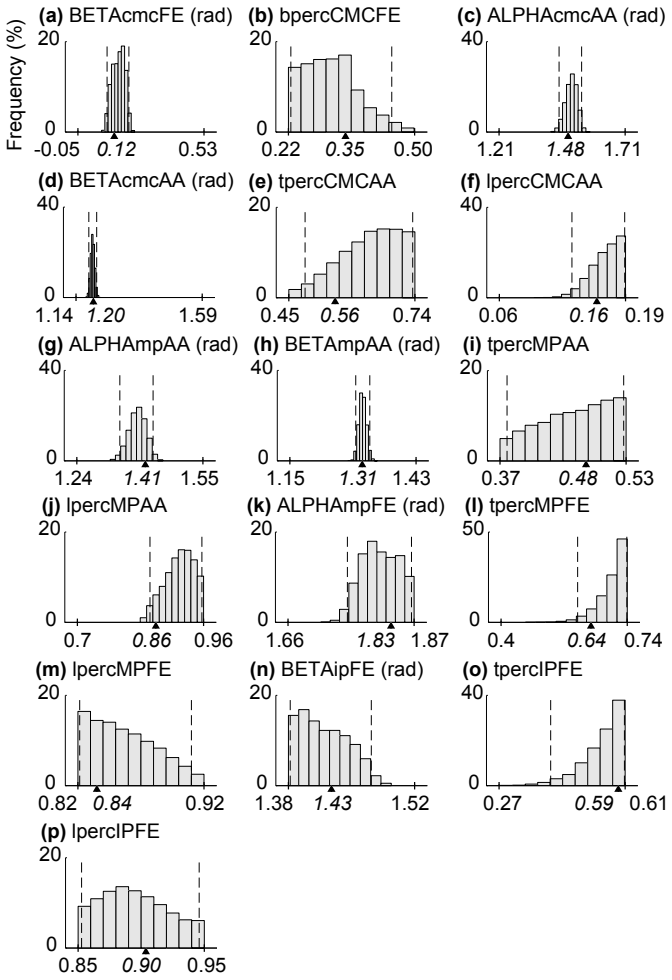


Fig. 8. Marginal distributions are shown for the 16 axis of rotation parameters [27]–[29] for the 36D test case 6. Dashed lines indicate the 95% confidence intervals while tick marks and triangles on the x-axes indicate the parameter bounds and true values, respectively.

transformations at the proximal base and distal tip of the thumb had extremely sharp marginal distributions, highlighting the sensitivity of the model to these parameters (Figure 9).

For the 36D test case 6, the errors in the posterior predictive samples for the six chains were minimal (Fig. 10, Subplots *a* and *b*). Despite the noise in the simulated data (eqn. 10), the Metropolis-Hastings algorithm was able to estimate and account for the simulated measurement errors we added to the "clean" deterministic truth model outputs and successfully locate the true posterior distribution. The errors in the posterior predictions for the cross-validation data set were also minimal (Fig. 10, Subplots *c* and *d*), demonstrating that the algorithm did not simply over-fit the data given, but actually found the true parameters in the 36D space.

IV. Discussion

We have demonstrated a proof-of-concept for our Metropolis-Hastings sampling algorithm for a complex 36D model parameter space with practical application to experimental data. Contrary to the assumed notion that MCMC methods are computationally prohibitive for non-trivial mod-

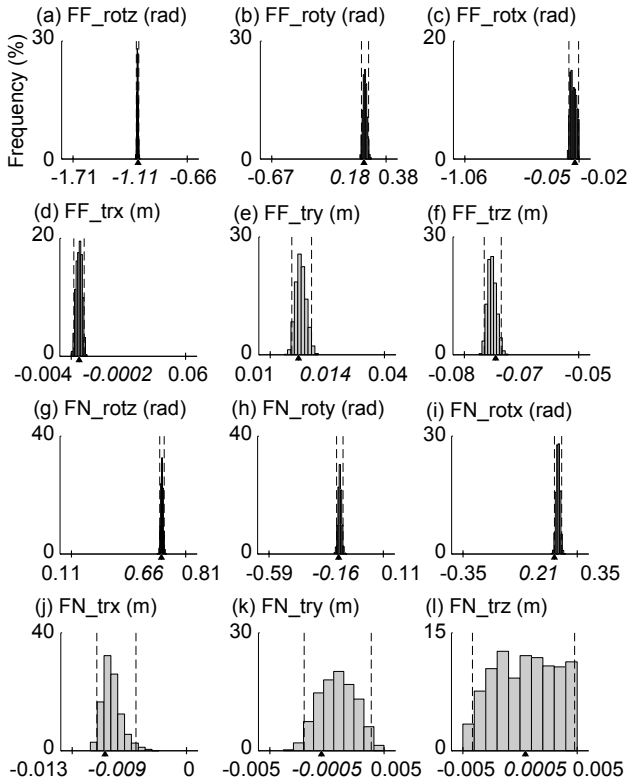


Fig. 9. Marginal distributions are shown for the 12 parameters for the frame transformations at the proximal base and distal tip of the thumb [31] for the 36D test case 6. Dashed lines indicate the 95% confidence intervals while tick marks and triangles on the x-axes indicate the parameter bounds and true values, respectively.

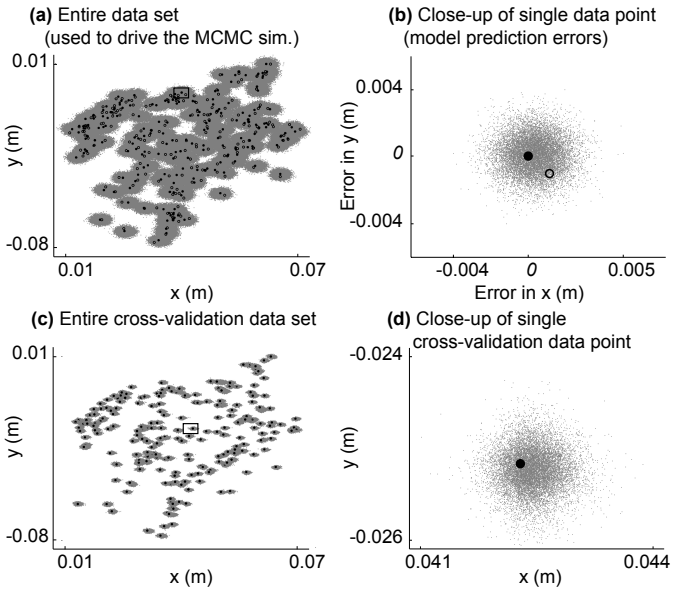


Fig. 10. Posterior prediction (Subplots *a*, *b*) and cross-validation (Subplots *c*, *d*) results are shown for the 36D test case 6 for two experimental outputs: x and y thumbnail coordinates. **(a)** Posterior predictive samples (gray points) are clustered around the true model outputs (solid circles) despite the use of noisy data (open circles; see eqn. 10). **(b)** Posterior predictive sample *errors* are shown for a single data point (see box in Subplot *a*). The solid and open circles indicate the true error value of zero and offset of the noisy simulated data point, respectively. **(c)** Model predictions (gray points) are clustered around the cross-validation data (solid circles). **(d)** A close-up view of a single data point (see box in Subplot *c*) is shown.

els, we have shown the feasibility of MCMC methods to improve the fitness of biomechanical models with large numbers of parameters. The advantages, limitations, implications, and applications of this work are described here.

A. Foundation for the Bayesian approach

MCMC methodology is often criticized for its subjective selection of the prior distribution. However, proponents of the Bayesian philosophy counter that (1) prior information should be used if available [22], (2) if the MCMC simulations are run to convergence, the use of possibly naïve prior distributions becomes less important as the data "speak for themselves" [22], [36], and (3) the subjective selection of the prior distribution is no different from the *a priori* subjective selection of a likelihood function for maximum likelihood estimation [22].

The Bayesian viewpoint is particularly appropriate for biomechanical modeling, where anatomical variability is the rule, because each parameter has its own probability distribution. A Frequentist approach yields a *single set* (point estimate) of model parameters that optimizes some objective function [41]. In contrast, a Bayesian approach such as MCMC yields *distributions* (cloud estimates) of model parameters and their outputs that have been informed by experimental data [37]. With distributions, revealing properties are immediately available: probability densities, parameter ranges, skewness, kurtosis, and multimodality. Confidence intervals and means are trivial to extract. Furthermore, the parameter marginal distributions that result can be used as informative prior distributions for future simulations, instead of the diffuse priors used at the start of such parameter estimation processes.

MCMC simulations are essentially stochastic, data-driven sensitivity analyses that enable us to determine the effects of parameter variability and uncertainty resulting from natural anatomical variability across a population and sparseness of experimental data, respectively. Latent parameter covariances that may not have been explicitly measured (and may not be measurable) can be extracted from the joint posterior, justifying a reduction in model complexity.

B. Effects of increasing parameter space complexity

During the validation of our Metropolis-Hastings algorithm, we gradually increased the dimensionality of the model parameter space from 3 to 8, 24, and 36 dimensions (Table I). As expected, convergence time (Table II) increased with the complexity of the model parameter space.

When landscape complexity increased, an increasing number of independent Markov chains located local minima. All ten Markov chains successfully located the true posterior distribution for the 3D and 8D scenarios (test cases 3 and 4, respectively). However, the number of chains to wander into local minima were two and four for the 24D and 36D scenarios (test cases 5 and 6, respectively). Importantly, in all cases the majority of the Markov chains successfully located the true posterior distribution. In theory [34], "trapped" chains should eventually "escape" local minima if run long enough. However, due to the stochastic nature of the sampling method, we cannot determine how long convergence will take *a priori*.

Guided by our Gelman-Rubin convergence diagnostic, we pooled the majority of the chains that converged and found that they had indeed located the true posterior distribution. As the complexity of the model parameter space increased, we found that the marginal distributions became more varied in range and shape (described below).

C. Marginal distributions

The model parameter marginal distributions (Figures 7, 8, and 9) provide insight into the sensitivity of the model predictions to variability in the model parameters. A sharp marginal distribution (e.g., Fig. 8, Subplot *d*) suggests high sensitivity of the model predictions to variation in the parameter while a flat marginal distribution (e.g., Fig. 7, Subplot *c*) can be due to a variety of reasons. The simplest reason is *insensitivity* of model predictions to variations in the parameter because the parameter values do not affect the model output [13], [42]. In this case, the parameter could be set to a constant and removed from the model parameter space altogether, thereby reducing the dimensionality of the parameter space and the number of update steps per iteration.

There could also be *redundancy* in the model structure itself. The flattest marginal distributions we observed were for model parameters that relate bone widths and lengths (Fig. 7, Subplots *c* and *e*). In our model structure, we represent bones by their bounding box dimensions. It may be that a "link" in our serial linkage model could fit within this bounding box in a variety of ways (e.g., perfectly aligned within the box or more diagonally oriented within the box). This redundancy could allow for greater variance of the model parameters that relate bone widths and lengths, but less variance in the model predictions. Finally, the flat marginal distributions could result from *unobservability* [41] because effects of the model parameters cannot be observed within the experimental data.

The range of the marginal distributions can highlight a need to reassess the prior distribution bounds. For instance, a marginal distribution that is pushed up against a parameter's upper bound (e.g., Fig. 7, Subplot *g*) may suggest a need to reevaluate the appropriateness of the bound. If the bound is nonnegotiable then the MCMC results point directly to limiting aspects of the model structure.

It is interesting that the marginal distributions for the 3D test case 3 had tight ranges and sharp distributions as compared to their counterparts for the 36D test case 6 (Fig. 7, Subplots *a*, *d*, *f*). This was not expected *a priori*. However, these results are not unreasonable given the increased chance for observing the effects of parameter insensitivity, model redundancy, and unobservability as the complexity of the model parameter space increased.

D. Correlations among bone lengths

From an anatomical perspective, it is reasonable to expect that larger people have larger hands and that bone lengths scale up, resulting in a positive correlation among bone lengths. However, this issue concerns trends across a population and cannot be addressed by the truth model in this study, which focused on modeling a single virtual subject. To ensure proper sampling of the parameter space within our single-site updating scheme, we removed constraints on relative bone lengths and allowed bone dimensions to be drawn independently of the others. While this does not explicitly enforce physiological realism, it ensures that our Metropolis-Hastings algorithm is properly constructed such that convergence to the posterior distribution is guaranteed. Simulation results can easily be checked for physiologically realistic relationships upon convergence. In addition, "block-updating" can be used to incorporate such parameter correlations and improve the efficiency of the search [37], [43]. With block updating, correlated model parameters are perturbed as a subset rather than individually, as with single-component Metropolis-Hastings. For instance, let model parameters 1-3 be correlated. At the first update step ($m = 1$) we perturb the first three parameters as a block. At the second update step, we continue the single-site updating from the 4^{th} parameter, shown here for the 36D test case.

$$\underline{\theta}_{cand,(m=1),(i)} = \begin{bmatrix} \begin{bmatrix} \theta_{1,cand} \\ \theta_{2,cand} \\ \theta_{3,cand} \end{bmatrix} \\ \theta_{4,(i-1)} \\ \vdots \\ \theta_{36,(i-1)} \end{bmatrix} \tag{11}$$

E. Gelman-Rubin convergence diagnostic

For the 36D test case 6, $\sqrt{\hat{R}}$ never fell below our threshold of 1.2 when calculating the shrink factor for all ten Markov chains together (Fig. 5, Subplot *a*). This apparent lack of convergence is not due to inadequacies of the MCMC approach, as it has been successful in other applications [23], [44], including problems with 200 dimensions [45]. The lack of convergence likely results because we are exploring a complex, nonlinear landscape with multiple local minima and, by construction, $\sqrt{\hat{R}}$ will not indicate convergence until all chains overlap in their sampling regions. Unfortunately, we have not found literature that explains how to deal with the failure of $\sqrt{\hat{R}}$ for multimodal posterior distributions.

Nonetheless, $\sqrt{\hat{R}}$ remains a popular, straightforward, and effective means to assess the progress of MCMC simulations. We took a common sense approach and calculated $\sqrt{\hat{R}}$ for the six-chain subset that was obviously sampling from the same subspace and found that $\sqrt{\hat{R}}$ satisfied our threshold (Fig. 5, Subplot *b*).

F. Posterior distribution multimodality: Real or artifact?

A valid concern with the MCMC approach is how to distinguish between modes that actually exist in a population and those that arise because of a complex high-dimensional model parameter space and/or the search methodology. With our truth model, we know that six of ten chains converged to the true posterior distributions while four chains wandered into local minima (Fig. 4). It may seem that we have found alternative solutions to the same problem, but further inspection reveals that the solutions are not equivalent and some outperform others (Fig. 4). If we had run a nonlinear least-squares

optimization (a Frequentist technique) and found two different solutions[2] (a likely occurrence given that convergence is also a challenge for these other methods), we would immediately look at the cost function for each solution and use the model that performs better. We could set a cost threshold relevant to the research question, objectively select the "best" mode, and use it to represent the true posterior distribution.

The "Achilles heel" of all search algorithms is that local minima can attract the algorithms to false solutions, often resulting in multimodal posterior distributions. Note that this only occurred for test cases 5 and 6, when the model parameter spaces were large (24D and 36D, respectively). One advanced technique for dealing with multimodality is Metropolis-coupled MCMC, or "hot-swapping" [46], [47], where one continuously monitors a user-defined metric (e.g., the sum of the squared residuals (Fig. 4)). A chain that is presumably trapped in a local minimum is given an opportunity to swap one or more states (model parameter values) with a chain that has located a better performing region in parameter space. Using a process analogous to the genetic algorithm concepts of recombination and crossover [48], this swapping enables chains to escape from their local attractors.

G. Application to clinical questions

The Bayesian concept that "true" parameters are random variables is well-suited to study how anatomical variability in musculoskeletal systems affects biomechanical function and contributes to the success or failure of a clinical method. The MCMC approach is not appropriate for quick subject-specific parameterizations (as in [49]) due to its computationally intensive nature. Rather, one can address whether a model structure will lead to meaningful clinical predictions. All too often, a model structure is assumed and its large-scale contributions to the success or failure of the model are overlooked and overshadowed by the need to specify its numerical details (i.e., the parameter estimation problem). With MCMC, the user can test a hypothesized model structure (e.g., anatomical kinematic constraints) against experimental data to elucidate the capabilities and limitations of such a structure [50].

The MCMC methods described here can be used to test whether a "one size fits all" model will suffice for the entire population, or if subject-specific models will be necessary. For instance, one can run separate sets of MCMC simulations each driven by various subsets of the experimental data. One could compare the posterior distributions from a group of subjects with presumably similar levels of impairment to those of a single member of that subpopulation.

The MCMC approach can highlight critical treatment-relevant model parameters and may be useful for defining clinically-relevant subpopulations (e.g., different modes) for susceptibility to impairment and/or response to treatment. By tracking posterior distribution characteristics for a population of interest over time, it may be possible to observe the evolution of a disease state in model parameter space. Such insights could be used to devise objective methods of grading impairment. Strong correlations between easily-measured model parameters and functional behavior could be incorporated into diagnostic tools. Furthermore, one can determine which model parameter(s) can afford to be noisy because of anatomical and/or surgical variability. We continue to address these issues to produce clinically-useful models for studying the functional consequences of orthopedic and neurological diseases and treatments, and their treatment outcomes.

[2]Note that a standard serial "hill-climber" optimizer could not have found multiple solutions without a Monte Carlo simulation on its initial conditions.

APPENDIX
COMPARISON OF MONTE CARLO AND MARKOV CHAIN MONTE CARLO METHODS

Unencumbered by experimental data, Monte Carlo simulations take an open-loop "shotgun" approach to exploration of the parameter space. These feedforward simulations work well for low-dimensional systems, but become impractical when the dimensionality of the problem is scaled up [16]. The main advantage of Markov chain Monte Carlo methods is that they allow the use of simple sampling techniques and experimental data to estimate model parameters $\underline{\theta}$ of complex multivariate systems that are difficult or impossible to solve for in closed form or using computationally expensive, open-loop Monte Carlo methods.

The concept of a *chain* comes from the fact that each search is akin to a biased random walk starting at a random location in parameter space, which "drifts" towards a favorable region in parameter space. In our implementation, the "drift" is governed by the Metropolis-Hastings algorithm (described below), which seeks parameter values that lead to predictions that are most compatible with the experimental data. By starting chains from dispersed locations, and running them long enough, the chains will converge to our region(s) of interest in parameter space [34].

DERIVATION OF OUR METROPOLIS-HASTINGS ACCEPTANCE PROBABILITY

With our *symmetric* proposal distribution (eqn. 3), the Hastings ratio $\frac{q(\underline{\theta}_{(i-1)}|\underline{\theta}_{cand})}{q(\underline{\theta}_{cand}|\underline{\theta}_{(i-1)})}$ equals one and our acceptance probability (eqn. 2) simplifies to the special Metropolis case [37].

$$\alpha(\underline{\theta}_{cand}, \underline{\theta}_{(i-1)}) = \min\left\{\frac{\pi(\underline{\theta}_{cand})}{\pi(\underline{\theta}_{(i-1)})}, 1\right\} \tag{12}$$

Combining eqn. 12 with Bayes' Rule [22], [40],

$$\pi(\underline{\theta}) = p(\underline{\theta}|\mathbf{X}) = \frac{p(\mathbf{X}|\underline{\theta})p(\underline{\theta})}{p(\mathbf{X})} \tag{13}$$

we obtain

$$\alpha(\underline{\theta}_{cand}, \underline{\theta}_{(i-1)}) = \min\left\{\frac{p(\mathbf{X}|\underline{\theta}_{cand})p(\underline{\theta}_{cand})}{p(\mathbf{X}|\underline{\theta}_{(i-1)})p(\underline{\theta}_{(i-1)})}, 1\right\} \tag{14}$$

When using uniform prior distributions, $p(\underline{\theta}_{cand})$ and $p(\underline{\theta}_{(i-1)})$ cancel in eqn. 14, leaving a ratio of likelihood functions.

$$\alpha(\underline{\theta}_{cand}, \underline{\theta}_{(i-1)}) = \min\left\{\frac{p(\mathbf{X}|\underline{\theta}_{cand})}{p(\mathbf{X}|\underline{\theta}_{(i-1)})}, 1\right\} \tag{15}$$

Assuming that the experimental data matrix $\mathbf{X}$ is composed of n i.i.d. observations, each likelihood function can be written as the product of the probabilities across all n observations [40].

$$p(\mathbf{X}|\underline{\theta}) = \prod_{j=1}^{n} p(\underline{\mathbf{X}}_j|\underline{\theta}, \underline{q}_j, \underline{\dot{q}}_j) \tag{16}$$

The arguments $\underline{q}_j$ and $\underline{\dot{q}}_j$ represent the constant input joint angles and joint angular velocities for the j^{th} observation. We assume each observation $\underline{\mathbf{X}}_j$ (the j^{th} row of data matrix $\mathbf{X}$, see eqn. 1) is i.i.d. from a multivariate Gaussian distribution $MVN(f(\underline{\theta}), \hat{\Sigma})$ with a vector of means $f(\underline{\theta})$ that is a deterministic function of $\underline{\theta}$ and a covariance matrix $\hat{\Sigma} = diag(\hat{\sigma}_k^2)$, where $k = 1 \ldots d_{expt}$, the dimensionality of the experimental measurements (number of columns in $\mathbf{X}$).

Combining the probability density function for a multivariate Gaussian distribution [36] and the assumed covariance matrix $\hat{\Sigma} = diag(\hat{\sigma}_k^2)$ (using eqn. 5), we can rewrite the likelihood in eqn. 16 as

$$p(\mathbf{X}|\underline{\theta}) = \prod_{k=1}^{d_{expt}} \prod_{j=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\hat{\sigma}_k^2}} \exp\left[-\frac{1}{2\hat{\sigma}_k^2} [x_{j,k} - f_k(\underline{\theta}, \underline{q}_j, \underline{\dot{q}}_j)]^2 \right] \right\} \tag{17}$$

Combining equations 15 and 17, assuming that $\hat{\sigma}_{k,cand}^2 = \hat{\sigma}_{k,(i-1)}^2$, and simplifying, we get

$$\alpha(\underline{\theta}_{cand}, \underline{\theta}_{(i-1)}) = \min \left\{ \frac{\exp\left\{ -\frac{1}{2} \sum_{k=1}^{d_{expt}} \sum_{j=1}^{n} \frac{[x_{j,k} - f_k(\underline{\theta}_{cand}, \underline{q}_j, \underline{\dot{q}}_j)]^2}{\hat{\sigma}_{k,(i-1)}^2} \right\}}{\exp\left\{ -\frac{1}{2} \sum_{k=1}^{d_{expt}} \sum_{j=1}^{n} \frac{[x_{j,k} - f_k(\underline{\theta}_{(i-1)}, \underline{q}_j, \underline{\dot{q}}_j)]^2}{\hat{\sigma}_{k,(i-1)}^2} \right\}}, 1 \right\} \tag{18}$$

ACKNOWLEDGMENT

We thank Cornell University's Prof. Donald L. Bartel, Prof. Mark Campbell, and Dr. Madhusudhan Venkadesan for constructive feedback, Prof. Hod Lipson for supercomputer resources, and Vic Anand for C programming tips.

REFERENCES

[1] F. J. Valero-Cuevas, M. E. Johanson, and J. D. Towles, "Towards a realistic biomechanical model of the thumb: the choice of kinematic description may be more critical than the solution method or the variability/uncertainty of musculoskeletal parameters," *J. Biomech.*, vol. 36, pp. 1019–1030, July 2003.

[2] V. J. Santos and F. J. Valero-Cuevas, "Reported anatomical variability naturally leads to multimodal distributions of Denavit-Hartenberg parameters for the human thumb," *IEEE T. Bio-Med. Eng.*, vol. 53, pp. 155–163, Feb. 2006.

[3] P. L. Davidson, D. J. Chalmers, and B. D. Wilson, "Stochastic-rheological simulation of free-fall arm impact in children: application to playground injuries," *Comput. Methods. Biomech. Biomed. Engin.*, vol. 7, pp. 63–71, Apr. 2004.

[4] J. E. Langenderfer, R. E. Hughes, and J. E. Carpenter, "A stochastic model of elbow flexion strength for subjects with and without long head biceps tear," *Comput. Meth. Biomech. Biomed. Eng.*, vol. 8, pp. 315–322, Oct. 2005.

[5] R. E. Hughes and K. N. An, "Monte Carlo simulation of a planar shoulder model," *Med. Biol. Eng. Comput.*, vol. 35, pp. 544–548, Sept. 1997.

[6] Y. W. Chang, R. E. Hughes, F. C. Su, E. Itoi, and K. N. An, "Prediction of muscle force involved in shoulder internal rotation," *J. Shoulder Elb. Surg.*, vol. 9, pp. 188–195, May 2000.

[7] J. E. Langenderfer, J. E. Carpenter, M. E. Johnson, K. N. An, and R. E. Hughes, "A probabilistic model of glenohumeral external rotation strength for healthy normals and rotator cuff tear cases," *Ann. Biomed. Eng.*, vol. 34, pp. 465–476, Mar. 2006.

[8] Y. Y. Dhaher, "Monte carlo-based musculoskeletal modeling suggests passive tissue afferents have optimal role in promoting knee stability," presented at the 2nd Joint EMBS/BMES Conf. 2002.

[9] S. G. McLean, A. Su, and A. J. van den Bogert, "Development and validation of a 3D model to predict knee joint loading during dynamic movement," *J. Biomech. Eng. - T. ASME*, vol. 125, p. 864874, Dec. 2003.

[10] S. G. McLean, X. Huang, A. Su, and A. J. van den Bogert, "Sagittal plane biomechanics cannot injure the ACL during sidestep cutting," *Clin. Biomech.*, vol. 19, p. 828838, Oct. 2004.

[11] G. A. Mirka and W. S. Marras, "A stochastic model of trunk muscle coactivation during trunk bending," *Spine*, vol. 18, pp. 1396–1409, Sept. 1993.

[12] P. Perrier, J. Perkell, Y. Payan, M. Zandipour, F. Guenther, and A. Khalighi, "Degrees of freedom of tongue movements in speech may be constrained by biomechanics," presented at the Int. Conf. on Spoken Language Processing, 2000.

[13] K. G. Keenan and F. J. Valero-Cuevas, "Experimentally valid predictions of muscle force and EMG in models of motor-unit function are most sensitive to neural properties," *J. Neurophysiol.*, vol. 98, p. 15811590, Sept. 2007.

[14] P. J. Laz, J. Q. Stowe, M. A. Baldwin, A. J. Petrella, and P. J. Rullkoetter, "Incorporating uncertainty in mechanical properties for finite element-based evaluation of bone mechanics," *J. Biomech.*, vol. 40, p. 28312836, 2007.

[15] M. J. Camilleri, "Response-surface mapping to generate distributions of forward dynamic simulations," presented at the American Society of Biomechanics Annual Meeting, 2007.

[16] R. E. Bellman, *Adaptive control processes: a guided tour*. Princeton, NJ: Princeton University Press, 1961.

[17] M. A. Beaumont and B. Rannala, "The Bayesian revolution in genetics," *Nature Rev. Gen.*, vol. 5, pp. 251–261, Apr. 2004.

[18] S. Corazza, L. Mündermann , A. M. Chaudhari, T. Demattio, C. Cobelli, and T. P. Andriacchi, "A markerless motion capture system to study musculoskeletal biomechanics: Visual hull and simulated annealing approach," *Ann. Biomed. Eng.*, vol. 34, pp. 1019–1029, 2006.

[19] T. D. Johnson, R. M. Elashoff, and S. J. Harkema, "A Bayesian change-point analysis of electromyographic data: detecting muscle activation patterns and associated applications," *Biostat.*, vol. 4, pp. 143–164, Jan. 2003.

[20] W. E. Kincses, C. Braun, S. Kaiser, W. Grodd, H. Ackermann, and K. Mathiak, "Reconstruction of extended cortical sources for EEG and MEG based on a Monte-Carlo-Markov-Chain estimator," *Hum. Brain Mapp.*, vol. 18, p. 100110, Feb. 2003.

[21] P. G. Ridall and A. N. Pettitt, "Motor unit number estimation using reversible jump Markov Chain Monte Carlo methods," *Appl. Stat.-J. Roy. St. C*, vol. 56, p. 126, 2007.

[22] J. Gill, *Bayesian Methods: A Social and Behavioral Sciences Approach*. Boca Raton, FL: Chapman & Hall/CRC, 2002.

[23] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, pp. 1087–1092, June 1953.

[24] W. K. Hastings, "Monte Carlo sampling methods using Markov Chains and their applications," *Biometrika*, vol. 57, pp. 97–109, Apr. 1970.

[25] G. Roberts, "Markov Chain concepts related to sampling algorithms," in *Markov Chain Monte Carlo in Practice*, W. Gilks, S. Richardson, and D. Spiegelhalter, Eds. Chapman & Hall/CRC, 1996, pp. 45–57.

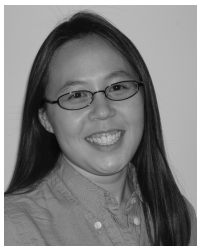
[26] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi, *GNU Scientific Library Reference Manual*, 2004.

[27] A. M. Hollister, W. L. Buford, L. M. Myers, D. J. Giurintano, and A. Novick, "The axes of rotation of the thumb carpometacarpal joint," *J. Orthopaed. Res.*, vol. 10, pp. 454–460, May 1992.

[28] A. M. Hollister, D. J. Giurintano, W. L. Buford, L. M. Myers, and A. Novick, "The axes of rotation of the thumb interphalangeal and metacarpophalangeal joints," *Clin. Orthop. Relat. R.*, vol. 320, pp. 188–193, Nov. 1995.

[29] D. J. Giurintano, A. M. Hollister, W. L. Buford, D. E. Thompson, and L. M. Myers, "A virtual five-link model of the thumb," *Med. Eng. Phys.*, vol. 17, pp. 297–303, June 1995.

[30] V. J. Santos and F. J. Valero-Cuevas, "Investigating the interaction between variability in both musculoskeletal structure and muscle coordination for maximal voluntary static thumb forces," presented at the Soc. Neural Control Movement Annu. Meeting, 2004.
[31] V. J. Santos, "A Bayesian approach to biomechanical modeling: A treatise on the human thumb," Mechanical and Aerospace Engineering Dept., Cornell University," Ph.D. Dissertation, May 2007.
[32] A. Schwab, "Quaternions, finite rotation and Euler parameters - lecture notes," Mechanical Engineering Dept., Delft University of Technology, Delft, Netherlands, Tech. Rep., May 2002.
[33] C. Hall, "Spacecraft attitude dynamics and control - lecture notes," Aerospace and Ocean Engineering Dept., Virginia Tech, Blacksburg, VA, Tech. Rep., Feb. 2003.
[34] R. B. Schinazi, *Classical and Spatial Stochastic Processes*. Boston, MA: Birkhäuser, 1999.
[35] J. C. Spall, "Estimation via Markov Chain Monte Carlo," *IEEE Contr. Syst. Mag.*, vol. 23, pp. 34–45, Apr. 2003.
[36] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC, 2003.
[37] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Boca Raton, FL: Chapman & Hall/CRC, 1996.
[38] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Stat. Sci.*, vol. 7, pp. 457–472, Nov. 1992.
[39] A. Gelman, "Inference and monitoring convergence," in *Markov Chain Monte Carlo in Practice*, W. Gilks, S. Richardson, and D. Spiegelhalter, Eds. Chapman & Hall/CRC, 1996, pp. 131–143.
[40] J. A. Rice, *Mathematical Statistics and Data Analysis*. Belmont, CA: Duxbury Press, 1995.
[41] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York, NY: John Wiley & Sons, 2001.
[42] F. J. Valero-Cuevas, "Predictive modulation of muscle coordination pattern magnitude scales fingertip force magnitude over the voluntary range," *J. Neurophys.*, vol. 83, pp. 1469–1479, Mar. 2000.
[43] M. A. Hurn, H. Rue, and N. A. Sheehan, "Block updating in constrained Markov Chain Monte Carlo sampling," *Stat. Probabil. Lett.*, vol. 41, pp. 353–361, Feb. 1999.
[44] J. O. Berger, "Bayesian analysis: A look at today and thoughts of tomorrow," *J. Am. Stat. Assoc.*, vol. 95, pp. 1269–1276, Dec. 2000.
[45] G. O. Roberts and J. S. Rosenthal, "Optimal scaling for various Metropolis-Hastings algorithms," *Stat. Sci.*, vol. 16, pp. 351–367, Nov. 2001.
[46] C. Geyer, "Markov Chain Monte Carlo maximum likelihood," presented at the Computing Science and Statistics, 23rd Symp. on the Interface, 1991.
[47] J. Gill and G. Casella, "Dynamic tempered transitions for exploring multimodal posterior distributions," *Polit. Anal.*, vol. 12, pp. 425–443, 2004.
[48] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA: Addison-Wesley, 1989.
[49] J. A. Reinbolt, J. F. Schutte, B. J. Fregly, B. I. Koh, R. T. Haftka, A. D. George, and K. H. Mitchell, "Determination of patient-specific multi-joint kinematic models through two-level optimization," *J. Biomech.*, vol. 38, pp. 621–626, Mar. 2005.
[50] F. J. Valero-Cuevas, V. V. Anand, A. Saxena, and H. Lipson, "Beyond parameter estimation: extending biomechanical modeling by the explicit exploration of model topology," *IEEE T. Bio-Med. Eng.*, vol. 54, pp. 1951–1964, Nov. 2007.



Veronica J. Santos (M'03) earned a B.S. in mechanical engineering with a music minor from the University of California at Berkeley in 1999, and an M.S. and Ph.D. in mechanical engineering with a biometry minor from Cornell University, Ithaca, NY in 2004 and 2007, respectively.

She is currently Assistant Professor at the Department of Mechanical and Aerospace Engineering at Arizona State University, Tempe, AZ. From 2007 to 2008, she was a Postdoctoral Research Associate at the A. E. Mann Institute for Biomedical Engineering at the University of Southern California. Her research interests include hand biomechanics, neural control of movement, robotics, stochastic modeling, and clinical applications of biomechanical modeling for prosthetics and rehabilitation technology.

Prof. Santos is a member of the IEEE Engineering in Medicine and Biology Society, the American and International Societies of Biomechanics, the American Society of Mechanical Engineers, the Biomedical Engineering Society, the Society for the Neural Control of Movement, and the Society of Women Engineers. She has received the Young Investigator Poster Presentation Award from the International Society of Biomechanics (2005), an Exceptional Teaching Assistant Award from the Sibley School of Mechanical and Aerospace Engineering at Cornell University (2005), and a National Science Foundation Graduate Research Fellowship (2001).

Carlos D. Bustamante earned a B.A. in biology (1997), an M.S. in statistics (2001), and a Ph.D. in biology (2001) from Harvard University.

He is currently Associate Professor in the Department of Biological Statistics and Computational Biology and in the Department of Statistical Sciences at Cornell University. His research interests focus on population genetics and genomics, statistical genetics and genomics, computational statistics, and application of Markov Chain Monte Carlo techniques to problems in biology.

Prof. Bustamante is a member of the Society for Molecular Biology and Evolution and the American Society of Human Genetics. He has received various awards including a Marshall-Sherfield fellowship to the University of Oxford (Oxford, UK) where he was in the Mathematical Genetics Group of the Department of Statistics from 2001 through 2002 as a post-doctoral fellow. He was a National Science Foundation Graduate Research Fellow (1998), and a Howard Hughes Graduate Fellow (1999-2001). He is currently an Alfred P. Sloan Research Fellow in Molecular Biology (2007-2009), and recipient the Cornell Provost's Award for Distinguished Research (2008).

Francisco J. Valero-Cuevas (M'99) earned a B.S. in engineering from Swarthmore College (1988), and an M.S. and Ph.D. in mechanical engineering from Queen's University in Kingston, Ontario, Canada (1991) and Stanford University (1997), respectively.

He is currently Associate Professor at the Department of Biomedical Engineering, and Division of Biokinesiology and Physical Therapy at the University of Southern California. His research interests focus on combining engineering, robotics, mathematics and neuroscience to understand organismal and robotic systems for basic science, engineering and clinical applications.

Prof. Valero-Cuevas is member of the IEEE Engineering in Medicine and Biology Society, the American and International Societies of Biomechanics, the American Society of Mechanical Engineers, the Society for Neuroscience, and the Society for the Neural Control of Movement. He has received a Research Fellowships from the Alexander von Humboldt (2005) Foundation, the Post-Doctoral Young Scientist Award from the American Society of Biomechanics (2003), the Faculty Early Career Development Program CAREER Award from the National Science Foundation (2003), the Innovation Prize from the State of Tyrol in Austria (1999), a Fellowship from the Thomas J. Watson Foundation (1988), and was elected Associate Member of the Scientific Research Society Sigma-Xi (1988). He served as Associate Editor for the IEEE Transactions on Biomedical Engineering from 2003-08.